Journal of Clinical Epidemiology 145 (2022) 29–38

Journal of Clinical Epidemiology

REVIEW

Systematic Review Reveals Lack of Causal Methodology Applied to Pooled Longitudinal Observational Infectious Disease Studies

Heather Hufstedler[a,*], Sabahat Rahman[b], Alexander M. Danzer[c,d,e], Hannah Goymann[a], Valentijn M.T. de Jong[f,g], Harlan Campbell[h], Paul Gustafson[h], Thomas P.A. Debray[f,g], Thomas Jaenisch[a,i,j], Lauren Maxwell[a], Ellicott C. Matthay[k], Till Bärnighausen[a,l]

[a] *Heidelberg Institute of Global Health, Heidelberg Medical School, Heidelberg University, Germany*
[b] *University of Massachusetts Medical School, University of Massachusetts, Worcester, Massachusetts, USA*
[c] *KU Eichstätt-Ingolstadt, Ingolstadt School of Management and Economics (WFI), Germany*
[d] *IZA Bonn, Germany*
[e] *CESifo Munich, Germany*
[f] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*
[g] *Cochrane Netherlands, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*
[h] *Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada*
[i] *Center for Global Health, Colorado School of Public Health, Aurora, Colorado, USA*
[j] *Department of Epidemiology, Colorado School of Public Health, Aurora, USA*
[k] *Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA*
[l] *Harvard T H Chan School of Public Health, Harvard University, Boston, Massachusetts, USA*

Accepted 13 January 2022; Available online 16 January 2022

ABSTRACT

Objectives: Among ID studies seeking to make causal inferences and pooling individual-level longitudinal data from multiple infectious disease cohorts, we sought to assess what methods are being used, how those methods are being reported, and whether these factors have changed over time.

Study Design and Setting: Systematic review of longitudinal observational infectious disease studies pooling individual-level patient data from 2+ studies published in English in 2009, 2014, or 2019. This systematic review protocol is registered with PROSPERO (CRD42020204104).

Results: Our search yielded 1,462 unique articles. Of these, 16 were included in the final review. Our analysis showed a lack of causal inference methods and of clear reporting on methods and the required assumptions.

Conclusion: There are many approaches to causal inference which may help facilitate accurate inference in the presence of unmeasured and time-varying confounding. In observational ID studies leveraging pooled, longitudinal IPD, the absence of these causal inference methods and gaps in the reporting of key methodological considerations suggests there is ample opportunity to enhance the rigor and reporting of research in this field. Interdisciplinary collaborations between substantive and methodological experts would strengthen future work.

Keywords: Infectious disease; Methodological systematic review; Causal inference; Reporting; Individual participant data meta-analysis; Pooled data

Competing Interests: None declared.

* Corresponding author.
E-mail address: heather.hufstedler@uni-heidelberg.de (H. Hufstedler).

https://doi.org/10.1016/j.jclinepi.2022.01.008

What is new?

Key findings

- A growing number of studies focusing on infectious disease (ID) outcomes involve pooling individual participant data (IPD) across multiple studies or cohorts.
- ID-focused studies that used IPD did not apply causal methods or adhere to best practices for reporting study design and findings.

What this adds to what is known?

- This is the first review of the application of causal methods to IPD from 2+ ID cohorts.

What is the implication and what should change now?

- We propose the development and implementation of guidelines for transparent comprehensive reporting of harmonization and statistical methods applied to ID focused studies that pool IPD from two or more studies, including a) harmonization efforts, b) approaches to account for clustering and heterogeneity, c) approaches to account for missing data, d) approaches to account for data quality, e) justification of methods used, and f) explicit discussion of assessment of testable assumptions and evaluation of untestable assumptions.
- With no evidence of causal methods implemented in longitudinal, ID-focused studies that pool IPD from two or more studies, we recommend that researchers develop practical guidelines to facilitate the application of causal methods to ID-focused studies that pool and analyse IPD.

1. Introduction

Randomized control trials (RCTs) have long been considered the gold standard over observational designs when endeavouring to infer a causal relationship in medicine. However, observational studies have inherent benefits over RCTs, including the potential for improved external validity [1,2], the possibility of incorporating a vast amount of data more quickly, and the increased number of potential hypotheses that can be tested [3]. In certain instances, RCTs may not be feasible or ethical, as in public health emergencies or when considering adverse exposures, making observational designs not only beneficial but also necessary.

Huge advances in aetiology were made with Rubin's extension of counterfactuals to include observational studies [4], as well as Robins' extension that demonstrated time-varying confounding created by time-dependent covariates which can both be confounders and intermediate variables [5], and the introduction of directed acyclic graphs (DAGs) [6–8]. Of the many ways to establish causality, conventional regression-based adjustments are, perhaps, the most well-known and widely implemented. However, regression adjustment may not appropriately account for certain types of confounding in longitudinal analyses, as with time-dependent confounders impacted by prior treatment [9]. In addition to the risks of time-varying confounding, adjusting for confounders through conventional regression does not account for unmeasured confounding. Therefore, for many longitudinal ID cohorts, conventional regression adjustment may not be the most suitable analytic tool.

In the last 30+ years, study designs and methods have been developed to better account for time-varying confounding and/or unmeasured confounding, and are hereafter referred to as 'modern causal methods'. These methods to account for time-varying confounding include G-computation, inverse probability treatment weighting and parametric G-formulas (collectively known as G-methods) [10]. Methods and study designs to account for varying levels of unmeasured confounding include difference-in-differences (DiD) [11], regression discontinuity designs (RDD) [12], interrupted time series (ITS) [13], and instrumental variable (IV) approaches [14]. Prior studies suggest that modern causal methods are being utilized widely in the analysis of single ID studies [15,16].

