

# UC Davis

## UC Davis Previously Published Works

### Title

Combining Effect Estimates Across Cohorts and Sufficient Adjustment Sets for Collaborative Research

### Permalink

<https://escholarship.org/uc/item/6nj7x65m>

### Journal

Epidemiology, 32(3)

### ISSN

1044-3983

### Authors

Hamra, Ghassan B

Lesko, Catherine R

Buckley, Jessie P

et al.

### Publication Date

2021-05-01

### DOI

10.1097/ede.0000000000001336

Peer reviewed



Published in final edited form as:

*Epidemiology*. 2021 May 01; 32(3): 421–424. doi:10.1097/EDE.0000000000001336.

## Combining Effect Estimates Across Cohorts and Sufficient Adjustment Sets for Collaborative Research: A Simulation Study

Ghassan B. Hamra<sup>\*,a,b</sup>, Catherine R. Lesko<sup>a</sup>, Jessie P. Buckley<sup>a,b</sup>, Elizabeth T. Jensen<sup>c</sup>, Daniel Tancredi<sup>d</sup>, Bryan Lau<sup>a,†</sup>, Irva Hertz-Picciotto<sup>e,†</sup> program collaborators for Environmental influences on Child Health Outcomes§

<sup>a</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

<sup>b</sup>Department of Environmental Sciences and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

<sup>c</sup>Wake Forest School of Medicine, Department of Epidemiology and Prevention, Winston Salem, NC

<sup>d</sup>Department of Pediatrics, University of California, Davis School of Medicine, Sacramento, CA

<sup>e</sup>Division of Environmental and Occupational Health, University of California, Davis School of Medicine, Davis, CA

### Abstract

**Background:** Collaborative research often combines findings across multiple, independent studies via meta-analysis. Ideally, all study estimates that contribute to the meta-analysis will be equally unbiased. Many meta-analyses require all studies to measure the same covariates. We explored whether differing minimally sufficient sets of confounders identified by a directed acyclic graph (DAG) ensures comparability of individual study estimates. Our analysis applied four statistical estimators to multiple minimally sufficient adjustment sets identified in a single DAG.

**Methods:** We compared estimates obtained via linear, log–binomial, and logistic regression and inverse probability weighting, and data were simulated based on a previously published DAG.

**Results:** Our results show that linear, log–binomial, and inverse probability weighting estimators generally provide the same estimate of effect for different estimands that are equally sufficient to adjust confounding bias, with modest differences in random error. In contrast, logistic regression often performed poorly, with notable differences in effect estimates obtained from unique minimally sufficient adjustment sets, and larger standard errors than other estimators.

**Conclusions:** Our findings do not support reliance of collaborative research on logistic regression results for meta-analyses. Use of DAGs to identify potentially differing minimally sufficient adjustment sets can allow meta-analyses without requiring the exact same covariates.

\*Correspondence: Ghassan B Hamra, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205; ghassanhamra@jhu.edu.

†Bryan Lau and Irva Hertz-Picciotto are co-senior authors.

§See acknowledgment for full listing of collaborators

**Conflicts of interest:** None declared.

## Keywords

directed acyclic graph; collaborative research; simulation

---

## Introduction

Collaborative epidemiologic science is vital to advancing knowledge to inform public health decisions.<sup>1</sup> Researchers may require that all participating studies measure the exact same variables for data pooling or meta-analysis.<sup>2</sup> When exposures and outcomes are measured similarly across the studies but the available covariates differ, meta-analysis across studies may still be feasible.

An important consideration when combining effect estimates with meta-analyses is whether or not individual study effect estimates are confounded. Directed acyclic graphs (DAGs) provide a useful means to evaluate the potential for confounding of adjusted effect estimates.<sup>3–5</sup> Notably, a single DAG may suggest different subsets of potential confounders that could be used to obtain an unbiased estimate of the causal or conditionally causal exposure–outcome relationship. While a DAG provides guidance for selection of appropriate adjustment sets, DAGs are qualitative and non-parametric;<sup>5</sup> hence, they do not inform the choice of statistical model.<sup>6</sup>

The present study evaluates performance of four statistical estimators when adjusting for different minimally sufficient confounder sets. Results have implications for confounder control in meta-analysis.

