PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2014; **23**: 165–177
Published online 5 December 2013 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3555

ORIGINAL REPORT

Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study[†]

Md. Jamal Uddin[1], Rolf H. H. Groenwold[1,2], Anthonius de Boer[1], Svetlana V. Belitser[1], Kit C. B. Roes[2], Arno W. Hoes[2] and Olaf H. Klungel[1,2]*

[1] *Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands*
[2] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands*

ABSTRACT

Purpose Instrumental variable (IV) analysis is becoming increasingly popular to adjust for confounding in observational pharmacoepidemiologic research. One of the prerequisites of an IV is that it is strongly associated with exposure; if it is weakly associated with exposure, IV estimates are reported to be biased. We aimed to assess the performance of IV estimates in various (pharmaco-) epidemiologic settings.

Methods Data were simulated for continuous/binary exposure, outcome and IV in cohort and nested case–control (NCC) designs with different incidences of the outcome. Pearson's correlation, point bi-serial correlation, odds ratio (OR), and F-statistic were used to assess the IV-exposure association. Two-stage analysis was performed to estimate the exposure effect.

Results For all types of IV and exposure in the cohort and NCC designs, IV estimates were extremely unstable and biased when the IV was very weakly associated with exposure (e.g. Pearson's correlation < 0.15 for continuous or OR < 2.0 for binary IV and exposure; although specific cut-off values depend on simulation settings). For stronger IVs, estimates were unbiased and become less variable compared with weaker IVs in the case of continuous and binary (risk difference scale) outcomes. For a similar IV-exposure association (e.g. OR = 1.4 and 5% incidence of the outcome), the variability of the estimates was more pronounced in the NCC (standard deviation = 2.37, case : control = 1:5) compared with the cohort design (standard deviation = 1.14). The variability was even more pronounced for rare (≤1%) outcomes. However, IV estimates from the NCC design became less variable with an increasing number of controls per case. Moreover, estimates were biased when the IV was related to confounders even with strong IVs.

Conclusions Instrumental variable analysis performs poorly when the IV-exposure association is extremely weak, especially in the NCC design. IV estimates in the NCC design become less variable when the number of control increases. As NCC does not use the entire cohort, in order to achieve stable estimates, this design requires a stronger IV-exposure association than the cohort design. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS—instrumental variable; cohort; nested case-control; rare outcome; simulation; bias; variability; weak IV; confounding in epidemiology; pharmacoepidemiology

Received 3 January 2013; Revised 29 October 2013; Accepted 5 November 2013

INTRODUCTION

Instrumental variable (IV) analysis is becoming increasingly popular to adjust for confounding in observational pharmacoepidemiologic research.[1–10] An IV is a variable that is associated with the exposure under study and only related to the outcome through exposure. Hence, an IV should neither directly nor indirectly (e.g. through confounders) be associated with the outcome. In that case, IV analysis controls for observed and unobserved confounding and provides consistent and asymptotically unbiased estimates of the exposure on the outcome.[5,11–14]

However, when the association between an IV and exposure is weak, the IV itself is called a weak instrument or weak IV.[15] In that situation, the statistical models of IV analysis are 'weakly identified',[16] and

*Correspondence to: O. H. Klungel, Pharmacoepidemiology and Clinical Pharmacology, University of Utrecht, Universiteitsweg 99, PO Box 80082, 3584CG, Utrecht, the Netherlands. E-mail: O.H.Klungel@uu.nl

[†] The manuscript in this form has not been published and is not being considered for publication elsewhere, in whole or in part, in any language. It should be noted that the abstract of this manuscript has been presented at the 28th International Conference on Pharmacoepidemiology & Therapeutic Risk Management, Barcelona, Spain, 22–26 August 2012 and the annual epidemiology conference WEON, Rotterdam, Netherlands, June 14–15, 2012.

the exposure effects obtained by IV analysis are often inconsistent and biased, with too large and unstable confidence intervals (CIs).[5,11,17–21]

Over the last two decades, there have been a number of studies[5,8–11,14,16,18,22–28] that have addressed the impact of weak IVs. These articles focused, for example, on inconsistent and biased estimates, small sample bias, the impact of weak IVs on CIs in the case of continuous or binary outcomes, and all focused on a cohort design only. These studies did not cover the entire range of common pharmacoepidemiologic settings, for example, cohort as well as nested case–control (NCC) designs, different types of exposure, IV and outcome (i.e. both continuous and binary), different incidences of the outcome (especially rare outcomes) and slight deviations of the IV assumptions due to imbalances of confounder distributions between IV levels. The aim of the present study was to assess the performance of IV estimates in different realistic pharmacoepidemiological settings using simulated data.

METHODS

We used simulated data to assess the performance of IV estimates in a cohort and an NCC design. Different combinations of continuous or binary IV, exposure and outcome were examined in both designs (obviously in the NCC design, only a binary outcome). The following notations were used: Y denotes the outcome, X denotes the exposure/treatment, Z denotes the IV and C denotes a set of confounding variables (some possibly unmeasured). For simplicity, we assumed time-independent exposure, IV and confounders. The simulation settings are presented in Table 1. We used statistical software R (Windows, version 2.15.1) to simulate and analyse our data.[29]

Simulation of the data

Step 1: basic setup. We studied several sample sizes in both designs. For the cohort data, sample sizes were 1000, 5000 and 10 000. For the NCC data, samples were drawn from large cohorts with sizes of 10 000 and 50 000. The incidences of the outcome we studied were 1%, 5%, 10% and 25% for the cohort and 1% and 5% for the cohorts from which the cases and controls in the NCC design were extracted. All scenarios were simulated 10 000 times.

Table 1. Overview of simulation settings

	Scenarios
Exposure (X)	Continuous: $X \sim N(0,\sigma^2)$ Binary: X ~ $Bernoulli(p)$; prevalence (Px) = 0.50
Outcome (Y)	Continuous: $Y \sim N(0,\sigma^2)$ Binary: $Y \sim Bernoulli(p)$ Incidence of Y in the cohort = 1%, 5%, 10%, and 25% Incidence of Y in the NCC = 1% and 5%
Instrumental variable (Z)	Continuous: $Z \sim N(0,1)$ Binary: $Z \sim Bernoull(p)$; prevalence (P_z) = 0.40
Confounding factors	Continuous: $C \sim N(0,1)$ Confounding effects: $\beta_C = 0.50$ to 2.0 Confounding effects (RD model): $\beta_C = 0.005$
Correlation between exposure and IV	PC: 0.01 to 0.60 (both X and Z are continuous) PBC: 0.01 to 0.60 (X continuous and Z binary or vice versa) OR: 1.0 to 6.0 (both X and Z are binary)
Correlation between IV and confounder	PC = 0.00 (i.e. IV is independent of confounders, valid IV) PC = 0.10 and 0.40 (i.e. IV is not independent of confounders, invalid IV)
Combination of exposure and IV either outcome is continuous or binary	1. Continuous exposure and continuous IV 2. Continuous exposure and binary IV 3. Binary exposure and continuous IV 4. Binary exposure and binary IV
Sample size (n)	Cohort: 1000, 5000, and 10 000 NCC: cohort sizes are 10 000 and 50000
Number of simulations ($n.sim$)	10 000
Case: control in the NCC design	1:1, 1:5 and 1:10
True exposure effect	Continuous outcome: $\beta_x = 1$ Binary outcome (OR scale): $\beta x = \log 2$ and $\beta x = \log 1$, i.e. OR = 2 and OR = 1, respectively Binary outcome (RD scale): $\beta_x = 0.005$ Intercept: $\beta_{0X} = 0.10\ to\ 1.0$ (continuous exposure) $\beta_{0X} = -1.0\ to\ 1.0$ (binary exposure)
Intercepts of the outcome models	Continuous outcome: $\beta_{0Y} = 1$ Binary outcome (OR scale): $\beta_{0Y} = -1\ to\ -6$ Binary outcome (RD scale): $\beta_{0Y} = 0.0998$
Nominal coverage probability	0.95

