The handle http://hdl.handle.net/1887/37763 holds various files of this Leiden University dissertation.

**Author:** Boef, Anna Gunnel Christina  
**Title:** Obtaining causal estimates of therapeutic effects in observational studies: the usefulness and validity of physician’s preference as an instrumental variable  
**Issue Date:** 2016-02-10
Chapter 9

General discussion
In this thesis we aimed to investigate the validity of instrumental variable analysis, in particular using physician’s preference as an instrument, to evaluate beneficial or adverse effects of interventions and we aimed to identify the settings and types of questions for which it most useful. In this chapter we present a summary of our main findings, discuss strengths and limitations of our research and consider its implications.

Summary of the principal findings

In Chapter 2 we found substantial variation amongst general practitioners in their treatment decisions when presented with the same set of eight fictitious cases of patients with subclinical hypothyroidism, supporting the existence of physician’s prescribing preference. Further, we found that the deterministic monotonicity assumption (that the instrument is related to treatment monotonically in one direction for all patients) did not hold even in a relatively simple setting, suggesting that this assumption is not plausible when physician’s preference is used as an instrumental variable. In contrast, we found that the stochastic monotonicity assumption (that the instrument is related to treatment monotonically across subjects within strata of a sufficient set of measured and unmeasured common causes of treatment and the outcome) held in the survey data when a different ‘prescription’ of the same general practitioner was used as the instrument. This suggests that a more relaxed version of the monotonicity assumption may be plausible when physician’s preference is used as an instrumental variable.

We subsequently applied physician’s preference based instrumental variable analysis in a clinical epidemiological study of typical size (a few hundred patients) in Chapter 3. We found that estimates of the effect of preoperative corticosteroids on mechanical ventilation time, duration of intensive care and hospital stay, occurrence of infections, atrial fibrillation, heart failure and delirium in elective cardiac surgery patients were similar in direction to estimates from a randomised controlled trial. However, the estimated effects were much larger with uninformative wide confidence intervals. We concluded that the lesser statistical precision of instrumental variable analysis limits its usefulness in a study that might be of sufficient size for conventional analyses - even if a strong and plausible instrument is available.

In Chapter 4 we showed through simulations how the performance of instrumental variable analysis in comparison to conventional analyses, a bias-variance trade-off, depends substantially on sample size. Other determinants are the strength of the instrument and the strength of confounding. We derived an equation that can be used to approximate a ‘threshold’ sample size above which the mean squared error (a summary measure of bias and variance) of instrumental variable analyses will be lower than that of conventional analyses. Further, we showed that substantial sample
sizes will generally be needed for the bias-variance trade-off to be in favour of an instrumental variable analysis in epidemiologic studies.

In Chapter 5, we investigated whether instrumental variable analysis is useful as a sensitivity analysis in studies of adverse effects to assess the presence of confounding. The topic of the study was the comparison of the occurrence of venous thrombosis in users of third generation oral contraceptives vs. second generation contraceptives. In principle, this is an unpredictable unintended effect and we therefore investigated whether an instrumental variable analysis would yield the same estimates as an analysis using standard statistical methods to adjust for confounding (as in principle we would expect little confounding). The study population consisted of new users of second or third generation oral contraceptives, derived from a very large primary care database. We showed that the instrumental variable estimates (using general practitioner’s preference as an instrument) of the effect of third versus second generation oral contraceptives on occurrence of venous thromboembolism were similar to estimates from conventional analyses. If anything, the conventional analysis seemed to be more conservative than the instrumental variable analysis. However, even in this very large study population the variance of the instrumental variable estimates was very large, due to the relatively rare outcome. Further, the analysis was complicated because changes in both prescribing preference and patient characteristics over time resulted in violation of the independence assumption and necessitated an adjusted instrumental variable analysis. We concluded that major confounding was unlikely due to the similarity of the estimates obtained under different sets of assumptions.

In Chapter 6 we recommended to report the association between the instrument and the outcome before performing any formal instrumental variable analysis, which amounts to a conventional epidemiologic analysis. This is important because it shows the association which is subsequently extrapolated in the formal instrumental variable analysis. This recommendation was proposed in response to a paper outlining guidelines for the reporting of instrumental variable analysis.

We then shifted our focus from physician’s preference based instrumental variable analysis to Mendelian randomisation studies. In Chapter 7 we reviewed methodological approaches used in Mendelian randomisation studies and found that the specific methods used vary widely, falling broadly into three categories: 1) using genetic information as a proxy for the exposure without further estimation, 2) performing an instrumental variable analysis; 3) comparing the observed with the expected genotype-outcome association. Further we found that Mendelian randomisation studies often insufficiently discuss underlying assumptions and
report statistical methods for IV analysis. In a certain sense the fact that Mendelian randomisation studies are also instrumental variable studies (with or without a formal instrumental variable analysis) is often disregarded. We therefore devised a checklist for the reporting of Mendelian randomisation studies. Finally, in Chapter 8 we explained that collider-stratification bias may exist if Mendelian randomisation studies are performed in elderly populations, as both the genetic variant used as an instrument and other causes of the outcome may be causally related to survival up to the age at which the population is selected.

