Journal of Clinical Epidemiology 67 (2014) 1258−1264

Journal of Clinical Epidemiology

# Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding

Anna G.C. Boef[a,*], Olaf M. Dekkers[a,b], Jan P. Vandenbroucke[a], Saskia le Cessie[a,c]

[a]*Department of Clinical Epidemiology, C7-P, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands*
[b]*Department of Endocrinology and Metabolic Diseases, Leiden University Medical Centre, Leiden, The Netherlands*
[c]*Department of Medical Statistics and Bioinformatics, Leiden University Medical Centre, Leiden, The Netherlands*

Accepted 26 May 2014; Published online 12 August 2014

## Abstract

**Objectives:** Instrumental variable (IV) analysis is promising for estimation of therapeutic effects from observational data as it can circumvent unmeasured confounding. However, even if IV assumptions hold, IV analyses will not necessarily provide an estimate closer to the true effect than conventional analyses as this depends on the estimates' bias and variance. We investigated how estimates from standard regression (ordinary least squares [OLS]) and IV (two-stage least squares) regression compare on mean squared error (MSE).

**Study Design:** We derived an equation for approximation of the threshold sample size, above which IV estimates have a smaller MSE than OLS estimates. Next, we performed simulations, varying sample size, instrument strength, and level of unmeasured confounding. IV assumptions were fulfilled by design.

**Results:** Although biased, OLS estimates were closer on average to the true effect than IV estimates at small sample sizes because of their smaller variance. The threshold sample size above which IV analysis outperforms OLS regression depends on instrument strength and strength of unmeasured confounding but will usually be large given the typical moderate instrument strength in medical research.

**Conclusion:** IV methods are of most value in large studies if considerable unmeasured confounding is likely and a strong and plausible instrument is available. 

*Keywords:* Instrumental variable; Observational studies; Confounding; Variance; Simulation study; Therapeutic effects

## 1. Introduction

Conventional methods to estimate therapeutic effects from observational data are often inherently affected by residual confounding because of unmeasured patient risk factors for which they cannot adjust. A potentially promising tool for estimation of therapeutic effects from observational data that may circumvent this problem is instrumental variable (IV) analysis. This method requires the identification of a variable that determines the probability of treatment but is not in other ways associated with the outcome under study and thereby mimics randomization. Expressed more formally, an instrument must fulfill three main assumptions: (1) the instrument is associated with the exposure (treatment), (2) the instrument does not affect the outcome in any other way other than through the exposure (exclusion restriction), and (3) the instrument and outcome do not share causes (independence assumption) [1−4]. The aforementioned assumptions allow estimation of bounds of the treatment effect [3,5]. One additional assumption that allows a point estimate to be obtained is the assumption of no heterogeneity of treatment effects, in which case, the IV analysis estimates the average treatment effect in the population [3,5]. Note that in the case of heterogeneity, alternative assumptions can be made, but this is beyond the scope of this article. Examples of instruments used in studies of therapeutic effects include regional variation in treatment rates (i.e., probability of treatment depends on area of residence) [6] and physician prescribing preference [7−9]. In etiologic studies, Mendelian randomization, which uses genetic information as an IV, is increasingly used [10].

Violations of the exclusion restriction and independence assumption will lead to biased IV estimates [5,7,11]. If those assumptions hold, the IV estimator will be asymptotically unbiased [1,11]. In contrast, the bias of ordinary least squares (OLS) linear regression depends on the amount of residual confounding. However, whether IV analysis effect estimates can be expected to be closer to the true effect than estimates from conventional analysis depends on not only the bias but also the variance of the estimates (larger variances leading

Conflict of interest: The authors declare no conflict of interest.

* Corresponding author. Tel.: +31-71-5261510; fax: +31-715266994.
*E-mail address*: a.g.c.boef@lumc.nl (A.G.C. Boef).

0895-4356/$ - see front matter 
http://dx.doi.org/10.1016/j.jclinepi.2014.05.019

**What is new?**

- Very large sample sizes are usually required for instrumental variable (IV) analysis to provide an estimate closer on average to the true effect than estimates from standard regression.
- We derived an equation that can approximate the threshold sample size above which IV estimates are closer on average to the true effect than standard regression.
- In deciding between IV analysis and conventional analyses, researchers should take into account not only the strength of the instrument and the level of confounding but also the available or feasible sample size.

to higher probability of deviating estimates). The variance of estimates from IV methods like two-stage least squares (2-SLS) regression is much larger than from linear regression at a given sample size because IV methods involve two estimation stages instead of one [12].

