

# Instrumental Variable Analysis in Epidemiologic Studies: An Overview of the Estimation Methods

Uddin MJ<sup>1</sup>, Groenwold RH<sup>1,2</sup>, Ton de Boer<sup>1</sup>, Belitser SV<sup>1</sup>, Roes KC<sup>2</sup> and Klungel OH<sup>1,2\*</sup>

<sup>1</sup>Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, The Netherlands

<sup>2</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

## Abstract

Instrumental variables (IV) analysis seems an attractive method to control for unmeasured confounding in observational epidemiological studies. Here, we provide an overview of the estimation methods of IV analysis and indicate their possible advantages and limitations. We found that two-stage least squares is the method of first choice if exposure and outcome are both continuous and show a linear relation. In case of a nonlinear relation, two-stage residual inclusion may be a suitable alternative. In settings with binary outcomes as well as nonlinear relations between exposure and outcome, generalized method of moments (GMM), structural mean models (SMM), and bivariate probit models perform well, yet GMM and SMM are generally more robust. The standard errors of the IV estimate can be estimated using a robust or bootstrap method. All estimation methods are prone to bias when the IV assumptions are violated. Researchers should be aware of the underlying assumptions of the estimation methods as well as the key assumptions of the IV when interpreting the exposure effects estimated through IV analysis.

**Keywords:** Instrumental variables; Estimation method; Unobserved confounding; Epidemiology; Statistical methods; Observational studies; Causal inference

## Introduction

Instrumental variable (IV) analysis has primarily been used in economics and social science research, as a tool for causal inference, but has begun to appear in epidemiologic research over the last decade to control for unmeasured confounding [1-6]. An IV is a variable that can be considered to mimic the treatment assignment process in a randomized study [7-10]. IV analysis generally involves in a two-stage modelling approach to estimate the exposure effects. In the first stage, the effect of the IV on exposure is estimated, whereas in the second stage, outcomes are compared in terms of predicted exposure rather than the actual exposure [11]. To value the estimates obtained through IV analysis, it is important to understand the underlying methodology of the estimation methods in the IV analysis.

Over the last decade several reviews of IV analysis were published, covering various aspects including the key assumptions, estimating parameters, possible IVs, estimation methods, reporting of the results, and the use of IVs in comparative effectiveness research [3,4,12-23]. We summarized these reviews in Table 1. However, none of these articles included all possible estimation methods of IV analysis. Hence, we aimed to provide an overview of the estimation methods and to indicate their possible advantages and limitations. After a general introduction to the assumptions underlying IV analysis, we will describe the methods that have been used in IV studies in medical research.

## Instrumental variables

The IV is an observed variable, which is related to exposure and only related to the outcome through exposure. This resembles a randomized trial, in which treatment allocation typically almost perfectly coincides with the actual treatment received and (in case of a double blind trial) treatment assignment only affects the outcome through the received treatment (hence the term pseudo-randomisation that is used for IV methods). This implies that an IV is neither directly nor indirectly (e.g. through observed or unobserved confounders) associated with the outcome [6,18,24]. Therefore, all observed and unobserved

confounders should on average be equally distributed among different levels of the IV (similar to a randomized trial). These assumptions are illustrated in Figure 1. Along with these basic assumptions, there are other assumptions (i.e., homogeneous treatment effects, monotonicity) that are needed for point identification of IV estimates [14,19].

## Notation

Throughout this article, we use the following notation: Y denotes the outcome, X denotes exposure, and Z denotes the IV. C and U denote the (one or more) observed and unobserved confounding variables, respectively.  $X_0$  denotes the predicted value of exposure. Finally,  $\hat{\beta}_v$  indicates the IV estimator, i.e., the estimator of the causal relation between exposure and outcome.