PC = Pearson's correlation coefficient; PBC = point bi-serial correlation; OR = odds ratio; RD = risk difference; NCC = nested case–control; IV = instrumental variable.

We imposed the following restrictions to the simulated data in order to meet the assumptions underlying IV analysis: the IV was independent of confounders; the IV was independent of the outcome given exposure and confounders; and in order to identify a point estimate of exposure on the outcome (such as the average causal effect), the effect of exposure on the outcome was the same for all subjects (homogeneous exposure effect).

Step 2: generation of instrumental variable and confounding variables. We assumed that the IV and confounding factor followed a multivariate normal distribution. For simplicity, we simulated a single confounding variable. For a valid IV, the correlation between IV and confounder was zero ($PC = 0$). We checked this in our simulated data by assessing the empirical association between these variables and observed that the simulated IV was indeed uncorrelated with the confounder. In a separate simulation, we also assessed the impact of violation of this assumption by imposing a correlation between the IV and the confounding factor ($PC = 0.10$ and 0.40). In the case of a binary IV, the continuous IV was dichotomized to create binary variables.[30] A cut-off value was used for dichotomization that resulted in a prevalence of the binary IV of 40%.

It is noted that in pharmacoepidemiological studies, the use of IVs is still rare but it has increased over the last years.[31] One of the most often used IVs in pharmacoepidemiology is physician preference.[32] This IV can be operationalized in different ways, for example, by using the previous prescription choice (e.g. drug A or drug B) made by a physician or by using the proportion of prescriptions (e.g. of drug A versus B) by a certain physician. This proportion is a continuous variable with values in the range (0, 1), which could follow approximately a truncated normal distribution. We therefore assumed that our simulated binary IV can be comparable with physician prescribing preferences based on the previous prescription, whereas the continuous IV can be comparable with the proportion of a prescription for a certain drug.

Although in the pharmacoepidemiological studies binary IVs are mostly used, one can imagine continuous IVs in pharmacoepidemiological studies. For example, in a study of COX-2 inhibitor use, physician preference (i.e. the historical proportion of a physician's NSAID prescriptions that were for a COX-2 inhibitor) was used as an IV.[1] Another example is the amount of drug copayment by patients, which served as an (continuous) IV in a study of the effects of adherence to beta-blocker therapy on blood pressure.[33] These IVs approximately follow a normal distribution, and we therefore assumed that our continuous IV is a realistic IV in the pharmacoepidemiological settings.

Step 3: generation of exposure variables. The continuous and binary exposures were generated on the basis of a linear model (see in the succeeding text; Equation (1)) and logistic model (Equation (2)), respectively.

$$X = \beta_0 + \beta_z Z + \beta_c C + \varepsilon \tag{1}$$

$$\text{logit}[\text{Prob}(\text{X} = 1|\text{Z}, \text{C})] = \beta_0 + \beta_z Z + \beta_c C \text{ and } X\text{~}Bernoulli\ (p) \tag{2}$$

where Z indicates the IV generated in *step 2*, the variable C denotes the confounding factor, β_0, β_z and β_c denote the intercept, IV and confounder effects on the exposure, respectively. In Equation (1), ε is the error term for the exposure, which follows a standard normal distribution (mean zero, variance 1). p is the probability of exposure, from the logistic model in Equation (2).

In order to assess the impact of the strength of the association between IV and exposure, the value of β_z was varied over a range of values. In addition, to assess the impact of various confounding effects, different values of β_c were considered (details in Table 1).

Step 4: generation of outcome variables. The continuous and binary outcomes (estimates in odds ratio (OR) and risk difference (RD) scales) were generated by using the following models, where Equation (3) is a linear model (continuous outcome), Equation (4) is a logistic model (OR scale) and Equation (5) is a linear RD model (RD scale).

$$Y = \beta_0 + \beta_x X + \beta_c C + \varepsilon \tag{3}$$

$$\text{logit}[\text{Prob}(\text{Y} = 1|\text{X}, \text{C})] = \beta_0 + \beta_x X + \beta_c C \text{ and } Y\text{~}Bernoulli\ (p) \tag{4}$$

$$\text{Prob}(\text{Y} = 1|\text{X}, \text{C}) = \beta_0 + \beta_x X + \beta_c C \text{ and } Y\text{~}Bernoulli\ (p) \tag{5}$$

where X indicates the exposure variable generated in *step 3*, the variable C denotes the confounding factor generated in *step 2*, β_0 and β_x denote the intercept and true exposure effect on the outcome, respectively, and β_c denotes the effect of the confounder on the outcome. In the linear model 3, ε is the error term for the outcome, which follows a normal distribution with mean zero and variance 1. p is the probability of the outcome, based on the models 4 as well as 5. Because bias of the IV estimate is invariant to the value of the parameter β_x, the exposure effects were $\beta_x = 1$ for the continuous outcome, $\beta_x = \log 2$ and $\beta_x = 0.005$ for the binary outcome on the OR and RD scale, respectively.

Step 5: study designs. Instrumental variable analysis in a cohort design is relatively well known in pharmacoepidemiological studies. The case–control design is popular in pharmacoepidemiological studies, which is a more efficient design than a cohort design, particularly in situations where outcomes are rare or information on key variables is hard or expensive to obtain (e.g. study on gene expression data). Therefore, we also evaluated performance of IV estimates in an NCC design.

