Online Supplement

This online Supplement Includes

I. eAppendix 2, with 1) Sketch of proof for main result; 2) Notation and Definitions; 3) The Main Theorem and proof; 4) Other Standard Populations; 5) Lower Bound; 6) Supporting theorems (statement only).

II. eAppendix 3, with statement and proofs of all supporting Theorems, Propositions and Lemmas.

III. eAppendix 4. Simulation Experiment to Evaluate Bias, compare Bounds and the R code for simulations

IV. eAppendix 5. Comparison of Bd with a bound assuming homogeneity.

V. eAppendix 6. R code to calculate bias (implements Equation 3), and bounds.

I. eAppendix 2:

eAppendix 2 contains: 1) a sketch of the main ideas of our proof that Bd in expression (3) bounds the bias; 2) key definitions and notation; 3) the main Theorem and its proof; and 4) a statement of Theorems 1-3 used to prove it. Proofs of Theorems 1-3 and the Propositions are in eAppendix 3.

1) Sketch of proof for main result:

The main ideas of our proof that Bd (expression 3, main text) bounds the bias are as follows. We characterize the strength of association between the variables E, V, C, N, D whose effects create the collider bias using three limits R_E, R_D and R_C. Using B_{max} to denote the maximum bias for all situations wherein the strengths of association don't exceed these limits, we argue (main Theorem) that: regularity conditions assure that we can approximate continuous or mixtures of continuous and categorical variables by their categorical counterparts so that the maximum bias for *categorical* variables is within ε of B_{max}. More precisely, $B_1 > B_{max} - \varepsilon$ where B_1 is the bias for some set of categorical variables E_1, V_1, C_1, N_1, D_1. We then show by construction (Theorems 1 and 2 and supporting propositions) that we can modify the probability distributions for these categorical variables to obtain new categorical variables E_2, V_2, C_2, N_2, D_2 with bias B_2 such that: *i*) $B_2 > B_{max} - 7\varepsilon$; and *ii*) the contribution to B_2 is negligible from all sets, other than two specific sets we explicitly define. This simplifies the bias expression, since we only have to consider v in one of these two sets. We next collect the approximation errors (Theorem 3) and show by construction that we can replace E_2, V_2, C_2, N_2, D_2 with variables E_3, V_3, C_3, N_3, D_3 such that *i*) all variables are dichotomous; and, *ii*) $B_3 > B_{max} - 9\varepsilon$. The resulting bias expression is further simplified and now involves only five parameters. We then use optimization techniques to find the maximum bias $\max(B_3)$ for such dichotomous variables. These steps establish that $B_{max} \geq \max(B_3) >$

$B_{max} - 9\varepsilon$. Since ε is arbitrary, $\max(B_3)$ must equal B_{max} justifying use of $\max(B_3)$ as the bound: $Bd = \max(B_3)$.

2) Notation and Definitions:

We only consider values of $R_E, R_C, R_D > 1$.

ϕ is the empty set;

$|x|$ has two meanings: if x is a number, $|x|$ is its absolute value; if x is a set, $|x|$ is its cardinality.

$\wp(\vec{V})$ is the set of variables $\vec{V} = E, V, C, N, D$ *and* their associated probability functions; that is $\wp(\vec{V}) = \{\vec{V};\ p_1(v), p_0(v), p_c(v,1), p_c(v,0), p_D(n), p_N(n)\}$ where E, N, D are dichotomous, V is categorical, but otherwise unrestricted, and the causal relationships of Figure 1C hold. Here, $p_e(v) = P(V = v|E = e) = p_{v,e}$, for $e = 0,1$, $p_c(v,n) = P(C = c|V = v, N = n)$ for $n = 0,1$ where c is the value of C that is selected or conditioned on in Figure 1C, $p_D(n) = p_D(D = 1|N = n)$ and $p_N(n) = P(N = n)$ with $O_N = \frac{p_N(N=1)}{p_N(N=0)}$ for $n = 0,1$. If $|V| < \infty$, we can parameterize these functions as: $p_e(v) = \pi_{v,e}$, $p_c(v,n) = \pi_{c,v,n}$, $p_D(n) = \pi_{D,n}$ and $p_N(n) = \pi_{N,n}$ for $v = 1,2,\ldots,|V|; e = 0,1; n = 0,1$.

$\wp(\vec{V})$ "satisfies constraint $CR(R_E, R_C, R_D)$" if and only if: $\max_{v,e} \frac{p_e(v)}{p_{1-e}(v)} \le R_E$, $\frac{\max_{v,n} p_c(v,n)}{\min_{v,n} p_c(v,n)} \le R_C$, $\max_n \frac{p_D(n)}{p_D(1-n)} \le R_D$

$\wp(\vec{V})$ "satisfies constraint $CL(mn, mx, mn_c, mx_c, O_N)$" if and only if:

$$mn \le p_e(v) \le mx,\ \text{for } e = 0,1;\ \max_{v,n} p_c(v,n) \le mx_c,\ \min_{v,n} p_c(v,n) \ge mn_c, \frac{p_N(N=1)}{p_N(N=0)} = O_N.$$

$B\left(\wp(\vec{V})\right)$ is the bias B for $\wp(\vec{V}) = \{\vec{V}; p_1(v), p_0(v), p_c(v,1), p_c(v,0), p_D(n), p_N(n)\}$.

$\mathcal{H} = \mathcal{H}(\mathcal{n}, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, \varepsilon)$ denotes the collection of all sets $\wp(\vec{V}) = \{\vec{V};\ p_1(v), p_0(v), p_c(v,1), p_c(v,0), p_D(n), p_N(n)\}$ such that: $|V| = \mathcal{n}$; $\wp(\vec{V})$ satisfies constraints $C_R(R_E, R_C, R_D)$ and $C_L(mn, mn, mn_c, mx_c, O_N)$; and, the bias $B(\wp(\vec{V}))$ satisfies $B(\wp(\vec{V})) \ge B_{max} - \varepsilon$.

Let B_{max} be the (overall) maximum of the bias $B(\wp(\vec{V}))$ defined in Equation (2) of the main text, for all probability functions that satisfy constraint $C_R(R_E, R_C, R_D)$. Our goal is to prove that $B_{max} \le Bd$.

Define the four sums by:

$$\Sigma_1 = \textstyle\sum_v p_1(v)(P(D = 1|E = 0, N = 1) \cdot O_N \cdot p_c(v,1) + P(D = 1|E = 0, N = 0) \cdot p_c(v,0))$$

$$\Sigma_2 = \textstyle\sum_v p_1(v)(O_N \cdot p_c(v,1) + p_c(v,0))$$

$$\Sigma_3 = \textstyle\sum_v p_0(v)(P(D = 1|E = 0, N = 1) \cdot O_N \cdot p_c(v,1) + P(D = 1E = 0, |N = 0) \cdot p_c(v,0))$$

$\Sigma_4 = \sum_v p_0\,(v)(O_N \cdot p_c(v,1) + p_c(v,0))$

With this notation, $B(\wp(\vec{V}))$ can be written as $B(\wp(\vec{V})) = \frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$.

It is useful to note that we can write Σ_1 to Σ_4 in several different forms, yet still have $B(\wp(\vec{V})) = \frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$. For example, since N is dichotomous, we can multiply Σ_1 to Σ_4 by $\frac{p_N(N=0)}{\min_n P(D=1|E=0,N=0)}$ and write $\frac{P(D=1|E=0,N=1)}{P(D=1|E=0,N=0)} = R_D$, $q_N = 1 - p_N$ and re-express Σ_1 to Σ_4 as:

$$\Sigma_1 = \sum_v p_1\,(v)(R_D \cdot p_N \cdot p_c(v,1) + q_N \cdot p_c(v,0)); \qquad \Sigma_2 = \sum_v p_1\,(v)(p_N \cdot p_c(v,1) + q_N \cdot p_c(v,0))$$
$$\Sigma_3 = \sum_v p_0\,(v)(R_D \cdot p_N \cdot p_c(v,1) + q_N \cdot p_c(v,0)); \qquad \Sigma_4 = \sum_v p_0\,(v)(p_N \cdot p_c(v,1) + q_N \cdot p_c(v,0))$$

The derivative with respect to R_D is: $\frac{\partial \ln{(B)}}{\partial R_D} = O_N[\frac{\sum_v p_1(v)p_c(v,1)}{\Sigma_1} - \frac{\sum_v p_0(v)p_c(v,1)}{\Sigma_3}]$. If this derivative is negative, we can relabel the levels of N so that the bias is increasing as a function of R_D. Thus, we still have $B(\wp(\vec{V})) = \frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$, if we take R_D to be the maximum R_D, with $R_D \geq 1$ and write:

$\Sigma_1 = \sum_v p_1\,(v)(R_D O_N p_c(v,1) + p_c(v,0))$ and $\Sigma_3 = \sum_v p_0\,(v)(R_D O_N p_c(v,1) + p_c(v,0))$.

Define the collections of subsets of V: $A_i = \{v: p_i(v) = mx\}$, $C_i = \{v: p_i(v) = mn\}$, $D_i = \{v: mx > p_i(v) > mn\}$ $i = 0{,}1$. The collection of subsets, (A_e, C_e, D_e) for $e = 0{,}1$ is mutually exclusive and collectively exhaustive for V. Note that A_e, C_e, D_e depend on $p_e(v) \in \wp(V)$, but for notational simplicity, we suppress that dependency here and elsewhere. There are 9 combinations of subsets $(A_1, C_1, D_1) \times (A_0, C_0, D_0)$.

3) *Theorem (Main)*. Fix $R_E, R_D, R_C > 1$ and let $\varepsilon > 0$. Assume that inequalities (5) of the main text hold for E, M, N, C and D and that structural selection bias is present (Figure 1C). Then the bias (Equation 3, main text) in the conditional RR doesn't exceed $Bd = \frac{R_E(\sqrt{R_D}R_C+1)+(\sqrt{R_D}+R_C)}{R_E(R_C+\sqrt{R_D})+(1+\sqrt{R_D}R_C)} \times \frac{(R_C+\sqrt{R_D})+R_E(1+\sqrt{R_D}R_C)}{(\sqrt{R_D}R_C+1)+R_E(\sqrt{R_D}+R_C)}$.

Proof: Under mild regularity conditions (e.g., similar to those used by Ding and VanderWeele [1], R_E, R_D and R_C are defined for continuous variables, categorical variables or a mixture of continuous and categorical variables. Furthermore, we need consider only categorical variables. This restriction is justified since, if the bias exceeds a particular number x for some R_E, R_D, R_C, and continuous variables E, V, C, N and D with distributions satisfying the inequalities in (5), mild regularity assumptions imply that the continuous variables and distributions can be approximated by categorical variables and distributions also satisfying the inequalities (5) so that the bias also exceeds x for the approximating categorical variables and distributions. Therefore, establishing the bound Bd in the categorical case also establishes it in the continuous case and justifies subsequent restriction of the proof to categorical variables.

By Theorem 3 (stated below), given fixed $mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D$ and $\varepsilon > 0$, there exists $\wp(\vec{V}_2) \in \mathcal{H} = \mathcal{H}(2, mn, mx, mn_c, mx_c, O_N, R_E R_C, R_D, 9\varepsilon)$ such that $B(\wp(\vec{V}_2))$ can be written as:

$$B(\wp(\vec{V}_2)) = \frac{R_E(R_D O_N r_c+1)+(R_D O_N+r_c'')}{R_E(O_N r_c+1)+(O_N+r_c'')} \times \frac{(O_N r_c+1)+R_E(O_N+r_c'')}{(R_D O_N r_c+1)+R_E(R_D O_N+r_c'')}$$

$\wp(\vec{V}_2) \in \mathcal{H}$ implies $\max_{v,e} \frac{p_e(v)}{p_{1-e}(v)} \leq R_E$, $\frac{\max_{v,n} p_c(v,n)}{\min_{v,n} p_c(v,n)} \leq R_C$, $\max_n \frac{p_D(n)}{p_D(1-n)} \leq R_D$ so that E, M, N, C and D satisfy inequalities 5 (main text) and $B(\wp(\vec{V}_2)) > B_{max} - 9\varepsilon$; all terms are defined in Theorem 3.

Subject to the constraints: $1 \leq r_c \leq R_C$, $1 \leq r_c'' \leq R_C$ and $0 \leq O_N$, $B(\wp(\vec{V}_2))$ attains its maximum at $r_c = R_C$, $r_c'' = R_C$, $O_N = 1/\sqrt{R_D}$. We found the maximum using the optimization method of Byrd et al. [2], an approach designed to solve non-linear problems using a gradient projection method and implemented in the optimr package in R (version 3.3.1). We used a grid of 100 starting values to provide reassurance that the maximum was identified and also verified the solution using the Maximize function in Wolfram Mathematica 11 [3].

Substituting these values of r_c, r_c'' and O_N into Equations (10), yields the bound:

$$Bd = \frac{R_E(\sqrt{R_D}R_C+1)+(\sqrt{R_D}+R_C)}{R_E(R_C+\sqrt{R_d})+(1+\sqrt{R_D}R_C)} \times \frac{(R_C+\sqrt{R_d})+R_E(1+\sqrt{R_D}R_C)}{(\sqrt{R_d}R_C+1)+R_E(\sqrt{R_D}+R_C)}$$

Since $Bd \geq B_{max} - 9\varepsilon$, $B_{max} \geq Bd$ by definition and ε is arbitrary, we must have $Bd = B_{max}$, proving the main Theorem.

Note: the solution above uses $r_c \leq R_C$ and $r_c'' \leq R_C$. However, if strong substantive knowledge indicates that $r_c \leq R_C$ and $r_c'' \leq R_C''$, with $R_C \neq R_C''$, a tighter bound may be obtained by optimizing $B(\wp(\vec{V}_2))$ subject to these alternative constraints.

4) Other Standard Populations.

In the main text we considered the exposed, selected population as the standard, but it is straightforward to also consider the unexposed or the entire selected population as the standard. Let B be the bias and sRR be the (causal) standardized risk ratio, with the exposed as the standard (Equation 2, main text), i.e. the effect of treatment on the treated. Denote the bias and the (causal) standardized risk ratio, if the *unexposed* were the standard as, say B_0 and s_0RR, where B_0 is the ratio of the oRR divided by s_0RR. In direct analogy to the case with the exposed as the standard (main text, Equation 3), B_0 is given by:

$$B_0 = \frac{\sum_{v,u} P(V=v|E=1)P(U=u)P(D=1|E=1,U=u)P(C=c|V=v,U=u)}{\sum_{v,u} P(V=v|E=0)P(U=u)P(D=1|E=1,U=u)P(C=c|V=v,U=u)} \cdot \frac{\sum_{v,u} P(V=v|E=0)P(U=u)P(C=c|V=v,U=u)}{\sum_{v,u} P(V=v|E=1)P(U=u)P(C=c|V=v,U=u)}$$

This Equation shows that B_0 has just the same structure as B, except that $P(D=1|E=0,U=u)$ in Equation (2) is replaced by $P(D=1|E=1,U=u)$. Thus, our arguments justifying bound Bd for B also hold for B_0, provided $R_D^1 = \frac{\max_n P(D=1|E=1,N=n)}{\min_n P(D=1|E=1,N=n)}$ replaces R_D in Equation (4) of the main text.

If the distribution of U among *all* those selected is the standard, much as in Ding and VanderWeele{Ding, 2016 #326} , we have:

$$ACE = w \cdot sRR + (1-w) \cdot s_0RR, \quad \text{where}$$

$$w = \frac{P(E=1|C=c)\sum_u P(D=1|E=0,U=u,C=c)P(U=u|E=1,C=c)}{P(E=1|C=c)\sum_u P(D=1|E=0,U=u,C=c)P(U=u|E=1,C=1) + P(E=0|C=c)\sum_u P(D=1|E=0,U=u,C=c)P(U=u|E=0,C=1)},$$ and ACE

is the standardized risk ratio with all in the selected population as the standard (the average causal effect for this population). If we define R_D^{mx} as max (R_D, R_D^1) and calculate Bd_{ACE} using R_E, R_D^{mx} and R_C in Equation (4, main text), we obtain a bound for the bias in the average causal effect B_{ACE}:

$$B_{ACE} = (\text{oRR}/ACE) \leq Bd_{ACE}.$$

5) Lower Bound.

