Instrumental variable analysis in randomized trials with non-compliance and observational pharmacoepidemiologic studies

RHH Groenwold[1,2*], MJ Uddin[1], KCB Roes[2], A de Boer[1], E Rivero-Ferrer[3], E Martin[4], NM Gatto[5], OH Klungel[1,2]

Abstract

Introduction

Instrumental variable (IV) analysis potentially accounts for unmeasured confounding in observational studies, but it can also control for non-compliance in randomized trials.

IV analysis requires that the IV is related to treatment status, yet independent of confounders of the treatment-outcome relation. This implies that in pharmaco-epidemiologic scenarios where IV analysis is needed the most (because of strong unmeasured confounding), IVs will typically be weakly associated with treatment. Furthermore, IV analysis assumes that the IV affects the outcome only through the treatment under study. A common IV in pharmacoepidemiological studies is the physician prescribing preference, which for the latter assumption implies that physicians only differ in their preference for the treatment under study, but they do not differ with respect to e.g. preferences for concomitant treatments, skills, organization of their practice, etc. Assumptions underlying IV assumptions need thorough evaluation before proceeding with IV analyses. Here, IV analysis is illustrated, its key assumptions are illustrated by a randomized trial with non-compliance, and the utility of IVs for observational pharmaco-epidemiologic studies is discussed.

Conclusion

The validity and applicability of IV analysis in observational pharmacoepidemiologic studies still have to be established, which requires more applications of IV analysis and debate on the likelihood of the assumptions underlying IV analysis.

*Corresponding author
Email: r.h.h.groenwold@umcutrecht.nl

[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
[3] Global Clinical Epidemiology, Novartis Farmaceutica S.A., Barcelona, Spain
[4] BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria), Pharmacoepidemiology and Pharmacovigilance Division, Medicines for Human Use Department Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), Madrid, Spain
[5] Epidemiology, Worldwide Safety & Regulatory, Pfizer Inc, New York, NY, USA

Introduction

Observational studies of the effects of medical interventions (e.g. pharmacological treatment) are prone to confounding. Different methods are available to control for confounding, including restriction, matching, multivariable regression analysis, propensity score analysis, and inverse probability weighting.[1]

What these methods have in common, is that they can control for measured confounders, but not for unmeasured confounders.

Instrumental variable (IV) analysis, on the other hand, has been proposed as a method to control for unmeasured confounding in observational studies. In this review, we first described IV analysis conceptually and illustrate it by non-compliance in a randomized trial. Next, we will discuss the limitations of commonly used IVs to control for unmeasured confounding in pharmacoepidemiology.

Discussion

The authors have referenced some of their own studies in this review. The protocols of these studies have been approved by the relevant ethics committees related to the institution in which they were performed.

Causal diagrams of observational studies and randomized trials

Figure 1 shows several directed acyclic graphs (DAGs), also referred to as causal diagrams. For a detailed explanation of DAGs, we refer to the literature.[2,3] Here, it suffices that causal relations between variables are represented by directed arrows from cause to effect and all causal relations of the treatment and outcome are represented. The DAG in figure 1a shows the typical structure of confounding. The allocation of treatment (T, e.g. treatment with an ACE-inhibitor) and outcome (Y, e.g. myocardial infarction) share a common cause (C, e.g. pre-treatment blood pressure). There is a so-called 'back-door path' from treatment to outcome, via the confounder C. Ignoring this back-door path when estimating the relation between treatment and outcome may result in a bias (i.e., confounding).

However, if one of the arrows from the confounder to either treatment or outcome is absent, there is no back-door path and hence no confounding. This is depicted in figure 1b, which represents an ideal randomized controlled trial. Because treatment allocation is random, it is independent of subject characteristics and hence there is no arrow from C to T.

In reality, in a randomized trial, adherence to the randomly allocated treatment may not be perfect. Hence, treatment allocation (A) and actual treatment status (T) may not be identical (Figure 1c). Note that treatment allocation is still a random process (hence independent of C), yet treatment use need not be a random process. The latter is reflected by the arrow between C and T. An analysis of actual treatment may therefore be biased (due to confounding by C), yet an analysis of treatment allocation (i.e.,

intention-to-treat analysis) will on average be unbiased.