## Estimation method of IV analysis

### Ratio estimator (RE)

In a study with a single binary IV, the RE (also called Wald [25] or grouping estimator) can be applied and which is expressed as:

$$\hat{\beta}_{IV} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0} \quad (1)$$

$$\hat{\beta}_{IV} = \frac{E[Y | Z = 1] - E[Y | Z = 0]}{p(X = 1 | Z = 1) - p(X = 1 | Z = 0)} \quad (2)$$

$$\hat{\beta}_{IV} = \frac{p(Y = 1 | Z = 1) - p(Y = 1 | Z = 0)}{p(X = 1 | Z = 1) - p(X = 1 | Z = 0)} \quad (3)$$

**\*Corresponding author:** Klungel OH, Division of Pharmacoepidemiology and Clinical Pharmacology, University of Utrecht, Utrecht, Netherlands, Tel: +31685384692, Fax +31-30 253 9166; E-mail: [O.H.Klungel@uu.nl](mailto:O.H.Klungel@uu.nl)

**Received** January 06, 2015; **Accepted** March 08, 2015; **Published** March 15, 2015

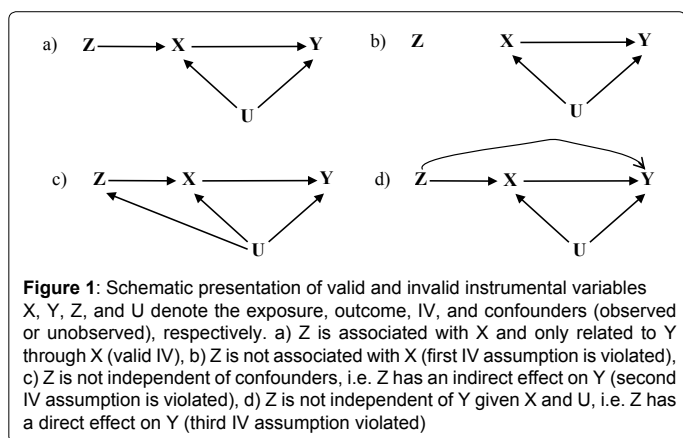
**Citation:** Uddin MJ, Groenwold RH, Ton de Boer, Belitser SV, Roes KC, et al. (2015) Instrumental Variable Analysis in Epidemiologic Studies: An Overview of the Estimation Methods. Pharm Anal Acta 6: 353. doi:10.4172/2153-2435.1000353

**Copyright:** © 2015 Uddin MJ, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Author	Publication year	Journal name	Title	Main features
Greenland	2000	International Journal of Epidemiology	An introduction to instrumental variables for epidemiologists	-basic introduction with an empirical example -link with randomized studies with non-compliance -estimated bound for the exposure effects
Martens et al.	2006	Epidemiology	Instrumental variables: application and limitations	-fundamental issues are described with several practical details using graphical representation
Hernan and Robins	2006	Epidemiology	Instruments for causal inference: an epidemiologists dream?	-overview of IV analysis with explanation of several key assumptions -highlights limitations and emphasis on estimating parameters of IV analysis
Rassen et al.	2009	Journal of Clinical Epidemiology	Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships	-demonstrates how IV analysis arises from an analogous but potentially impossible RCT design -shows estimation of effects with an empirical example
Rassen et al.	2009	Journal of Clinical Epidemiology	Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance	-assesses the overall relationship between strength and imbalance of confounders between IV categories with an empirical example -assesses several possible IVs
Rassen et al.	2009	American Journal of Epidemiology	Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes	-reviews commonly used IV estimation methods for binary outcome and compared them in empirical examples
Brookhart et al.	2010	Pharmacoepidemiology and Drug Safety	Instrumental variable methods in comparative safety and effectiveness research	-guidance on reporting of IV analysis with an empirical example
Clarke and Windmeijer	2010	Journal of American Statistical Association	Instrumental variable estimators for binary outcomes	-estimation methods of IV analysis for binary outcome with mathematical descriptions
Chen and Briesacher	2011	Journal of Clinical Epidemiology	Use of instrumental variable in prescription drug research with observational data: a systematic review	-review of practice of IV analysis in epidemiology
Palmer et al.	2011	American Journal of Epidemiology	Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization	-overview of commonly used IV estimation methods for continuous exposure -empirical example of Mendelian randomization study
Davies et al.	2013	Epidemiology	Issues in the reporting and conduct of instrumental variable studies: a systematic review	- review of practice of IV analysis in epidemiology -focus on target parameter (e.g. RD, OR) -reviews methods used to estimate standard errors - proposes a checklist of information to be reported by studies using instrumental variables
Swanson and Hernan	2013	Epidemiology	Commentary: How to report instrumental variable analyses (suggestions welcome)	-provided flow chart for reporting of IV analyses