The pooling of studies, first documented in 1904 by Pearson [17], is often achieved by combining published aggregate data (AD), though the pooling of individual participant data (IPD) has a number of statistical advantages over AD analyses [18,19]. Implementing modern causal methodologies with pooled IPD from several different cohorts differs importantly from multi-center cohort studies, and can be more challenging, particularly when accounting for differences in types of variables that are captured and the ways they are measured, more extreme heterogeneity in cohort enrolment and follow-up procedures, and data missing at both the study and the participant levels [20–22]. Though this is true, pooling IPD across cohorts can offer important benefits, including greater variability in exposure and outcome measures, thereby increasing statistical power to study rare diseases and exposures [20,23]. Access to IPD allows researchers to adjust for subject-level covariates and to investigate sources of between-study heterogeneity [23–25]. The application of modern causal inference methods to pooled longitudinal (2+ timepoints) observational data therefore has the potential to offer important insights into the causes and consequences of IDs [26]. Considering the COVID-19 pandemic, where the absence of systematic randomization of exposures or interventions can clearly be seen on a large scale, the saliency of these methods for ID-related studies becomes even more apparent.

Previous reviews of how modern causal methods are being applied include one study that reviewed the application of modern causal inference methods to deal with time-dependent confounding with non-randomized expo-

sure data from RCTs [9], and one study which examined methods used to adjust for unmeasured confounding in nonrandomized longitudinal studies [11]. Still, it is not well understood if modern causal methods are being used in ID studies pooling IPD: if so, whether they are being applied rigorously; if there are gaps in reporting or application of these methods; and if and/or how these factors have changed over time. To our knowledge, there exists no methodological review of the application and reporting of modern causal methods to pooled longitudinal (2+ timepoints), observational ID studies. Among ID studies seeking to make causal inferences with pooled individual-level longitudinal data from multiple cohorts, we sought to assess what methods are being used, how those methods are being reported, and whether these factors have changed over time.

2. Methods

2.1. Search strategy

Four researchers (HH, LM, EM, SR) developed a search strategy in collaboration with librarians from Universitätsklinikum Heidelberg (UKHD), University of California San Francisco (UCSF), and Harvard University. The search strategy combined the terms: 'cohort', 'longitudinal', and related terms; 'pool*,' 'aggregat*,' and 'harmoniz*'; terms that refer to IPD; and terms that indicate ID applications. A list of IDs were compiled from the CDC's list of IDs [27] and PubMed's MeSH terms for communicable diseases. These elements were combined to increase sensitivity in the search results. We excluded RCTs, animal studies, and other unrelated publication types, including grey literature (conference abstracts and dissertations, e.g.). The search did not select for specific types of modern causal methods. We restricted our search to English-language articles published online over the last decade (2009-2019) in five-year increments for increased screening feasibility (2009, 2014, 2019). The search strategy for each database can be found in Supplementary Material 1. We applied the search strategy to the following databases in October, 2020: EBSCO (including Academic Search Complete, Business Source Premier, CINAHL, EconLit, and PsycINFO), EMBASE, PubMed, and Web of Science. The protocol for the systematic review was registered with PROSPERO (CRD42020204104) and published prior to initiating the search and subsequently published [28].

Citations were exported to EndNote, deduplicated, and uploaded to Covidence systematic review software [29]. Two reviewers (HH, SR) screened articles using Covidence's double-blind screening tool, and a third reviewer (EM) resolved discrepancies using the following inclusion criteria: 1) used individual-level data, 2) pooled data from at least two studies with at least two timepoints; 3) observational studies focused on ID health outcomes, 4) reported estimates referring to a pre-stated causal question, and 5) had a full-text version accessible through open access, university license, another collaborator on the project, or sent to us by the authors. We also included studies that incorporated data from RCTs if at least one data source included in the analysis was drawn from an observational study or if the study included only RCTs but at least one exposure/treatment variable analysed was not randomized.

2.2. Causal inference survey with study authors

In order to see how our team's assessment aligned with the study authors' intention to infer causality regarding their research question and results, we sent a short survey to each study's authors. Survey questions included: 1) When you began your research, did you set out to infer a causal relationship?; 2) When you submitted your paper, did you intend to imply a causal relationship?; and 3) Do you expect any source of bias in your effect sizes?.

2.3. Data extraction

Two reviewers (HH, HG) extracted data, and one reviewer (AD) reviewed the extracted data for completeness. For each relevant paper, we assessed the following four criteria related to reporting of methods and assumptions.

Reporting of Methods

i if they discussed issues related to variable definitions, data quality, and missing data, and how they accounted for them

ii reporting of methods to account for clustering and heterogeneity

iii reporting of methods implemented to causally analyse the data, and

Discussion of Assumptions

i discussion of assumptions related to the estimation methods, selected modern causal methods (confounding-specific methods, e.g., ignorability, positivity; or model-specific assumptions) (see Data Extraction Form in Supplementary Materials 2)

To communicate the quality of the reporting within and across studies, we developed a point system with three values for each of the aforementioned criteria: if the item in question was clearly discussed, 1 point; if the description did not entirely address the item but touched on the topic, 0.5 points; if the topic was not mentioned, 0 points. When two reviewers disagreed on the criterion quality, we selected the higher score. Since we analysed 16 studies, the maximum score across all studies for each item was 16 points.