## METHODS

### Estimands, -mators, and -mates

We assessed the ability of four different estimators to return similar or equal estimates of an estimand when conditioning on different adjustment sets. To clarify these terms: An estimand is the effect measure we aim to quantify; the estimator is the tool or statistical model we use to obtain this parameter; and the resultant quantity is the estimate.

A challenge for combining results from different studies may arise when estimands differ. Our Figure was informed by a previously published DAG developed for the association between pre-pregnancy body mass index and cesarean delivery.<sup>7</sup> This DAG contains nine potential confounders: maternal age (continuous), race (categorical), education (categorical), height (continuous), and poverty index (dichotomous), gestational weight gain (continuous), estimated fetal weight (continuous), presence of pre-eclampsia (dichotomous) and chronic hypertension (dichotomous). The previously published DAG supported four minimally sufficient adjustment sets and, thus, four possible estimands from which a presumably unconfounded effect of the relationship between the exposure and the outcome was possible. From Figure 2 in reference 7, these four minimally sufficient adjustment sets are as follows: #1 [C1,C2,C4,C5,C7], #2 [C1,C2,C4,C5,C8], #3 [C1,C2,C4,C6,C7], and #4 [C1,C2,C4,C6,C8].

## Structure of the simulated data

The exposure and outcome of interest were both simulated as normally distributed variables with a mean of zero and variance of one. Dichotomous versions of the exposure and outcome were simulated with prevalence of 30% and 10% by categorizing the simulated, continuous variables to obtain the desired exposure or outcome prevalence. All of the covariates of interest were simulated based on the structure of each from the previously published DAG as noted above. The direction of the confounding effect was based on expectations from previous research. We also structured the simulations to avoid sources of bias including: measurement error, differential selection, positivity, causal consistency, interference, and incorrect model specification.

We evaluated four covariate-adjusted statistical estimators, each estimating a distinct quantity. Linear regression quantifies a risk difference; the conditional effect estimated by linear regression equals the marginal effect. Log-binomial and logistic regression estimate a risk and odds ratio, respectively, which are conditional exposure effects. Because of known challenges of the conditional odds ratio obtained from logistic regression, namely noncollapsibility,<sup>8</sup> we also applied inverse probability weighting to estimate a marginal odds ratio for comparison with the typical conditional odds ratio.

We used linear regression for the continuous outcome; for dichotomous outcomes, other models were used. Because log-binomial regression models often fail to converge, we estimated (conditional) risk ratios assuming a Poisson outcome distribution and applied a robust standard error correction.<sup>9</sup> The value for the true effect of continuous exposure on the continuous outcome was directly parameterized in the simulation. However, for the dichotomous outcome, the true value was estimated by running the simulation 100 times for a sample size of 1 million. This allowed us to fix the outcome prevalence, which otherwise would not be possible. We emphasize that the true value differs across the different estimands. As an alternative, we compared each estimand to a reference estimand from a model that included all potential confounders.

The results for minimally sufficient adjustment sets #1 and #3 were very similar; this was also the case for minimally sufficient adjustment sets #2 and #4. Therefore, we present results for minimally sufficient adjustment sets #1 and #2 only. Confounder #8 (C8) was based on a continuous quantity: gestational weight gain. Thus, to evaluate the potential influence of residual confounding, we dichotomized this confounder at the median, which is referred to as minimally sufficient adjustment set with error, or 2e.

We ran each simulation 1000 times for samples sizes of 1000 and 300. We conducted simulations in which the target estimate was positive and near zero. We evaluated estimator performance by calculating average standard error (average of standard errors estimated for each simulation), MSE (average squared difference between reference values and simulated values), and 95% confidence interval coverage (proportion of confidence intervals that contain the reference value). The code necessary for recreating all simulations and results using R statistical software is provided in eAppendix 1 and eAppendix 2.