An NCC study is a case–control study that is nested within a larger cohort of known size; hence, the sampling fraction of cases and controls is known. In this setting, validity of the IV analysis may be equivalent to that of IV analysis in a cohort study, given that the IV is valid and strongly associated with the exposure and the control subjects are sampled appropriately.[34]

We simulated data for the NCC design in the following way. Firstly, a large cohort with 1% and 5% incidence of the outcome was generated. Secondly, all cases were selected from the cohort data, and control subjects were randomly selected from that cohort (a cross-sectional approach, which is valid given the low cumulative incidence of the outcome over the study period). Although matching in NCC studies is typically carried out in order to increase efficiency of control for confounding, we did not consider matching, because we aimed to control for confounding by means of IV analysis. We considered a ratio of case to control of 1:1, 1:5 and 1:10. This means that, for example, in case of a cohort of 10 000 subjects in which the case–control study is nested, and an incidence of the outcome of 1% with 10 controls per each case, the effective sample size for the IV analysis based on NCC data is 1100 subjects.

Analysis of simulated data

Cohort data. We analysed the data using a two-stage IV method. In all settings, the first-stage model was a linear regression model, irrespective of whether the exposure was continuous or binary.[19,35] In this model, the exposure was the dependent and the IV was the independent variable.

The second-stage model was a linear regression model in the case of a continuous outcome (Equation (6); see succeeding text) as well as in case of a binary outcome (IV estimates on RD scale; Equation (7)) and a logistic regression model[36] (LRM) in the case of a binary outcome (IV estimates on OR scale, Equation (8)). In the second-stage model, the dependent variable was the outcome and the independent variable was the predicted value of the exposure (obtained from the first-stage model) rather than the actual exposure. It should be noted that in all settings of IV analysis, the confounder 'C' was considered as unobserved so that in all analyses, the variable 'C' was omitted from the IV models.

The statistical models for the second-stage are given in Equations (6), (7), and (8).

$$\text{Second-stage } (\textit{continuous outcome}) : Y_i = \beta_0 + \beta_{IV}\hat{X}_i + \varepsilon_i; \text{for } i{=}1{,}2, \ldots\ldots n \tag{6}$$

$$\text{Second-stage}(\textit{binary outcome}, RD) : \text{Prob}(\mathrm{Y} = 1) = \beta_0 + \beta_{IV}\hat{X}_i; \text{for } i{=}1{,}2, \ldots\ldots n \tag{7}$$

$$\text{Second-stage}(\textit{binary outcome}, OR) : logit(p_i) = \beta_0 + \beta_{IV}\hat{X}_i; \text{for } i = 1{,}2, \ldots\ldots n \tag{8}$$

where $\hat{X}_i$ denotes the predicted value of the exposure, for binary exposure (X), the predicted value of X represents the probability of X = 1 (conditional on IV) estimated from the first-stage IV model, Y_i in Equation (6) denotes the continuous outcome and $\mathrm{p_i}$ is the probability of having the binary outcome. ε_i follows a normal distribution with mean zero and constant variance σ^2. The regression coefficient (β_{IV}) estimated from the second-stage model denotes the IV estimator. For binary outcome, IV estimates were on the RD scale and OR scale (Equations (7) and (8)), respectively. As a comparison, we also estimated exposure effects using conventional models (regression of the outcome on exposure), in which the confounder was considered unobserved.

The IV estimators, $\hat{\beta}_{IV}$, from Equations (6) and (7), provide estimates of the average causal effect of the exposure for all subjects of the study population given that the assumptions of IV were fulfilled.[12] However, $\hat{\beta}_{IV}$ from Equation (8) does not generally provide a consistent estimate of the causal OR.[36] In order to achieve consistent estimates with LRM, we also simulated data under the 'null' hypothesis,[12] that is, the exposure effect was set to zero (OR = 1).

Nested case–control data. Data were analysed in a similar way in the NCC design with the first-stage model weighted by the inverse of the sampling fraction of cases (i.e. 1/1 = 1) and controls (i.e. 1/(no. controls/[cohort size − no. cases]) = [cohort size − no. cases]/number of controls). The second-stage model was the (unweighted) LRM given in Equation (8).

Strength of the instrumental variable

Different measures were used to assess the strength of the IV-exposure association. In order to assess the association between a continuous IV and a continuous exposure, the Pearson's correlation (PC) was used. When a binary IV and continuous exposure (or vice versa) were present, the point bi-serial correlation (PBC)[37,38] was used, and in case of a binary exposure and binary IV, the OR was used. Additionally, the strength of the IV was also verified by the partial F-statistic value of the first-stage regression model although this statistic is highly affected by the sample size. Throughout the article, we refer to this as the F-statistic.

Bias, standard error, root mean square error and coverage probability

Each scenario was simulated 10 000 times. Bias of IV estimates was defined as the difference between the mean of IV estimates based on 10 000 simulation runs and the true exposure effect. Two types of confidence intervals (CIs) were estimated and reported: (i) to identify the precision of estimating the bias of the IV estimates and (ii) to assess the variability between estimates from 10 000 simulation runs (i.e. the variation between different studies). The first CIs were estimated using the standard errors of the mean of the estimates (i.e. standard deviation of the IV estimates divided by the square root of the number of simulations), and the second CI was estimated by 2.5 and 97.5 percentiles of the 10 000 estimates.

In case of a binary outcome, we also estimated bias with no treatment effect (OR = 1) to achieve consistent estimates and evaluate the impact of non-collapsibility[39] of the ORs in the LRM (second-stage model for binary outcomes; Equations (8)). Because the second-stage model (Equation (8)) is not conditional on the confounder, it does not estimate a conditional exposure effect. As a comparison for the IV estimates, we also estimated marginal exposure effects based on marginal structural models (MSMs) in the cohort design.[40] For these MSMs, the confounding factor (C) was considered an observed variable. As previously mentioned, the data generation process was under a homogeneous treatment effect, which allows for a comparison between IV estimates and MSMs estimates.[41] In the MSMs, inverse probability of treatment weighting (IPTW) was applied to estimate the marginal treatment effect, including the observed confounder in the treatment model. For a binary exposure, the IPTW was estimated by logistic regression and stabilized. For a continuous exposure, the IPTW was estimated by density functions of a linear model.[40] We assessed the sensitivity of the estimates to weight truncation at 0.5 and 99.5 percentiles, at 1.25 and 98.75 percentiles, and at 2.5 and 97.5 percentiles.[42] To evaluate the performance of IV estimates in the context of bias, accuracy and coverage for different scenarios, root mean square error (RMSE) and coverage probability were estimated for settings with a continuous outcome.[30] The 95% coverage probability was estimated on the basis of CIs that were estimated by 2.5 and 97.5 percentiles of the estimates from 1000 bootstrap samples in each simulation run.

RESULTS

In our simulations, results from conventional analyses were biased because of unobserved confounding. For example, the association between a continuous exposure and a continuous outcome was estimated to be 1.50 to 1.23 (for different settings) using conventional linear regression model instead of a true exposure effect of 1 (Table 2). Likewise, the association between a binary exposure and a binary outcome was estimated to be OR 4.01 to OR 3.72 (for different settings) using conventional LRM instead of a true exposure effect of OR = 2.0 (Table A1).