Baiocchi et al.	2014	Statistics in Med	Instrumental variable methods for causal inference	-generic tutorial and guidelines of IV analysis with an empirical example
Garabedian et al.	2014	Annals of Internal Medicine	Potential Bias of Instrumental Variable Analyses for Observational Comparative Effectiveness Research	-this review found that the results of IV analyses may be biased substantially if the IV and outcome are related through an unadjusted third variable: an "IV-outcome confounder"  - the authors caution against overreliance on IV studies comparative effectiveness research

Table 1: Introductory and review articles of instrumental variable analysis in Epidemiologic studies (2000-2014)



where  $\bar{x}_1$  and  $\bar{x}_0$  are the mean of  $y$  and  $x$ , respectively, when  $Z=0$  and  $\bar{y}_0$  and  $\bar{x}_0$ , when  $Z=0$ ;  $p(Y = 1 | Z = 1) - p(Y = 1 | Z = 0)$  is the difference in probability of being exposed for  $Z=1$  and  $Z=0$ ; and  $p(Y = 1 | Z = 1) - p(Y = 1 | Z = 0)$  is the risk difference of an event between  $Z=1$  and  $Z=0$ . Equation (1) is suitable for settings with continuous exposure and continuous outcome, equation (2) for binary exposure and continuous outcome [26,27], and equation (3) for binary exposure and binary outcome.

The RE is a simple estimation method to estimate the exposure effects from the IV analysis. However, it is not suitable for multiple IVs or in a situation when measured confounders need to be adjusted for in the analysis.

### Two-stage least squares method (2SLS)

The best known two-stage method for IV analysis is the 2SLS method which is traditionally used in IV analyses [10,28,29]. Unlike ratio estimators, this method is able to adjust any possible measured confounders. The 2SLS estimator can be obtained by the following models:

$$X_i = \alpha_0 + \alpha_z Z_i + \alpha_c C_i + \varepsilon_{1i}; \text{ for } i = 1, 2, \dots, n \quad [4]$$

$$Y_i = \beta_0 + \beta_{IV} \hat{X}_i + \beta_c C_i + \varepsilon_{2i}; \text{ for } i = 1, 2, \dots, n \quad [5]$$

The first model estimates the effect of the IV on exposure, whereas in the second model outcomes are compared in terms of predicted exposure rather than the actual exposure. The latter model yields the estimated parameter,  $\hat{\beta}_v$ , which is the IV estimator. For a single IV, the  $\hat{\beta}_v$  is equivalent to the estimators in the equations (1), (2), and (3). In case of multiple IVs, information on these IVs can be simultaneously

incorporated in model (4). Then,  $\hat{\beta}_v$  is the weighted average of the ratio estimators [30]. For multiple IVs, 2SLS provides biased estimates [30-32] and another method, e.g., limited information maximum likelihood (LIML), [33] can be an alternative. One of the conditions of this method is that the error term should be homoscedastic (homogeneity of variance). However, in case of heteroscedasticity, other methods (e.g., generalized method of moments) can be considered [34]. Moreover, the 2SLS may produce biased results in the case of binary variables or non-linear relation between exposure and outcome (Table 2).