Here we use our main results to derive a lower bound, say $\underline{Bd}$, for the bias (Equation 3, main text) when oRR is used to estimate effects of exposure E. We establish that: $\underline{Bd} = \frac{1}{Bd}$. To derive this result, we define a new variable $\bar{E} = 1 - E$. With those "unexposed" to $\bar{E}$ as the standard ($\bar{E} = 0$) in the definition of the effect of $\bar{E}$, the bias is:

$$B_0(\bar{E}) = \frac{\sum_{v,u} P(V=v|\bar{E}=1)P(U=u)P(D=1|\bar{E}=1,U=u)P(C=c|V=v,U=u)}{\sum_{v,u} P(V=v|\bar{E}=0)P(U=u)P(D=1|\bar{E}=1,U=u)P(C=c|V=v,U=u)} \times \frac{\sum_{v,u} P(V=v|\bar{E}=0)P(U=u)P(C=c|V=v,U=u)}{\sum_{v,u} P(V=v|\bar{E}=1)P(U=u)P(C=c|V=v,U=u)}$$

$$= \frac{\sum_{v,u} P(V=v|E=0)P(U=u)P(D=1|E=0,U=u)P(C=c|V=v,U=u)}{\sum_{v,u} P(V=v|E=1)P(U=u)P(D=1|E=0,U=u)P(C=c|V=v,U=u)} \times \frac{\sum_{v,u} P(V=v|E=1)P(U=u)P(C=c|V=v,U=u)}{\sum_{v,u} P(V=v|E=0)P(U=u)P(C=c|V=v,U=u)}$$

Then $B_0(\bar{E}) = \frac{1}{B}$, where B is given by Equation (2), main text. Applying our result for the upper bound on the bias in estimating the effects of $\bar{E}$, with those "unexposed" to $\bar{E}$ as the standard ($\bar{E} = 0$), gives:

$$\frac{1}{B} = B(\bar{E}) \leq \left(\frac{R_E\left(\sqrt{R_D^1}R_C+1\right)+\left(R_C+\sqrt{R_D^1}\right)}{\left(\sqrt{R_D^1}R_C+1\right)+R_E\left(R_C+\sqrt{R_D^1}\right)} \right)^2,$$ where $R_D^1 = \frac{\max_n P(D=1|\bar{E}=1,N=n)}{\min_n P(D=1|\bar{E}=1,N=n)} = \frac{\max_n P(D=1|E=0,N=n)}{\min_n P(D=1|\bar{E}=0,N=n)} =$

R_D.

Combining results: $\frac{1}{Bd} = \underline{Bd} \leq B \leq Bd.$

Simulations like those described in eAppendix 4 where the exposed, selected population was the standard, yield similar results with the bias approaching but never exceeding the bound when the unexposed, selected population is the standard and also for the lower bound (data not shown).

5) Supporting Theorems:

We now state Theorems 1-3. Their proofs and supporting Propositions and Lemmas are in eAppendix 3.

***Theorem* 1.** Let R_E, R_C, R_D and $\varepsilon > 0$ be given. Set $\varepsilon' = \frac{R_E-1}{140R_D}\varepsilon$ and let B_{max} be the maximum of the bias B defined in Equation (3; main text) for all probability functions that satisfy constraint $C_R(R_E, R_C, R_D)$ and the causal relationships of Figure 1C. Then $\exists$ numbers mn, mx, mn_c, mx_c and O_N and $\aleph < \infty$ and variables with associated probability functions $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 4\varepsilon)$ such that: (i) $p_e(\{v: v \in A_1 \cap A_0\}) \leq \varepsilon'$, for $e = 0{,}1$. Proof. eAppendix 3.

***Theorem* 2** Fix $\varepsilon > 0$, $R_E, R_C, R_D > 1$, mn, mx, mn_c, mx_c with $\frac{mx}{mn} \leq R_E$ and $\varepsilon' = \frac{R_E-1}{140R_D}\varepsilon$. If $\mathcal{H}(\mathcal{n}', mn, mx, mn_c, mx_c, R_E, R_C, R_D, 5\varepsilon) \neq \phi$, then $\exists$ a number $\aleph < \infty$ and $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 7\varepsilon)$ such that conclusions of Proposition 3 hold (including those of Theorem 1); and, $p_e(\{v: v \in C_1 \cap C_0\}) \leq \varepsilon'$, for $e = 0{,}1$. Proof: eAppendix 3.

***Theorem* 3**: Fix $mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D$, For every $\varepsilon > 0$ there exists $\wp(\vec{V}_2) \in \mathcal{H}(2, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 9\varepsilon)$ with $B(\wp(\vec{V}_2)) = \frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$ such that: $B(\wp(\vec{V}_2)) > B_{max} - 8\varepsilon$, where: $\Sigma_1 = R_E(R_D O_N r_c + 1) + (R_D O_N + r_c'')$; $\Sigma_2 = R_E(O_N r_c + 1) + (O_N + r_c'')$

$\Sigma_3 = (R_D O_N r_c + 1) + R_E(R_D O_N + r_c'')$; $\Sigma_4 = (O_N r_c + 1) + R_E(O_N + r_c'')$

Proof: eAppendix 3.

II. eAppendix 3:

eAppendix 3 contains proofs of Theorems, Propositions, and Lemmas used to prove the main Theorem, assumptions of which are used throughout. We first prove three Lemmas that will be used in subsequent proofs of Theorems 1-3 and Propositions.

Note: Theorems 1 and 2 have analogous proofs, with similar construction steps; this is also true of Propositions 1, 2, 4 and 5.

***Lemma* 1**. Fix $mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D$ and ε. Consider $\mathcal{H} = \mathcal{H}(\mathcal{n}, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, \varepsilon)$. If $\mathcal{H} \neq \phi$, then for some $\mathcal{n}' < \infty$, $\exists$ variables and functions $\wp(\vec{V}') \in \mathcal{H}(\mathcal{n}', mn', mx', mn_c, mx_c, O_N, R_E, R_C, R_D, 2\varepsilon)$ such that:$|V'| = \mathcal{n}'$, $p_e(v')$ has bounds mn' and mx' such that: $mn' \leq p_e(v') \leq mx$' and $\frac{mx\prime}{mn\prime} \leq R_E$.

Proof: Let $\wp(\vec{V}) \in \mathcal{H}$, with $B(\wp(\vec{V})) > B_{max} - \varepsilon$. Under mild regularity, $p_e(v),\ e = 0{,}1$ can be approximated by $p_e{}'(v) \in \wp(\vec{V}')$ with range $\mathbb{Q}$ (rational numbers) so that: $B(\wp(\overrightarrow{V'})) > B_{max} - 2\varepsilon$. For simplicity, relabel V' as V. Use the following algorithm to recursively define a new variable and functions $\wp(\vec{V}')$ with the following Steps:

0. Let $mn = \min_{e,v} p_e(v)$. Define $G = \{v\!:\ \max_e p_e(v) > R_E \cdot mn\}$. If G is empty, *stop*. Otherwise,

1. Pick $v_1 \in G$ such that $\max_e p_e(v_1) = \max_{e,v\in G} p_e(v)$. Let $mx_{v_1} = \max_e p_e(v_1)$. If $\max_e p_e(v_1) \leq R \cdot mn$, *stop*. Otherwise: $\exists$ integers $K_1 > K_2$ such that $mx_{v_1}/K_1 = R_E \cdot mn/K_2$. Label the categories of V: $v_1, \dots., v_{\mathcal{n}}$.
2. Create a new variable V' with $|V'| \leq K_1 + (\mathcal{n} - 1) \times K_2$ labeled as $v'_{i,j}$ for $i = 1,2, \dots, \mathcal{n}$, $j = 1,2, \dots, J_i$ where $J_i = K_1$ if $\max_e p_e(v_i) = mx_{v_1}$ and $J_i = K_2$ otherwise.
3. Variables E, C, N, D and the probability functions $p_D(n)$ and $p_N(n)$ are unchanged.
4. Define new probability functions $p'_e(v'_{i,j}), p'_{1-e}(v'_{i,j})$ and $p'_c(v'_{i,j}, n)$ as:
   $p'_e(v'_{i,j}) = \frac{p_e(v_i)}{J_i}$; $p'_c(v'_{i,j}, n) = p_c(v_i, n)$, for $i = 1,2, \dots, \mathcal{n};\ j = 1,2, \dots, J_i;\ e, n = 0{,}1$.
5. Define: $mn_{new} = \frac{mn}{K_2}$; $mx_{new} = \max_{v',e} p'_e(v')$.
6. Relabel: $mn = mn_{new}$, $mx = mx_{new}$, drop the primes from V' and the probability functions $\wp(V')$. Return to Step 0.

Each iteration reduces the number of unique values of $p_e(v)$ outside the interval $[mn_{new},\ mx_{new}]$, assuring that the algorithm eventually stops. With each iteration, the values of $p_c(v, n)$ and, because of duplication, $B(\wp((\overrightarrow{V'}))$ don't change. Therefore at completion, the final variables and functions $\wp(\vec{V}')$ still satisfy constraint $C_R(R_E, R_C, R_D)$, and the construction assures $mx = R_E \cdot mn$.

Lemma 2: If $B = B(\wp(\vec{V})) > 1$ and $R_E > 1$, then $\frac{R_E\Sigma_4 - \Sigma_3}{R_E\Sigma_2 - \Sigma_1} > \frac{\Sigma_3 - \Sigma_4}{\Sigma_1 - \Sigma_2}$.

Proof: by contradiction. If the conclusion is false, then we would have $\frac{R_E\Sigma_4 - \Sigma_3}{R_E\Sigma_2 - \Sigma_1} \le \frac{\Sigma_3 - \Sigma_4}{\Sigma_1 - \Sigma_2}$. Dividing both sides by $\Sigma_3 - \Sigma_4$ and multiplying both sides by $R_E\Sigma_2 - \Sigma_1$ would give:

$\frac{R_E\Sigma_4 - \Sigma_3}{\Sigma_3 - \Sigma_4} < \frac{R_E\Sigma_2 - \Sigma_1}{\Sigma_1 - \Sigma_2}$. Dividing numerator and denominator on the left (right) by Σ_3 (Σ_1) would imply $\frac{R_E\frac{\Sigma_4}{\Sigma_3} - 1}{1 - \frac{\Sigma_4}{\Sigma_3}} < \frac{R_E\frac{\Sigma_2}{\Sigma_1} - 1}{1 - \frac{\Sigma_2}{\Sigma_1}}$. Cross multiplying by $1 - \frac{\Sigma_4}{\Sigma_3}$ and $1 - \frac{\Sigma_2}{\Sigma_1}$ then would imply:

$R_E\frac{\Sigma_4}{\Sigma_3} - 1 - R_E\frac{\Sigma_4}{\Sigma_3}\frac{\Sigma_2}{\Sigma_1} + \frac{\Sigma_2}{\Sigma_1} < R_E\frac{\Sigma_2}{\Sigma_1} - 1 - R_E\frac{\Sigma_4}{\Sigma_3}\frac{\Sigma_2}{\Sigma_1} + \frac{\Sigma_4}{\Sigma_3}$. Canceling common terms gives:

$\frac{\Sigma_4}{\Sigma_3}(R_E - 1) < \frac{\Sigma_2}{\Sigma_1}(R_E - 1)$, which would imply $B < 1$ if $R_E > 1$, a *contradiction*, completing the proof.

***Lemma* 3**. For fixed $n < \infty$, mn, mx, mn_c, mx_c R_E, R_C, R_D, O_N and ε, let $\mathcal{H} = \mathcal{H}(n, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, \varepsilon)$ where $\frac{mx}{mn} \le R_E$. If $\mathcal{H} \ne \phi$, $\exists\ \wp(\vec{V}') \in \mathcal{H}$ such that: $p_c(v, n)$ has the same value for $n = 0{,}1$ and all $v \in A_1 \cap C_0$, and similarly for v in each combinations of sets $(A_1, C_1) \times (A_0, C_0)$.

Proof: Define $S = A_1 \cap C_0$. Create $\wp(\vec{V}') \in \mathcal{H}$ from $\wp(\vec{V}) \in \mathcal{H}$, where $\wp(\vec{V}') = \wp(\vec{V})$ *except* that: $p_c{}'(v, n) = \sum_{v\in S} p_c(v, n)/|S|$. Then: $B(\wp(\vec{V}')) = B(\wp(\vec{V}))$ since $p_e(v)$ is constant for all $v \in S$ in each of the defining sums; also $\wp(\vec{V}') \in \mathcal{H}$, $\frac{mx}{mn} \le R_E$ holds if true for $\wp(\vec{V})$ and $p_c{}'(v, n)$ has the same value for all $v \in S$ for $n = 0{,}1$. The corresponding result holds for the other combinations $(A_1, C_1) \times (A_0, C_0)$, concluding the proof.

***Theorem* 1.** Let R_E, R_C, R_D and $\varepsilon, \varepsilon' > 0$ be given and let B_{max} be the maximum of the bias B defined in Equation (3; main text) for all probability functions that satisfy constraint $C_R(R_E, R_C, R_D)$ and the causal relationships of Figure 1C. Then $\exists$ numbers mn, mx, mn_c, mx_c and O_N and $\aleph < \infty$ and variables with associated probability functions $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 4\varepsilon)$ such that $p_e(\{v\colon v \in A_1 \cap A_0\}) \le \varepsilon'$, for $e = 0{,}1$.

Proof: Recall $A_e = \{v\colon p_e(v) = mx\}$. By the first part of the proof of the main Theorem, $\exists$ categorical variables $\vec{V}_J$ and associated probability functions $\wp(\vec{V}_J)$ with $|V_J| = J < \infty$ that satisfy constraint $C_R(R_E, R_C, R_D)$ and $B_{max} \ge B(\wp(\vec{V}_J)) \ge B_{max} - \varepsilon$. Taking $mn = \min_{e,v} p_e(v)$, $mx = \max_{e,v} p_e(v)$, $mn_c = \min_{v,n} p_c(v, n)$, $mx_c = \max_{v,n} p_c(v, n)$ and $O_N = \frac{p_N(N=1)}{p_N(N=0)}$, $\wp(\vec{V}_J)$ satisfies constraint $C_L(mn, mx, mn_c, mx_c, O_N)$ and thus $\wp(\vec{V}_J) \in \mathcal{H}(J, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, \varepsilon)$. By Lemma 1, for some $n' < \infty$, $\exists$ variables and functions $\wp(\vec{V}_{n'}^{*}) \in \mathcal{H}(n', mn', mx', mn_c, mx_c, O_N, R_E, R_C, R_D, 2\varepsilon)$ such that:$|V_{n'}^{*}| = n'$, $p_e(v')$ has bounds $mn' \leq p_e(v') \leq mx$' with $\frac{mx'}{mn'} \leq R_E$. By Lemma 3, we assume $p_c(v,1)$ and $p_c(v,0)$ each has the same value for all v in each combination $(A_1, C_1) \times (A_0, C_0)$,

For each value v of V, consider the partial derivatives $\frac{\partial ln(B)}{\partial p_c(v,1)}$ and $\frac{\partial ln(B)}{\partial p_c(v,0)}$. If both were nonnegative, we would have:

$$\frac{\partial ln(B)}{\partial p_c(v,1)} = O_N[\frac{R_D mx}{\Sigma_1} - \frac{mx}{\Sigma_2} - \frac{R_D mx}{\Sigma_3} + \frac{mx}{\Sigma_4}] \geq 0 \quad \text{so } 1 \geq \frac{\Sigma_1 \Sigma_2}{\Sigma_3 \Sigma_4} \frac{R_D \Sigma_4 - \Sigma_3}{R_D \Sigma_2 - \Sigma_1} \text{; and,}$$

$$\frac{\partial ln(B)}{\partial p_c(v,0)} = \frac{mx}{\Sigma_1} - \frac{mx}{\Sigma_2} - \frac{mx}{\Sigma_3} + \frac{mx}{\Sigma_4} \geq 0 \quad \text{so } 1 \leq \frac{\Sigma_1 \Sigma_2}{\Sigma_3 \Sigma_4} \frac{\Sigma_3 - \Sigma_4}{\Sigma_1 - \Sigma_2}.$$

This would imply: $\frac{\Sigma_1 \Sigma_2}{\Sigma_3 \Sigma_4} \frac{R_D \Sigma_4 - \Sigma_3}{R_D \Sigma_2 - \Sigma_1} \leq \frac{\Sigma_1 \Sigma_2}{\Sigma_3 \Sigma_4} \frac{\Sigma_3 - \Sigma_4}{\Sigma_1 - \Sigma_2}$, a contradiction if $B(\wp(\vec{V})) > 1$ and $R_E > 1$ by Lemma 2. Thus, we must have either case (*i*) or (*ii*):

(*i*) $\frac{\partial ln(B)}{\partial p_c(v,1)} \geq 0$ and $\frac{\partial ln(B)}{\partial p_c(v,0)} < 0$. In this case, we can decrease $p_c(v,0)$ until $p_c(v,0) = mn_c$, without decreasing $B(\wp(\vec{V}))$; or we have:

(*ii*) $\frac{\partial ln(B)}{\partial p_c(v,1)} < 0$. In this case, we can decrease $p_c(v,1)$ until $p_c(v,1) = mn_c$, without decreasing $B(\wp(\vec{V}))$. Thus, we can assume either (*i*) $p_c(v,1) = mn_c$; or (*ii*) $p_c(v,0) = mn_c$.