Figure 1 shows that in a cohort design with a continuous outcome, the IV estimates were highly unstable and also be biased if the association between IV and exposure was extremely weak. For example, in the case of both a continuous IV and exposure or both a binary IV and exposure and a cohort size of 10 000, the IV estimates were biased if the correlation between IV and exposure was smaller than 0.15 or the OR was smaller than 2.0, respectively. The specific cut-off values differed between simulation settings. Similar patterns were observed for other types (binary and continuous) of IV and exposure. Although the bias was within 5% of the true exposure effect of 1, this magnitude depends on simulation settings.

Furthermore, the variation in IV estimates between different simulations increased with weaker associations

Table 2. Estimates, RMSE and coverage probability of simulation of a continuous outcome in the cohort design

	IV continuous X continuous Y continuous							IV binary X continuous Y continuous						
	IV method ¥					Conventional method**		IV method					Conventional method	
Effect of IV on exposure (β_Z)	PC	Estimate	RMSE	Coverage* probability	Width CI†	Estimate	RMSE	PBC	Estimate	RMSE	Coverage probability	Width CI	Estimate	RMSE
0.01	0.01	2.201	179.136	0.989	20.270	1.500	0.500	0.00	1.317	65.772	0.992	21.793	1.500	0.500
0.05	0.04	0.930	2.167	0.962	2.686	1.499	0.500	0.02	1.018	12.234	0.980	12.966	1.500	0.500
0.10	0.07	0.991	0.148	0.947	0.608	1.498	0.498	0.04	0.937	0.824	0.962	2.785	1.499	0.500
0.15	0.11	0.996	0.096	0.947	0.383	1.494	0.495	0.05	0.979	0.213	0.950	0.946	1.499	0.499
0.60	0.39	1.000	0.024	0.947	0.092	1.424	0.424	0.20	0.999	0.048	0.949	0.189	1.479	0.479
1.50	0.73	1.000	0.009	0.947	0.037	1.235	0.235	0.46	1.000	0.019	0.949	0.075	1.394	0.394
	IV continuous X binary Y continuous							IV binary X binary Y continuous						
	IV method					Conventional method		IV method					Conventional method	
	PBC	Estimate	RMSE	Coverage probability	Width CI	Estimate	RMSE	OR	Estimate	RMSE	Coverage probability	Width CI	Estimate	RMSE
0.01	0.00	2.476	160.125	0.997	66.228	1.826	0.827	1.00	12.946	1505.620	0.996	68.099	1.826	0.827
0.05	0.02	1.144	33.540	0.978	27.578	1.826	0.827	1.04	2.110	162.576	0.992	54.763	1.826	0.827
0.10	0.04	0.944	0.836	0.952	4.266	1.825	0.826	1.09	0.753	21.343	0.978	28.370	1.826	0.826
0.15	0.05	0.976	0.481	0.949	2.016	1.824	0.824	1.13	0.675	16.677	0.963	10.728	1.826	0.826
0.60	0.20	0.998	0.120	0.949	0.471	1.785	0.786	1.64	0.994	0.238	0.949	0.937	1.816	0.816
1.50	0.46	1.000	0.059	0.949	0.229	1.643	0.644	3.48	0.999	0.096	0.949	0.376	1.762	0.763

IV, instrumental variable; X, exposure; Y, outcome; PC, Pearson's correlation; PBC, point bi-serial correlation; OR, odds ratio; RMSE, root mean square error.
Sample size: $n = 10\,000$.
Number of simulations = 10 000.
¥IV estimate from two-stage linear IV models, and the IV is independent of the confounders (valid IV).
*Nominal coverage probability is 0.95, and the coverage probabilities for conventional model were almost close to zero.
**Conventional method: linear regression model.
†Width of CI: mean width of the 95% CIs (calculated across 10 000 simulated samples).

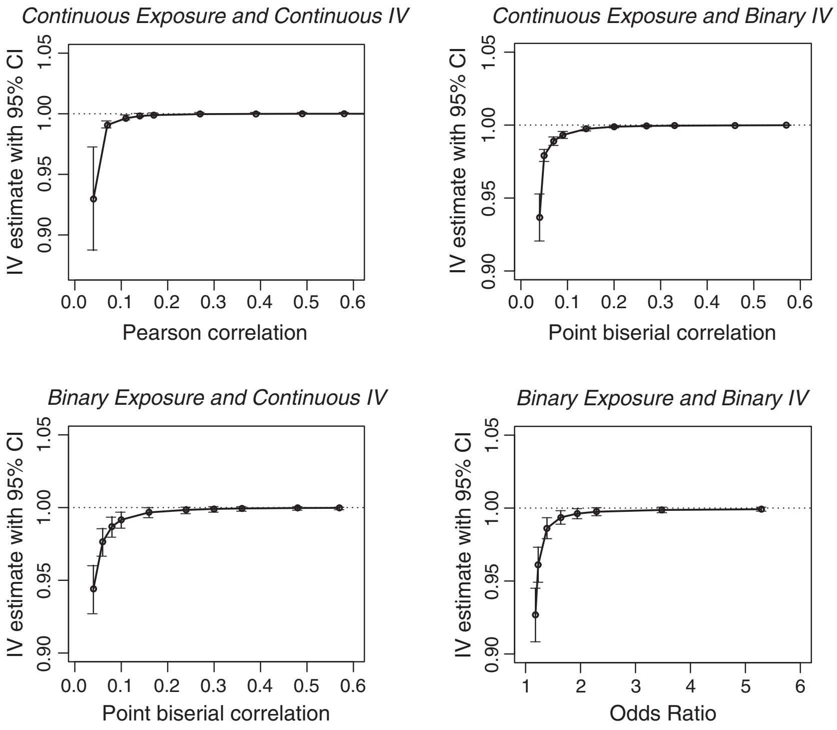

Figure 1. Mean of instrumental variable (IV) estimates from simulations of a cohort design with a continuous outcome X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect = 1). Vertical bars indicate 95% confidence intervals around the mean of the estimates. Results are based on simulations of a cohort with sample size 10 000, and each scenario was simulated 10 000 times