### Linear probability model (LPM)

This method is a particular form of the 2SLS in which the outcome, exposure, and IV are binary and provides exposure effects on the risk difference scale. When there is a single binary IV, the estimator can be expressed as in equation (3) [13,35-37].

LPM is a simple technique to estimate the parameter and interpret as the regression coefficients based on linear regression. However, in linear IV analysis, LPM may provide ambiguous results because the common technique of linear IV is designed for a continuous response [38]. It should be noted that the LPM of binary exposure and outcome may produce predicted values outside of the 0-1 range [28]. Hence, for rare binary outcomes, some predicted probabilities may become negative [39]. In addition, the probability of success increases linearly with exposure, that is, the marginal or incremental effect of exposure remains constant [37], which is logically impossible for binary outcomes [14].

### Two-stage predictor substitution (2SPS)

The two-stage predictor substitution is an extension of the 2SLS to nonlinear models, which targets a marginal (population-averaged) odds ratio [36,40-42]. In the first-stage, a nonlinear least squares method (NLS) or any other consistent estimation technique is used to estimate the relation between the IV and exposure [43]. Then, the predicted exposure status from the first-stage model replaces the observed exposure as the principal covariate in the second-stage model on the outcome [43,44]. For a continuous exposure and outcome, 2SPS and 2SLS show similar results [24,36].

### Two-stage residual inclusion (2SRI)

2SRI (also called control function estimator) [45] is another two-stage method and was first suggested by Hausman [46]. The general notion of the 2SRI is to include the error terms (residuals) from the first-stage model as an additional variable along with the exposure in the second-stage model [47]. The models in the first and second-stage can be either linear or nonlinear models. In case of linear models, the 2SRI estimate is equivalent to the 2SLS and 2SPS estimates [44,48].