(*i*) Proposition 1 (below) shows that we can find $\aleph < \infty$ and variables with associated probability functions $\wp(\vec{V}_{\aleph}) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 3\varepsilon)$ such that: $p_e(\{v: v \in A_1 \cap A_0,$ $p_c(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$.

(*ii*) Proposition 2 (below) shows that there exists $\aleph < \infty$ and variables $\vec{V}_{\aleph}$ with associated probability functions $\wp(\vec{V}_{\aleph}) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 4\varepsilon)$ such that: conclusions of proposition 2 continue to hold; and $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,0) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$.

Parts (*i*-*ii*) cover both possible combination of values for $p_c(v,1)$ and $p_c(v,0)$ with $v \in A_1 \cap A_0$, and together show that $p_e(\{v: v \in A_1 \cap A_0\}) \leq \varepsilon$, for $e = 0,1$, completing the proof of Theorem 1.

Proposition 1. Fix $\varepsilon, \varepsilon' > 0$, $R_E, R_C, R_D > 1$, and mn, mx, mn_c, mx_c, O_N with $\frac{mx}{mn} \leq R_E$. If $\exists\ \wp(\vec{V}_{n'}) \in \mathcal{H}(n', mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 2\varepsilon)$, then $\exists$ a number $\aleph < \infty$ and $\wp(\vec{V}_{\aleph}) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 3\varepsilon)$ such that $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$.

Proof. Recall that $A_e = \{v: p_e(v) = mx_c\}$ for $e = 0,1$.

1) By the first part of Theorem 1, $\exists\ \wp(\vec{V}_{\aleph}) \in \mathcal{H}(n', mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 2\varepsilon)$ with $\frac{mx'}{mn'} \leq R_E$, and $B(\wp(\vec{V}_{\aleph})) > B_{max} - 2\varepsilon$.

2) With error at most ε in $B(\wp(\vec{V}_{\aleph}))$, we can approximate $p_e(v)$, so that mn and mx are rational, and in particular $\exists$ integers $b > a$: $a \cdot mx = b \cdot mn$ and even after approximation, $B(\wp(\vec{V}_{\aleph})) > B_{max} - 3\varepsilon$.

3) If $p_e(\{v: v \in A_1 \cap A_0,\ p_{c,J}(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0{,}1$, we are done. If not, then we show how to construct $\wp(\vec{V}_{\aleph})$ with the desired properties without decreasing $B(\wp(\vec{V}_{\aleph}))$. Choose an integer $K > \frac{a+b}{\varepsilon'}$.

4) Label the categories of V_J by v_i for $i = 1,2,\ldots,J = |V_J|$. For each category v_i of V_J, define categories of a variable $V'_{J\cdot K}$ by: $v_{i,k},\ i = 1,\ldots,J;\ \ k = 1,2,\ldots,K$ and associated probability functions: $\wp(\vec{V}'_{J\cdot K}) = \{\vec{V}'_{J\cdot K}; p'_{1,K}(v_{i,k}) = p_1(v_i)/K,\ p'_{0,K}(v_{i,k}) = p_0(v_i)/K,\ p'_{c,K}(v_{i,k},1) = p_c(v_i,1),\ p'_{c,K}(v_{i,k},0) = p_c(v_i,0)\}$. By construction K cancels from the bias expressions, so $B(\wp(\vec{V}'_{J\cdot K}))$ is unchanged, $\frac{mx_K}{mn_K} \leq R_E$ and $\wp(\vec{V}'_{J\cdot K}) \in \mathcal{H}(J \cdot K, mn_K, mx_K, mn_c, mx_c, O_N, R_E, R_C, R_D, 3\varepsilon)$.

5) If $p_{e,K}(\{v: v \in A_1 \cap A_0,\ p_c(v,1) = mn_c\}) \leq \varepsilon$' for $e = 0{,}1$, we are done. If not then we construct other variables and probability functions $\wp(\vec{V}_{\aleph}) \in \mathcal{H} = \mathcal{H}(\aleph, mn_K, mx_K, mn_c, mx_c, O_N, R_E, R_C, R_D, 3\varepsilon)$, with: $\frac{mx_K}{mn_K} \leq R_E$, an equal or larger value of $B(\wp(\vec{V}_{\aleph}))$ and a smaller value of $p_{e,K}(\{v: v \in A_1 \cap A_0,$ $p_c(v,1) = mn_c\})$.

6) Let $AA_K = \{v: v \in A_1 \cap A_0,\ p'_{c,N}(v,1) = mn_K,\}$. Then $|AA_K|$ must be greater than $a + b$. Otherwise: $p'_{e,K}(AA_K) \leq |AA_K| \cdot mx_K \leq (a+b) \cdot mx_K = (a+b) \cdot \frac{mx}{K} < (a+b)\frac{mx}{(a+b)}\varepsilon' \leq \varepsilon$', by choice of N and since $mx \leq 1$.

7) Thus, choose $a + b$ points in AA_K. To simplify notation, we drop K and primes ($'$) from variables and functions in $\wp(\vec{V}_{J\cdot K})$ and the limits: e.g., $p'_{e,K}(v) = p_e(v)$ for $e = 0{,}1$ and $mx = mx_K$.

We now show that we could decrease $|AA_K| = |\{v: v \in A_1 \cap A_0,\ p_c(v,1) = mn_c\}|$ without decreasing $B(\wp(\vec{V}_{J\cdot K}))$ by assigning new values of $p_e(v)$ and $p_c(v,n)$ for sets of $a + b$ points. The construction will assure, that with these new variables and functions $\wp(\vec{V}')$, we have: $B(\wp(\vec{V}')) > B_{max} - 3\varepsilon$, $p_e(|AA_K'|) \leq \varepsilon'$ and $\wp(\vec{V}') \in \mathcal{H}(n, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 3\varepsilon)$.

For each set of $n = a + b$ points in $A_1 \cap A_0$ with $p_c(v,1) = mn_c$, the current ("old") values and proposed, new (to be assigned) values are shown in eTable A1. In general, $p_c(v,0)$ can vary, but we can replace $p_c(v,0)$ with the average $md = \sum_{v \in A_1 \cap A_0} p_c(v,0)/(a+b)$ by Lemma eTable A1: Old and reassigned ("new") values of $p_0(v)$, $p_1(v)$, $p_c(v,1)$ and $p_c(v,0)$ for each of $n = a + b$ points in $A_1 \cap A_0$ with $p_c(v,1) = mn_c$

eTable A1

Type, number of points n	$p_1(v)$	$p_0(v)$	$p_c(v,1)$	$p_c(v,0)$
Old	Old values of probability functions for the n selected points			
$n = a + b$	mx	mx	mn_c	$md^{\dagger}$
Total: $n = a + b$	$(a+b)mx$	$(a+b)mx$		
New	New values of probability functions for the n selected points			
Type 1, $n = b$	mx	mn	w	x
Type 2, $n = b$	mn	mx	y	z
Total, $n = 2b$	$(a+b)mx$	$(a+b)mx$		

† $p_c(v,0)$ is replaced with the average (see text above).

Recall that $B((\wp\vec{V}_{\aleph})) = \frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$. The contributions of the n points to each of the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ before (old) and after using the new values are shown in eAppendix Table A2.

eTable A2: Old and "New" contribution of the $n = a + b$ points to $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$

Sum	Old Value	New Value: Type 1 + Type2
Σ_1	$(a+b)\cdot mx[R_D O_N mn_c + md]$	$b\cdot mx\cdot(R_D O_N w + x) \quad + b\cdot mn\cdot(R_D O_N \mathrm{y} + z)$
Σ_2	$(a+b)\cdot mx[O_N mn_c + md]$	$b\cdot mx\cdot(\quad O_N w + x) \quad + b\cdot mn\cdot(\quad O_N \mathrm{y} + z)$
Σ_3	$(a+b)\cdot mx[R_D O_N mn_c + md]$	$b\cdot mn\cdot(R_D O_N w + x) \quad + b\cdot mx\cdot(R_D O_N \mathrm{y} + z)$
Σ_4	$(a+b)\cdot mx[O_N mn_c + md]$	$b\cdot mn\cdot(\quad O_N w + x) \quad + b\cdot mx\cdot(\quad O_N \mathrm{y} + z)$

By subtracting the old contribution from the new contribution, we get the change in each of the sums – the difference that will occur by reassigning $p_c(v,1)$ and $p_c(v,0)$ as proposed: eAppendix Table A3.

eAppendix Table A3. The changes in the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ due to changing $p_c(v,1)$ and $p_c(v,0)$.

Δ_i	Difference: $\Delta_i =$ New$_i$ – Old$_i$
Δ_1	$b\cdot mx(R_D O_N w + x) + b\cdot mn(R_D O_N \mathrm{y} + z) - (a+b)R_D O_N mx\cdot mn_c - (a+b)mx\cdot md$ $= R_E\cdot R_E\cdot R_D O_N(w - mn_c) + R_E\cdot R_E(x - md) + R_E\cdot R_D O_N(\mathrm{y} - mn_c) + R_E\cdot(z - md)$
Δ_2	$b\cdot mx(O_N k + x) \quad + b\cdot mn(O_N \mathrm{y} + z) - (a+b)\cdot mx\cdot[O_N mn_c + md]$ $= R_E\cdot RO_N(k - mn_c) + R_E\cdot R(x - md) + R_E\cdot O_N\cdot(\mathrm{y} - mn_c) + R_E(z - md)$
Δ_3	$b\cdot R_D mn\cdot O_N w + b\cdot mn\cdot x + b\cdot mx\cdot(R_D O_N \mathrm{y} + z) - (a+b)\cdot mx[R_D O_N mn_c + md]$ $= R_E\cdot R_D O_N(w - mn_c) + R_E(x - md) + R_E\cdot R_E\cdot R_D O_N(\mathrm{y} - mn_c) + R_E\cdot R_E(z - md)$
Δ_4	$= b\cdot mn\cdot(O_N w + x) \quad + b\cdot mx\cdot(O_N \mathrm{y} + z) - (a+b)\cdot mx\cdot[O_N mn_c + md]$ $= R_E O_N(w - mn_c) + R_E(x - md) + R_E\cdot R_E O_N(\mathrm{y} - mn_c) + R_E\cdot R_E(z - md)$

The second line in each cell of the right column of Table A3 follows by rearranging and then using $b = a\cdot mn, mx = R_E\cdot mn$ and cancelling a and mn. (Cancelation simplifies the expression but does not affect the value of $B = B(\wp(\vec{V}'_{J\cdot K}))$). Setting $w = mn_c, x = md, y = mn_c$ and $z = md$, doesn't change

B (all Δ_i are 0) and the constraints are still satisfied. [Many other possible substitutions would increase B.]

This construction decreases $|AA_K|$ by $a+b$: after redefining the distribution functions $p_e(v)$ in $\wp(\vec{V}_{\aleph})$ as in Table A1, $p_e(v) \in \{mn, mx\}$ so that none of the $a+b$ selected points remain in $A_1 \cap A_0$. We can repeat the process of reassigning probabilities, starting at step 6 and as reflected in eAppendix Tables A1-A3, until $|\{v: v \in A_1 \cap A_0,\ p_c(v,1) = mn_c\}| < a+b$ so that $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0{,}1$, completing proof of Proposition 1.

***Proposition* 2**: Fix $\varepsilon > 0$, $R_E, R_C, R_D > 1$, and mn, mx, mn_c, mx_c with $\frac{mx}{mn} \leq R_E$. If $\exists\ \wp(\vec{V}_{n'}) \in \mathcal{H}(n', mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 3\varepsilon)$ satisfying conclusions of Proposition 1, then $\exists$ a number $\aleph < \infty$ and $\wp(\vec{V}_{\aleph}) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 4\varepsilon)$ such that conclusions of Proposition *1* hold; and $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,0) = mn_c\}) \leq \varepsilon'$ for $e = 0{,}1$.

<u>Proof</u>. Proposition 2 follows by showing it would be possible to redefine the functions in $\wp(V_{\aleph})$ so that $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,0) = mn_c\}) \leq \varepsilon'$, without increasing $B(\wp(\vec{V}_{\aleph}))$ so that $\wp(\vec{V}_{\aleph}) \in \mathcal{H}' = \mathcal{H} \cap \{\wp(V_{\aleph}): p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,0) = mn_c\}) \leq \varepsilon' \text{ for } e = 0{,}1\}$.

Let $\wp(\vec{V}_{\aleph}) \in \mathcal{H}(n', mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 3\varepsilon)$ denote the variables and functions whose existence is proven in Proposition 1. Suppose, for the functions in $\wp(\vec{V}_{\aleph})$, that $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0$ or 1. Paralleling the proof of Proposition 1, if $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v,0) = mn_c\}) > \varepsilon'$, modify $\wp(\vec{V}_{\aleph})$ to increase the number of points if needed and approximate $p_e(v)$ by rational numbers so that $B(\wp(\vec{V}_{\aleph})) > B_{max} - 4\varepsilon$ and $\aleph > \frac{a+b}{\varepsilon}$, where $\aleph = |V|$ and $b > a$ are integers with $a \cdot mx = b \cdot mn$. Select $a+b$ points in $A_1 \cap A_0$ such that $p_c(v,0) = mn_c$. Replace $p_c(v,1)$ by the average md, without affecting $B(\wp(\vec{V}'))$ or the constraints ($md = \sum_{v \in A_1 \cap A_0} p_c(v,1)/|A_1 \cap A_0|$).

We now sketch how to modify $\wp(\vec{V}_{\aleph})$, analogous to Proposition 1, with the values for $p_0(v)$ and $p_c(v,n)$ in Table A4, so that $B(\wp(\vec{V}_{\aleph}))$ won't decrease, but $p_{e,K}(\{v: v \in A_1 \cap A_0,\ p_c(v,0) = mn_c\}) \leq \varepsilon$'.

eAppendix Table A4: Old and reassigned ("new") values of $p_0(v)$, $p_1(v)$, $p_c(v,1)$ and $p_c(v,0)$ for each set of $n = a+b$ points in $A_1 \cap A_0$ with $p_c(v,0) = mn_c\}$

Type, number of points n	$p_1(v)$	$p_0(v)$	$p_c(v,1)$	$p_c(v,0)$
Old	Old values of probability functions for the selected points			
$n = a+b$	mx	mx	$md^{\dagger}$	mn_c
Total: $n = a+b$	$(a+b)mx$	$(a+b)mx$		
New	New values of probability functions for the selected points			
Type 1, $n = b$	mx	mn	w	x

Type 2, $n = b$	mn	mx	y	z
Total, $n = 2b$	$(a + b)mx$	$(a + b)mx$		

† $p_c(v, 1)$ can vary, but is replaced by the average by Lemma 3.

The contributions of these points to each of the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ before (old) and after (using new values) are shown in eTable A5.

eTable A5: Old and "New" contribution of the $n = a + b$ points to $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$

Sum	Old Value	New Value: Type 1 + Type2
Σ_1	$(a + b) \cdot mx[R_D O_N md + mn_c]$	$b \cdot mx \cdot (R_D O_N w + x) \quad + b \cdot mn \cdot (R_D O_N \mathrm{y} + z)$
Σ_2	$(a + b) \cdot mx[O_N md + mn_c]$	$b \cdot mx \cdot (\quad O_N w + x) \quad +b \cdot mn \cdot (\quad O_N \mathrm{y} + z)$
Σ_3	$(a + b) \cdot mx[R_D O_N md + mn_c]$	$b \cdot mn \cdot (R_D O_N w + x) \quad + b \cdot mx \cdot (R_D O_N \mathrm{y} + z)$
Σ_4	$(a + b) \cdot mx[O_N md + mn_c]$	$b \cdot mn \cdot (\quad O_N w + x) \quad +b \cdot mx \cdot (\quad O_N \mathrm{y} + z)$

By subtracting the old contribution from the new contribution, we get the change (Δ_i) in each of the sums (difference that will occur by reassigning $p_c(v, n)$ as proposed), as summarized in eTable A6.

eTable A6. Changes in the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ due to changing $p_c(v, 1)$ and $p_c(v, 0)$.