Implications and recommendations

Instrumental variable analysis as a primary analysis

Instrumental variable analysis for estimation of therapeutic effects has the greatest potential in situations in which very large datasets are available and in which there is substantial confounding by indication and also little available information on these confounding factors (see conclusion of Chapter 4). However, if the direction of this confounding is predictable, sensitivity analyses could alternatively be used to derive a plausible range of the treatment effect. Instrumental variable analysis would therefore more specifically be suited to situations in which there is confounding with such complexity that the direction and magnitude of the resulting bias is unpredictable. Further we remark that although instrumental variable analysis is of most value in situations with substantial unmeasured confounding, paradoxically, substantial unmeasured confounding limits the potential strength of any instrument (as this confounding will determine a substantial part of the variation in exposure).

Instrumental variable analysis as a sensitivity analysis

In case of unpredictable adverse effects, such as the increased risk of venous thromboembolism of 3rd generation oral contraceptives in comparison to 2nd generation oral contraceptives, confounding by (contra-)indication is unlikely and instrumental variable analysis would therefore not be particularly suitable as a primary analysis. However, it may be used as a sensitivity analysis: if results of an instrumental variable analysis are similar to those of the conventional analyses, this supports the notion that there is little confounding by contra-indication (provided, of course, that instrumental variable assumptions hold, and that suitably large databases exist). Instrumental variable analysis may also have a role as a sensitivity analysis in studies of intended effects. Looking at the same data under different sets of assumptions can contribute to the understanding of causal effects.

Types of therapeutic question for which instrumental variable analysis is most useful

Instrumental variable analysis for estimation of therapeutic effects seems primarily
suited to therapeutic decisions in which there are two clearly defined alternatives. Davies et al compare 5 different treatment options (2 selective COX-2 inhibitors and 3 non-specific NSAIDs) as a sensitivity analysis in a study primarily comparing the 2 drug classes. However, examples of instrumental variable studies in which more than two treatment options are compared are rare, although there are many situations in which there are more than two alternatives to consider in a treatment decision. Even if only two options are primarily of interest for the research question at hand, exclusion of patients who received other treatment options may result in inclusion of different subsets of the patient population for different physicians, threatening the validity of the independence assumption (i.e. that there is no confounding of the instrument-outcome relation). This may have been an issue in Chapter 5, in which the introduction of drospirenone-containing oral contraceptives resulted in an additional option besides 2nd and 3rd generation oral contraceptives. Whereas formerly the comparison of 2nd and 3rd generation oral contraceptives will have included (nearly) all women who started using a combined hormonal oral contraceptive, the introduction of an additional option will have reduced this population in a manner which is not necessarily random.

Instrumental variable analysis of randomised controlled trials
In the ideal randomised controlled trial with complete compliance and complete follow-up, the treatment effect estimated is the average treatment effect in the population. This changes when compliance is incomplete, in which case an intention-to-treat effect is usually estimated. This estimates the effect of assigning the treatment rather than the effect of taking the treatment, and is therefore not always the effect of interest. The intention-to-treat effect is a conservative estimate of the effect of taking treatment, i.e. biased towards the null. This can be particularly problematic in studies of adverse effects or in non-inferiority trials. An as-treated analysis or a per-protocol analysis on the other hand essentially negates the randomisation, resulting in incomparable populations and an estimate without a clear causal interpretation. The analysis of RCTs with non-compliance using methods common to observational studies has been advocated, in order to obtain a valid estimate of the effect of taking treatment (besides the intention-to-treat estimate of the effect of assigned treatment). One such way is to perform an instrumental variable analysis, using treatment arm as an instrumental variable. This will usually be a strong instrumental variable, as treatment arm strongly predicts treatment unless compliance is very low. In the context of a randomised trial, the deterministic monotonicity assumption, i.e. the absence of defiers (subjects who would take the opposite of what they are assigned to in either treatment arm) is usually reasonable. Under this assumption the instrumental variable analysis estimates the effect within the compliers: those who take the treatment to which they
are assigned. Importantly, the interpretation of this estimate therefore differs from
the average treatment effect in the population and from the intention-to-treat effect.
The estimate can for example be obtained using the Wald estimator, which divides the
intention-to-treat effect by the difference in probability of being treated with the study
treatment. For time-varying treatments (time-varying adherence) more complex
instrumental variable methods such as g-estimation can be used. Examples of
instrumental variable analysis in RCTs are as yet rare. One example is a randomised
trial investigating the effect of yoga (in addition to usual general practitioner care) on
chronic low back pain. The intention-to-treat estimate of the effect of assignment to
yoga in addition to usual care on 3-month Roland Morris Disability Questionnaire
score was -2.17 (95% CI -3.31; -1.03), whereas the complier-average-causal-effect
estimate of attending at least one yoga class was -2.45 (-3.67; -1.24) and the complier-
average-causal-effect estimate of attending all 12 offered yoga classes was -3.30 (-4.90;-
1.70).