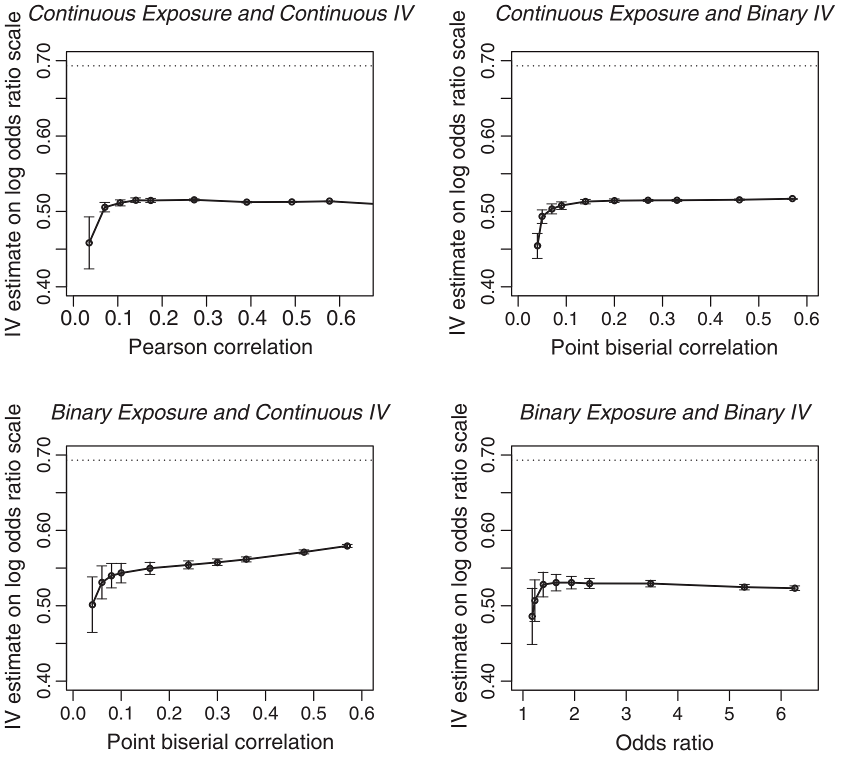

Figure 2. Mean of instrumental variable (IV) estimates from simulations of a cohort design with a binary outcome, using a logistic model in the second stage of IV analysis X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect = log(2)). Vertical bars indicate 95% confidence intervals around the mean of the estimates. Results are based on simulations of a cohort with sample size 10 000, incidence of the outcome 10%, and each scenario was simulated 10 000 times

DOI: 10.1002/pds

between IV and exposure (details in Figure A1). This is also reflected by the large RMSE in case of a weak IV-exposure association (Table 2). Both bias and variability of IV estimates increases with decreasing strength of the IV-exposure association (Table 2). In case of an IV that is not weakly associated with exposure (and hence no bias), the coverage rates of IV analysis were indeed close to the nominal level. However, for extremely weak IV-exposure associations, the coverage rate exceeded the nominal level, due to the increased standard errors for weak IV. Moreover, in this case, the mean width of the 95% CIs (calculated across 10 000 simulated samples) was very large (Table 2).

In case of a binary outcome that was simulated on the basis of a logistic model in the cohort design, a similar pattern was observed as for the situation with a continuous outcome: a weak IV-exposure association resulted in biased IV estimates (Figure 2). In addition, estimates were also systematically biased for strong IV-exposure associations in case of an LRM for the second-stage of IV analysis (Figure 2). This is partly due to noncollapsibility of the OR, but it can also be a result of model misspecification. Moreover, when in case of a nonzero exposure effect comparing the IV estimates with estimates from MSMs, the IV estimates for binary exposure were closer to the MSM estimates than to the conditional effects from the data generating model (Table A1). For continuous exposure, we examined the MSM estimates with and without truncation of weights. In the first case, the estimates were stable, whereas in the latter, the MSM estimates were highly unstable because of extreme weights and far from the IV estimates (Table A1). Moreover, we assessed the sensitivity of exposure effect estimates to weight truncation, and we observed that the estimates were more close to IV estimates when 1% of the extreme weights were truncated (truncation at 0.5 and 99.5 percentiles) than 2.5% or 5% (truncation at 1.25 and 98.75 percentiles and at 2.5 and 97.5 percentiles, respectively). In all cases, the RMSE of MSM estimates was lower than the RMSE of IV estimates.

There was no evidence of bias for the IV estimates on RD scale, as the 95% CIs for the mean estimate always included the true exposure effect (Figure 3). The variability of the estimates in the case of extremely weak IVs was the same as described before for other settings (data not shown).

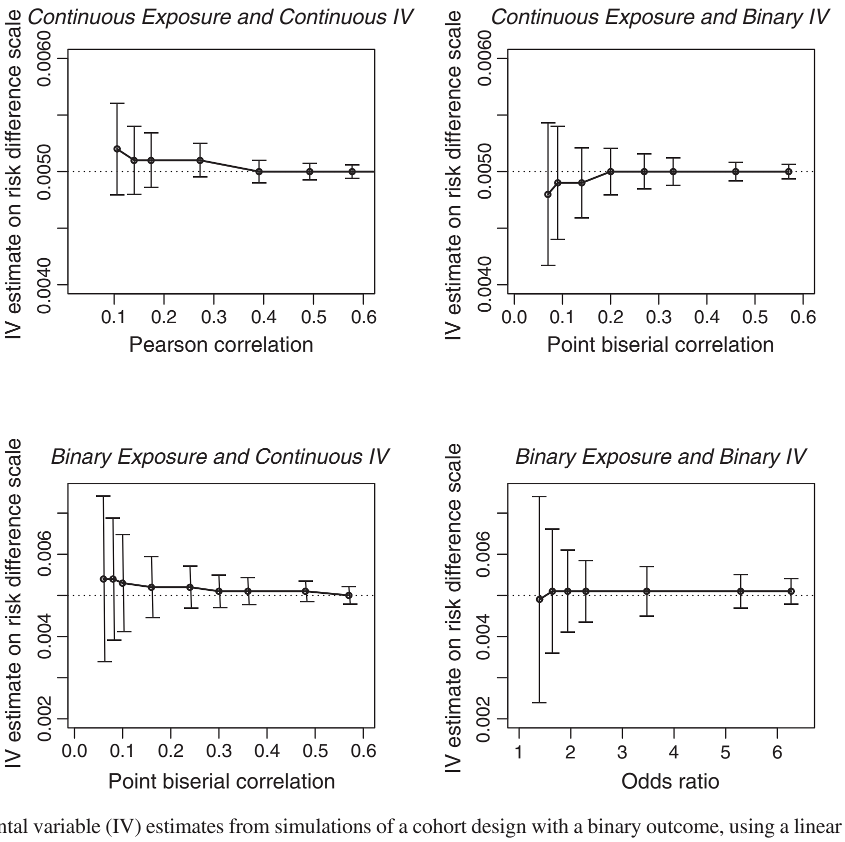

Figure 3. Mean of instrumental variable (IV) estimates from simulations of a cohort design with a binary outcome, using a linear model in the second stage of IV analysis X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect on RD scale=0.005). Vertical bars indicate 95% confidence intervals around the mean of the estimates. Results are based on simulations of a cohort with sample size 10 000, incidence of the outcome 10%, and each scenario was simulated 10 000 times

In Figure 4, results are presented for the NCC design. Because in the NCC design the LRM was applied as the second-stage IV model, the IV estimates in the NCC design were biased irrespective of the weak or strong IVs. However, there was much more variation between estimates in the NCC design than in the cohort design (Table A2). This variation in IV estimates between simulations decreased with increasing strength of the IV-exposure association and also with an increasing number of controls per case (Figure 4).