Δ_i	Difference: $\Delta_i =$ New$_i$ – Old$_i$
Δ_1	$b \cdot mx(R_D O_N w + x) + b \cdot mn(R_D O_N \mathrm{y} + z) - (a + b)R_D O_N mx \cdot md - (a + b)mx \cdot mn_c$ $= R_E \cdot R_E \cdot R_D O_N(w - md) + R_E \cdot R_E(x - mn_c) + R_E \cdot R_D O_N(\mathrm{y} - md) + R_E \cdot (z - mn_c)$
Δ_2	$b \cdot mx(O_N k + x) \quad +b \cdot mn(O_N \mathrm{y} + z) - (a + b) \cdot mx \cdot [O_N md + mn_c]$ $= R_E \cdot R_E O_N(k - md) + R_E \cdot R_E(x - mn_c) + R_E \cdot O_N \cdot (\mathrm{y} - md) + R_E(z - mn_c)$
Δ_3	$b \cdot R_D mn \cdot O_N w + b \cdot mn \cdot x + b \cdot mx \cdot (R_D O_N \mathrm{y} + z) - (a + b) \cdot mx[R_D O_N md + mn_c]$ $= R_E \cdot R_D O_N(w - md) + R_E(x - mn_c) + R_E \cdot R_E \cdot R_D O_N(\mathrm{y} - md) + R_E \cdot R_E(z - mn_c)$
Δ_4	$= b \cdot mn \cdot (O_N w + x) \quad +b \cdot mx \cdot (O_N \mathrm{y} + z) - (a + b) \cdot mx \cdot [O_N md + mn_c]$ $= R_E O_N(w - md) + R_E(x - mn_c) + R_E \cdot R_E O_N(\mathrm{y} - md) + R_E \cdot R_E\big(z - mn_c\big)$

If we set $w = md$; $x = mn_c$; $\mathrm{y} = md$ and $z = mn_c$, then $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ do not change ($\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$), the constraints are still satisfied, $B(\wp(\vec{V}_{\aleph}))$ does not decrease, conclusions of Proposition 1, continue to hold except $B(\wp(\vec{V}_{\aleph})) \geq B_{max} - 4\varepsilon$. However, with the changes $p_e(v)$ no longer equals mx for the $a + b$ selected points so that $|\{v \in A_1 \cap A_0,\}|$ is reduced by $a + b$. We can repeat the process of reassigning probability functions (Tables A4-A6) until $|A_1 \cap A_0 \cap \{v: mn_c \leq p_c(v, 0) = mx_c\}| < a + b$ implying that $p_e(\{v: v \in A_1 \cap A_0,\ p_c(v, 0) = mn_c \}) \leq \varepsilon'$ for $e = 0,1$, concluding the proof of Proposition 2.

***Proposition* 3**: Fix $\varepsilon, \varepsilon' > 0$, $R_E, R_C, R_D > 1$, and mn, mx, mn_c, mx_c with $\frac{mx}{mn} \leq R_E$. If $\exists\ \wp\big(\vec{V}_{n'}\big) \in \mathcal{H} = \mathcal{H}(n', mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 4\varepsilon)$ satisfying conclusions of Theorem 1, then $\exists$ a number

$\aleph < \infty$ and $\wp(\vec{V}_\aleph) \in \mathcal{H}' = \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, R_E, R_C, R_D, 5\varepsilon)$ such that conclusions of Theorem 1 hold; and $p_e(V \in D_e) \le \varepsilon'$ for $e = 0{,}1$.

Proof. Recall $D_e = \{v: mn < p_e(v) < mx\}$ for $e = 0{,}1$. First, consider $e = 1$. Let $\wp(\vec{V}_{n'}) \in \mathcal{H}$ be the variables and functions whose existence is proven in Theorem 1. By the construction in Propositions 1 and 2, $\exists\ \wp(\vec{V}_\aleph) \in \mathcal{H}$ such that $p_e(\{v: v \in A_1 \cap A_0\}) \le \varepsilon'; e = 0{,}1$. If $p_1(V_\aleph \in D_1) > mx$, D_1 must include at least two points, say v_1 and v_2 labeled so $p_1(v_1) > p_1(v_2)$. Let $B(\wp(\vec{V}_\aleph))(\alpha, \Delta)$ represent the bias, after changing $p_1(v_1)$ by a small amount $\Delta > 0$ and $p_1(v_2)$ by $-\alpha\Delta$: $p_1(v_i) \to p_1(v_i) + \Delta(-\alpha)^{i-1}$, $0 < \alpha, p_1(v_i)$. We show we can increase $p_1(v_1)$ by Δ and decrease $p_1(v_2)$ $\alpha\Delta$ without changing $B\left(\wp(\vec{V}_\aleph)\right)$; if so, the proportional change, $PC = \frac{B(\wp(\vec{V}_\aleph))(\alpha,\Delta) - B(\wp(\vec{V}_\aleph))(0,0)}{B(\wp(\vec{V}_\aleph))(0,0)}$ must be 0. Direct evaluation of PC leads to:

$0 = \frac{\Sigma_1 + \Delta(R_D O_N p_c(v_1,1) + p_c(v_1,0)) - \alpha\Delta(R_D O_N p_c(v_2,1) + p_c(v_2,0))/\Sigma_1}{\Sigma_2 + \Delta(O_N p_c(v_1,1) + p_c(v_1,0)) - \alpha\Delta(O_N p_c(v_2,1) + p_c(v_2,0))/\Sigma_2} - 1$ which has the solution

$\alpha = \frac{\Sigma_2(R_D O_N p_c(v_1,1) + p_c(v_1,0)) - \Sigma_1(O_N p_c(v_1,1) + p_c(v_1,0))}{\Sigma_2(R_D O_N p_c(v_2,1) + p_c(v_2,0)) - \Sigma_1(O_N p_c(v_2,1) + p_c(v_2,0))}$. Set $\Delta = \min(mx - p_1(v_i),\ \frac{1}{\alpha}(p_1(v_2) - mn))$. If the denominator is 0, set up the problem to change $p_1(v_i) \to p_1(v_i) - \Delta(-\beta)^{2-i}$, $0 < \beta, \Delta$. Reasoning as before identifies the solution:

$\beta = \frac{\Sigma_2(R_D O_N p_c(v_2,1) + p_c(v_2,0)) - \Sigma_1(O_N p_c(v_2,1) + p_c(v_2,0))}{\Sigma_2(R_D O_N p_c(v_1,1) + p_c(v_1,0)) - \Sigma_1(O_N p_c(v_1,1) + p_c(v_1,0))}$. $\Delta = \min\left(\frac{1}{\beta}(mx - p_1(v_i),\ (p_1(v_2) - mn)\right)$

If this denominator is also 0, then choose $0 < \alpha, \Delta$ so that $p_1(v_1) + \Delta = mx$ and $p_1(v_2) - \Delta\alpha = mn$. After the changes, v_1 or $v_2 \notin D_1$ $B(\wp(\vec{V}_\aleph)) \ge B_{mx} - 3\varepsilon$ as the offsetting changes don't affect $B(\wp(\vec{V}_\aleph))$, constraints or the conclusions of Theorem 1 or Proposition 1 or 2. Repeat the process until $|D_1| \le 1$. The analogous construction applies for $e = 0$, proving Proposition 3.

***Theorem* 2**. Fix $\varepsilon, \varepsilon' > 0$, $R_E, R_C, R_D > 1$, and mn, mx, mn_c, mx_c with $\frac{mx}{mn} \le R_E$. If $\exists\ \wp(\vec{V}_{n'}) \in \mathcal{H}(n', mn, mx, mn_c, mx_c, R_E, R_C, R_D, 5\varepsilon)$ satisfying conclusion of Proposition 3, then $\exists$ a number $\aleph < \infty$ and $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 7\varepsilon)$ such that conclusions of Proposition 3 hold (including those of Theorem 1); and, $p_e(\{v: v \in C_1 \cap C_0\}) \le \varepsilon'$, for $e = 0{,}1$.

Proof: Recall $C_e = \{v: p_e(v) = mn\}$ for $e = 0{,}1$.The proof parallels that of Theorem 1. Let $\wp(\vec{V}_J)$ be the categorical variables and functions satisfying conclusions of Proposition 3, under assumptions stated in the main Theorem. Denote the probability functions in $\wp(\vec{V}_J)$ by: $\{p_{1,J}(v), p_{0,J}(v),\ p_{c,J}(v,1), p_{c,J}(v,0)\}$ and the associated limits by mn_J, mx_J. To simplify notation, we drop the subscript J for V_J, mn_J, mx_J and the probability functions.

For each value v of V, consider $\frac{\partial ln(B)}{\partial p_c(v,1)}$ and $\frac{\partial ln(B)}{\partial p_c(v,0)}$. If both were non-negative, then we would have:

$$\frac{\partial ln(B)}{\partial p_c(v,1)} = O_N[\frac{R_D mn}{\Sigma_1} - \frac{mn}{\Sigma_2} - \frac{R_D mn}{\Sigma_3} + \frac{mn}{\Sigma_4}] \geq 0 \quad \text{so} \quad 1 \geq \frac{\Sigma_1\Sigma_2}{\Sigma_3\Sigma_4}\frac{R_D\Sigma_4 - \Sigma_3}{R_D\Sigma_2 - \Sigma_1} \text{; and,}$$

$$\frac{\partial ln(B)}{\partial p_c(v,0)} = \frac{mn}{\Sigma_1} - \frac{mn}{\Sigma_2} - \frac{mn}{\Sigma_3} + \frac{mn}{\Sigma_4} \geq 0 \quad \text{so} \quad 1 \leq \frac{\Sigma_1\Sigma_2}{\Sigma_3\Sigma_4}\frac{\Sigma_3 - \Sigma_4}{\Sigma_1 - \Sigma_2}.$$

This would imply: $\frac{\Sigma_1\Sigma_2}{\Sigma_3\Sigma_4}\frac{R_D\Sigma_4 - \Sigma_3}{R_D\Sigma_2 - \Sigma_1} \leq \frac{\Sigma_1\Sigma_2}{\Sigma_3\Sigma_4}\frac{\Sigma_3 - \Sigma_4}{\Sigma_1 - \Sigma_2}$, a contradiction if $B(\wp(\vec{V})) > 1$ by Lemma 2. Thus, we must have either case (*i*) or (*ii*):

(*i*) $\frac{\partial ln(B)}{\partial p_c(v,1)} \geq 0$ and $\frac{\partial ln(B)}{\partial p_c(v,0)} < 0$. In this case, we can decrease $p_c(v,0)$ until $p_c(v,0) = mn_c$, without decreasing $B(\wp(\vec{V}))$; or we have:

(*ii*) $\frac{\partial ln(B)}{\partial p_c(v,1)} < 0$. In this case, we can decrease $p_c(v,1)$ until $p_c(v,1) = mn_c$, without decreasing $B(\wp(\vec{V}))$. Thus, we can assume either (*i*) $p_c(v,1) = mn_c$; or (*ii*) $p_c(v,0) = mn_c$.

i). We prove in Proposition 4 (below) that, if $\mathcal{H}(n', mn, mx, mn_c, mx_c, R_E, R_C, R_D, 5\varepsilon) \neq \phi$, then $\exists$ a number $\aleph < \infty$ and $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, R_E, R_C, R_D, 6\varepsilon)$ such that: conclusions of Proposition 3 continue to hold; and $p_e(\{v: v \in C_1 \cap C_0,\ p_c(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$;.

ii). We prove in Proposition 5 (below) that if $\mathcal{H}(n', mn, mx, mn_c, mx_c, R_E, R_C, R_D, 6\varepsilon) \neq \phi$, then $\exists$ a number $\aleph < \infty$ and $\wp(\overrightarrow{V_\aleph}) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, R_E, R_C, R_D, 7\varepsilon)$ such that: conclusions of Proposition 4 continue to hold; (*ii*) $p_e(\{v: v \in C_1 \cap C_0,\ p_c(v,0) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$.

Cases (*i*-*ii*) cover the possible combination of values for $p_c(v,1)$ and $p_c(v,0)$ with $v \in C_1 \cap C_0$, and together show that $p_e(\{v: v \in A_1 \cap A_0) \leq \varepsilon'$, for $e = 0,1$, completing the proof of Theorem 2.

***Proposition* 4**. Fix $\varepsilon, \varepsilon' > 0$, $R_E, R_C, R_D > 1$, and mn, mx, mn_c, mx_c with $\frac{mx}{mn} \leq R_E$. If $\wp(\vec{V}_{n'}) \in \mathcal{H}(n', mn, mx, mn_c, mx_c, R_E, R_C, R_D, 5\varepsilon)$ satisfying conclusion of Proposition 3, then $\exists$ a number $\aleph < \infty$ and $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, R_E, R_C, R_D, 6\varepsilon)$ such that: conclusions of Proposition 3 continue to hold; and $P(\{v: v \in C_1 \cap C_0,\ p_c(v,1) = mn_c\}) \leq \varepsilon'$.

<u>Proof</u>. Recall $C_i = \{v: p_i(v) = mn\}$ for $i = 0,1$.

1) By assumptions of the main Theorem including mild regularity conditions, $\exists$ a categorical variable V_J and probability functions $\wp(\vec{V}_J) \in \mathcal{H}(n', mn, mx, mn_c, mx_c, R_E, R_C, R_D, 5\varepsilon)$.

2) With error at most ε in $B(\wp(\vec{V}_\aleph))$, we can approximate $p_c(v)$ so that mn and mx are rational, and in particular $\exists$ integers $b > a$: $a \cdot mx = b \cdot mn$ and $B(\wp(\vec{V}_\aleph)) > B_{max} - 6\varepsilon$. ..

3) If $p_e(\{v: v \in C_1 \cap C_0,\ p_{c,J}(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$, we are done. If not, we show how to construct $\wp(\vec{V}_{\aleph})$ with the desired properties without decreasing $B(\wp(\vec{V}_{\aleph}))$. Choose an integer $K > \frac{a+b}{\varepsilon'}$.

4) For each value v_i of $V_J, i = 1,2,\ldots,J = |V_J|$, define K new points $v_{i,k},\ i = 1,\ldots,J;\ \ k = 1,2,\ldots,K$ and associated probability functions: $\wp'(V_{J\cdot K}) = \{p'_{1,K}(v_{i,k}) = p_1(v_i)/K,\ p'_{0,K}(v_{i,k}) = p_0(v_i)/K,\ p'_{c,K}(v_i,1) = p_c(v_i,1)/K,\ p'_{c,K}(v_{i,k},0) = p_c(v_i,0)/K\}$. By construction K cancels from the bias expression, so $B(\wp(\vec{V}'_{J\cdot K}))$ is unchanged, $\wp(\vec{V}'_{J\cdot K}) \in \mathcal{H}(J \cdot K, mn_K, mx_K, mn_c, mx_c, O_N, R_E, R_C, R_D, 6\varepsilon)$ and $\frac{mx_K}{mn_K} \leq R_E$.

5) If $p_e(\{v: v \in C_1 \cap C_0,\ p_{c,K}(v,1) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$, *stop*. If not we show how to construct other variables and probability functions $\wp(\vec{V}_{\aleph}) \in \mathcal{H}(\aleph, mn_K, mx_K, mn_c, mx_c, O_N, R_E, R_C, R_D, 6\varepsilon)$, with $\frac{mx_K}{mn_K} = R_E$ but with $B(\wp(\vec{V}_{\aleph}))$ equal or larger, and $p_{e,K}(\{v: v \in A_1 \cap A_0,\ p_c(v,1) = mn_c\}) \leq \varepsilon'$.

6) $|CC_K| \geq a + b$, where $CC_K = \{v: v \in C_1 \cap C_0,\ p_{c,K}(v,1) = mn_c\}$ since: $p_{e,K}(CC) \leq mx \cdot mx_K \leq (a+b) \cdot mx_K = (a+b) \leq (a+b)\frac{mx}{(a+b)}\varepsilon' \leq \varepsilon'$ and $mx \leq 1$.