3. Results

3.1. Search results

The search retrieved 2,073 studies, 611 of which were duplicates. Of the 1,462 unique records, 1,433 were excluded during the title and/or abstract screening. At the

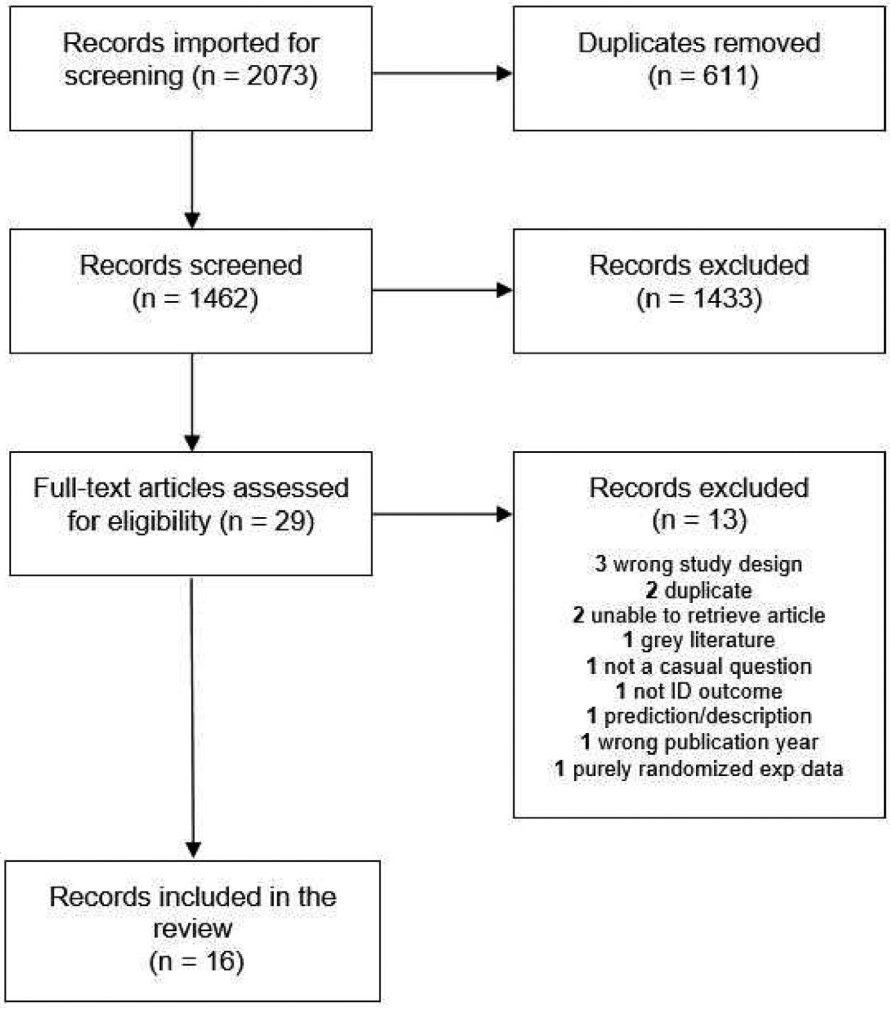

Fig. 1. Flow chart

full-text review stage, 13 were excluded: nine were excluded based on previously stated selection criteria; two were excluded because we could not access the articles; and two were duplicates. The final sample included 16 relevant studies [30–45] (see the Flow Chart in Fig. 1). The publication of pooled cohort analyses appears to be rapidly trending upwards, with two studies in 2009, four in 2014, and ten in 2019.

3.2. Study characteristics

Of the studies included, the IDs of interest were HIV (4/16), Influenza (3/16), Malaria (3/16), Tuberculosis (3/16), H. pylori Infection (1/16), HPV-2 (1/16), and Varicella Zoster (1/16). The number of studies included in each pooled analysis ranged from 2 to 81, with a median 22.4. The number of included countries that contributed data to the pooled analyses ranged from 1 to 39, with a median 12.6.

Study populations included adolescents, patients with multidrug-resistant (MDR) and/or extensively drug-resistant (XDR) tuberculosis, patients with uncomplicated *P.vivax* infection, and HIV-negative men who have sex with men (MSM), to name a few. Pooled studies predominantly included cohort studies (96), followed by case control (21), RCT (5), nested case control (2), and surveillance (2). Three studies did not report the breakdown of contributing study designs. See Supplementary Material 3 for a full characteristics table.

3.3. Study author survey results

Authors from nine of the sixteen articles responded to the survey. Most respondents indicated that they initially set out to infer a causal relationship (7/9) and intended to imply causality in their results (6/9). The only respondent whose intent changed after analysis stated that the shift was due to a negative result. All respondents said that they expected some level of bias to affect their effect sizes, with the most reported bias concerns potentially deriving from selection bias and unknown (unmeasured) factors.

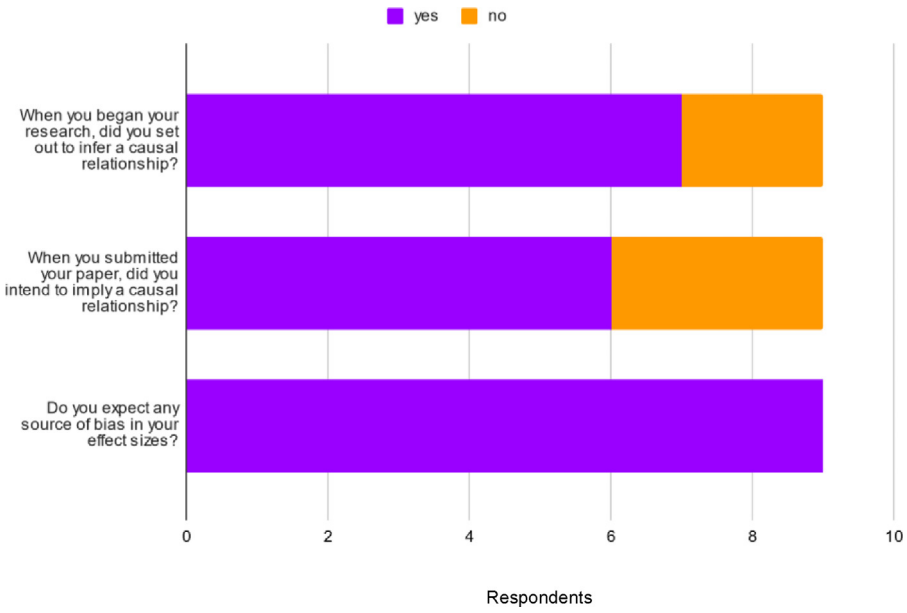

Fig. 2. Causality survey

3.4. Reporting of methods: variable definitions, data quality, missing data, account for clustering and heterogeneity, and causal analysis