The magnitude of bias varied for different incidences of the outcome in both designs (Table A3 for cohort design). Our simulations also revealed that the IV estimates were highly unstable when the incidence of the outcome was very rare (e.g. 1% in our simulations of a cohort of size 10 000), which was more pronounced in the NCC design than in the cohort design. Table A4 shows the impact of sample sizes on the IV estimates in the cohort design. For each simulation setting, the RMSE decreased as sample sizes increased.

A comparison between conventional and IV analysis revealed that the IV analysis yielded unbiased results within the reasonable range of PC (>0.15) or OR (>2.0), which is also reflected by the lower RMSE of the IV estimates for continuous outcome and binary outcome in RD scale even if the variance of conventional estimates was lower than for IV estimates. Furthermore, the coverage rate of conventional estimates was very low due to their substantial bias, in contrast to correct coverage rates for IV estimates.

Our simulation study showed that in many situations when the IV was extremely weak (e.g. PC < 0.10 or OR < 1.5), the F-statistic value was more than 10 (a commonly used cut-off value for a weak IV), yet IV estimates were still significantly biased and highly variable. For example, when exposure and IV were continuous and sample size was 10 000, the bias was −0.070 with an F-value = 14 and $PC = 0.04$. Similarly, when both exposure and IV were binary, the bias was −0.073 with F-value = 17 and OR = 1.2. In these situations, the F-statistics values were misleading for assessing the strength of the IV as this is mostly affected by the sample sizes.

Figure 5 shows the impact of violation of the assumption that the IV is independent of confounders. It shows the patterns of bias in the cohort design for different combinations of exposure and IV with invalid IV (i.e. correlation between the IV and the confounding variable, $PC = 0.10$ and $PC = 0.40$) and estimates from the conventional regression model. In all settings, when the IV was weak and related to the

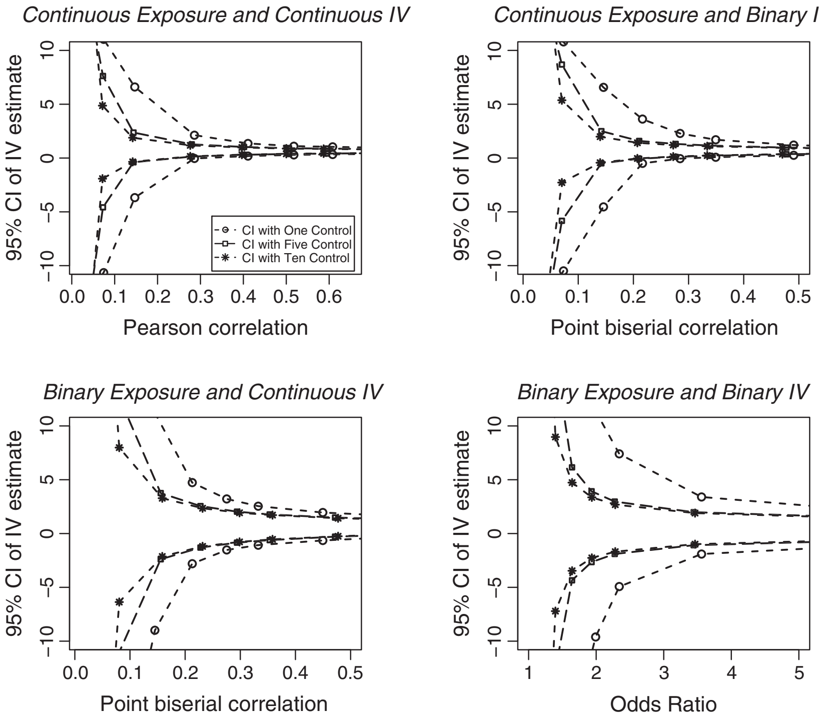

Figure 4. Variation in instrumental variable (IV) estimates in the nested case-control design X-axis represents the association between exposure and IV. Lines indicate the 2.5 and 97.5 percentiles of the estimates in nested case-control design for different combinations of exposure and IV with case:controls ratio of 1:1, 1:5, and 1:10. Results are based on simulations of a case-control study nested within a cohort with sample size 10,000 and an incidence of the outcome of 1%. Each scenario was simulated 10,000 times

DOI: 10.1002/pds

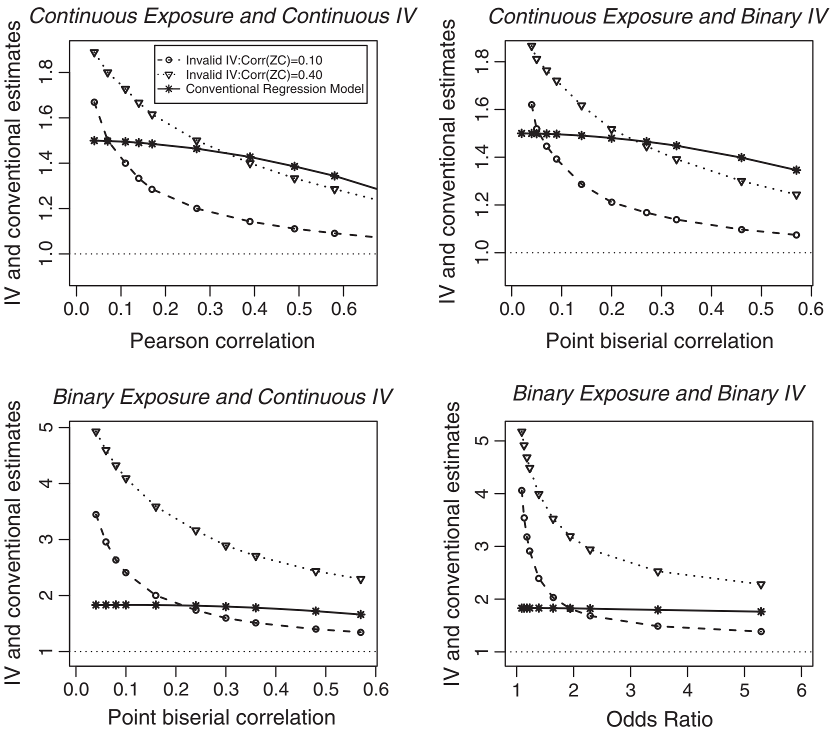

Figure 5. Mean of instrumental variable (IV) estimates and conventional estimates from simulations of a cohort design with a continuous outcome and an invalid IV (IV correlated with confounders) X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect - 1). Results are based on simulations where the correlation between IV and confounder was PC=0.10 and PC=0.40 and a sample of size 10 000 with 10 000 replications

confounding variable, the biases were significantly larger than in the conventional model. This pattern was more pronounced for $PC = 0.40$. However, for a stronger IV with $PC = 0.10$, the bias was lower than conventional estimates.

In all simulation settings, the pattern of bias was similar for different confounding effects, but the magnitude of bias increased with increasing the effects of confounder (values of β_c) on the exposure and outcome. Because these results are expected, we do not show these results in detail; all presented results for $\beta_c = 1$ and $\beta_c = 0.005$ (RD models).