7) Choose $a + b$ points in CC_K. To simplify notation, we drop K from the subscripts of functions in $\wp(\vec{V}_{J\cdot K})$ and from the limits: e.g., $mx = mx_K$ and $p_{e,K}(v) = p_e(v)$ for $e = 0,1$.

We now show that we could decrease $|CC_K|$ without decreasing $B(\wp(\vec{V}_{J\cdot K}))$ by assigning new values of $p_e(v)$ and $p_c(v,n)$ for sets of $a + b$ points. The construction will assure, that with these new variables and functions $\wp(\vec{V}')$ we have: $B(\wp(\vec{V}')) > B_{max} - 6\varepsilon$; $p_e(|CC_K|) \leq \varepsilon$';
$\wp(V') \in \mathcal{H}(n, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 6\varepsilon)$. The old values and the new assignment are shown in eTable A7 for the $n = a + b$ points in $C_1 \cap C_0$ with $p_c(v,1) = mn_c,\ p_c(v,0) = md$.

eTable A7: Old and reassigned ("new") values of $p_0(v), p_1(v), p_c(v,1)$ and $p_c(v,0)$ for each set of $n = a + b$ points in $C_1 \cap C_0$ with $p_c(v,1) = mn_c$

Type, number of points n	$p_1(v)$	$p_0(v)$	$p_c(v,1)$	$p_c(v,0)$
Old	Old values of probability functions for the selected points			
$n = a + b$	mn	mn	mn_c	$md^{\dagger}$
Total: $n = a + b$	$(a+b)mn$	$(a+b)mn$		
New	New values of probability functions for the selected points			
Type 1, $n = a$	mx	mn	w	x
Type 2, $n = a$	mn	mx	y	z
Total, $n = 2a$	$(a+b)mn$	$(a+b)mn$		

†In general, $p_c(v,0)$ will vary for $v \in C_1 \cap C_0$, but can be replaced by the average by Lemma 3.

The contributions of these n points to each of the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ before (old) and after (using new values) are shown in eTable A8.

eTable A8: Old and “New” contribution of the $n = a + b$ points to $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$

Sum	Old Value	New Value: Type 1 + Type2
Σ_1	$(a+b) \cdot mn[R_D O_N mn_c + md]$	$a \cdot mx \cdot (R_D O_N w + x) \quad + a \cdot mn \cdot (R_D O_N \mathrm{y} + z)$
Σ_2	$(a+b) \cdot mn[O_N mn_c + md]$	$a \cdot mx \cdot (\quad O_N w + x) \quad +a \cdot mn \cdot (\quad O_N \mathrm{y} + z)$
Σ_3	$(a+b) \cdot mn[R_D O_N mn_c + md]$	$a \cdot mn \cdot (R_D O_N w + x) \quad + a \cdot mx \cdot (R_D O_N \mathrm{y} + z)$
Σ_4	$(a+b) \cdot mn[O_N mn_c + md]$	$a \cdot mn \cdot (\quad O_N w + x) \quad + a \cdot mx \cdot (\quad O_N \mathrm{y} + z)$

By subtracting the old from the new contribution, we get the change in each of the sums – the difference that will occur by reassigning $p_c(v, 1)$ and $p_c(v, 0)$ as proposed: eTable A9.

eTable A9. Changes in the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ due to changing $p_c(v, 1)$ and $p_c(v, 0)$.

Δ_i	Difference: $\Delta_i =$ New$_i$ – Old$_i$
Δ_1	$a \cdot mx(R_D O_N w + x) + a \cdot mn(R_D O_N \mathrm{y} + z) - (a+b) R_D O_N mn \cdot mn_c - (a+b) mn \cdot md$ $= R_E \cdot R_D O_N (w - mn_c) + R_E(x - md) + R_D O_N(\mathrm{y} - mn_c) + (z - md)$
Δ_2	$a \cdot mx(O_N k + x) \quad +a \cdot mn(\ O_N \mathrm{y} + z) - (a+b) \cdot mn \cdot [O_N mn_c + md]$ $= R_E O_N (k - mn_c) + R_E(x - md) + O_N \cdot (\mathrm{y} - mn_c) + (z - md)$
Δ_3	$a \cdot R_D mn \cdot O_N w + a \cdot mn \cdot x + b \cdot mx \cdot (R_D O_N \mathrm{y} + z) - (a+b) \cdot mn[R_D O_N mn_c + md]$ $= R_D O_N (w - mn_c) + (x - md) + \ R_E \cdot R_D O_N (\mathrm{y} - mn_c) + R_E(z - md)$
Δ_4	$= a \cdot mn \cdot (\ O_N w + x) \quad +a \cdot mx \cdot (\ O_N \mathrm{y} + z) - (a+b) \cdot mn \cdot [O_N mn_c + md]$ $= O_N(w - mn_c) + (x - md) + R_E O_N (\mathrm{y} - mn_c) + R_E(z - md)$

If we set $w = mn_c, x = md, y = mn_c$ and $z = md$, then $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ don’t change, the constraints are still satisfied, and conclusions of Propositions 1-3 still hold. However, with the changes $p_e(v) \neq mx$ so that $|\{v \in C_1 \cap C_0,\}|$ is reduced by $a + b$. Many other possible substitutions would increase B.

We can continue the construction until $|C_1 \cap C_0 \cap \ \{v: p_c(v, 1) = mn_c\}| < a + b$ implying that $p_k(\{v: v \in C_1 \cap C_0, \ p_c(v, 1) = mn_c\}) \leq \varepsilon'$ for $k = 0,1.$, concluding the proof of Proposition 4.

Proposition **5**: Fix $\varepsilon, \varepsilon' > 0$, $R_E, R_C, R_D > 1$, and mn, mx, mn_c, mx_c with $\frac{mx}{mn} \leq R_E$. If $\mathcal{H}(n', mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 6\varepsilon) \neq \phi$ then $\exists$ a number $\aleph < \infty$ and $\wp(\overrightarrow{V_\aleph}) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 7\varepsilon)$ such that conclusions of Propositions 1-4 continue to hold; and, $p_e(\{v: v \in C_1 \cap C_0, \ p_c(v, 0) = mn_c\}) \leq \varepsilon'$ for $e = 0,1$.

Proof. Proposition 5 follows by showing it would be possible to redefine the functions in $\wp(V_\aleph)$ so that $p_e(\{v: v \in C_1 \cap C_0, \ p_c(v, 1) = mn_c\}) \leq \varepsilon'$ without increasing $B\big(\wp(V_\aleph)\big)$ and so that $B\big(\wp(V_\aleph)\big) > B_{max} - 7\varepsilon$ for the new functions.

Let $\wp(\vec{V}_\aleph)$ be the variables and functions whose existence is proven in Proposition 4. As in the proofs of Propositions 1, 2 and 4, increase the number of points if needed so that $\aleph > \frac{a+b}{\varepsilon'}$, select $a+b$ points in $C_1 \cap C_0$ such that $p_c(v,0) = mn$. If $p_e(\{v: v \in C_1 \cap C_0,\ p_c(v,0) = mn_c\}) > \varepsilon'$, as in the proof of Propositions 1 and 4, we can modify $\wp(\vec{V}_\aleph)$ as indicated in eTable A10.

eTable A10: Old and reassigned ("new") values of $p_0(v)$, $p_1(v)$, $p_c(v,1)$ and $p_c(v,0)$ for each set of $n = a+b$ points in $C_1 \cap C_0$ with $p_c(v,0) = mn_c$

Type, number of points n	$p_1(v)$	$p_0(v)$	$p_c(v,1)$	$p_c(v,0)$
Old	Old values of probability functions for the selected points			
$n = a+b$	mn	mn	$\overline{md}^{\dagger} \to md$	mn_c
Total: $n = a+b$	$(a+b)mn$	$(a+b)mn$		
New	New values of probability functions for the selected points			
Type 1, $n = a$	mx	mn	w	x
Type 2, $n = a$	mn	mx	y	z
Total, $n = 2a$	$(a+b)mn$	$(a+b)mn$		

† In general, $p_c(v,0)$ will vary for $v \in C_1 \cap C_0$, but can be replaced with the average by Lemma 3.

The contributions of these points to each of the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ before (old) and after (using new values) are shown in eTable 11.

eTable A11: Old and "New" contribution of the $n = a+b$ points to $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$

Sum	Old Value	New Value: Type 1 + Type2
Σ_1	$(a+b) \cdot mn[R_D O_N md + mn_c]$	$a \cdot mx \cdot (R_D O_N w + x) \quad + a \cdot mn \cdot (R_D O_N \mathrm{y} + z)$
Σ_2	$(a+b) \cdot mn[O_N md + mn_c]$	$a \cdot mx \cdot (\ \ O_N w + x) \quad + a \cdot mn \cdot (\ \ O_N \mathrm{y} + z)$
Σ_3	$(a+b) \cdot mn[R_D O_N md + mn_c]$	$a \cdot mn \cdot (R_D O_N w + x) \quad + a \cdot mx \cdot (R_D O_N \mathrm{y} + z)$
Σ_4	$(a+b) \cdot mn[O_N md + mn_c]$	$a \cdot mn \cdot (\ \ O_N w + x) \quad + a \cdot mx \cdot (\ \ O_N \mathrm{y} + z)$

By subtracting old contributions from the new ones, we get the change in each of the sums – the difference that will occur by reassigning $p_c(v,1)$ and $p_c(v,0)$ as proposed: eTable A12, where Δ_i denotes the change in Σ_i for *i*=1,2,3,4 due to modifications of the functions in $\wp(\vec{V}_\aleph)$.

eTable A12. The change in each of the sums $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ due to changing $p_c(v,n)$.

Δ_i	Difference: $\Delta_i =$ $\mathrm{New}_i - \mathrm{Old}_i$
Δ_1	$a \cdot mx(R_D O_N w + x) + a \cdot mn(R_D O_N \mathrm{y} + z) - (a+b) R_D O_N mn \cdot md - (a+b) mn \cdot mn_c$ $= R_E \cdot R_D O_N (w - md) + R_E (x - mn_c) + R_D O_N (\mathrm{y} - md) + (z - mn_c)$
Δ_2	$a \cdot mx(\ O_N k + x) \quad + a \cdot mn(\ O_N \mathrm{y} + z) - (a+b) \cdot mn \cdot [O_N md + mn_c]$ $= R_E O_N (w - md) + R_E (x - mn_c) + O_N \cdot (\mathrm{y} - md) + (z - mn_c)$
Δ_3	$a \cdot R_D mn \cdot O_N w + a \cdot mn \cdot x + b \cdot mx \cdot (R_D O_N \mathrm{y} + z) - (a+b) \cdot mn[R_D O_N md + mn_c]$ $= R_D O_N (w - md) + (\ x - mn_c) +\ R_E \cdot R_D O_N (\mathrm{y} - md) + R_E (\ z - mn_c)$
Δ_4	$= a \cdot mn \cdot (\ O_N w + x) \quad + a \cdot mx \cdot (\ O_N \mathrm{y} + z) - (a+b) \cdot mn \cdot [O_N md + mn_c]$ $= O_N (w - md) + (x - mn_c) + R_E O_N (\mathrm{y} - md) + R_E (z - mn_c)$

If we set $w = md$; $x = mn_c$; y $= md$ and $z = mn_c$, then $\Sigma_1, \Sigma_2, \Sigma_3$, Σ_4 and B do not change since $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$, and the constraints are still satisfied. However, after the changes $p_e(v) \neq mx$ so that $|\{v \in C_1 \cap C_0\}|$ is reduced by $a + b$. Many other possible substitutions would increase B.

We repeat the construction until $|CC_K \cap \{v: p_c(v, 0) = mn_c\}| < a + b$. The construction does not affect the conclusions of Propositions 1-4, concluding the proof.

Theorem 3: Fix $mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D$, For every $\varepsilon > 0$ there exists a dichotomous variable V and $\wp(\vec{V}_2) \in \mathcal{H}(2, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 9\varepsilon)$ with $B(\wp(\vec{V}_2)) = \frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$, $B(\wp(\vec{V}_2)) > B_{max} - 9\varepsilon$ where: (1) $\Sigma_1 = R_E(R_D O_N r_c + 1) + (R_D O_N + r_c'')$; $\Sigma_2 = R_E(O_N r_c + 1) + (O_N + r_c'')$

$$\Sigma_3 = (R_D O_N r_c + 1) + R_E(R_D O_N + r_c''); \qquad \Sigma_4 = (O_N r_c + 1) + R_E(O_N + r_c'')$$

<u>Proof</u>: Given R_E, R_C, R_D, $\varepsilon > 0$ and B_{max}, set $\varepsilon' = \frac{(R_E-1)\varepsilon}{(1+R_D)^2[(14R_E+42)R_C R_D - 7(R_E+1)\varepsilon]}$ .Theorems 1 and 2 imply there exists $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 7\varepsilon)$ such that $B(\wp(\vec{V}_\aleph)) \geq B_{max} - 7\varepsilon$ and $p_e(X) < 7\varepsilon'$ where $X = \{v: v \notin (A_1 \cap C_0) \cup (A_1 \cap C_0)\}$. Using $\Sigma_{1,\aleph}$ to $\Sigma_{4,\aleph}$ to denote the four sums in $B\big(\wp(\vec{V}_\aleph)\big) = \frac{\Sigma_{1,\aleph}\Sigma_{4,\aleph}}{\Sigma_{2,\aleph}\Sigma_{3,\aleph}}$, we have $\frac{\Sigma_{1,\aleph}\Sigma_{4,\aleph}}{\Sigma_{2,\aleph}\Sigma_{3,\aleph}} \geq B_{max} - 7\varepsilon$. We can write $\Sigma_{1,\aleph}$ to $\Sigma_{4,\aleph}$ as:

(2) $\Sigma_{1,\aleph} = mx \sum_v^A (p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + mn \sum_v^B (p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + er_1$

$\Sigma_{2,\aleph} = mx \sum_v^A (p_N\, p_c(v,1) + q_N p_c(v,0)) + mn \sum_v^B (p_N p_c(v,1) + q_N p_c(v,0)) + er_2$

$\Sigma_{3,\aleph} = mn \sum_v^A (p_D p_N p_c(v,1) + q_D q_N p_c(v,0) + mx \sum_v^B (p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + er_3$

$\Sigma_{4,\aleph} = mn \sum_v^A (p_N p_c(v,1) + q_N p_c(v,0)) + mx \sum_v^B (p_N p_c(v,1) + q_N p_c(v,0) + er_4$

where: $er_1 = \sum_{v \epsilon X} p_1(v)(p_D p_N p_c(v,1) + q_D q_N p_c(v,0))$ with corresponding definitions for er_2, er_3, er_4; $p_N = P(N = 1)$, $q_N = 1 - p_N$, $p_D = P(D = 1|E = 1, N = 1)$, $q_D = P(D = 1|E = 0, N = 0)$; and where Σ_v^A and Σ_v^B are sums over $v \in A_1 \cap C_0$ and $v \in C_1 \cap A_0$. The factor mx that appears in $\Sigma_{1,\aleph}$ represents $p_1(v)$, since $p_1(v) = mx$ for all $v \in A_1 \cap C_0$; justification of other factors is similar. We have:

$$0 \leq er_1 = \sum_{v \epsilon X} p_1(v)[p_D p_N p_c(v,1) + q_D q_N p_c(v,0)] \leq 7\varepsilon' R_D R_C (O_N + 1) q_N q_D mn_c;$$

the same bound holds for er_2, er_3 and er_4.