In reviewing the methods that authors used in their analysis to produce what we identified as causal claims, none of the studies used the modern causal methods we were looking for; 15 implemented traditional regression-based adjustment, and one used bivariate statistical tests. For describing the methods implemented to account for variations in variable definitions, 12 studies were awarded a full point and one study received a half point (for a total of 12.5 across studies). Four studies received a full-point for addressing their approaches to dealing with data quality (methods included cross-tabulation; intercepts and slopes which were allowed to vary; efforts to decrease miscategorisation, and a sensitivity analysis to test data quality while extreme values were either translated to missing or accepted as they were).

In evaluating methods reporting across studies, 8 points were given to studies for describing how they dealt with missing data across studies: excluding cases (6/16), single imputation (2/16), multiple imputation (1/16); of these, one study described a combination of methods: single imputation and excluding studies. The majority (13/16) of studies discussed their approaches to account for clustering or heterogeneity, utilizing either random effects models (7/16) or stratification (5/16). Three studies provided justification for the study design and statistical methods chosen to analyze their pooled data, with nearly all studies opting for covariate adjustment over weighting or matching. Only half (8/16) of the studies discussed the impact of heterogeneity on the generalizability of results, and one study was given one half point for very loosely touching on the subject, for a total of 8.5 points.

3.5. Reporting of assumptions for methods and justification of methods

Four studies received full points for describing the testing of assumptions related to the estimation methods— all used Schoenfeld residuals (to test for the proportional hazards assumption). Only one full point was awarded to one study that clearly discussed the assumptions required for the methods they chose to analyse the data—no uncontrolled confounding.

Two studies received a full point for reporting that they tested any testable assumption(s) required for the analysis methods they selected; one of the studies receiving this point did not receive a point for the item in the previous sentence ‘clearly discussing assumptions for analysis methods’ because they did not state or discuss the assumption but only the test they used. We surmised they meant to test this assumption and therefore awarded them a full-point here based on the generosity of the rating system.

Six and a half points were awarded to studies for discussing the plausibility of untestable assumptions: full points were awarded to 5 studies, and three studies were given 0.5 points for loosely discussing the issue. See Fig. 4 for Reporting of Assumptions analysis.

3.6. Summary of reporting

Despite the generosity of the rating system, only four of 10 items were reported in at least 70% of the 16 studies

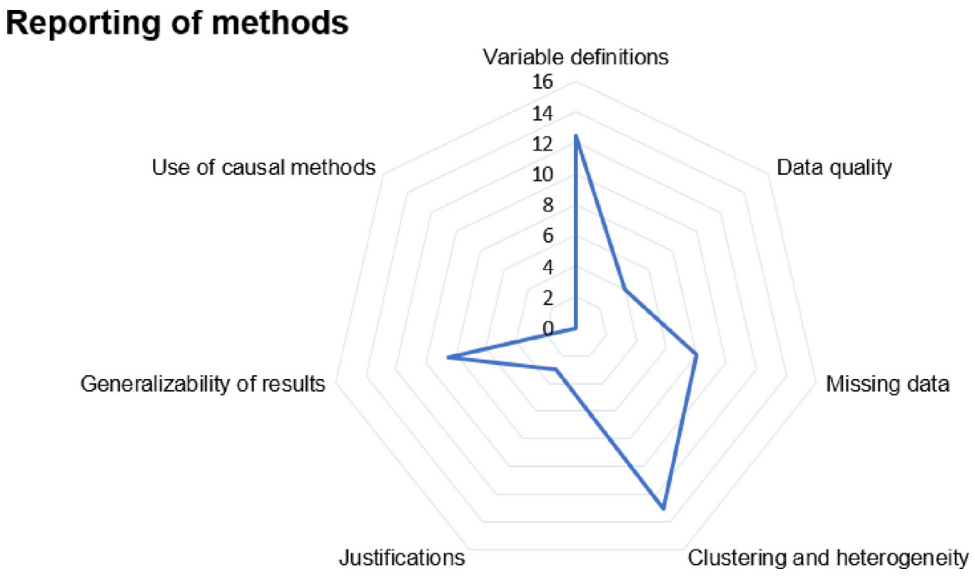

Fig. 3. Extent of reporting on methods and generalizability, according to point system