Non-compliance in a randomized trial

The intention-to-treat (ITT) analysis of a randomized trial provides an unbiased estimate of the effect of treatment allocation, rather than the effect of actual treatment use. If the treatment is effective, the ITT analysis underestimates the effect of treatment use, when there is considerable non-compliance.[4] However, by taking the extent of non-compliance into account, one can estimate what the treatment effect would be under perfect compliance. We illustrate this using numerical examples.

Table 1 shows three numerical examples of randomized trials. In the first scenario, there is perfect compliance: all subjects allocated to the experimental treatment actually receive the experimental treatment and all allocated to the control treatment receive the control treatment. Hence, the estimate of the effect of treatment allocation equals the effect estimate of actual treatment received: risk difference (RD) = -0.25.

The second scenario is that of a randomized trial with non-compliance: 60% and 80% of those assigned the control treatment and the experimental treatment, respectively, comply with the assigned treatment.

The ITT effect can be estimated as RD = 300/1000 – 400/1000 = -0.1, which underestimates the effect that would be observed under perfect compliance (scenario 1). To obtain the effect that would be observed under perfect compliance, the ITT effect needs to be extrapolated to a situation with full compliance.

This can be achieved by dividing the ITT effect by the difference in the observed probabilities of receiving experimental treatment between the two treatment allocation groups: -0.1 / (800/1000 – 400/1000) = -0.1 / 0.4 = -0.25, which indeed equals the effect that is observed under perfect compliance.[5,6]

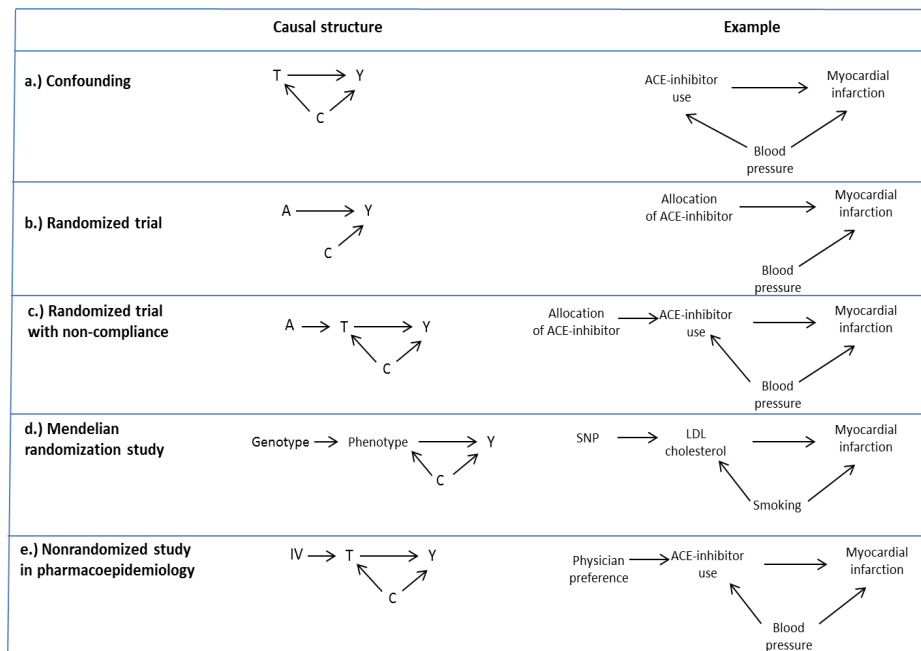A graphical representation of this procedure is given in figure 2. The

