

**Current Trends in the Application of Causal Inference Methods to Pooled Longitudinal  
Observational Infectious Disease Studies - Data Extraction Form**

<b>Title of the paper/article/report</b>
<b>Study ID</b> <i>(surname of first author and year first full report of study was published e.g. Smith 2001)</i>
<b>Report IDs of other reports of this study</b> <i>(e.g. duplicate publications, follow-up studies)</i>
Notes:

**1. General Information** *(Example entries)*

<b>1. Date form completed</b> (dd/mm/yyyy)	15/04/2020
<b>2. Name/ID of person extracting data</b>	Heather/Saba/Ellie/Lauren
<b>3. Title of article (title of paper/abstract/report that data are extracted from)</b>	"Smoking and bmi in 21 twin birth cohorts"
<b>4. First author</b>	Donegan, L. M.
<b>5. Contact Details</b>	lisa.donegan@fake-uni.edu
<b>6. Publication type (e.g. full report, abstract, letter)</b>	Full report
<b>7. Possible conflicts of interest (for study authors)</b>	None claimed
Notes:	

**2. Eligibility to be included in this systematic review (not inclusion/exclusion criteria for the study participants)**

Study Characteristics	Copy and paste related descriptions as stated in report/paper	Location in text (pg and paragraph number/fig/table)
<b>8. Is the parent study composed of 2+ separate observational and longitudinal studies?</b> <i>If yes, include. if only single-site or multi-site, single-cohort, exclude; if includes RCTs, go to next question.</i>		
<b>9. Are any of the included studies RCTs?</b> <i>If yes, is the parent study using non-randomized variables from at least one RCT or observational cohort in the analysis? If yes, include. If not, exclude.</i>		
<b>10. Does the paper deal with infectious disease?</b> <i>If yes, include. If not, exclude.</i>		
<b>11. Are study subjects human?</b> <i>If animals, or human tissue samples or similar, exclude.</i>		
<b>12. Is it focused on health outcomes?</b> <i>If not, exclude. No wages, salaries, etc.</i>		
<b>13. Does the paper include case studies?</b> <i>If yes, exclude.</i>		
<b>14. Is it a methods paper with the primary goal of describing, developing, or summarizing a statistical analytical method, with no original analysis of observational, longitudinal data?</b> <i>If yes, exclude. If with an applied real-life example, include.</i>		

<b>15. Is the paper a protocol, review, commentary correction, editorial, erratum, or similar? <i>If yes, exclude</i></b>		
<b>16. Is the paper in English? <i>If not, exclude</i></b>		
<b>17. Is the data pooled on an individual level? <i>If not, exclude.</i></b>		
<b>18. Does the study attempt to establish a causal relationship? <i>E.g. by controlling for confounding variables or using an IV approach. If the goal of the study is descriptive, prediction, or prognostication, exclude. If unsure, flag as unsure.</i></b>		
<b>19. Is an effect size estimated?</b>		
<b>20. Is the effect size directly related to the causal question? <i>If the effect size does not correspond to the stated causal research question, exclude</i></b>		
<b>21. Is the study published in either 2009, 2014, or 2019? <i>If not, exclude.</i></b>		
<b>22. Decision to include</b>		
<b>23. Notes</b>		

**DO NOT PROCEED IF STUDY IS EXCLUDED FROM REVIEW**

### **3. Study characteristics**

	<b>Copy and paste related descriptions as stated in report/paper</b>	<b>Location in text (pg &amp; paragraph/fig/table)</b>
<b>24. List number &amp; types of <u>individual</u> studies or cohorts/clusters included in the pooled analysis <i>E.g. 3 cohort studies, 2</i></b>		

<i>case-control studies</i>		
<b>25. Study populations of each individual study or cohort included in the pooled analysis</b>		
<b>26. Number of participants in each individual study or cohort included in the pooled analysis</b>		
<b>27. Recruitment period of each individual study or cohort included in the pooled analysis</b>		
<b>28. Location of data collection of each individual study or cohort included in the pooled analysis</b>		
<b>Parent study</b>		
<b>29. Discipline of parent study?</b> (e.g. economics, medicine, sociology)		
<b>30. What exposures are studied in parent study?</b> <i>List which are randomized and which are non-randomized</i>		
<b>31. What are the outcomes in the parent study?</b> (e.g. myocardial infarction, hypertension, remission)		
<b>32. In which journal was the parent study published?</b>		
<b>33. Funding Source</b> <i>Copy &amp; paste funding section of manuscript here</i>		
<b>34. Key conclusions of study authors (of pooled study data not single studies).</b> <i>From abstract section</i>		
<b>35. Notes</b>		