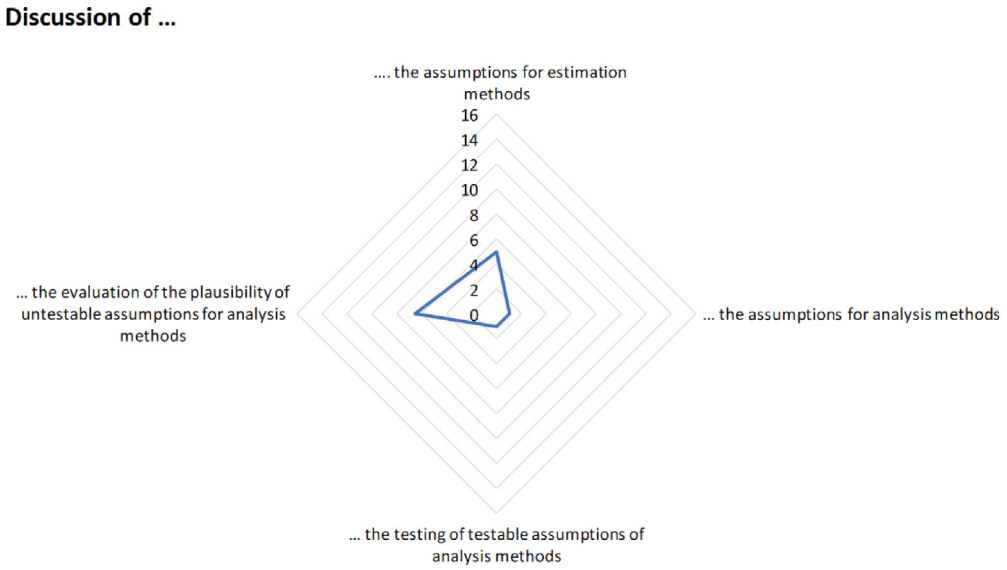

Fig. 4. Reporting of assumptions

(Figs. 3 and 4). If broken down by year, the average points awarded to studies for reporting these items were 4.5 of 10 for 2009, 6.3 of 10 for 2014, and 4.0 of 10 for 2019, indicating no consistent improvement over time. It is also worth noting that not a single study received a full 10 points. See Fig. 5 for a listing of reported methods and assumptions by year.

4. Discussion

4.1. Increased pooled analyses, transparent harmonization, and clustering and/or heterogeneity

Results of this systematic review show that the conducting of ID studies where data are pooled at the participant level have increased over time, and that included studies scored high on their reporting of harmonization/standardization methods (78%) and their approaches to account for clustering/heterogeneity (81%). With the recent publication of guidelines to retrospective harmonization [46] and the many guidance documents on best practices in accounting for clustering/heterogeneity [47,48], it is our hope that all future studies will implement rigorous harmonization efforts, and wisely select and describe their approaches to control for clustering and heterogeneity in pooled cohort data. Transparent reporting of methods for harmonization and accounting for cross-study heterogeneity methods will assist the reader to assess the rigor with which the analysis was conducted.

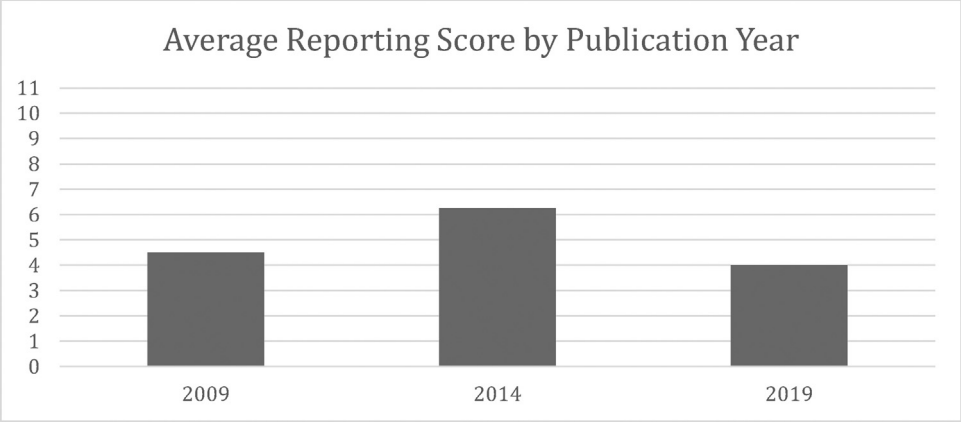

Fig. 5. Average reporting by publication year

4.2. Lack of causal methodology

The primary goal of this review was to examine the recent trends of the application of causal methods to ID-related analyses that pooled IPD across studies. Despite a rigorously-developed search strategy, the results of the review were unvaried, as no study identified in this review implemented modern causal methods. Yet, survey respondents frequently indicated concerns about unmeasured confounding. Because the results did not reveal *any* pooled study utilizing causal methods, no comparison was possible.

In single cohort ID studies, causal analyses of longitudinal data have been widely adopted (particularly in HIV research)[49,50] and have helped to rule out specific threats to internal validity, relaxed the assumptions required by the analytic approach, and enhanced understanding of causal effects. Pooled observational studies (i.e., multi-cohort analyses) also stand to benefit from this diversification of methodology. All methods involve tradeoffs and untestable assumptions [51] Thus, studies with differing methodologies cover each other's weaknesses, and a study applying a novel methodology may yield more scientific benefit than a study applying an existing method that relies on the same untestable assumptions as prior research.

Barriers to implementation must be evaluated, and may include: the complexity of implementing modern causal inference methods; the need for analytic tools or implementation guides specifically tailored to pooled studies [52–54]; or a process of methods selection in which multiple research teams must understand and agree upon a single analytic strategy. Interdisciplinary collaborations between statistical methodologists and ID-focused researchers through a coalition or consortium, as seen in ReCoDID (www.recodid.eu), could break down these barriers and promote innovation.

4.3. Reporting applied methods and assumptions

Any causal claim requires a 'defense' of assumptions and justifications, whether one utilizes regression-based adjustment or *modern* causal methods. However, authors who responded that they set out to infer a causal relationship at the beginning of their study were not more likely to meet the four criteria for reporting methods and assumptions than those who did not intend to imply a causal relationship at the beginning of their study. There was also a lack of reporting on the rationale for selecting regression-based adjustment methods over modern causal methodologies. Furthermore, most key methodological points that we felt were important for pooled longitudinal (2+ timepoints) observational studies to report on were lacking, including methods used to account for missing data or data quality, justification for methods used, and explicit discussion of the testing of testable assumptions and the evaluation of untestable assumptions.