IV methods have been applied in large pharmacoepidemiological databases, typically exceeding 10,000 patients. However, study populations in clinical research practice are often much smaller. Although in principle, the large variance of the IV estimate at smaller sample sizes may not influence the validity of the IV estimates, it does affect how informative and useful the IV estimate is. It translates into a very wide confidence interval (CI), and the mean squared error (MSE) of the IV estimate may be much larger than that of the biased conventional estimate [1]. The influence of sample size on the error of IV estimates has been investigated in conjunction with violations of IV assumptions [13]. In contrast, we will focus on the ideal scenario in which the exclusion restriction and independence assumptions hold to focus on the role of sample size, confounding, and strength of instrument. Using theoretical derivations and simulations, we will investigate the influence of sample size on how OLS linear regression estimates and 2-SLS IV regression estimates compare in terms of MSE (which incorporates both the bias and the variance of the estimates), depending on instrument strength and level of confounding.

## 2. Two-SLS IV analysis

Two-SLS IV regression involves two linear regression steps. The first-stage linear regression is used to obtain predicted probabilities of treatment for each patient, based on the instrument. Covariates can be included, giving predicted probabilities of treatment conditional on the instrument and these observed covariates. The independence assumption then states that the instrument is not related to patient prognosis given these covariates [2]. The second stage is a regression of the outcome on these predicted treatment probabilities (and covariates if included), thereby providing an estimate of the effect of the treatment on the outcome [2,7,14]. For continuous outcomes, the obtained effect estimate is a mean difference and for binary outcomes a risk difference.

The variance of the 2-SLS estimate is

$$\mathrm{var}\left(\widehat{\beta}_{\mathrm{IV}}\right) = \frac{\sigma^2_{Y.X,C}}{n\sigma^2_{X.C}\cdot\rho^2_{X,Z.C}},$$

where $\rho_{X,Z.C}$ is the partial correlation between the instrument $Z$ and the exposure $X$ given covariates $C$, that is, the strength of the instrument, and $Y$ is the outcome [11,15]. The variance is therefore $1/\rho^2_{X,Z.C}$ times larger than the variance of an OLS linear regression estimate. This implies that the CI for the 2-SLS estimator is $1/\rho_{X,Z.C}$ times wider than the CI of the OLS estimator. For example, for a moderately strong instrument with a correlation between instrument and exposure of 0.2, the CI for the 2-SLS estimator will be fivefold wider than the CI of the OLS estimator. If IV assumptions hold that the 2-SLS estimates are asymptotically unbiased: bias will exist in finite samples and depends on the sample size and strength of the instrument. This is known as small sample bias [5,11], finite sample bias [1], or weak instrument bias [16]. The partial $F$-statistic of the first-stage regression provides an indication of the magnitude of the small sample bias: generally small sample bias is negligible at an $F$-statistic above 10 [11].

### 2.1. MSE: a summary measure for bias and variance

The MSE measures the squared average deviation of an estimated effect from the true effect. It is equal to

$$\mathrm{MSE} = E\left[\left(\widehat{\beta} - \beta\right)^2\right] \quad (1)$$

in which $E$ denotes expectation, $\widehat{\beta}$ is the estimated treatment effect, and $\beta$ is the true treatment effect. It can be shown that the MSE is the sum of the variance and the squared bias of an estimate. It is a measure of how far on average the effect estimate is from the true effect. Comparison of the MSEs of the different analysis methods therefore indicates which estimate is closest on average to the true effect.

### 2.2. Calculation of a sample size at which IV outperforms OLS on MSE

The trade-off between the larger bias of the OLS estimates and the larger variance of the IV estimates means that OLS estimates will be closer on average to the true effect at small sample sizes, but IV estimates will eventually be closer on average to the true effect as sample size increases. We derived Equation (2) (the derivation is provided in eAppendix 1 of the Appendix at www.jclinepi.com) to calculate