Figure 1: Caption: Directed acyclic graphs of randomized and observational studies. Legend: Abbreviations: T: treatment; Y: outcome; C: confounders; A: random treatment assignment, IV: instrumental variable. SNP: single-nucleotide polymorphism.
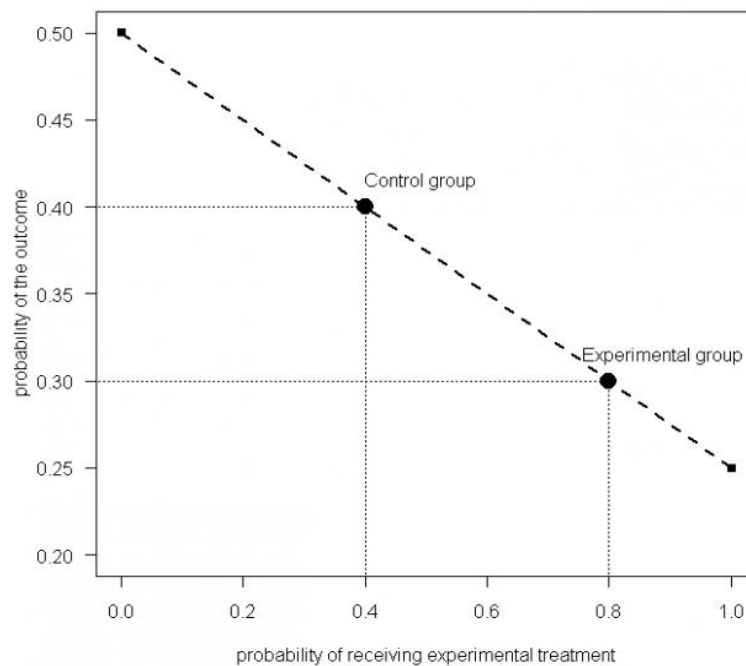

Figure 2: Caption: Graphical representation of IV analysis of a randomized trial with non-compliance.

observed risks among the two treatment allocation groups (0.4 for the control group and 0.3 for the experimental group) are plotted against the probabilities of actually receiving experimental treatment among those two groups (0.4 and 0.8 for the control and experimental treatment groups, respectively). These two points are then connected. The risk difference that would be observed under perfect compliance can be obtained by extrapolating this line to the point at which the probability of receiving experimental treatment is either 0 or 1. The difference between those two extremes can be read off the y-axis and is the risk difference that

OA Epidemiology
Open Access

Table 1: Caption: Numerical example of trials with no or partial non-compliance.				
Scenario 1: no non-compliance				
Treatment assigned	Treatment received		No. subjects	No. events
T = 0	T = 0		1000	500
T = 1	T = 1		1000	250
Scenario 2: partial non-compliance, unrelated to any cause of the outcome				
Treatment assigned	Treatment received		No. subjects	No. events
T = 0	T = 0		600	300
	T = 1		400	100
T = 1	T = 0		200	100
	T = 1		800	200
Scenario 3: partial non-compliance, related to blood pressure at baseline				
Treatment assigned	Blood pressure	Treatment received	No. subjects	No. events
T = 0	High	T = 0	300	225
		T = 1	300	150
	Low	T = 0	320	160
		T = 1	80	20
T = 1	High	T = 0	120	90
		T = 1	480	240
	Low	T = 0	200	100
		T = 1	200	50
Abbreviations: T = 0: control treatment; T = 1: experimental treatment.				

would be observed under perfect compliance.

In scenario 2, compliance differs between treatment arms, but within treatment arms it is a random process. However, the method to account for non-compliance that was described also works if actual treatment status depends on random treatment allocation as well as risk factors for the outcome (the DAG in figure 1c), which is illustrated by scenario 3 (Table 1). The actual treatment received now depends on treatment allocation, but also on blood pressure: regardless of treatment allocation, those with a high pre-treatment blood pressure are more likely to use experimental treatment compared to those with a low blood pressure. Note that the distribution of blood pressure is the same in the two randomization groups. The ITT effect (RD = 480/1000 – 555/1000 = -0.075) again underestimates the treatment effect that would be observed under perfect compliance. However, the ITT effect can be adjusted by the difference in the observed probabilities of receiving experimental treatment between the two treatment allocation groups to obtain the treatment effect under perfect compliance: -0.075 /

(680/1000 – 380/1000) = -0.075 / 0.3 = -0.25.

In scenario 3, an analysis that is stratified by blood pressure will also yield an unbiased estimate of the treatment effect under perfect compliance. However, this obviously requires that blood pressure is actually measured, whereas the analysis outlined in scenario 2 above can also be conducted when blood pressure is unmeasured; hence in the presence of unmeasured confounding.