Large variance in reporting is not a rare finding [55] Currently, there are no overarching reporting guidelines in the field when it comes to pooled IPD across studies. Even the PRISMA IPD Statement [56], which is only concerned with the reporting of IPD-MAs, does not include specific recommendations for reporting the quantitative assessment of testable model assumptions or concerns specific to longitudinal, observational studies– ID-related or otherwise; these include recommended sensitivity or statistical approaches to addressing measured confounders affected by prior exposure in the context of cross-cohort analyses, amongst myriad other causal inference-related concerns. To better enable readers inside and outside the ID field to determine the rigor of these pooled studies, we recommend that standards in reporting of analytical methods be developed, and that future IPD ID studies report on a) harmonization efforts, b) approaches to account for clustering and heterogeneity, c) approaches to account

for (differences in) data quality (e.g., missing data, measurement error), e) justification for methods used, and f) explicit discussion of testable assumptions and untestable assumptions. Given that the average points awarded to articles published in 2019 (4 points) were lower than the average of those published in 2014 (6.3 points), it is our hope that researchers do not merely increase the quantity of IPD meta-analyses but consistently improve the quality of reporting of such studies, as well.

4.4. Strengths and limitations

Strengths of this systematic review were the rigor applied in following best practice for the development and execution of systematic reviews, including the development of a comprehensive, tailored search strategy and use of Covidence's blinded review. One possible limitation is the small number of papers we found. Another limitation is that our study only investigated the quality of *reporting,* but could not determine the quality of *conduct* (including appropriateness of statistical methods, validity of model assumptions, etc.). Such an evaluation would ideally require access to patient-level data from the included pooled cohort studies, which is rarely straightforward.

5. Conclusions

Based on the small number of pooled ID-focused IPD analyses and the lack of modern causal method implementation found in this systematic review, we recommend increased interdisciplinary collaborations between statistical methodologists and ID-focused researchers. We also suggest the development of guidance documents related to the implementation and reporting of causal methods in the analysis of pooled participant-level data from longitudinal observational ID-focused studies which could be similar to the guidance produced for a recent series on quasi-experimental methods in health [57–69].

Ethical statement

This study did not require any ethical approval because we used only publicly available data.

Funding

This work is supported by the ReCoDID study, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 825746 and the Canadian Institutes of Health Research, Institute of Genetics (CIHR-IG) under Grant Agreement N.01886-000.

Acknowledgments

We would like to acknowledge Leonard Levin, the associate director of research, instruction, and collection resources at the Francis A. Countway Library of Medicine at Harvard Medical School, for his time and effort in this work. We would also like to sincerely thank all of the authors, research assistants and data managers who responded to our survey. And special thanks to the editors and reviewers for their valuable comments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2022.01.008.

REFERENCES

[1] Cartwright N. Are RCTs the Gold Standard? BioSocieties 2007;2(1):11–20. doi:10.1017/s1745855207005029.
[2] Cartwright N, Munro E. The limitations of randomized controlled trials in predicting effectiveness. J Eval Clin Prac 2010;16(2):260–6. doi:10.1111/j.1365-2753.2010.01382.x.
[3] Port FK. Role of observational studies versus clinical trials in ESRD research. Kidney Internat, Suppl 2000;57(74):3–6. doi:10.1046/j.1523-1755.2000.07402.x.
[4] Rubin DB. Estimating causal effects of treatment in randomized and nonrandomized studies. J Edu Psycholy 1974;66(5):688–701.
[5] Robins M, Hernan A. Marginal Structural Models and Causal Inference in Epidemiology.:550-560.
[6] Pearl Judea. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, CA, USA: Morgan Kaufmann Publishers; 1988.
[7] Pearl J. Causal Diagrams for Empirical Research. Biometrika 1995;82(4):669–88. doi:10.2307/2337329.
[8] Pearl J, Robins JM. Probabilistic evaluation of sequential plans from causal models with hidden variables. Uncertainty in Artificial Intelligence. Published online 1995:444-453.
[9] Farmer RE, Kounali D, Walker AS, et al. Application of causal inference methods in the analyses of randomized controlled trials: a systematic review. Trials 2018;19(1):1–14. doi:10.1186/s13063-017-2381-x.
[10] Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research interest are time varying, and the. Published online 2017:1-6. doi:10.1136/bmj.j4587
[11] Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodologi-