Physician’s preference as an instrumental variable

Physician’s preference can specifically be useful as an instrumental variable in situations
in which there are no stringently applied medical practice guidelines, i.e. when there
is room for preference to play a role in treatment decisions. An example of such a
situation is the decision whether to treat patients with subclinical hypothyroidism. The
Dutch general practitioners’ guideline, for example, does not recommend treatment of
subclinical hypothyroidism in general, but states that general practitioners may choose
to try levothyroxine treatment and evaluate whether symptoms improve. We showed
in Chapter 2 that there was substantial variation in treatment decisions among general
practitioners presented with the same subclinical hypothyroidism cases. In these
survey data, this variation remained after adjusting for characteristics of the GP and
their patient population, which was reassuring with regard to the main instrumental
variable assumptions.

However, a note of caution is warranted with regard to the exclusion restriction
assumption. In applications of physician’s preference based instrumental variable
analysis, the treatment choice by the physician for (one or more) previous patients
is usually used as an estimate of physician’s preference, because this preference is not
a directly measurable characteristic. Yet situations may occur in which instrumental
variable assumptions hold for the underlying preference, but not for an estimate of this
preference based on previous prescriptions. We will explain this using the directed
acyclic graph depicted in Figure 1. In this figure the treatment of the previous patient
Z* is used as a proxy for underlying preference of Z. If there is variation in the case-mix
M of the different physicians in the study, the current patient and the previous patient
are likely to be more similar than any random two patients from the entire study population. The instrument $Z^*$ will then be related to characteristics of the current patient $C$ and $U$, through case-mix $M$ and characteristics of the previous patient $V$, violating the independence assumption. Note that underlying preference $Z$ is still a valid instrument, because $Z^*$ acts as a collider, blocking the path between $Z$ and patient characteristics $C$ and $U$. In the setting of Chapter 3, we think this is unlikely to have occurred, as patients will have been treated by the anaesthesiologist on duty and systematic differences in case-mix are therefore unlikely to exist. For studies in which preference of general practitioners is used as an instrument, differences in case-mix are likely (e.g. due to geographical variation in socio-economic status and demographic characteristics) and the confounding described may therefore threaten the validity of previous prescriptions of the GP as an instrument.

What does an instrumental variable analysis estimate?
As stated previously, in the ideal randomised controlled trial with complete compliance and complete follow-up, the treatment effect estimated is the average treatment effect in the study population. If the study population is representative of the population of interest, the estimate represents the average treatment effect in the population of interest. However, due to inclusion and exclusion criteria, RCT populations are often
not representative of the population of interest, which limits the generalisability of results. An often stated advantage of observational studies over randomised controlled trials (RCTs) is the greater generalisability of the results, because subjects who are unlikely to be included in RCTs can be part of the study population in an observational study. Conventional methods to adjust for confounding in observational studies results in estimates which represent average effects in the study population conditional on the confounders for which they have adjusted. For results of an instrumental variable study the question to whom the results apply and how to interpret the estimate is more complicated. It depends on the assumption which is made in order to obtain a point estimate (the “fourth” assumption, in addition to the three main instrumental variable assumptions). Under the assumption of homogeneity of treatment effects the results apply to the entire study population. However, if homogeneity of treatment effects is not realistic, another assumption is necessary in order to obtain an interpretable point estimate. This is often some form of the monotonicity assumption, in which case the question to which population the results apply is more complex.