Method	Basic notion	Exposure effects	Strength	Limitation
Ratio estimator (RE)	-the RE is appropriate when only one IV	-RD, RR, OR	-simple estimation method -with a single binary IV and no other confounders, 2SLS = RE	-not suitable for multiple IVs -does not allow to adjust confounders - may not consistent for the causal OR
Two-stage least squares (2SLS)	-linear models without making parametric assumptions on the error terms -for multiple IVs, IV estimator is the weighted average of the ratio estimators	-estimator similar as classical regression	-natural starting point of IV analysis -the estimate asymptotically unbiased -widely used for binary exposure and outcome and provides the exposure effect on risk difference scale -unlike RE, it is able to adjust the possible measured confounders	-show biased results in binary cases or in the case of non-linear models -for multiple IVs, 2SLS estimator is biased and hence limited information of maximum likelihood method would be an alternative -for smaller sample sizes, limited information maximum likelihood estimator is more efficient and consistent than 2SLS -IV and 2SLS are a special case of GMM; however both yield the same results in the case of homoscedastic errors variance
Linear probability models (LPM)	-applied for binary outcome, exposure, and IV, the data are modelled using linear functions -for a single binary IV, the estimator equivalent to the RE	RD	-simple to estimate and interpret as the regression coefficients -the RD is consistent for the ACE	- sometimes predicted probabilities outside of the 0–1 range and for rare outcomes this may become negative - assumes the marginal/incremental effect of exposure remains constant which is logically impossible for binary outcome
Two-stage predictor substitution (2SPS)	-the rote extension to nonlinear models of the linear IV models -targets a marginal (population-averaged) odds ratio -it is the mimic of 2SLS -non-linear least squares is used to estimates the parameter -for a linear model, 2SPS = 2SLS	-RD, RR, OR	-suitable for non-linear association between exposure and outcome	-in practice, 2SPS in non-linear model does not always yield consistent exposure effects on the outcome - parameter estimation process is more difficult than 2SLS -under a logistic regression model, 2SPS may not provide causal OR
Two-stage residual inclusion (2SRI)	-include the estimated unobservable confounder (residual) from the first-stage as an additional variable along with the exposure in the second-stage model - also called control function estimator -under a linear model, 2SRI = 2SLS = 2SPS	-RD, RR, OR	-yields consistent estimates for linear and non-linear models -performs better than 2SPS -possible to apply in the specific case of a binary exposure with a binary or count outcome -for a log-linear model in the stage-two, 2SRI estimator provides CRR	-it may give biased estimates when there is strong unmeasured confounding, as is usually the case in an IV analysis -under a logistic regression model, 2SRI estimator may not provide causal OR -generally require the exposure to be continuous, rather than binary, discrete, or censored
Two-stage logistic regression (2SLR)	-when outcome and exposure are binary and interest to estimate OR -fully parametric, maximum likelihood technique is used to estimate the parameters	-OR	-parallel to 2SLS using LRM in both stages instead of linear models	-if the first-stage logistic model is not correctly specified then second-stage parameter estimates might be biased -estimator does not provide COR
Three-stage least squares (3SLS)	-an extension of 2SLS but unlike the 2SLS, all coefficients are estimated simultaneously, requires three steps -in 2SLS, if the errors in the two equations are correlated, the 3SLS can be a suitable alternative	-RD, RR	-more information is used and hence the estimators are likely to be more efficient than 2SLS	-more vulnerable to a misspecification of the error terms -very rarely applied in epidemiologic studies -estimation process is more complicated than 2SLS -3SLS becomes inconsistent if errors are heteroskedastic
Structural mean models (SMM)	-SMMs use IVs via G-estimation and involves the assumption of conditional mean independence -additive SMMs use continuous outcome and multiplicative SMMs use positive-valued outcomes -MSMM assumed log-linear model to measure the risk ratio -LSMM assumes logistic regression model which is fitted by maximum likelihood technique	RD, RR, OR	-it relaxes several of the modelling restrictions (constant treatment effects) required by ratio estimator/ two-stage methods -can be used in the case of time-dependent instruments, exposures, and confounders -provides average treatment effects for the treated subjects	-the assumption of no effect modification is impossible to verify -with a binary outcome, additive SMMs and MSMM suffer from the limitations of linear and log-linear models (e.g., predicted response probabilities may outside of the interval [0, 1])

Generalized method of moments (GMM)	<ul style="list-style-type: none"> <li>-a non-linear analogue of 2SLS</li> <li>-the standard IV (2SLS) estimator is a special case of a GMM estimator</li> <li>-making assumptions about the moments of the error term</li> <li>-allows estimation of parameters in over-identified model (number of IV greater than number of exposure variable)</li> <li>-the parameters are estimated in an iterative process</li> </ul>	RD, RR, OR	<ul style="list-style-type: none"> <li>-it requires specification only of certain moment conditions</li> <li>-applicable for the linear and non-linear models</li> <li>-non-linear GMM estimator is asymptotically more efficient than 2SLS</li> <li>-more robust and less sensitive to parametric conditions</li> <li>-works better than 2SLR when exposure and outcome are binary</li> <li>-in case of heteroskedasticity, this is more efficient than the linear IV estimators</li> </ul>	-GMM estimator with logistic regression model is not consistent for the COR due to non-collapsibility of the OR
Bivariate probit models (BPM)	<ul style="list-style-type: none"> <li>-two-stage method, but as different to 2SLS and model the probabilities directly and are restricted on [0,1]</li> <li>-full information maximum likelihood is used to estimate the parameter</li> <li>-accounts for the correlation between the errors</li> </ul>	Probit coefficient*	<ul style="list-style-type: none"> <li>-for binary outcome and exposure, BPM perform better than linear IV methods</li> <li>-the estimator of BPM have no interpretation like OR. However, by multiplying a probit coefficient by approximately 1.6, the estimator can be made to approximate OR</li> </ul>	-when the distribution of error terms are not normal or the average probability of the outcome variable is close to one or zero, the BPM estimator may not be consistent for ACE