## RESULTS

All results describe simulations with a sample size of 1000. Simulations for a sample size of 300 provided no additional insight and, thus, are not presented. Table 1 summarizes results for simulations exploring the impact of a continuous exposure on both continuous and dichotomous outcomes. For linear and log–binomial models, estimates were approximately equal to the reference value, regardless of outcome prevalence. Performance of log–binomial regression improved with increasing outcome prevalence, as reflected by the reduced average standard error, MSE, and 95% confidence interval coverage. When applying logistic regression, estimates noticeably differed from the reference value. Average standard errors slightly improved when outcome prevalence was increased. Interestingly, with logistic regression and continuous exposure, confidence interval coverage improved with decreasing outcome prevalence, even as both average standard error and MSE increased. Inducing residual confounding in the parameterization of confounder #8 (minimally sufficient adjustment set 2e) did not materially impact results.

Table 2 summarizes results for simulations exploring a dichotomous exposure with prevalence of 30% and a positive effect estimate. Linear, log–binomial, and inverse probability weighting models generally returned estimates that approximate the reference value. Conventional logistic regression models returned mean effect estimates that were notably biased away from the reference values across all simulations. Confidence interval coverage was closer to optimal for logistic regression compared to Table 1 simulations for a continuous exposure; however, we believe this is a consequence of increased average standard error rather than improved performance.

We further simulated scenarios where the outcome prevalence was reduced to 10% and where the reference value was near null. The performance of each estimator was largely consistent with Table 2 results, so these alternative scenarios are presented as in eAppendix 3.

## DISCUSSION

We explored the ability of common statistical estimators to provide similar quantitative estimates of effect when estimands were different and non-null, but theoretically equally unbiased in terms of control for confounding. We found that linear, log–binomial, and inverse probability weighted estimators obtained estimates that were similar across different minimally sufficient adjustment sets. Conventional logistic regression was unsuccessful in returning comparable estimates for positive effect estimates. However, for null effect estimates, all estimators performed relatively accurately.

It is, perhaps, not surprising that the logistic regression estimates for different estimands were not quantitatively similar. Others have demonstrated that the conditional odds ratio is a non-collapsible quantity; that is, the marginal estimate is not equal to a weighted average of conditional estimates.<sup>8,10,11</sup> Our findings further illustrate the distinction between principles of confounding and non-collapsibility. In other words, it cannot be assumed that odds ratio values from individual studies subject to a meta-analysis differ as a result of random error or

residual confounding. Rather, these quantities are fundamentally different, regardless of adequate confounder control.

## CONCLUSION

We conclude that collaborative research should not rely on conventional covariate-adjusted (conditional) logistic regression when combining information across cohorts.<sup>12</sup> Rather, after articulating a research question, investigators should choose among alternative estimators, including inverse probability weighted models which, thanks to advances in statistical software, are more accessible.

Meta-analysts may assume that the same set of covariates must be available in all individual cohorts, and as a result, reduce the set of required covariates to accommodate more studies or exclude cohorts whose covariate sets do not match. What this study demonstrates is that it is *not* necessary for individual studies to measure the same covariates in most cases. Instead, a carefully constructed DAG may identify multiple sufficient confounder adjustment sets and thereby allow meta-analyses to include more studies while maintaining strong confounder control, so long as researchers avoid a conditional estimator.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

The authors wish to thank our ECHO colleagues, the medical, nursing and program staff, as well as the children and families participating in the ECHO cohorts. We also acknowledge the contribution of the following ECHO program collaborators:

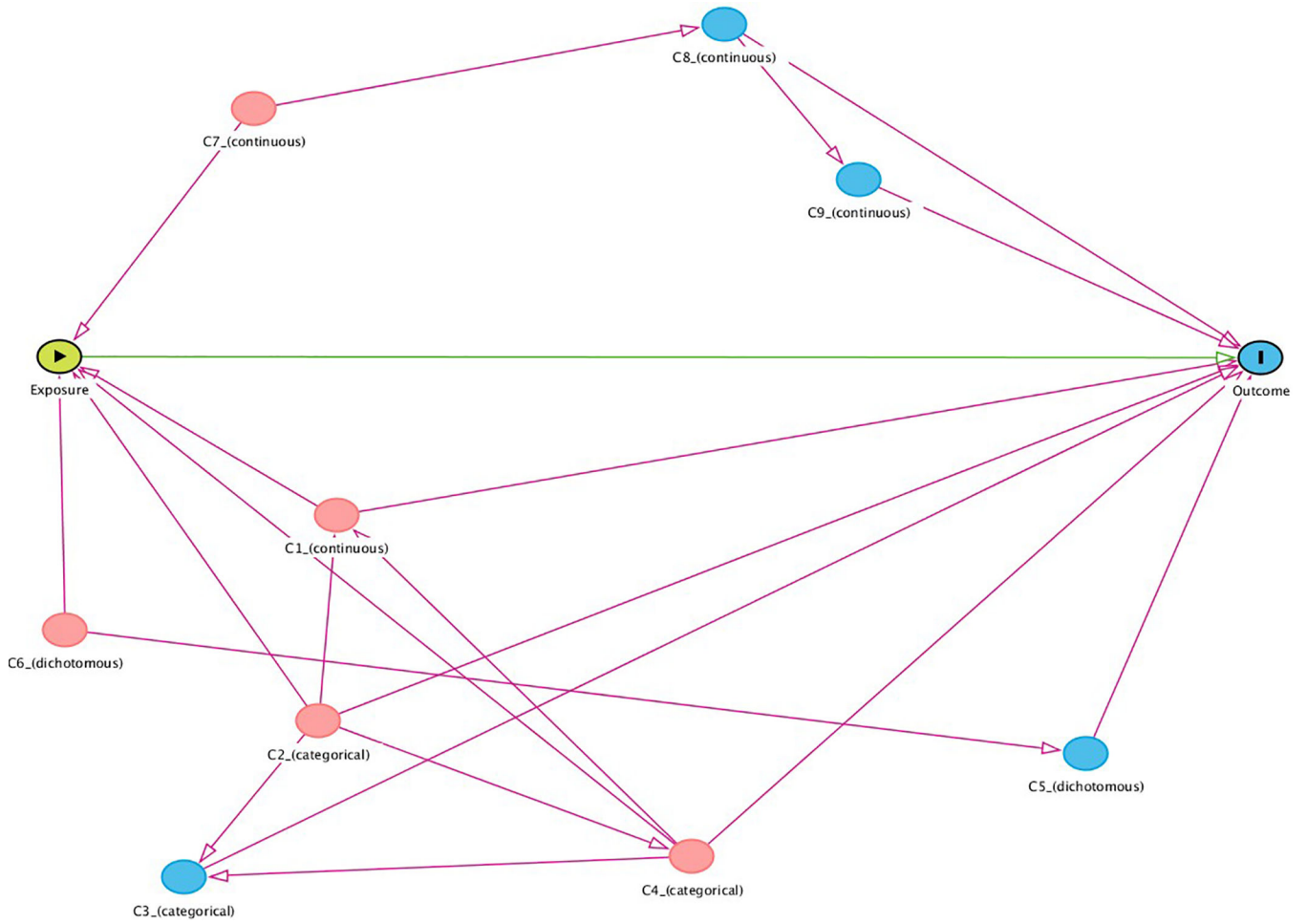
**ECHO Coordinating Center:** *Duke Clinical Research Institute, Durham, North Carolina:* Smith PB, Newby KL, Benjamin DK

Research reported in this publication was supported by the Environmental influences on Child Health Outcomes (ECHO) program, Office of The Director, National Institutes of Health, under Award Numbers U2COD023375 (Coordinating Center), U24OD023382 (Data Analysis Center), UH3OD023348-04 (Environment, epigenetics, neurodevelopment and health of extremely preterm children), and UG3OD023365 (Pre-adolescent and Late-adolescent Follow-up of the CHARGE Study Children). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

1. Lesko CR, Jacobson LP, Althoff KN, et al. Collaborative, pooled and harmonized study designs for epidemiologic research: challenges and opportunities. *Int J Epidemiol.* 2018;47(2):654–668. [PubMed: 29438495]
2. Voerman E, Santos S, Patro Golab B, et al. Maternal body mass index, gestational weight gain, and the risk of overweight and obesity across childhood: An individual participant data meta-analysis. *PLoS Med.* 2019;16(2):e1002744. [PubMed: 30742624]
3. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol.* 2008;8:70. [PubMed: 18973665]
4. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol.* 2002;155(2):176–184. [PubMed: 11790682]

5. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37–48. [PubMed: 9888278]
6. Neuhaus JM, Jewell NP. A Geometric Approach to Assess Bias Due to Omitted Covariates in Generalized Linear Models. *Biometrika*. 1993;80(4):807–815.
7. Hamra GB, Kaufman JS, Vahratian A. Model Averaging for Improving Inference from Causal Diagrams. *Int J Environ Res Public Health*. 2015;12(8):9391–9407. [PubMed: 26270672]
8. Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. *Epidemiol Perspect Innov*. 2009;6:4. [PubMed: 19732410]
9. Zou G A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159(7):702–706. [PubMed: 15033648]
10. Hernán MA, Clayton D, Keiding N. The Simpson’s paradox unraveled. *Int J Epidemiol*. 2011;40(3):780–785. [PubMed: 21454324]
11. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science*. 1999;14:29–46.
12. Robinson LD, Jewell NP. Some Surprising Results about Covariate Adjustment in Logistic Regression Models. *International Statistical Review / Revue Internationale de Statistique*. 1991;59(2):227–240.



**Figure.** Directed acyclic graph on which simulated data are based. This DAG was constructed with DAGitty online software (<http://www.dagitty.net/>).



Table 1.

Simulations Based on a Continuous Exposure.

Model	MSAS	Reference value	Mean	Outcome prevalence: 30%				Outcome prevalence: 10%					
				Average standard error	MSE	95% CI coverage	Reference value	Mean	Average standard error	MSE	95% CI coverage		
Linear	1	0.300	0.302	0.038	0.001	95%							
	2		0.301	0.035	0.001	95%							
	2e		0.299	0.036	0.001	96%							
Logistic	1	0.519	0.416	0.076	0.016	70%	0.559	0.471	0.110	0.020	87%		
	2		0.465	0.080	0.009	89%		0.516	0.116	0.015	93%		
	2e		0.445	0.077	0.011	83%		0.494	0.114	0.017	91%		
Log-binomial	1	0.242	0.242	0.042	0.002	95%	0.366	0.371	0.084	0.007	95%		
	2		0.243	0.040	0.002	95%		0.374	0.083	0.007	94%		
	2e		0.243	0.040	0.002	95%		0.370	0.083	0.007	94%		

Sample size = 1000. MSAS=minimally sufficient adjustment set; MSE=mean squared error; CI=confidence interval.

**Table 2.**

Simulations Based on a Binary Exposure: Exposure Prevalence 30%.

Model	MSAS	Reference value	Mean	Outcome prevalence: 30%				Outcome prevalence: 10%					
				Average standard error	MSE	95% CI coverage	Reference value	Mean	Average standard error	MSE	95% CI coverage		
Linear	1	0.300	0.302	0.087	0.008	96%							
	2		0.301	0.079	0.006	97%							
	2e		0.298	0.085	0.007	94%							
Logistic	1	0.514	0.407	0.157	0.036	92%	0.556	0.459	0.236	0.065	93%		
	2		0.459	0.170	0.032	95%		0.505	0.251	0.066	94%		
	2e		0.438	0.164	0.032	93%		0.470	0.237	0.063	94%		
Log-binomial	1	0.241	0.243	0.096	0.009	95%	0.384	0.382	0.200	0.040	94%		
	2		0.244	0.094	0.009	96%		0.385	0.198	0.039	94%		
	2e		0.242	0.093	0.009	95%		0.373	0.192	0.037	96%		
IPW	1	0.378	0.380	0.153	0.023	96%	0.438	0.436	0.226	0.051	95%		
	2		0.379	0.147	0.021	97%		0.433	0.222	0.049	96%		
	2e		0.378	0.146	0.021	97%		0.426	0.219	0.048	97%		

Sample size = 1000. IPW=inverse probability weighting. MSAS=minimally sufficient adjustment set; MSE=mean squared error; CI=confidence interval.