DISCUSSION

Our simulation study shows that, in all binary/continuous combinations of IV, exposure and outcome in the cohort and NCC designs, the validity of IV analysis strongly depends on the strength of the association between IV and exposure. IV estimates are unbiased in case of a strong IV and a continuous outcome, whereas conventional estimates can be extremely biased (because of unobserved confounding). For binary outcomes, the IV estimates on the RD scale are also unbiased (again, the conventional estimates are biased) but on the OR scale, which are systematically biased even with strong IVs. In all cases, the variance of estimates is lower for the conventional estimates compared with IV estimates, which is partly due to bias-variance trade-off.[5] In addition, for weak IVs, large variation between IV estimates from different studies was observed.[8] In fact, the extreme variability of the estimates for very weak IVs makes it difficult to assess the accuracy of the point estimates and makes these estimates practically useless. This pattern was observed for different types (binary/continuous) of IV, exposure and outcome in the cohort as well as the NCC design. Because the effective sample size in the NCC design is smaller than in a cohort design, this behaviour of IV analysis was more pronounced in the NCC design but could partly be remedied by increasing the number of controls per case.

Although the set-up of our simulations was such that the IV was independent of confounders and the IV had no direct effect on the outcome, in finite-samples random variation may cause the actual data to deviate from these assumptions, which can result in some amount of bias (conditional on the observed data) for weak IVs.[11,13,18,43] Furthermore, in finite samples, even a small bias of IV estimates can be amplified considerably by a weak IV, resulting in large variation between IV estimates.[6,7,12] Another aspect of our simulations that deserves attention is that the bias in the conventional estimates became smaller with increasing strength of the IV-exposure association. It is inherent to our set-up that there is an inverse relation between the strength of the IV and the magnitude of the unobserved confounding. For details, we refer to the work by Martens *et al.*[11]

Variability of IV estimates not only depends on the strength of the IV-exposure association but also on the outcome. We observed that if the outcome is rare (e.g. 1% in a cohort of 10 000), IV estimates are highly variable. This is even more pronounced in the NCC design because of relatively small effective sample size. Additionally, the instability of the IV estimates could also be increased because of the two-step analysis: variability in the first-stage model can result in even more variation in the second modelling step.[22] This problem can be reduced by increasing the sample size, the number of cases and increasing the number of controls in the NCC design, as long as the IV-exposure association is sufficiently strong. These findings suggest that IV analysis should be carefully performed in the case of rare outcomes especially in the NCC design.

Our NCC design findings suggest that selection and construction of instrumental variables in the NCC study is in line with the cohort study if the control subjects are sampled appropriately. However, as selection of appropriate controls is a major challenge in practical settings, cautions should be taken when selecting the controls. Our findings also suggest increasing the number of controls in order to increase effective sample size and reduce variability of the IV estimates.

In simulations with a binary outcome, we used logistic and linear regression models in the second-stage model, and observed that the IV estimates on the OR scale are biased even with a strong IV. Others also stated that IV estimates are biased with nonlinear models (e.g. logistic regression).[44–47] Reasons for this include noncollapsibility of the OR,[39,48] model misspecification,[19] and mean and variance dependency of the LRM.[47] We compared the IV estimates on the OR scale with a (reference) marginal exposure effect instead of the conditional exposure effect; still, IV estimates were different from MSM estimates, which might be due to the fact that MSMs estimates are sensitive to weights especially for continuous exposure[49] or due to the random variation or model misspecification. However, the estimates on the RD scale are unbiased when IV-exposure association is strong enough. This finding is in line with previous literature.[5] Obviously, this result strongly relies on the fact that the linear model was used for generating and analyzing the data (i.e. no model misspecification) and generalization of this result is conditional on appropriate specification of the second-stage model in empirical data.

Other studies have previously shown that if the F-statistic value of the first-stage regression model is 10, the IV is strong enough and the bias of IV estimates is negligible.[11,16,26,50,51] However, in our simulations, the IV estimates were still biased, and estimates were extremely variable when the values of the F-statistic were around 10.[8] This was observed in all combinations of the exposure and IV, suggesting that a cut-off value for the F-statistic value around 10 is inadequate to identify a weak IV. Hence, only the F-statistic should not be considered as a reliable criterion for assessing the strength of the IV. We recommend researchers to report both the F-statistic value and the association between exposure and IV, by means of the Pearson's correlation, point bi-serial correlation or OR.[8,10] We stress, however, that interpretation of these measures of association depends on the effective sample size.

We also considered a situation where the IV is invalid (i.e. correlated with the confounding variable). We found that the amount of bias can be substantial even with a strong IV if the IV is associated with unmeasured confounders. Additionally, for a weak IV with a small association with a confounder, the bias is significantly higher, but for a strong IV, the bias is smaller than the bias of a conventional model. Although in the latter situation the IV estimates are less biased, a small association between IV and confounder may lead to the violation of an assumption that the IV has no direct effect on the outcome. In that case, there is no guarantee that the IV analysis consistently estimates the average effect of exposure on outcome.[12] Martens *et al.*[11] and Rassen *et al.*[14] already reported that violation of any IV assumption can magnify the bias because of confounding. The results of our simulations confirm that an IV analysis is not a valid analysis (and estimated exposure effects are biased) if the IV is associated with confounders (distribution of confounders does not balance between IV levels) even when the IV is strongly related to exposure.

We did not assess the bias and variability of IV estimates with heterogeneous exposure effects as well as time-varying IV, exposure and confounder. Therefore, it may be interesting to assess the trend of bias and variability due to a weak IV in these settings in future research.

In conclusion, for different binary/continuous combinations of IV, exposure and outcome in the cohort and NCC designs, IV estimates performed uniformly poor in case of weak IV-exposure associations. When the IV is valid and has a strong association with exposure, IV methods provide unbiased exposure effect (for continuous outcome as well as binary outcome on a RD scale), whereas conventional analysis is substantially biased due to unobserved confounding. For weaker IVs, estimates become highly variable. This variability was more pronounced for rare outcomes, especially in the NCC design. To some extent, this can be remedied by increasing the sample size or by increasing the number of controls per case in a case–control study. Because the effective sample size in the NCC design is smaller than in the cohort design, in order to achieve stable estimates, the association between exposure and IV should even be stronger in NCC studies than in cohort studies. When the IV is not independent of confounders, this may result in severe bias even with strong IVs. We recommend researchers to routinely evaluate and report the strength of the IV in order to evaluate the potential for bias of IV estimates due to a weak IV.

CONFLICT OF INTEREST

Olaf Klungel had received unrestricted funding for pharmacoepidemiological research from the Dutch private–public funded Top Institute Pharma.