Assumptions of instrumental variable analysis

The procedure outlined above to account for non-compliance in randomized trials (in scenario 2 above) is a particular form of IV analysis. IV analysis can account for non-compliance in a randomized trial to the extent that there is some contrast in the probability of experimental treatment use between the two randomization groups.[7]

This is the first main assumption of IV analysis, which can be summarized as: an IV predicts treatment status (assumption 1). Several statistical measures are available to quantify the relation between IV and treatment,

including correlation, odds ratio, and proportion of explained variance.[8,9]

The importance of this assumption can be easily understood by looking at figure 2. If the two points are very close to each other, extrapolating the line between the two points will become a very inaccurate process. The further away these points are, the more precise the extrapolation will be. A weak association between IV and treatment becomes less influential in larger samples. Although the two point are close to each other, they have a large precision (due to the large sample size) which will attenuate the instability of the extrapolation. The relation between IV and treatment status is reflected by the arrow between these variables in the DAGs in figure 1.

There are two other main assumptions underlying IV analysis: an IV is independent of confounders of the treatment-outcome relation (assumption 2); and an IV affects the outcome only through the treatment (assumption 3). For the DAGs in figures 1b and 1c, this implies that the observed effect of treatment assignment on the outcome runs completely through the indicated arrows, i.e., there are no unrepresented associations (arrows) between the IV

and the outcome (assumption 3), nor are there any back-door paths from the IV to treatment status or from the IV to the outcome (assumption 2). In a randomized trial, blinding is used in an attempt to meet assumption 3 (which ensures that treatment arms will remain comparable during follow-up), whereas randomization is used to the meet assumption 2.

A causal effect can still be estimated if the assumption of 'no relation between IV and confounders' (assumption 2) can be relaxed. There should be no unmeasured confounders of relation between IV and treatment status and the relation between IV and outcome. In the numerical examples above (scenario 2), the treatment effect was estimated as the ratio of the ITT effect and the relation between IV status and treatment status. If either risk difference is biased (by unmeasured confounding), their ratio may be biased as well. However, in the absence of other biases, if both elements of the ratio are adjusted for measured confounders of those risk differences, the ratio may yield an unbiased estimate of the treatment effect. The assumption of no unmeasured confounders cannot be proven, but it may be falsified in the data.[10]

An observed imbalance in measured confounders within IV strata may suggest that unmeasured confounders are imbalanced as well, thus invalidating IV analysis. Importantly, the bias due to unmeasured confounding can be much larger when conducting IV analysis compared to conventional analysis.[10]

The ratio method outlined above is just one of many possible statistical approaches to IV analysis.

Furthermore, additional assumptions are required to interpret the estimated IV effect as causal, for example assumptions related to homogeneity of the treatment effect.

These additional assumptions as well as more flexible IV analytical methods are beyond the scope of this review and we refer to the literature for more details.[11,12,13,14,15,16,17]

Physician	Prescriptions	IV status based on all prescription $(n_A / n_A + n_B)$	IV status based on previous prescription
Phys 1	A	7/10	-
Phys 1	A	7/10	A
Phys 1	B	7/10	A
Phys 1	A	7/10	B
Phys 1	B	7/10	A
Phys 1	B	7/10	B
Phys 1	A	7/10	B
Phys 1	A	7/10	A
Phys 1	A	7/10	A
Phys 1	A	7/10	A
Phys 2	B	2/7	-
Phys 2	B	2/7	B
Phys 2	B	2/7	B
Phys 2	A	2/7	B
Phys 2	B	2/7	A
Phys 2	B	2/7	B
Phys 2	A	2/7	B

Figure 3: Caption: Definitions of common ways of building the instrumental variable physician's prescribing preference.

Instrumental variables in observational studies

The application of IV analysis can be extended beyond non-compliance in randomized trials. In fact, randomized trials are just one of many possible fields of application. In observational studies, a variable that is not randomly allocated by the investigator, yet fulfils the assumptions of an IV, can act in the same way as random treatment allocation in a randomized trial.