Remarks for all methods:

-all basic IV assumptions are needed for all estimation methods and violation of any IV assumption, all methods provide biased results  
 -under the constant exposure effect, all methods provide ACE; in case of a heterogeneous treatment effects, under the monotonicity and no effect modification assumptions, all methods (except SMMs) provides LATE and SMMs provides ATT, respectively  
 \*the bivariate probit model is fully parametric, all of the treatment parameters such as risk difference, odds-ratio or risk ratio, can be derived from the probit coefficients as marginal effects.

Table 2: Overview of commonly used estimation methods for IV analysis (basic notions, estimator, strengths, and limitations)

However, for logistic regression model (LRM), 2SRI estimator may not provide causal odds ratio due to non-collapsibility of the odds ratio.

2SRI yields consistent estimates for both linear and nonlinear models [49,50]. The advantage of 2SRI over 2SLS is that 2SLS is only consistent when the second-stage model is linear, whereas this restriction does not hold for 2SRI [43,51]. Moreover, this method shows more precise estimates than 2SPS [52].

Two-stage logistic regression (2SLR)

When both the outcome and exposure are binary and the interest is to use IV to estimate odds ratios, 2SLR can be applied. It is similar to 2SLS, but instead of linear models using logistic models in both stages [4,53]. This method is fully parametric and maximum likelihood estimation is used to estimate the parameters. If the first-stage logistic model is not correctly specified, the estimates from the second-stage can be biased [54,55]. Also, note that this method may not provide the causal odds ratio due to the non-collapsibility of the OR [19].

Three-stage least squares method (3SLS)

The 3SLS generalizes the 2SLS. Possible correlation of the errors ( $\mathcal{E}_1$  and  $\mathcal{E}_2$ ) in equations (4) and (5) is not taken into account by 2SLS. 3SLS accounts for the possible correlations between errors and may improve the efficiency of the estimator [56,57]. Unlike 2SLS, in which the coefficients of the two equations are estimated separately, in 3SLS all coefficients are estimated simultaneously. This requires three steps. The first-stage is similar to the 2SLS, i.e., a linear regression of X on Z to get X. In the second-stage, the residuals of the second-stage 2SLS model are obtained to estimate the cross-model correlation matrix (correlation between error terms in both models). Finally, in the third-stage the estimated correlation matrix is used to obtain the IV estimator. When there is no correlation between the error terms of the 2SLS models, the 3SLS reduces to a 2SLS. However, 3SLS is more vulnerable to misspecification error since misspecification of one of the

models in the first or second will affect the third stage model [58].

Structural mean models (SMMs)

SMMs explicitly use counterfactuals or potential outcomes [52], which were originally proposed by Robins [59] in the context of randomized trials with non-compliance to estimate the causal effects for the treated (exposed) individuals. SMMs are semi-parametric models and use IVs via G-estimation for identification and estimation of the causal parameter. This method involves the assumption of a conditional mean independence [14,19,60-62] and does not make distributional assumptions about the exposure [19]. SMMs with an identity link is sometimes called additive SMMs and can be used for continuous outcomes and multiplicative SMMs with log-linear model can be used for positive-valued/binary outcomes in order to estimate the causal risk ratio [19,63]. Additionally, the logistic structural mean model (LSMM) developed by Vansteelandt and Goetghebeur [64] and Robins and Rotnitzky [65] can also be used for binary outcome in order to estimate causal odds ratio [19,63].

To handle continuous outcome data, the IV estimator from the additive SMMs can be expressed as equation (2) given that the assumptions of CMI and no effect modification by Z are fulfilled [14,62,66,67]. This estimator provides the average treatment effect (ATT) for the treated individuals [19,68].