The exact interpretation of the instrumental variable estimate depends on the version of the monotonicity assumption. The situation with a binary instrument and the deterministic monotonicity assumption is relatively easy: the point estimate represents the average effect among the ‘compliers’. However, the compliers in the study population cannot be identified: all subjects only experienced the treatment they received at the actual value of the instrument and the treatment they would have received at the counterfactual value of the instrument is unknown. A description of the distribution of the characterisation of the compliers is possible however, as described by Angrist and Pischke. Under the stochastic monotonicity assumption the question to which patients the results apply becomes more difficult, as the estimate is a weighted average of treatment effects. Clearly, the results do not apply to those subjects who evidently would have received the same treatment regardless of the value of the instrument (i.e. regardless of the preference of their physician – e.g. because of overriding medical reasons). The degree to which results apply to other patients depends not only to the proportion of the study population which consisted of that type of patient, but also on the strength of the instrument for the specific type of patient: the stronger the instrument, the higher the relative contribution of that specific type of patients to the estimate. A characterisation of the strength of instrumental variable weighted average treatment effect (SIVWATE) population would be rather more difficult than the characterisation of the compliers described previously.
In part because of the difficulty in the interpretation of the point estimate from an instrumental variable analysis and the additional assumption required, the reporting of bounds of the instrumental variable estimate has been advocated in reporting guidelines for instrumental variable analysis. These bounds represent the upper and lower limits of the average causal effect in the study population. Balke and Pearl describe the calculation of these bounds and how they can be narrowed under different assumptions. In practice these bounds will generally be so disparate that they are uninformative. The main value in calculating bounds lies in the subsequent explicit decision on which fourth assumption is plausible and how the point estimate consequently should be interpreted.

Reflections on Mendelian randomisation and subsequent considerations for other forms of instrumental variable analysis

There is discussion on whether Mendelian randomisation should be viewed as instrumental variable analysis with genetic instrumental variables. One point of discussion is that some studies qualify as Mendelian randomisation studies but do not perform a formal instrumental variable analysis. More recently the question if and when formal instrumental variable analysis should be performed in Mendelian randomisation studies has been addressed by VanderWeele. Most importantly this depends on the definition of the exposure and the consequences of this definition for the validity of the main instrumental variable assumptions: in some situations an estimate of the effect of the genetic instrument on the outcome may be valid, while no valid IV estimate of the effect of the exposure on the outcome can be obtained. One example discussed by VanderWeele is that the effect of certain genetic variants on smoking behaviour will not be entirely captured by the effect on the number of cigarettes smoked per day: IV analysis estimating the effect of the number of cigarettes per day on lung cancer using these variants as instruments will then be biased because there are additional pathways from the variants to lung cancer (via other aspects of smoking behaviour). Investigating the association between the variants and lung cancer is a valid test of the presence of an effect of smoking behaviour in a more general sense on lung cancer.

The question if and when a formal instrumental variable analysis should be performed applies not just to MR studies, but also to other instrumental variable studies. In case of studies of therapeutic effects, adequately capturing all aspects of the association between the instrument and the exposure may not be as difficult as in Mendelian randomisation studies. However, the outcome of interest in a study of therapeutic effects is often a binary or survival outcome. Formal IV analysis methods for such outcomes are not yet well-established and obtaining a valid quantitative estimate of
effect of the therapy on the outcome is therefore difficult, but estimation of the effect of the exposure on the outcome can serve as a test of causality. For many therapeutic questions quantification of the therapeutic effect will be important however, and without a formal instrumental variable analysis only the existence and direction of the effect can be evaluated.

Based on our findings in Chapter 7 we would argue that for some aspects of Mendelian randomisation studies the instrumental variable perspective can be valuable. For example, reporting guidelines for instrumental variable analysis largely apply to Mendelian randomisation studies and can aid in improving the reporting of future MR studies. Furthermore, many Mendelian randomisation studies obtain an instrumental-variable type estimate of the effect of the exposure on the outcome using statistical methods which are not well-established (e.g. estimation of an odds ratio using a method similar to the Wald estimator). It is important that those who perform such studies are aware that the problems and limitations of instrumental variable analyses with binary or survival outcomes also apply to MR studies using these methods.

Conclusions
We set out to investigate the validity and usefulness of instrumental variable analysis, in particular using physician’s preference as an instrument, in clinical epidemiological studies. We aimed to expose potential problems and limitations of this method and to identify the settings and types of questions for which it is most useful. By exploring several aspects of the method in both applications using existing data and in simulation studies we came to a number of conclusions. Instrumental variable analysis can be of value as a primary analysis in very large epidemiological studies with substantial unmeasured confounding (of unpredictable direction and magnitude) and a strong instrument. For studies of a more typical size in clinical epidemiology, e.g. several hundreds of patients an instrumental variable estimate will generally be uninformatively imprecise. A more broadly suitable role for instrumental variable analysis could be as a sensitivity analysis, for example to assess the presence of confounding (by contra-indication) in studies of adverse effects. Treatment preference differences exist between physicians, independent of characteristics of these physicians and their patient populations. Ascertaining that IV assumptions hold not only for underlying physician’s preference, but also for the estimate of preference used as a proxy instrument, is paramount. For physician’s preference as an instrument the stochastic monotonicity assumption may be a plausible assumption for obtaining a point estimate. Viewing Mendelian randomisation studies as instrumental variable studies can aid in improving the reporting of Mendelian randomisation studies, even if no formal instrumental variable analysis is performed.
Reference List