For example, in a study of the relation between HDL-cholesterol levels and myocardial infarction, a genetic polymorphism that increases HDL-cholesterol levels (and does not affect LDL-cholesterol or other cardiovascular risk factors) was used as IV.[18] Genetic polymorphisms are randomly distributed in populations and are in that respect similar to random treatment allocation in a randomized trial (Figure 1d). Studies that make use of this phenomenon are called Mendelian randomization studies.[18,19,20]

In contrast to randomized trials, however, the relation between the genotype (e.g. polymorphisms) and phenotype (e.g. cholesterol-levels) is typically weak in Mendelian randomization studies, therefore requiring (very) large sample sizes,[21,22] as explained above.

Full knowledge of the biological mechanism by which the genetic polymorphism acts (e.g. does the polymorphism only affect HDL-cholesterol levels or also other biomarkers which may affect the risk of the myocardial infarction) is necessary to be confident that the assumptions of IV analysis hold.[23]

Instrumental variables in pharmacoepidemiology

Pharmacoepidemiologic studies are often conducted in large databases of electronic healthcare records, which provide detailed information about for example co-morbidity and co-medication, but often have limited information about health behaviour (e.g., smoking, exercise, and dietary habits). The latter leads to a potential for unmeasured confounding, which may be overcome by IV analysis.

A review of IV analysis in pharmacoepidemiology, published in 2011, identified 5 types of instrumental variables that are typically used: regional variation, facility prescribing patterns, physician preference, patient history / financial status, calendar time.[24] Facility prescribing patterns together with physician preference are together the most commonly used IV in pharmacoepidemiology.

In the remainder we focus on the IV physician prescribing preference (or physician preference).

Figure 1e shows the assumed causal structure of a study using physician preference as an IV The IV physician preference can be defined in different ways (Figure 3). For example, in a study in which two drugs are compared against each other (A vs. B), all subjects treated with either A or B, from a number of participating practices, are enrolled in the study. For each physician the preference can then be defined as the number of prescriptions of drug A (nA) compared to all prescriptions (nA + nB) made by that physician (third column in figure 3). If the preference changes over time, one overall preference per physician may not be appropriate.[25,26]

Instead, for each physician the percentage of prescriptions of drug A can be determined per year, or per quarter, to better account for possible changes over time. Ultimately, the prescription that was issued for the last patient before the current one could be used as a proxy for the preference of a physician at that moment in time.[25,27] If the last patient was prescribed drug A, then apparently the physician's preference at that moment is in favor of drug A (fourth column in figure 3).

Interplay between IV assumptions

When applying IV analysis, researchers must demonstrate, or explicitly argue why, the assumptions of IV analysis hold. It is straightforward to check whether physician preference is indeed related to actual treatment. For example, IV status should predict to a considerable extent the actual treatment status. It is hard to provide cut-points that universally apply, but simulations suggest that for example the odds ratio between a binary IV and a binary treatment in a typical pharmacoepidemiologic study should exceed 2 (note that this value depends on sample size, but not on statistical significance).[9]

The assumption of independence between IV and confounders can be checked at least for the measured confounders, by making a comparison of confounders between levels of physician preference.[10]

According to the DAGs in Figure 1e, both the IV and potential confounders of the treatment-outcome relation affect actual treatment status. This means that if the proportion of explained variation in the treatment due to the IV is relatively large, there is little variation in treatment left that can be attributed to the confounders.[28] And vice versa, if the proportion of explained variation in the treatment due to confounders is relatively large, there is little variation in treatment left that can be attributed to the IV. Hence, in case of strong confounding, any IV that is independent of the confounders will only be weakly related to treatment. Only if the amount of confounding is limited, one may identify a strong IV.

An exception may be a situation in which the confounder-treatment association is relatively weak, yet the amount of confounding is nevertheless substantial due to a very strong confounder-outcome association. Thus, particularly in those situations where IV analysis is needed the most to deal with (strong) unmeasured confounding, IVs will typically be weakly associated with treatment and thus require large sample sizes.

When treatment options are clearly spelled-out in clinical guidelines, any variation in prescribing rates between physicians will likely be small. In those situations in which the preference of a physician can lead to large variations in prescribing behaviour, apparently guidelines aren't that strict, which may be because there is clinical equipoise; i.e. there are no apparent risks or benefits related to one drug compared to the other. Consequently, treatment will probably not be prescribed very selectively, which means that the potential for confounding will be small.