By Lemma 3, we can assume $p_c(v, n) = p_c(v', n)$, for all $v, v' \in A_1 \cap C_0$ and similarly for $C_1 \cap A_0$. Thus, letting N_A ($N_{A'}$) be the number of points in $A_1 \cap C_0$ ($C_1 \cap A_0$), we can write Equations (2) as:

(3) $\Sigma_{1,\aleph} = mx N_A (p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + mn N_{A'} (p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + er_1$

$\Sigma_{2,\aleph} = mxN_A(p_N p_c(v,1) + q_N p_c(v,0)) + mnN_{A'}(p_N p_c(v,1) + q_N p_c(v,0)) + er_2$

$\Sigma_{3,\aleph} = mnN_A(p_D p_N p_c(v,1) + q_D q_N p_c(v,0) + mxN_{A'}(p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + er_3$

$\Sigma_{4,\aleph} = mnN_A(p_N p_c(v,1) + q_N p_c(v,0)) + mxN_{A'}(p_N p_c(v,1) + q_N p_c(v,0) + er_4$

$\sum_v p_e(v) = 1$, for $e = 0,1$, implying: $mxN_A + mnN_{A'} + \varepsilon_1 = 1$ and $mnN_A + mxN_{A'} + \varepsilon_0 = 1$, where $\varepsilon_e = p_e(X) \leq 7\varepsilon'$. Lemma 4 quantifies $|N_A - N_{A'}|$ and provides a bound for the error if N_A replaces $N_{A'}$:

(4) $\Sigma_{1,\aleph} = mxN_A(p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + mnN_A(p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + er_1'$

$\Sigma_{2,\aleph} = mxN_A(p_N p_c(v,1) + q_N p_c(v,0)) + mxN_A(p_N p_c(v,1) + q_N p_c(v,0)) + er_2'$

$\Sigma_{3,\aleph} = mnN_A(p_D p_N p_c(v,1) + q_D q_N p_c(v,0) + mxN_A(p_D p_N p_c(v,1) + q_D q_N p_c(v,0)) + er_3'$

$\Sigma_{4,\aleph} = mnN_A(p_N p_c(v,1) + q_N p_c(v,0)) + mxN_A(p_N p_c(v,1) + q_N p_c(v,0) + er_4'$

where $er_i' = er_i + \delta_i$ and δ_i is the additional error due to replacing $N_{A'}$ with N_A. δ_i is defined in Lemma 4 which shows $|\delta_i| \leq \frac{28\varepsilon'}{(R_E-1)} R_D R_C (O_N + 1) q_D q_N mn_c$. Thus, $|er_i'| \leq (7\varepsilon' + \frac{28\varepsilon'}{R_E-1}) R_D R_C (O_N + 1) q_D q_N mn_c$.

Define $\Sigma_i = \Sigma_{i,\aleph} - er_i'$. Lemma 5 shows that we can replace $\Sigma_{i,\aleph}$ with Σ_i (dropping er_i') to write $B(\wp(\vec{V}_\aleph))$ with error bounded by $\left|\frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3} - \frac{\Sigma_{1,\aleph}\Sigma_{4,\aleph}}{\Sigma_{2,\aleph}\Sigma_{3,\aleph}}\right| \leq 2\varepsilon$. Thus $\frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3} \geq B_{max} - 9\varepsilon$.

We now re-express Σ_1 to Σ_4. Divide Σ_1 and Σ_3 by $mnN_A q_D q_N mn_c$ and divide Σ_2 and Σ_4 by $mnN_A q_N mn_c$ without changing $\frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$. Replace $\frac{p_D}{q_D}, \frac{p_N}{q_N}$ and $\frac{max}{min}$ with R_D, O_N and R_E (defined previously). Replace $\frac{p_c(v,1)}{mn_c}$ and $\frac{p_c(v,0)}{mn_c}$ with r_c and r for $v \in A_1 \cap C_0$ and $\frac{p_c(v,1)}{mn_c}$ and $\frac{p_c(v,0)}{mn_c}$ with r_c' and r_c'' for $v \in C_1 \cap A_0$). By the construction in Theorems 1-2 and Propositions 1-5, $1 \leq r, r_c, r_c', r_c'' \leq \frac{mx_c}{mn_c} = R_C$. By Lemma 6, one of r, r_c and one of r_c', r_c'' must be at its boundary (1 or R_C). We take $r = 1$ and $r_c' = 1$, but each of the other choices is handled similarly. These changes don't affect the value of $\frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$ (or $B(\wp(\vec{V}_\aleph))$). After the divisions and substitutions, taking $r = 1$ and $r_c' = 1$, we can express Σ_1 to Σ_4 as:

$\Sigma_1 = R_E(R_D O_N r_c + 1) + (O_N + r_c'')$; $\Sigma_2 = R_E(O_N r_c + 1) + (O_N + r_c'')$

$\Sigma_3 = (R_D O_N r_c + 1) + R_E(R_D O_N + r_c'')$; $\Sigma_4 = (O_N r_c + 1) + R_E(O_N + r_c'')$

Now $\frac{\Sigma_{1,\aleph}\Sigma_{4,\aleph}}{\Sigma_{2,\aleph}\Sigma_{3,\aleph}}$ is an approximation (within 7ε) to $B(\wp(\vec{V}_\aleph))$, and $\frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$ approximates $\frac{\Sigma_{1,\aleph}\Sigma_{4,\aleph}}{\Sigma_{2,\aleph}\Sigma_{3,\aleph}}$. However, $\frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3}$ has exactly the structure of the bias for $\wp(\vec{V}_2)$, where $\vec{V}_2 = E, V, C, N, D$ with all variables dichotomous; $P(V=1|E=1) = \frac{R_E}{R_E+1}$; $P(V=1|E=0) = \frac{1}{R_E+1}$; $R_D = \frac{P(D=1|N=1)}{P(D=1|N=0)}$; $O_N = \frac{P(N=1)}{P(N=0)}$; and, $p_c(v,n) = P(C=c|V=v, N=n)$. Reflecting the definitions and construction, $\wp(\vec{V}_2) \in \mathcal{H}(2, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, 9\varepsilon)$, with $B(\wp(\vec{V}_2)) = \frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3} > B_{max} - 9\varepsilon$, proving Theorem 3.

Lemma 4: Define $\Sigma_{i,\aleph}$ as in Theorem 3, and let $\Sigma'_{i,\aleph} = \Sigma_{i,\aleph}$, but with $N_{A'}$ replaced by N_A; e.g.,
$\Sigma'_{1,\aleph} = mxN_A\big(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)\big) + mnN_A\big(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)\big) + er_1$.
If both: i) $R_EmnN_A + mnN_{A'} + \varepsilon_1 = 1$; and, ii) $mnN_A + R_EmnN_{A'} + \varepsilon_0 = 1$ hold with $|\varepsilon_0|$, $|\varepsilon_1| \leq p_e(X) \leq 7\varepsilon'$, then: $|N_A - N_{A'}| \leq \frac{28\varepsilon'}{(R-1)mn}$ and $\delta_i = |\Sigma'_{i,\aleph} - \Sigma_{i,\aleph}| \leq \frac{28\varepsilon'}{(R-1)} R_DR_C(O_N+1)q_Dq_Nmn_c$.

<u>Proof</u>: Multiplying both sides of Equation (ii) by R_E gives

(1) $R_EmnN_A + {R_E}^2mnN_{A'} + R_E\varepsilon_0 = R_E$. Subtracting Equation ($i$) from (1) and solving gives:

(2) $N_{A'} = \frac{1}{(R_E+1)mn} - \frac{R_E\varepsilon_0 - \varepsilon_1}{({R_E}^2-1)mn}$ and $N_A = \frac{1}{(R_E+1)mn} + \frac{\varepsilon_0 - R_E\varepsilon_1}{({R_E}^2-1)mn}$. Subtracting N_A from $N_{A'}$

(3) $N_{A'} - N_A = \frac{\varepsilon_1 - \varepsilon_0}{({R_E}^2-1)mn}$; since $|\varepsilon_i| \leq 7\varepsilon'$: $|N_{A'} - N_A| \leq \frac{28\varepsilon'}{(R_E-1)mn}$, and $N_Amn \geq \frac{1}{R_E+1} - \frac{7\varepsilon'}{R_E-1}$.

We now prove $\delta_i \leq \frac{28\varepsilon'}{(R_E-1)} R_DR_C(O_N+1)q_Dq_Nmn_c$. Proof for $i = 2,3,4$ is similar. By definition

$$\begin{aligned}
|\Sigma'_{1,\aleph} - \Sigma_{1,\aleph}| &= |mxN_A(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)) + mnN_A(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)) \\
&\quad - mxN_A(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)) + mnN_{A'}(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0))| \\
&= |N_A - N_{A'}|mn[p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)] \\
&= |N_A - N_{A'}|mn[R_DO_Nr_c' + r_c''](q_Dq_Nmn_c) \\
&\leq \frac{28\varepsilon'}{(R_E-1)} R_DR_C(O_N+1)q_Dq_Nmn_c,
\end{aligned}$$

as claimed. The last inequality follows from (3) and because $R_DR_C(O_N+1) \leq [R_DO_Nr_c' + r_c'']$. The same bound holds for δ_2, δ_3, and δ_4.

Lemma 5: Let $\Sigma_i, \Sigma_{i,\aleph}, er_i'$ and ε be as defined in Theorem 3. If $|er_i'| \leq (7\varepsilon' + \frac{28\varepsilon'}{R_E-1})R_DR_C(O_N + 1)\, q_Dq_Nmn_c$, then $|\frac{\Sigma_1\Sigma_4}{\Sigma_2\Sigma_3} - \frac{\Sigma_{1,\aleph}}{\Sigma_{2,\aleph}}\frac{\Sigma_{4,\aleph}}{\Sigma_{3,\aleph}}| \leq 2\varepsilon$. To simplify notation, set $er_i = er_i'$.

<u>Proof</u>: The following equalities and inequalities follow from the definitions and algebraic manipulation, and inequalities proven at the end of Lemma 5:

$$\begin{aligned}
\Sigma_1 &= mxN_A(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)) + mnN_A(p_Dp_Np_c(v,1) + q_Dq_Np_c(v,0)) \\
&= mnN_A[R_E(R_DO_Nr + r_c)q_Dq_Nmn_c + (R_DO_Nr' + r_c'')q_Dq_Nmn_c]
\end{aligned}$$

$$\frac{\Sigma_{1,\aleph}}{\Sigma_{2,\aleph}} - \frac{\Sigma_1}{\Sigma_2} = \frac{\Sigma_1 + er_1}{\Sigma_2 + er_2} - \frac{\Sigma_1}{\Sigma_2} = \frac{\Sigma_1\Sigma_2 + er_1\Sigma_2 - \Sigma_1\,\Sigma_2 - er_2\Sigma_1}{\Sigma_2\,(\Sigma_2 + er_2)}$$

$$= \frac{er_1\Sigma_2 - er_2\Sigma_1}{\Sigma_2\,(\Sigma_2+er_2)} = \frac{er_1 - er_2\frac{\Sigma_1}{\Sigma_2}}{\Sigma_2+er_2} = \frac{\frac{er_1}{\Sigma_2} - er_2\frac{\Sigma_1}{(\Sigma_2)^2}}{1+\frac{er_2}{\Sigma_2}} = \frac{\frac{er_1}{\Sigma_2} - \frac{er_2\Sigma_1}{\Sigma_2\,\Sigma_2}}{1+\frac{er_2}{\Sigma_2}}.$$ Therefore:

$$\left|\frac{\frac{er_1}{\Sigma_2} - er_2\frac{\Sigma_1}{(\Sigma_2)^2}}{1+\frac{er_2}{\Sigma_2}}\right| \le \frac{\max(|er_1|,|er_2|)}{\Sigma_2}\frac{1+R_D}{1-\frac{|er_2|}{\Sigma_2}} \le (1+R_D)\max\left(\frac{|er_1|}{\Sigma_2},\frac{|er_2|}{\Sigma_2}\right)\left(1+\max\left(\frac{|er_1|}{\Sigma_2},\frac{|er_2|}{\Sigma_2}\right)\right)$$

$\le \frac{(1+R_D)\varepsilon}{2(1+R_D)^2}\left(1+\frac{\varepsilon}{2(1+R_D)^2}\right) < \frac{\varepsilon}{1+R_D}$, provided $\frac{\varepsilon}{2(1+R_D)^2} < \frac{1}{2}$. The penultimate inequality follows from $\max(\frac{|er_1|}{\Sigma_2},\frac{|er_2|}{\Sigma_2}) \le \frac{(7R_E+21)\varepsilon' R_D R_C}{R_E-1-(R_E+1)7\varepsilon'}$ shown below and choice of ε' so that $\frac{(7R_E+21)\varepsilon' R_D R_C}{R_E-1-(R_E+1)7\varepsilon'} \le \frac{\varepsilon}{2(1+R_D)^2}$. Similarly,

$$\frac{\Sigma_{4,\aleph}}{\Sigma_{4,\aleph}} - \frac{\Sigma_4}{\Sigma_3} = \frac{\Sigma_4+er_4}{\Sigma_3+er_3} - \frac{\Sigma_4}{\Sigma_3} = \frac{\Sigma_4\Sigma_3+er_4\Sigma_3-\Sigma_4\,\Sigma_3-er_3\Sigma_4}{\Sigma_{3,\aleph}\,(\Sigma_{3,\aleph}+er_3)} = \frac{er_4-er_3\frac{\Sigma_4}{\Sigma_3}}{\Sigma_3+er_3} = \frac{\frac{er_4}{\Sigma_3}-\frac{er_3\Sigma_4}{(\Sigma_3)^2}}{1+\frac{er_3}{\Sigma_3}}$$ so that

$\left|\frac{\frac{er_4}{\Sigma_3}-\frac{er_3}{(\Sigma_3)^2}}{1+\frac{er_3}{\Sigma_3}}\right| \le \frac{\max(|er_3|,|er_4|)}{\Sigma_3}\frac{1+R_D}{1-\frac{|er_3|}{\Sigma_3}} \le \frac{\varepsilon}{1+R_D}$, since $\max(\frac{|er_3|}{\Sigma_3},\frac{|er_4|}{\Sigma_3}) \le \frac{(7R_E+21)\varepsilon' R_D R_C}{R_E-1-(R_E+1)7\varepsilon'}$

, , the same bound as for $\max\left(\frac{|er_1|}{\Sigma_1},\frac{|er_2|}{\Sigma_2}\right)$. Finally:

$$\left|\frac{\Sigma_{1,\aleph}}{\Sigma_{2,\aleph}}\frac{\Sigma_{4,\aleph}}{\Sigma_{3,\aleph}} - \frac{\Sigma_1}{\Sigma_2}\frac{\Sigma_4}{\Sigma_3}\right| = \left|\left(\frac{\Sigma_{1,\aleph}}{\Sigma_{2,\aleph}}\frac{\Sigma_{4,\aleph}}{\Sigma_{3,\aleph}} - \frac{\Sigma_1}{\Sigma_2}\frac{\Sigma_{4,\aleph}}{\Sigma_{3,\aleph}}\right) + \left(\frac{\Sigma_1}{\Sigma_2}\frac{\Sigma_{4,\aleph}}{\Sigma_{3,\aleph}} - \frac{\Sigma_1}{\Sigma_2}\frac{\Sigma_4}{\Sigma_3}\right)\right|$$

$= \left|\left(\frac{\Sigma_{1,\aleph}}{\Sigma_{2,\aleph}} - \frac{\Sigma_{1,\aleph}}{\Sigma_{2,\aleph}}\right)\frac{\Sigma_{4,\aleph}}{\Sigma_{3,\aleph}} + \frac{\Sigma_1}{\Sigma_2}\left(\frac{\Sigma_{4,\aleph}}{\Sigma_{3,\aleph}} - \frac{\Sigma_4}{\Sigma_3}\right)\right| \le R_D\frac{\varepsilon}{1+R_D} + R_D\frac{\varepsilon}{1+R_D} < 2\varepsilon$, proving Lemma 5.

Inequalities used in proof of Lemma 5:

0) $|er_1| \le (7\varepsilon' + \frac{28\varepsilon'}{R-1})R_D R_C(O_N+1)q_D q_N mn_c$, proven in Theorem 3

1) $\Sigma_2 = mnN_A[R_E(p_N p_c(v,1) + q_N p_c(v,0)) + (p_N p_c(v,1) + q_N p_c(v,0))]$

$$\ge mnN_A(R_E+1)(O_N+1)q_N mn_c \ge \left(\frac{1}{R_E+1} - \frac{7\varepsilon'}{R_E-1}\right)(R_E+1)(O_N+1)q_N mn_c;$$

2) $\frac{|er_1|}{\Sigma_2} \le \frac{\max|er_1|}{\min\Sigma_2} \le \frac{(7\varepsilon'+\frac{28\varepsilon'}{R_E-1})R_D R_C(O_N+1)q_D q_N mn_c}{\left(\frac{1}{R_E+1}-\frac{7\varepsilon'}{R_E-1}\right)(R_E+1)(O_N+1)q_N mn_c} = \frac{(7\varepsilon'+\frac{28\varepsilon'}{R_E-1})R_D R_C}{\left(\frac{1}{R_E+1}-\frac{7\varepsilon'}{R_E-1}\right)(R_E+1)} = \frac{(7R_E+21)\varepsilon' R_D R_C}{R_E-1-(R_E+1)7\varepsilon'} \le \frac{\varepsilon}{2(1+R_D)^2}$ by choice of ε' and $|er_1| \le (7\varepsilon' + \frac{28\varepsilon'}{R-1})R_D R_C(O_N+1)q_D q_N mn_c$; $\frac{|er_2|}{\Sigma_2},\frac{|er_3|}{\Sigma_3},\frac{|er_4|}{\Sigma_3}$ are also bounded by $\frac{\varepsilon}{1+R_D}$.