#### 4. Methods and reporting standards

	Copy and paste related descriptions as stated in report/paper	Location in text (pg & paragraph/fig/table)
<b>36. Did they discuss issues such as difference in variable definitions, data quality, missing data?</b>		
<b>37. Did they take approach(es) to account for differences in variable definitions and data quality across individual cohorts?</b> <i>(e.g. any stated information about harmonization efforts (redefining variables) or statistical methods (adopting measurement error methods) studies/cohorts, etc), or did authors deal with missing data within and across studies? (e.g. multilevel multiple imputation, or separate imputation for each dataset, or complete case analysis)</i>		
<b>38. How did authors deal with missing data within and across studies?</b> <i>(e.g. multilevel multiple imputation, or separate imputation for each dataset, or complete case analysis)</i>		
<b>39. Do the authors report testing any of the assumptions required for the analysis methods they have chosen to pool the data? Which ones?</b> <i>Copy and paste relevant text describing the tests or reporting the results of those tests, if any, here. If not reported, write "not reported". If unclear, write "unclear".</i>		
<b>40a. Approach(es) to account for clustering/heterogeneity at the cohort or study level (whichever units are pooled across)</b> <i>(e.g. no adjustment,</i>		

<p><i>fixed effects (indicator variable for each cluster), random effects (random effect for each cluster), robust standard errors, etc). Did the authors adopt a one-stage or two-stage approach? (See e.g. Debray et al. PLOS ONE for a description of this)</i></p>		
<p><b>40b. Did the authors explicitly state assumptions for methods to account for clustering/heterogeneity?</b>  <i>(choice of fixed effects, random effects, ignoring of clustering: simply analyzing all data as if they come from one study)</i></p>		
<p><b>41a. Which causal methods were used with the pooled data to make causal inferences?</b> <i>(e.g. interrupted time series with a control group; comparative study without concurrent controls; instrumental variables; Mendelian randomization; regression discontinuity; interrupted time series, including difference-in-differences estimation; G-estimation; multiple regression adjusting for confounders; propensity score matching; inverse probability of treatment weighting; etc - this is not an exhaustive list)</i></p>		
<p><b>41b. Justification for method(s) used</b> <i>(e.g. “we selected a synthetic control approach because this method is well-suited to situations involving 1 intervention unit, and many controls and may better approximate counterfactual post-intervention outcomes than using any single control or an evenly weighted combination of controls” or “This approach is advantageous, because characteristics of each region, other than the occurrence of the treatment, are unlikely to change</i></p>		

<p><i>appreciably over so short a time period. Thus, each region serves as its own control, allowing us to control for other community-level characteristics that may also be associated with injuries.”)</i></p>		
<p><b>41c. Did the authors explicitly state the assumptions required for causal inference methods? Which ones?</b> (e.g. ignorability, positivity, stable unit treatment value, transitivity, “no unobserved confounding” ...)  <i>Copy and paste relevant text, if any, here. If not reported, write “not reported”. If unclear, write “unclear”</i></p>		
<p><b>41d. Do the authors report testing any of the assumptions required for the analysis methods they have chosen to deliver causal effects? Which ones?</b> <i>Copy and paste relevant text describing the tests or reporting the results of those tests, if any, here. If not reported, write “not reported”. If unclear, write “unclear”.</i></p>		
<p><b>41e. For untestable assumptions (e.g. unmeasured confounding), is there anything the authors do to evaluate the plausibility of those assumptions (e.g. negative control exposures or outcomes, quantitative bias analysis)? If yes, which ones?</b> <i>Copy and paste relevant text describing the tests or reporting the results of those tests, if any, here. If not reported, write “not reported”. If unclear, write “unclear”.</i></p>		

<b>42. Did the authors discuss heterogeneity of estimated causal effects and the possible impact on the generalizability of research findings?</b>		
--	--	--

<b>43. Notes</b>		
------------------	--	--