cal review. J Clin Epidemiol 2017;87:23–34. doi:10.1016/j.jclinepi.2017.04.022.
[12] Thistlewaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex-post facto experiment. J Edu Psychol 1960;51:309–17.
[13] Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomization is not an option: Interrupted time series analysis. BMJ (Online) 2015;350:1–4. doi:10.1136/bmj.h2750.
[14] Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Statistics in Medicine 2014;33(13):2297–340. doi:10.1002/sim.6128.
[15] Sasikumar M, Boyer S, Remacle-Bonnet A, Ventelou B, Brouqui P. The value of specialist care—infectious disease specialist referrals—why and for whom? A retrospective cohort study in a French tertiary hospital. Euro J Clin Microbiol and Infectious Dis 2017;36(4):625–33. doi:10.1007/s10096-016-2838-y.
[16] Quintero J, Pulido NR, Logan J, Ant T, Bruce J, Carrasquilla G. Effectiveness of an intervention for Aedes aegypti control scaled-up under an intersectoral approach in a Colombian city hyperendemic for dengue virus. PLoS ONE 2020;15(4):1–16. doi:10.1371/journal.pone.0230486.
[17] Pearson K. Report on Certain Enteric Fever Inoculation Statistics. The British Med Jl 1904;2(2288):1243–6.
[18] Tierney JF, Fisher DJ, Burdett S, Stewart LA, Parmar MKB. Comparison of aggregate and individual participant data approaches to meta-analysis of randomised trials: An observational study. PLoS Med 2020;17(1):1–22. doi:10.1371/JOURNAL.PMED.1003019.
[19] Tudor Smith C, Marcucci M, Nolan S, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data (Review). Published online 2016. doi:10.1002/14651858. Available at: MR000007.pub3.www.cochranelibrary.com
[20] Lesko CR, Jacobson LP, Althoff KN, et al. Collaborative, pooled and harmonized study designs for epidemiologic research : challenges and opportunities. 2018:654-668. doi:10.1093/ije/dyx283
[21] Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. Statist Methods Med Res 2018;27(6):1634–49 doi:. doi:10.1177/0962280216666564.
[22] Audigier V, White I, Jolani S, et al. Multiple imputation for multilevel data with continuous and binary variables. Statistical Sci 2018;3(2):160–83.
[23] Thompson A. Thinking big: Large-scale collaborative research in observational epidemiology. Euro J Epidemiol 2009;24(12):727–31. doi:10.1007/s10654-009-9412-1.
[24] Debray TPA, Moons KGM, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods 2015;6(4):293–309. doi:10.1002/jrsm.1160.
[25] Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: Rationale, conduct, and reporting. BMJ (Online) 2010;340(7745):521–5. doi:10.1136/bmj.c221.
[26] Halloran ME, Struchiner CJ. Causal inference in infectious diseases. Epidemiology 1995;6(2):142–51. doi:10.1097/00001648-199503000-00010.
[27] Prevention C for DC and. 2020 National Notifiable Infectious Diseases. Presented at: 2020.
[28] Hufstedler H, Matthay EC, Rahman S, et al. Current trends in the application of causal inference methods to pooled longitudinal observational infectious disease studies—A protocol for a methodological systematic review. PLoS ONE 2021;16(4 April):1–7. doi:10.1371/journal.pone.0250778.
[29] Covidence systematic review software.
[30] Martinez De Tejada B. Birth defects after exposure to efavirenz-based antiretroviral therapy at conception/first trimester of pregnancy: a multicohort analysis. J Acq Immune Defic Synd 2019;80(3):316–24. doi:10.1097/QAI.0000000000001922.
[31] ET M, Krantz E, SL G, et al. A pooled analysis of the effect of condoms in preventing HSV-2 acquisition. Arch Intern Med 2009;169(13 PG-1233-40):1233–40. doi:10.1001/archinternmed.2009.177.
[32] Chung JR, Flannery B, Ambrose CS, et al. Live attenuated and inactivated influenza vaccine effectiveness. Pediatrics 2019;143(2). doi:10.1542/peds.2018-2094.
[33] Ferro A, Morais S, Pelucchi C, et al. Sex differences in the prevalence of Helicobacter pylori infection: an individual participant data pooled analysis (StoP Project). Euro J Gastroenterol and Hepatol 2019;31(5):593–8. doi:10.1097/MEG.0000000000001389.
[34] Khaksarian M, Masoumi F, Ahmadi M, Ahmadi SAY, Taherikalani M. Accompaniment of multiple sclerosis with varicella zoster virus; a systematic review and individual participant data meta-analysis. Neurology Asia 2019;24(2 PG-147-156):147–156.
[35] Commons RJ, Simpson JA, Thriemer K, et al. The hematological consequences of Plasmodium vivax malaria after chloroquine treatment with and without primaquine: A World Wide Antimalarial Resistance Network systematic review and individual patient data meta-analysis. BMC Medicine 2019;17(1):1–13. doi:10.1186/s12916-019-1386-6.
[36] Toska E, Campeau L, Cluver L, et al. Consistent provisions mitigate exposure to sexual risk and hiv among young adolescents in South Africa. AIDS and Behav 2019(PG-). doi:10.1007/s10461-019-02735-x.
[37] ML B, Hussain H, Weyer K, et al. Treatment outcomes of patients with multidrug-resistant and extensively drug-resistant tuberculosis according to drug susceptibility testing to first- and second-line drugs: an individual patient data meta-analysis. Clin Infect Dis 2014;59(10 PG-1364-74):1364–74. doi:10.1093/cid/ciu619.
[38] Mertz D, Lo CKF, Lytvyn L, et al. Pregnancy as a risk factor for severe influenza infection: an individual participant data meta-analysis. Bmc Infectious Diseases 2019;19(1 PG-). doi:10.1186/s12879-019-4318-3.
[39] Chiappini E, Ene L, Galli L, et al. Severe haematologic toxicity is rare in high risk HIV-exposed infants receiving combination neonatal prophylaxis. HIV Medi 2019;20(5):291–307. doi:10.1111/hiv.12696.
[40] RJ C JA S, Thriemer K, et al. The efficacy of dihydroartemisinin-piperaquine and artemether-lumefantrine with and without primaquine on Plasmodium vivax recurrence: a systematic review and individual patient data meta-analysis. PLoS Med 2019;16(10 PG-e1002928):e1002928. doi:10.1371/journal.pmed.1002928.
[41] Venkatesan M, NB G, Stepniewska K, et al. Polymorphisms in Plasmodium falciparum chloroquine resistance transporter and multidrug resistance 1 genes: parasite risk factors that affect treatment outcomes for P. falciparum malaria after artemether-lumefantrine and artesunate-amodiaquine. Am J Trop Med Hyg 2014;91(4 PG-833-843):833–43. doi:10.4269/ajtmh.14-0031.
[42] Scott HM, Vittinghoff E, Irvin R, et al. Age, race/ethnicity, and behavioral risk factors associated with per contact risk of HIV infection among men who have sex with men in the united states. J Acq Immune Deficiency Synd 2014;65(1):115–21. doi:10.1097/QAI.0b013e3182a98bae.
[43] SG M, Venkatesan S, PR M, et al. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. Lancet Respir Med 2014;2(5 PG-395-404):395–404. doi:10.1016/S2213-2600(14)70041-4.
[44] JE H, MK G, Meinke L, EB K, SJ R. Tuberculosis in liver transplant recipients: a systematic review and meta-analysis of individual patient data. Liver Transpl 2009;15(8 PG-894-906):894–906. doi:10.1002/lt.21709.
[45] Aibana O, CC H, Aboud S, et al. Vitamin D status and risk of incident tuberculosis disease: a nested case-control study, systematic review, and individual-participant data meta-analysis. PLoS