The IV assumption that physician preference affects the outcome only through the treatment cannot be checked in the data. This assumption implies that physicians only differ in their preference for the treatment under study, but they do not differ with respect to all kinds of other aspects (e.g. preferences for concomitant treatments, skills, organization of their practice, etc) that may affect the outcome.[12] If they only differ in that respect, however, differences in preferences will likely be small and hence the IV will be weakly related to treatment status.

On the other hand, in case of really distinct preferences (strong IV), physicians will likely differ in more respects than only their preference for that particular treatment, which impairs the validity of the IV physician preference. Obviously, physicians can be different in terms of e.g. sex and age. But as long as the sex and age of a physician are not related to the outcome, such differences will not affect the validity of the IV.

Conclusion

The use of instrumental variable analysis is clearly indicated in trials with non-compliance and in Mendelian randomization studies.

However, its validity and applicability in observational studies of the effects of (pharmacological) treatments still have to be established. This requires more

applied studies using different types of possible IVs.

For each of these, the assumptions underlying IV analysis have to be thoroughly assessed and those assumptions that cannot be verified using the data have to be debated.[29]

Importantly, physician preference is not the only possible IV for pharmacoepidemiologic studies. Differential implementation of guidelines between (similar!) regions, or evaluating the implementation of guidelines (before-after comparison) may provide valid IVs. We propose that new IVs are considered that allow for estimating unbiased treatment effects of safety and effectiveness in pharmacoepidemiology.

Conflict of interests

References

1. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BH, Leufkens HG, de Boer A. Methods to assess intended effects of drug treatment in observational studies are reviewed. J Clin Epidemiol. 2004; 57 (12):1223-31.
2. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
3. Pearl J. Causal diagrams for empirical research. Biometrika. 1995;82(4):669-688.
4. Hernan MA, Hernandez-Dıaz S. Beyond the intention-to-treat in comparative effectiveness research. Clin Trials. 2012;9(1):48-55.
5. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000; 29(4):722-9.
6. Sexton M, Hebel JR. A clinical trial of change in maternal smoking and its effect on birth weight. JAMA. 1984; 251(7):911-5.
7. Efron B, Feldman D. Compliance as an explanatory variable in clinical trials. J Am Stat Assoc. 1991; 86(413):9-26.
8. 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(430): 443-50.
9. Uddin MJ, Groenwold RH, de Boer A, Belitser SV, Roes KC, Hoes AW, Klungel OH. Performance of instrumental variable methods in cohort and nested case-control studies: a simulation study. Pharmacoepidemiol Drug Saf. 2014; 23(2):165-77.
10. Ali MS, Uddin MJ, Groenwold RHH, Pestman WR, Belitser SV, Hoes AW, de Boer A, Roes KCB, Klungel OH. Quantitative falsification of instrumental variables assumption using balance measures. Epidem-iology. 2014;25(5):770-2.
11. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006;17(4):360-72.
12. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med. 2014; 33(13):2297-340.
13. 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(12):1392-403.
14. Clarke PS, Windmeijer F. Identification of causal effects on binary outcomes using structural mean models. Biostatistics. 2010; 11(4):756-70.
15. Angrist J, Imbens G, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91(434):444-72.
16. Robins J, Rotnitzky A. Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. Biometrika. 2004;91(4):763-83.
17. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet. 2012; 380(9841):572-80.
18. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. BMJ. 2012;345:e7325.
19. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007; 16(4):309-30.
20. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63.
21. Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. Int J Epidemiol. 2013;42(4):1157-63.
22. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol. 2014 Mar 6
23. Vanderweele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. Epidemiology. 2014;25(3):427-35.
24. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. J Clin Epidemiol. 2011;64(6):687-700.
25. 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(3):268-75

26. 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(4):494-502.

27. 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(12):1233-41.

28. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology. 2006; 17(3):260-7.

29. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). Epidemiology. 2013; 24(3): 370-374.

For citation purposes: *Groenwold et al. Instrumental variable analysis in randomized trials with non-compliance and observational pharmacoepidemiologic studies. OA Epidemiology 2014 May 09;2(1):9.*