3) $1 < \frac{\Sigma_1}{\Sigma_2} < R_D$: $\frac{\Sigma_1}{\Sigma_2} = \frac{R_E(R_D O_N r_c+1)+(R_D O_N + r_c'')}{R_E(O_N r_c+1)+(O_N+r_c'')}$ is a weighted average of 1 and $R_D(a,b)$ over $a = 0,1; b = 0,1$: $R_D(1,b) = R_D$ & $R_D(0,b) = 1, b = 0,1,$ and weights: $w(1,1) = \frac{R_E}{R_E+1}O_N r_c$, $w(1,0) = \frac{1}{R_E+1}O_N, w(0,1) = \frac{R_E}{R_E+1}, w(0,0) = \frac{1}{R_E+1}r''_c$. The same reasoning shows $1 < \frac{\Sigma_3}{\Sigma_4} < R_D$.

Lemma 6: Consider $\wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, \varepsilon) \ne \phi$ such that conclusions of Theorem 2 hold. Then $\exists\, \wp(\vec{V}_\aleph) \in \mathcal{H}(\aleph, mn, mx, mn_c, mx_c, O_N, R_E, R_C, R_D, \varepsilon)$ such that for $v \in A_1 \cap C_0$ either $p_c(v,n) = mn_c$ or $p_c(v,n) = mx_c$ for all v and $n = 0$ or 1. The same claim holds for $v \in C_1 \cap A_0$.

Proof. This result was proven in the first part of the proofs of Theorems 1 and 2 by considering the partial derivatives $\frac{\partial ln(B)}{\partial p_c(v,1)}$ and $\frac{\partial ln(B)}{\partial p_c(v,0)}$; it is summarized and repeated here for convenience.

III. eAppendix 4. Simulation Experiment to evaluate bound and R program for simulations

We also illustrate empirically that bound Bd is valid. Thus, we conduct a series of simulations, each with its own set of randomly generated parameters. The purpose is to empirically illustrate that the actual bias approaches, but doesn't exceed Bd. We assess bound Bd's performance and "tightness", measured as the difference between the expected bias and Bd, under our simulation conditions. We consider scenarios in which V and N has either 2, 3, 4 or 5 categories; E, and D are dichotomous; and, we select or condition on $C = c$, consistent with the causal structure in Figure 1C. We randomly generate parameters for the probability functions in bias expression (3) (eTable A13; R program just below). For each set of parameters generated, we calculate the bias according to Equation 3 (main text), the values of R_E, R_D and R_C, the bound Bd (Equation 4), and compare the ratio of the bias to the bound. The purpose isn't to assess the magnitude of bias for plausible parameters (some combinations are extreme), but rather to empirically demonstrate validity – for any parameter combination. Examples 1-8 of the main text utilize only a restricted range of parameters.

eTable A13. Summary of random Parameter generation for simulation study

Distribution	Model for generating Parameters
$P(N = n) = \pi_{N,n}; n = 1, \dots, N; 0 \le \pi_{N,n};\ \sum_{n=1}^{N} \pi_{N,n} = 1$	Dirichlet (N, $\alpha_1, \dots, \alpha_N$); $\alpha_n = \frac{1}{N}$
$P(C = c \mid V = v, N = n) = \pi_{c,v,n};\ v = 1, \dots, V, n = 1, \dots, N\ ; 0 \le \pi_{c,v,n} \le 1$	$\pi_{c,v,n} = expit(\beta_v + \beta_n + \beta_{v,n})$ where $\beta_v, \beta_n, \beta_{V,N} \sim MVN$
$P(D = 1 \mid E = 0, N = n) = \pi_{D,n}; n = 1, \dots, N^{\dagger}; 0 \le \pi_{D,n} \le 1$	$\pi_{D,n} = expit(\gamma_n)$, where $\gamma_n \sim N(0, \dim{(N)}^{-1})$
$P(V = v \mid E = e) = \pi_{v,e}; v = 1, \dots, V, E = 0,1;\ 0 \le \pi_{v,e}; \sum_{v=1}^{V} \pi_{v,e} = 1$	Dirichlet(V, $\alpha_{1,e}, \dots, \alpha_{V.e})^{\ddagger}$; $\alpha_{v,e} = \frac{1}{V}$

†bias in the standardized RR, with the exposed as the standard does not depend on $P(D = 1 \mid E = 1, N = n)$, see Equation (3); main text. ‡absent further constraints, $R_E = \max_{v,e} \frac{\pi_{v,e}}{\pi_{v,1-e}}$ was often exceedingly large (e.g., > 10^{20}), so we used constraints so that $\max_{v,e} \frac{\pi_{v,e}}{\pi_{v,1-e}}$ was less than 100.

As summarized in eTable A14, we simulate bias over a wide range of bound parameters (R_E, R_D and R_C). In ten scenarios collectively including over 1,000,000 simulations, the actual bias never exceeded Bd. Furthermore, the bias was close to the bound in each scenario for at least some parameter values. The maximum simulated bias was slightly less than Bd in scenarios seven and ten, perhaps because the parameter-space dimensions were larger than for other scenarios and our simulations may have left some parts incompletely explored. By design, large parameter combinations aren't excluded to evaluate empiric validity for a wide range of situations.

eTable A14 Description of Simulation Results, and Ratio† of Largest Bias to Bound Observed in each simulation Scenario (R_E, R_D, R_C as defined in Equations 5)⁋.

Scenario*	Dim M, N	Covariance (muCov in program)	R_E ‡ median, max	R_D median, max	R_C median, max	Bd median, max	Ratio † max(bias/Bd)
1	2, 2	4	7.204, 100.000	1.554, 58.469	1.045, 10.073	1.005, 2.582	1
2	2, 2	2	7.204, 100.000	1.554, 58.469	1.314, 53.854	1.028, 6.765	1
3	2, 2	1	7.204, 100.000	1.554, 58.469	1.766, 120.019	1.057, 9.412	0.9999999
4	2, 2	0	7.204, 100.000	1.554, 58.469	2.667, 227.381	1.095, 10.998	0.9999998
5	2, 2	-1	7.204, 100.000	1.554, 58.469	4.043, 457.327	1.130, 12.452	0.9999997
6	2, 2	-2	7.204, 100.000	1.554, 58.469	5.605, 841.442	1.154, 14.166	0.9999997
7	3, 3	0	13.111, 100.000	2.132, 57.857	4.861, 1172.693	1.438, 15.467	0.9993791
8	4, 2	0	15.594, 100.000	1.560, 56.406	4.403, 719.145	1.239, 14.439	0.9999902
9	2, 4	0	7.230, 100.000	2.600, 57.647	4.380, 1071.381	1.370, 18.981	0.9999985
10	5, 5	0	17.358, 100.000	2.987, 74.171	9.478, 809.786	2.084, 36.008	0.9707631

⁋R program for simulations in this online Supplement below. *Each Scenario included 100,000 simulations. ‡Max R_E ~100 in each Scenario, due to constraints. † Maximum ratio of bias to upper bound (Bd),.

```
#Supplemental Material,   R program to simulate performance of Bound Bd.
#M Bias:   E<--- V --->C<--- N --->D            #in equations use P(M=m|E=e)


#   Evaluate lower Bound as well (RatioL should be >= 1)
# Non-Null
library (MASS)  #for mvn
library (MCMCpack)   #Dirichlet density: rdirichlet(n, alpha)
expit=function(x){exp(x)/(1+exp(x))}  #expit(0)


set.seed(77373)         #<---Set the seed, if desired for repeatability
########## set the input parameters for scenarios 1-10  ############
 NSim = 100000               #<---Set Number of simulations (100,000 for Tables)
 W = 0                       #<---Set the standard: W=1 for E+,  W=0 for E-, or e.g. 0.5 for P(E)=0.5
 DimM=2                        #<---Set dimiensions of M, N
```

```
 DimN=2
 DimMx =max(DimM, DimN)      #<---used to create covariance matrix
 muCov =rep( 4, DimMx)       #<-- the Covariance for interaction parms,  M x N --> C
# Include Above: DimM, DimN,  muCov in Table  (muCov sets the average covariance of the M-effect
and N-effect on C, logit scale)

# Other parameters, no need to change for different scenarios
Cov1  =matrix(rep( 0, DimMx^2), nrow=DimMx) # CovarMatrix for interzction parms
muC1  =       rep( 0, DimM )   # Mean, parms for efffect of each M-category on C
muC2  =       rep( 0, DimN )   # Mean, parms for efffect of each N-category on E
muD   =       rep( 0, DimN )   # Mean, parms for efffect of each N-category  on D
L1 = sqrt(100)   #< -- Set Approx limit on simulated max P(M=m|E=1)/P(M=m|E=0). Can increase to
L1^2
                        #       after normalize.
SigC1= diag(DimM )
SigC2= diag(DimN )
SigD = diag(DimN )          #used w/ muD in mvrnorm to get D parms

alphM0 = rep(1/DimM, DimM)    #<-- use to generate M-parms for E=0,  M=V in m/s
alphM1 = rep(1/DimM, DimM)    #<-- use to generate M-parms for E=1,  M=V in m/s
alphN  = rep(1/DimN, DimN)

Save  =NULL                  #Initialize
SaveM =NULL
ratio =NULL
ratioL=NULL            #use for lower bound

# use following to describe simulation conditions
mxRe =0
mxRd=0
mxRc=0
mxOn=0
mxBd=0
for (k in 1:NSim){                              #simulation loop:  each simulation in Table = 100,000
```

```
# -- get parms for P(M=m|E=e) --
 parmM1=rdirichlet(1, alphM1)        #<-- parms for P(M=m|E=1),  M=V in m/s
 parmM0=rdirichlet(1, alphM0)        #<-- parms for P(M=m|E=0),  M=V in m/s
 for (kk in 1:length(parmM1)) {       #
   if (parmM1[kk] >= L1*parmM0[kk]) {parmM1[kk]=L1*parmM0[kk]}
   if (parmM0[kk] >= L1*parmM1[kk]) {parmM0[kk]=L1*parmM1[kk]}  # limit to ratio of 1000
  }   # rbind(parmM1[1:8], pM1[1:8], pM0[1:8], pM1[1:8]/pM0[1:8])

 parmM1=parmM1/sum(parmM1)          #normalize parmM1, parmM0
 parmM0=parmM0/sum(parmM0)

# -- get parms for P(N=n) -- #
 parmN=rdirichlet(1, alphN)                  #N parameters

# -- get parms for M-->C
 parmC1=mvrnorm(n = 1, muC1, SigC1, tol = 1e-5, empirical = FALSE, EISPACK = FALSE)

# -- get parms for N-->C
 parmC2=mvrnorm(n = 1, muC2, SigC2, tol = 1e-5, empirical = FALSE, EISPACK = FALSE)
# - COV
 parmC12=mvrnorm(n=DimMx, muCov, Cov1, tol = 1e-5, empirical = FALSE, EISPACK = FALSE)

# -- get parms for N-->D
 parmDN =mvrnorm(n = 1, muD, SigD, tol = 1e-5, empirical = FALSE, EISPACK = FALSE)
 parmDNI=mvrnorm(n = 1, muD, SigD, tol = 1e-5, empirical = FALSE, EISPACK = FALSE)
#interaction w/E
 parmDE=rnorm(n = 1, 1, 1)  # main effect of E
 S1 =0          #initialize sums
 S2 =0
 S3 =0
 S4 =0
 S5 =0          #used for SRR, E+ is std
 S6 =0          #used for SRR, E- is std
 mxM =0           #initialize max, min
```

```
 mnC =1
 mxC =0
 mxD1=0
 mnD1=1
 mxD0=0
 mnD0=1
for (i in 1:DimM){            ## i=1
 for (j in 1:DimN){
  M1=c(rep(0,max(0,i-1)), 1, rep(0, max(0,DimM-i)) )   #value M1, or =rmultinom(1, 1, alphM)
  pM0=M1 %*% t(parmM0)                                  #Prob (M=i|E=0)
  pM1=M1 %*% t(parmM1)                                  #Prob (M=i|E=1)

  N=c(rep(0,max(0,j-1)), 1, rep(0, max(0,DimN-j)) )     #value N;  or =rmultinom(1, 1, alphN)
  pN=dmultinom(N, size = 1, parmN)    #Prob (N=j)

# -- get probs for M,N-->C,  includes  interaction terms for M,N-->C  (parmC12)
  pC1 = dbinom(1, 1, expit(0+ M1%*%parmC1 + N%*%parmC2 + parmC12[i,j]))   #p(C=1| M=i, N=j)

  pD0 = dbinom(1, 1, expit(0 + t(N)%*%parmDN ) )                                #p(D=1| N=j,E=0)
  pD1 = dbinom(1, 1, expit(0 + t(N)%*%parmDN + parmDE + t(N)%*%parmDNI))  #p(D=1| N=j,E=1)

 S1 = S1+ pM1*pN*pC1*pD1
 S2=  S2+ pM1*pN*pC1
 S3 = S3+ pM0*pN*pC1*pD0
 S4=  S4+ pM0*pN*pC1
 S5 = S5+ pM1*pN*pC1*pD0
 S6 = S6+ pM0*pN*pC1*pD1

 #Use Equation 5:  Re   = max(m,e) {P(M=m|E=e)} /min(m,e) {P(M=m|E=e)}
 #                      Rd = max(P(D=1|n=n)/min(P(D=1|n=n')
 #                      Rc = max(m,n {P(C=1|M=m, N=n)} /min(v,n {P(C=1|M=m, N=n)}
 mxM=max(mxM, pM1/pM0, pM0/pM1 )
 mnC=min(pC1,  mnC)
 mxC=max(pC1,  mxC)
```

```
 mxD1=max(pD1, mxD1)
 mnD1=min(pD1, mnD1)
 mxD0=max(pD0, mxD0)
 mnD0=min(pD0, mnD0)
}}  # for i,j                        #end loops over i,j

Re = mxM

if (0 < W && W< 1) {Rd= max(mxD0/mnD0, mxD1/mnD1)} #all selected, pE=W
if (W == 1 ) {Rd= mxD0/mnD0}                    #exposed is std
if (W == 0 ) {Rd= mxD1/mnD1}                    #unexposed is std

Rc=mxC/mnC

if ( Re <= 1000000000000000000){  #only do calculations if R not very large
  On = sqrt( 1/Rd )
  Bd =
(Re*(Rd*On*Rc+1)+(Rd*On+Rc))*((On*Rc+1)+Re*(On+Rc))/(((Rd*On*Rc+1)+Re*(Rd*On+Rc))*(R
e*(On*Rc+1)+(On+Rc)))
  #sRR = S1/S5                     #Standardized Risk Ratio, E+ = std
  #sRR = S6/S3                     #Standardized Risk Ratio, E- = std
  sRR = (W*S1+ (1-W)*S6)/(W*S5+(1-W)*S3)

  B=(S1*S4/(S2*S3))/sRR              #true Bias, calculated directly from expected values
  ratio = rbind(ratio, B/Bd)          #ratio: true bias/bound  = ?distance? between Bound & actual bias
  ratioL= rbind(ratioL, B/ (1/Bd))     #look at lower bound too, should be >= 1
  Save=rbind(Save, c(Bd, Re, Rd, Rc))  #describe the conditions simulated
  mxRe =max(mxRe, Re)
  mxRd =max(mxRd, Rd)
  mxRc =max(mxRc, Rc)
  mxBd=max(Bd, mxBd)                   # keep track of largest Bd
 } #end if Re not too large
}  #end Simulation loop
c(mxBd, mxRe, mxRd, mxRc, length(ratio))
```

```
#histogram, if desired
#hist(ratio, breaks=c(0,.1,.2,.3,.4, .5,.6,.7,.8,.9,1.0, 1.1))  #, plot=TRUE, title=Title)
 Bdd = Save[,1]
 Bd_q=quantile(Bdd, probs=c(0.25, 0.50, 0.95))   #25, 50 and 95 percentile for Bdd
 Re = Save[,2]
 Re_q=quantile(Re, probs=c(0.25, 0.50, 0.95))   #25, 50 and 95 percentile for Re
 Rd = Save[,3]
 Rd_q=quantile(Rd, probs=c(0.25, 0.50, 0.95))  #25, 50 and 95 percentile for Rd
 Rc = Save[,4]
 Rc_q=quantile(Rc, probs=c(0.25, 0.50, 0.95))  #25, 50 and 95 percentile for Rc

#--------------- output, e.g., Table 2, main text  (along with dim M x N, muCov) ---------------------------#
Results =rbind(c(Bd_q, max(Bdd)), c(Re_q, max(Re)), c(Rd_q, max(Rd)), c(Rc_q, max(Rc)))       #Row 1=Bd, Row 2=Rd, Row 3 =Rc
colnames(Results)[4]='max'
rownames(Results)=c('Bd','Re','Rd','Rc')
Results
max(ratio)                        #maximum of Ratio, should not be > 1
min(ratioL)                        #for Lower Bound, should not be < 1
#getOption("digits")
#options(digits=8)
```

IV. eAppendix 5. Comparison of Bd with Bound Assuming Homogeneity.

We compare bound Bd with the bound derived by Greenland{Greenland, 2003 #46} under a homogeneity assumption (Bd_H). Here, we consider only dichotomous variables, causal relationships consistent with Figure 1C (main text) and no true exposure effect, the situation considered by Greenland. Greenland used odds ratios to measure the strength of association between variables (Figure 1C) in Bd_H, whereas we use risk ratios (Equations 5, main text). To compare Bd_H with Bd and evaluate robustness, we calculate odds ratios for each risk ratio in Equation 5, assuming specific frequencies for D, C and V: $P(C = 1|V = 1, N = n) = 0.000001$, $P(V = 1|E = 1) = \frac{\sqrt{R_E}}{1+\sqrt{R_E}}$ and $P(D = 1|N = 0, E = 0) = 0.000001$, choosing this last value as it illustrates the largest differences we identified between Bd_H and Bd. [Differences for other values tend to be smaller.] We set parameters for the strength of association between variables in Figure 1C using risk ratios (Equation 5) and then calculate the corresponding odds ratios (Equation 6, main text; Table A13, line 1) used in Bd_H.