The advantage of this method is that it relaxes several of the modelling restrictions such as homogeneous treatment effects required by more classical methods such as RE/two-stage IV methods [14,19]. One of the key assumptions of this method is no effect modification, which is difficult to verify in practical situations [67].

SMMs have been extended by Robins [60] to a general setting of structural nested mean models (SNMM) for repeated measures at multiple time points. The SMMs are a subclass of the SNMM [59,69]. When instruments, exposures, and confounders are time-dependent,

SNMM can be used to estimate causal effects of exposure on the outcome [14]. Details and mathematical formulations of SMMs are described elsewhere [14,19,63].

### Generalized method of moments (GMM)

When applying the GMM a system of equations is set up, which is then solved numerically using computer algorithms. This technique was formalized by Hansen [70] and is a broad class of estimation methods that allow for a larger number of equations (moment conditions) than parameters [4,53,71] that are not possible in the MSMM and LSMM [19]. More clearly, the GMM allows for estimation of parameters in an over-identified model (number of IVs greater than the number of exposures). GMM with linear model can be similar to the ones used in 2SLS [72] but GMM is also a non-linear analogue of 2SLS [17], which is called multiplicative GMM. Detailed explanations can be found elsewhere [4,19,53].

In general, the nonlinear optimum GMM estimator is asymptotically more efficient than 2SLS [73]. Since GMM is a moment based method without parametric assumptions, it is less prone to model misspecification than 2SLR or bivariate probit models when exposure and outcome are binary [4]. In case of a linear model and single IV, the GMM estimator is equivalent to 2SLS, additive SMM, and LIML [53,66,74]. On the other hand, with log-linear model, (i.e., MGMM) [19], it is equivalent with MSMM and provides the population causal risk ratio [19]. However, this estimator with logistic regression model is not consistent for the causal odds ratio due to non-collapsibility of the odds ratio [17].

In case of a binary or count outcome, Palmer et al. [75] suggested a two-stage IV method where the first-stage is a linear regression and the second stage-model is a logistic or log-linear model [19]. Since IV analysis with logistic regression may not provide a consistent exposure effect, in order to estimate causal risk ratio, GMM with log-linear model is preferable. Moreover, 2SRI [48] is also applicable in the setting of count outcome.

### Bivariate probit models (BPM)

When the outcome of interest is binary, so-called probit models can be applied for IV analysis. In contrast to 2SLS, probit models directly model probabilities (i.e., are restricted on (0, 1)) [4,30]. BPM can be applied in two-stages, but unlike common two-stage estimation methods, this method is estimated via full-information maximum likelihood, which takes into account the correlation between the error terms in the two equations [24]. A more detailed model description can be found elsewhere [4,30].

The interpretation of BPM parameters is not like those of ordinary regression model parameters (e.g., logarithm of odds ratio from a logistic model). However, by multiplying a probit coefficient by approximately 1.6 or 1.8, probit coefficients approximate the coefficients obtained through logistic regression [4].

In case of binary outcome, linear IV methods may yield biased results and BPM may be preferable [30,47]. Furthermore, the estimates are more efficient than 2SLS, whereas 2SLS models are more robust to incorrect modelling assumptions regarding the bivariate normal distribution of the error terms [76,77]. However, when the distribution of error terms is not normal or the average probability of the outcome variable is close to one or zero, or if there is more than one exposure, the estimates from the BPM are generally not consistent for the average causal effect [30,77].

### Other estimation methods

Apart from the methods discussed above, the outcome variable in epidemiologic research may also be a time-to-event. Also in case of these outcome variables, IV analysis has been applied with two-stage method. In that case, the second-stage model could be a Cox proportional hazards model [78-80]. However, Brookhart et al. [3] stated that this approach for IV analysis is not motivated by a theoretical model and, therefore, parameters that are obtained from this approach may not be causally interpretable. Examples of this approach are a study of the effect of rosiglitazone on (time to) cardiovascular hospitalization and all-cause mortality using facility-prescribing patterns as an IV [78], and a study of the effect of adjuvant chemotherapy on (time to) breast cancer recurrence using physician preference as an IV [79].