KEY POINTS

- Extensive simulation studies were performed to evaluate instrumental variable (IV) analysis for cohort and NCC designs in (pharmaco-) epidemiologic settings.
- For all types of IV and exposure in the cohort and NCC designs, IV estimates performed uniformly poor in case of weak IV-exposure associations, particularly for smaller sample sizes.
- The variability of the estimates was more pronounced for rare outcomes, especially in NCC studies. In NCC studies the effective sample size is smaller, and consequently the variability of IV estimates was larger compared to cohort studies. This can be partly remedied by increasing the number of control subjects per case.
- As the F-statistic value strongly depends on the effective sample size, the F-value from the first-stage model of IV analysis should not be considered as a reliable criterion for assessing the strength of the IV. It is recommended to report the F-statistic value in combination with other measures, including e.g. Pearson's correlation, point bi-serial correlation, or odds ratio.

ACKNOWLEDGEMENTS

The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium), which is a public–private partnership coordinated by the European Medicines Agency. The PROTECT project is supported by Innovative Medicine Initiative Joint Undertaking (www.imi.europa.eu) under Grant Agreement no 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. In the context of the IMI Joint Undertaking (IMI JU), the Department of Pharmacoepidemiology, Utrecht University, also received a direct financial contribution from Pfizer. The views expressed are those of the authors only and not of their respective institution or company.

REFERENCES

1. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**: 268–275.
2. Wang PS, Schneeweiss S, Avorn J, et al. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. *N Engl J Med* 2005; **353**: 2335–2341.
3. Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. *Arthritis Rheum* 2006; **54**: 3390–3398.
4. Schneeweiss S, Setoguchi S, Brookhart A, Dormuth C, Wang PS. Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. *Can Med Assoc J* 2007; **176**: 627.
5. Ionescu-Ittu R, Delaney JAC, Abrahamowicz M. Bias–variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol Drug Saf* 2009; **18**: 562–571.
6. Shetty KD, Vogt WB, Bhattacharya J. Hormone replacement therapy and cardiovascular health in the United States. *Med Care* 2009; **47**: 600.
7. Suh HS, Hay JW, Johnson KA, Doctor JN. Comparative effectiveness of statin plus fibrate combination therapy and statin monotherapy in patients with type 2 diabetes: use of propensity-score and instrumental variable methods to adjust for treatment-selection bias. *Pharmacoepidemiol Drug Saf* 2012; **21**(5): 470–484.
8. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; **65**: 155–162.
9. Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care* 2007; **45**: S116–S122.
10. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: an instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; **19**: 699–707.
11. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology* 2006; **17**: 260–267.
12. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**: 360–372.
13. Angrist J, Imbens G, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **91**: 444–472.
14. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 2009; **62**: 1226–1232.
15. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; **19**: 537–544.

16. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 2002; **20**: 518–529.
17. Zivot E, Startz R, Nelson CR. Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 1998; **39**: 1119–1144.
18. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc* 1995; **90**: 443–450.
19. Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Perspect* 2001; **15**: 69–85.
20. Chao J, Swanson NR. Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction. *J Econ* 2007; **137**: 515–555.
21. Stock JH, Yogo M. Testing for weak instruments in linear IV regression. 2002; NBER Technical Working Paper No. 284.
22. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; **62**: 1233–1241.
23. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011; **40**: 740–752.
24. Crown WH, Henk HJ, Vanness DJ. Some cautions on the use of instrumental variables estimators in outcomes research: how bias in instrumental variables estimators is affected by instrument strength, instrument contamination, and sample size. *Value Health* 2011; **14**: 1078–1084.
25. Chiba Y. Bias analysis of the instrumental variable estimator as an estimator of the average causal effect. *Contemp Clin Trials* 2010; **31**: 12–17.
26. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; **65**: 557–586.
27. Andrews DWK, Stock JH. Inference with weak instruments. 2005 ;NBER Technical Working Paper No. 313.
28. Abrahamowicz M, Beauchamp M, Ionescu-Ittu R, Delaney JAC, Pilote L. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. *Am J Epidemiol* 2011; **174**: 494–502.
29. R Development Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing: Vienna, Austria*, 2010; ISBN 3-900051-07-0.
30. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med* 2006; **25**: 4279–4292.
31. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology* 2013; **24**(3): 363–369.
32. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 2011; **64**: 687–700.
33. Cole JA, Norman H, Weatherby LB, Walker AM. Drug copayment and adherence in chronic heart failure: effect on cost and outcomes. *Pharmacotherapy* 2006; **26**: 1157–1164.
34. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. Developing a protocol for observational comparative effectiveness research: a user's guide. 2013.
35. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat* 2001; **19**: 2–28.
36. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011; **173**: 1392–1403.
37. Hennessy S, Leonard CE, Palumbo CM, Shi X, Ten Have TR. Instantaneous preference was a stronger instrumental variable than 3- and 6-month prescribing preference for NSAIDs. *J Clin Epidemiol* 2008; **61**: 1285–1288.
38. Klugh HE. *Statistics: The Essentials for Research*. Lawrence Erlbaum: New Jersey, USA; 1986.
39. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; 29–46.
40. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
41. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 2011; **175**: 60–65.
42. Xiao Y, Moodie EE, Abrahamowicz M. Comparison of approaches to weight truncation for marginal structural Cox models. *Epidemiological Methods* 2013; 1–20.
43. Murray MP. Avoiding invalid instruments and coping with weak instruments. *The journal of economic perspectives* 2006; **20**: 111–132.
44. Cai B, Small DS, Have TRT. Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Stat Med* 2011; **30**: 1809–1824.
45. Henneman TA, Van Der Laan MJ, Hubbard AE. *Estimating Causal Parameters in Marginal Structural Models with Unmeasured Confounders Using Instrumental Variables*. U.C. Berkeley Division of Biostatistics Working Paper Series, Paper 104. The Berkeley Electronic Press: Berkeley, CA; 2002.
46. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 2008; **37**: 1161–1168.
47. Didelez V, Meng S, Sheehan N. On the bias of IV estimators for Mendelian randomisation. *submitted manuscript*, http://137.222.80.3/research/stats/reports/2008/0820.pdf 2008;.
48. Burgess S. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Stat Med* 2013; **32**(27): 4726–4747.
49. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. *Stat Med* 2012; **32**(9): 1584–1618.
50. Han C, Schmidt P. The asymptotic distribution of the instrumental variable estimators when the instruments are not correlated with the regressors. *Econ Lett* 2001; **74**: 61–66.
51. Angrist JD, Pischke JS. *Mostly Harmless Econometrics: An empiricist's Companion*. Princeton Univ Pr: New Jersey, USA; 2008.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Figure A1. Mean and variation of IV estimates from simulations of a cohort design with a continuous outcome and different combinations of exposure and IV

Table A1. Estimates and RMSE of simulation of a binary outcome in the cohort design (IV estimates, MSMs, and Conventional Estimates)

Table A2. Comparison of bias and standard deviation of IV estimates based on full cohort versus NCC design (with various case:control ratios)

Table A3. Impact of incidence of the outcome on the IV estimates in a cohort design with continuous outcome

Table A4. Impact of sample size on the IV estimates in a cohort design with continuous outcome

DOI: 10.1002/pds