(6) $OR = \frac{P(V=1|E=1)}{P(V=0|E=1)} \cdot \frac{P(V=0|E=0)}{P(V=1|E=0)}$; $\quad OR_{CV,n} = \frac{P(C=1|V=1,N=n)}{P(C=0|V=1,N=n)} \cdot \frac{P(C=0|V=0,N=n)}{P(C=1|V=0,N=n)},\ n = 0,1;$

$$OR_D = \frac{P(D=1|N=1)}{P(D=0|N=1)} \cdot \frac{P(D=0|N=0)}{P(D=1|N=0)}; \qquad OR_{CN,v} = \frac{P(C=1|V=v,N=1)}{P(C=0|V=v,N=1)} \cdot \frac{P(C=0|V=v,N=0)}{P(C=1|V=v,N=0)},\ v = 0,1$$

Finally, we calculate the bias expected in the observed risk ratio using Expression (3) and the theoretical bound Bd (Expression 4) which gives an approximate bias and bound for the risk odds ratio (ROR) for rare disease.

Under Greenland's homogeneity assumption, $OR = OR_D = OR_{CV,n} = OR_{CN,v}$, for $n, v = 0,1$. To evaluate sensitivity to this assumption, we consider an extreme violation (less extreme situations are considered in examples in the main text) wherein $OR = OR_D = OR_{CN,v=1} = OR_{CV,n=1} > 1$, but $OR_{CN,v=0} = OR_{CVn=0} = 1/OR_{CVn=1}$, and used the *maximum* of these odds ratios in the expression for Bd_H. As summarized in eTable A15, the expected bias (Equation 3) in the observed ROR can be substantially larger than Bd_H if the homogeneity assumption is incorrect. For example, suppose all the odds ratios in expression (6) were between ¼ and 4: $OR = OR_D = OR_{CN,v=1} = OR_{CV,n=1} = 4$, but $OR_{CN,v=0} = OR_{CVn=0} = 1/4$. The maximum bias if all ORs were 4 (homogeneity) is 1.05, but the bias is as large as 1.31 absent homogeneity (all ORs between ¼ and 4) - a potentially important exceedance of the bound. The actual bias did not exceed Bd.

eTable A15 Summary Comparison of bound incorrectly assuming homogeneity with Bd

Value of OR (maximum strength of effects of biasing variables, Equation 6)	2.00	4.00	8.00	16.00$
Actual Bias in ROR (Equation 3)	1.04	1.31	2.05	3.73
Bound for ROR (Bd; Equation 4†)	1.04	1.31	2.05	3.73
Bound for ROR under incorrect assumption of homogeneity‡	1.003	1.05	1.23	1.69

†Equation 4 gives bound for the bias in the standardized risk ratio which is an approximate bound for the standardized risk odds ratio because disease is rare ($P(D = 1|N = 0, E = 0) = 0.000001$). ‡Bound derived by Greenland under homogeneity assumption that $OR = OR_d = OR_{CV,n} = OR_{CN,v}$ (Equation 6). $Example of probabilities yielding maximum odds ratio of 16.00: P(V=1|E=1) ≈ 0.8, P(V=1|E=0) ≈ 0.2, P(D=1|E=0, N=1) ≈ 0.000016, P(D=1|E=0, N=1) ≈ 0.000001, P(C=c|V=1, N=1) = P(C=c|V=0, N=0) ≈ 0.000016, P(C=c|V=0, N=1) = P(C=c|V=1, N=0) ≈ 0.000001, P(N=1) = 0.2; homogeneity fails, since $OR_{CV,n=1} \approx 16$ but $OR_{CV,n=0} \approx 1/16$ yielding Bound ≈ 1.69 if the maximum OR is used. Here, $R_E \approx R_C \approx R_D \approx 16$ so $Bd \approx 3.73$.

V. eAppendix 6. R code to Calculate Bias (Equation 3) and Bounds.

```
# Program to Calulate: 1) Expeect Bias, based on Equation, and 2) Bound Bd and Bd_H
  logit= function(x){log(x/(1-x))}
```

```
  expit= function(x)(exp(x)/(1+exp(x)))
#  Variables, Coding (Liu's  SSRI example):
#     SSRI = selective seritonin reuptake inhibitor
#     V =1 for depressed, 0 otherwise; it affects CAD, and SSRI use
#     C= 1 if no CAD & volunteer to participate, 0 otherwise
#     N= 1 if smoker, 0 otherwise. It that affects participation & Disease, like Lung Cancer
     al =rep(0,4)     #initialize certain paramters to be input
     bet=rep(0,3)    #initialize  certain parms
 #Parameter Input
  p_V = 0.25   #< --- marginal prevalence of Depression
  p_E = 0.10   #< --- marginal prevalence of SSRI (Exposure)
  OR = 27      # odds ratio for SSRE - Depression Association (Liu uses OR's, convert to RRs

  p_N = 0.10   #< --- marginal prevalence of Smoking
  #logistic model parms for CAD, the collider (as per Liu's example)
  al[1] = logit(0.076)     # logit( Prob(CAD=1| V=0, N=0) )
  al[2] = log(1.6)         # Log(OR), CAD - Depression| Smoking = 0
  al[3] = log(3.)        # Log(OR), CAD - Smoking| Depression = 0
  al[4] = log(1.)      # Log(OR), Depression, Smoking "interaction", logistic model
  #Refusals: P(Refuse|V=0, N=0)/ P(Refuse|V=1, N=0)/  P(Refuse|V=0, N=1)/ P(Refuse|V=1, N=1)/
  q=c(.50,.51,.51,.85)   #set q to c(0,0,0,0)  if no refusals
  q = c(0,0,0,0)
  #Parms for outcome D= LungCancer
  bet[1] =  logit(0.0037)
  bet[2] = log(1.0)   #SSRI
  bet[3] = log(15.0)  #smoking


# -------- calculate RRs from OR for Depression(V) - SSRI (E) --------- #
 T  = 1000
 M1 = p_V*T  # Depression
 N1 = p_E*T  # SSRI (Exposure)
 N2 = T- N1
 M2 = T- M1
```

```
b = OR*(M1+N1) + M2-N1    #from 2x2 Table: total 1000, margins: SSRI+=100, Dep=250
a= -OR +1
c= -OR*M1*N1

x1 = (-b + sqrt(b^2-4*a*c))/(2*a)
x2 = (-b - sqrt(b^2-4*a*c))/(2*a)
 c(x1,x2)
aa = x1            #use the solution
#aa =  70
bb = M1 - aa
cc = N1 - aa
dd = N2-M1+aa

aa*dd/(bb*cc)
aa+bb
aa+cc
c(aa,bb,cc,dd, aa*dd/(bb*cc))

RREU1 = (aa/N1)/(bb/N2)  #P(V=1|SSRI=1) / P(V=1|SSRI=0
RREU2 = (cc/N1)/(dd/N2)  #P(V=0|SSRI=1) / P(V=0|SSRI=0)
c( aa/(aa+cc), bb/(bb+dd), cc/(aa+cc), dd/(bb+dd) )

######## Estimate the Bias  ##############



##pV=function(v){ p_V*v+ (1-v)*(1-p_V)}
pN=function(n){ p_N*n+ (1-n)*(1-p_N)}
c(pN(1),pN(0))

pV_E=function(v, e){           #Prob (V=1 | E=e),  for convenience
 if(v==1 && e ==1) {p= aa/N1}
 if(v==1 && e ==0) {p= bb/N2}
 if(v==0 && e ==1) {p= cc/N1}
```

```
 if(v==0 && e ==0) {p= dd/N2}
 p}
c(pV_E(1,1),  pV_E(1,0), pV_E(0,1),  pV_E(0,0))


pC_VN= function(c,v,n){
  p=expit(al[1]+al[2]*v+al[3]*n+al[4]*v*n)
 #Now account for possible Refusals: prob refuse| no CAD, & Depr, Smoking status
 if (v==0 && n==0){p=p+q[1]*(1-p)}
 if (v==1 && n==0){p=p+q[2]*(1-p)}
 if (v==0 && n==1){p=p+q[3]*(1-p)}
 if (v==1 && n==1){p=p+q[4]*(1-p)}
 p=c*p + (1-c)*(1-p)
 p}
c(pC_VN(1,1,1), pC_VN(1,1,0), pC_VN(1,0,1), pC_VN(1,0,0) )
c(pC_VN(0,1,1), pC_VN(0,1,0), pC_VN(0,0,1), pC_VN(0,0,0))


pD_EN= function(e,n){
   expit(bet[1]+bet[2]*e +bet[3]*n) }
 c(pD_EN(0,1),  pD_EN(0,0))

#get Bias B
#stratum w/ C=1
S11=0
S12=0
S13=0
S14=0
for (v in c(0,1)) {
 for (n in c(0,1)) {
   S11= S11 + pN(n)*pV_E(v,1)*pC_VN(1,v,n)*pD_EN(0,n)
   S12= S12 + pN(n)*pV_E(v,0)*pC_VN(1,v,n)*pD_EN(0,n)
   S13= S13 + pN(n)*pV_E(v,1)*pC_VN(1,v,n)
   S14= S14 + pN(n)*pV_E(v,0)*pC_VN(1,v,n)
```

```
}}
c(S11,S13,S12,S14)
B1=S11*S14/(S12*S13)
OB1 = ((S11/S13)/(1-(S11/S13)))/((S12/S14)/(1-(S12/S14)))   #Odds Ratio
B1 #Bias in statrum: C=1

#Now, stratum w/ C=0
S1=0
S2=0
S3=0
S4=0
for (v in c(0,1)) {
 for (n in c(0,1)) {
   S1= S1 + pN(n)*pV_E(v,1)*pC_VN(0,v,n)*pD_EN(0,n)
   S2= S2 + pN(n)*pV_E(v,0)*pC_VN(0,v,n)*pD_EN(0,n)
   S3= S3 + pN(n)*pV_E(v,1)*pC_VN(0,v,n)
   S4= S4 + pN(n)*pV_E(v,0)*pC_VN(0,v,n)
}}
B0=S1*S4/(S2*S3)
OB0 = ((S1/S3)/(1-(S1/S3)))/((S2/S4)/(1-(S2/S4)))   #Odds Ratio
B0

# -------  now get parms needed to calculate bounds --------- #
Re = max(RREU1, RREU2, 1/RREU1, 1/RREU2)
c(RREU1, RREU2, 1/RREU1, 1/RREU2)
Rd =   max(pD_EN(1,1),pD_EN(0,1),pD_EN(1,0),pD_EN(0,0))  #Smoking Effect on outcome(lung
cancer)
Rd =Rd/min(pD_EN(1,1),pD_EN(0,1),pD_EN(1,0),pD_EN(0,0))
Rc1=max(pC_VN(1,1,1),pC_VN(1,0,1),pC_VN(1,1,0),pC_VN(1,0,0))
Rc1=Rc1/min(pC_VN(1,1,1),pC_VN(1,0,1),pC_VN(1,1,0),pC_VN(1,0,0))
Rc0=max(pC_VN(0,1,1),pC_VN(0,0,1),pC_VN(0,1,0),pC_VN(0,0,0))
Rc0=Rc0/min(pC_VN(0,1,1),pC_VN(0,0,1),pC_VN(0,1,0),pC_VN(0,0,0))
Rc= max(Rc1, Rc0)       #make sure apply relevant Rc: Rc0 for C=0
Rc=Rc0
```

```
c(Re, Rd, Rc1, Rc0, Rc)

####### Calculate the Bound
 Bd =  (Re*(sqrt(Rd)*Rc +1)+ (sqrt(Rd)+Rc) )*   ((Rc/sqrt(Rd) +1)+ Re*(1/sqrt(Rd)+Rc) )
 Bd=Bd/((  (sqrt(Rd)*Rc +1)+ Re*(sqrt(Rd)+Rc) )* (Re*(Rc/sqrt(Rd) +1)+ (1/sqrt(Rd)+Rc) ) )
 Bd     ##NB: Bound for RR, if disease rare, approx Bound for OR
 #now, get Bd_H, use largest OR if not uniform
 R=aa*dd/(bb*cc)    #Star with OR for SSRI, depression (V)
 ORCV1= (pC_VN(1,1,1)/pC_VN(0,1,1))/(pC_VN(1,0,1)/pC_VN(0,0,1))
 ORCV0= (pC_VN(1,1,0)/pC_VN(0,1,0))/(pC_VN(1,0,0)/pC_VN(0,0,0))
 ORCN1= (pC_VN(1,1,1)/pC_VN(0,1,1))/(pC_VN(1,1,0)/pC_VN(0,1,0))
 ORCN0= (pC_VN(1,0,1)/pC_VN(0,0,1))/(pC_VN(1,0,0)/pC_VN(0,0,0))
 ORD =  (pD_EN(1,1)/(1-pD_EN(1,1)))/(pD_EN(1,0)/(1-pD_EN(1,0)))
 R= max(R, 1/R, ORCV1, 1/ORCV1, ORCV0, 1/ORCV0, ORCN1, 1/ORCN1, ORCN0, 1/ORCN0,
ORD)
 OR_H=((R^2+6*R+1)^2)/(16*R*(R+1)^2)
 c(OR_H,R,aa*dd/(bb*cc),ORD, 1/ORD, ORCV1, 1/ORCV1, ORCV0, 1/ORCV0, ORCN1, 1/ORCN1,
ORCN0, 1/ORCN0)

c(Re, Rd, Rc, Rc1, Rc0, R)     #Parameters used to calculate Boutnd and Bd_H
c(B1, OB1, B0, OB0, Bd, OR_H)  #Bias for stratum =1, for Stratum=0, Bound for stratum=0 and Bd_H


out=as.list( c(B0, 1/Bd, Bd, 1/OR_H, OR_H) )
names(out)=c("bias,C=0", "Lo Bound", "Up Bound", "Lo Bd_H", "Hi Bd_H")
out
```

References

[1] Ding P, VanderWeele TJ: Sensitivity analysis without assumptions. Epidemiology (Cambridge, Mass) 2016, 27:368.
[2] Byrd RH, Lu P, Nocedal J, Zhu C: A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 1995, 16:1190-208.
[3] Wolfram Research I: Mathematica. Version 11.2.0 ed. Champaign, IL, 2017.