Med 2019;16(9 PG-e1002907):e1002907. doi:10.1371/journal.pmed.1002907.
[46] Fortier I, Raina P, Heuvel ER Van Den, et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. 2017;(June 2016):103-115. doi:10.1093/ije/dyw075
[47] de Jong VMT, Moons KGM, Riley RD, et al. Individual participant data meta-analysis of intervention studies with time-to-event outcomes: A review of the methodology and an applied example. Res Synth Methods 2020;11(2):148–68. doi:10.1002/jrsm.1384.
[48] Abo-Zaid G, Guo B, Deeks JJ, et al. Individual participant data meta-analyses should not ignore clustering. J Clin Epidemiol 2013;66(8):865–73 e4. doi:10.1016/j.jclinepi.2012.12.017.
[49] Clare PJ, Dobbins TA, Mattick RP. Causal models adjusting for time-varying confounding — a systematic review of the literature. 2019;(October 2018):254-265. doi:10.1093/ije/dyy218
[50] Lange T, Rod NH. Commentary: Causal models adjusting for time-varying confounding -please send more data. Intern J Epidemiol 2019;48(1):265–7. doi:10.1093/ije/dyy300.
[51] Matthay EC, Hagan E, Gottlieb LM, et al. Alternative causal inference methods in population health research: Evaluating tradeoffs and triangulating evidence. SSM - Population Health 2020;10:100526. doi:10.1016/j.ssmph.2019.100526.
[52] Debray TPA, Moons KGM, Abo-Zaid GMA, Koffijberg H, da Riley R. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? PLoS ONE 2013;8(4). doi:10.1371/journal.pone.0060650.
[53] Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. Statistics in Med 2017;36(5):855–75. doi:10.1002/sim.7141.
[54] Stewart GB, Altman DG, Askie LM, Duley L, Simmonds MC, Stewart LA. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. PLoS ONE 2012;7(10). doi:10.1371/journal.pone.0046042.
[55] Oldenburg CE, Moscoe E, Bärnighausen T. Regression discontinuity for causal effect estimation in epidemiology. Current Epidemiol Reports 2016;3(3):233–41. doi:10.1007/s40471-016-0080-x.
[56] Page MJ, Mckenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement : an updated guideline for reporting systematic reviews. J Clin Epidemiol 2021(xxxx):1–12. doi:10.1016/j.jclinepi.2021.03.001.
[57] Bärnighausen T, Røttingen JA, Rockers P, Shemilt I, Tugwell P. Quasi-experimental study designs series—paper 1: introduction: two historical lineages. J Clin Epidemiol 2017;89:4–11. doi:10.1016/j.jclinepi.2017.02.020.
[58] Geldsetzer P, Fawzi W. Quasi-experimental study designs series-paper 2: complementary approaches to advancing global health knowledge. J Clin Epidemiol 2017;89:12–16. doi:10.1016/j.jclinepi.2017.03.015.
[59] Bärnighausen T, Tugwell P, Røttingen JA, et al. Quasi-experimental study designs series—paper 4: uses and value. J Clin Epidemiol 2017;89:21–9. doi:10.1016/j.jclinepi.2017.03.012.
[60] Frenk J, Gómez-Dantés O. Quasi-experimental study designs series-paper 3: systematic generation of evidence through public policy evaluation. J Clin Epidemiol 2017;89:17–20. doi:10.1016/j.jclinepi.2017.03.013.
[61] Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels. J Clin Epidemiol 2017;89:30–42. doi:10.1016/j.jclinepi.2017.02.016.
[62] Waddington H, Aloe AM, Becker BJ, et al. Quasi-experimental study designs series—paper 6: risk of bias assessment. J Clin Epidemiol 2017;89:43–52. doi:10.1016/j.jclinepi.2017.02.015.
[63] Bärnighausen T, Oldenburg C, Tugwell P, et al. Quasi-experimental study designs series—paper 7: assessing the assumptions. J Clin Epidemiol 2017;89:53–66. doi:10.1016/j.jclinepi.2017.02.017.
[64] Glanville J, Eyers J, Jones AM, et al. Quasi-experimental study designs series—paper 8: identifying quasi-experimental studies to inform systematic reviews. J Clin Epidemiol 2017;89:67–76. doi:10.1016/j.jclinepi.2017.02.018.
[65] Aloe AM, Becker BJ, Duvendack M, Valentine JC, Shemilt I, Waddington H. Quasi-experimental study designs series—paper 9: collecting data from quasi-experimental studies. J Clin Epidemiol 2017;89:77–83. doi:10.1016/j.jclinepi.2017.02.013.
[66] Becker BJ, Aloe AM, Duvendack M, et al. Quasi-experimental study designs series—paper 10: synthesizing evidence for effects collected from quasi-experimental studies presents surmountable challenges. J Clin Epidemiol 2017;89(1):84–91. doi:10.1016/j.jclinepi.2017.02.014.
[67] Lavis JN, Bärnighausen T, El-Jardali F. Quasi-experimental study designs series—paper 11: supporting the production and use of health systems research syntheses that draw on quasi-experimental study designs. J Clin Epidemiol 2017;89:92–7. doi:10.1016/j.jclinepi.2017.03.014.
[68] Rockers PC, Tugwell P, Grimshaw J, et al. Quasi-experimental study designs series–paper 12: strengthening global capacity for evidence synthesis of quasi-experimental health systems research. J Clin Epidemiol 2017;89:98–105. doi:10.1016/j.jclinepi.2016.03.034.
[69] Rockers PC, Tugwell P, Røttingen JA, Bärnighausen T. Quasi-experimental study designs series—paper 13: realizing the full potential of quasi-experiments for health research. J Clin Epidemiol 2017;89:106–10. doi:10.1016/j.jclinepi.2017.03.016.