### Standard error and characteristics of IVestimators

Consider two-stage models for IV analysis, in which the predicted value of exposure from the first-stage model is included in the second-stage model. The uncertainty around this prediction is not taken into account in the latter model, which therefore may result in incorrect precision. Typically, standard errors (SEs) of the IV estimate from the second-stage model are too small [24,30,44,45]. An alternative method to estimate a correct SE is the so-called sandwich variance estimator (robust SE), which involves cross products of the predicted treatment and a dispersion factor based on the observed treatment [49]. Most statistical software packages provide this sandwich variance estimate [10]. Angrist and Krueger [10] noticed that these SEs are asymptotically valid, but in practice (with finite sample size) they are only approximately valid.

An alternative way of estimating SEs is the bootstrap method [81]. Here bootstrap samples of the original data can be used to estimate the variation in the IV estimates and hence its SE [4,6,82-84]. It should be stressed that one of the weaknesses of the IV estimator is that it tends to display large SEs relative to the conventional regression estimator [13,85]. It is also noted that the IV estimator can perform poorly in finite samples and show biased results [31] and this bias is amplified when the IV is weak [14,31].

### Interpretation of exposure effects from IV analysis

Researchers may be interested to estimate the average treatment effects over the entire study population [27]. However, it has been argued that the basic assumptions of IV analysis are not sufficient to achieve point estimates for the causal effect of exposure on the outcome, but only estimate upper and lower bounds of this parameter [14,86,87]. To achieve a point estimate of the average causal effect (ACE) over the entire study population, the additional strong assumption of homogeneity of exposure across levels of the IV should be satisfied [52]. Moreover, IV analysis captures the ATT under the assumption of no effect modification by IV [52]. When exposure effects are not homogeneous across IV levels, under the monotonicity assumption (i.e., the IV affects the treatment deterministically in one direction), the IV estimate quantifies the local average treatment effect (LATE) [88], which is only informative for a subset of the study population, namely those who comply with the IV [27,89-91].

### Assessment of IVassumptions

As noted, IV analysis must satisfy three basic assumptions and if these assumptions do not hold, results may be severely biased [3,13]. The first assumption (i.e., the IV is related to exposure) is generally easier to check using available statistical methods than the other two

assumptions. The second (IV has no direct effect on outcome) and third (IV is independent of confounders) assumptions are unverifiable or not directly testable as they involve unobservable variables [1,13,18,19,68,76,92]. Some authors proposed circumstantial evidence to support these assumptions [2,5,93,94]. Alternatively, for the third assumption a falsification test based on the standardized difference can be applied [95].

In order to check the first assumption, the F-statistic value from the first-stage linear regression model is widely used although this statistic highly affected by sample size [76,83,85]. There is a rule of thumb that if the F-statistic value is greater than 10, the first assumption holds [13,96,97]. Other measures for the strength of the association between IV and exposure include the first-stage regression coefficient of the IV [50,98] or the  $R^2$  of a linear first-stage model [15,78,83], the odds ratio [6,93], or pseudo-R-squared of the first-stage model [76]. When the correlation between IV and exposure is not strong enough, IV analysis is likely to be biased (weak IV bias, which increases with the weakness of the IV). A weak IV will provide large SEs for the IV estimator [3,13,31,47,99].

## Discussion

We provided an overview of estimation methods of IV analyses, highlighting their strengths and limitations for epidemiological research. These methods share aspects, yet also have some particularities. However, when the IV assumptions are violated, the sample size is small, or IV models are not correctly specified, all methods tend to perform poorly and show biased results.

The methods can be categorized as moments based and semi-parametric (e.g. 2SLS, GMM, SMM) or likelihood based (e.g. BPM, 2SLR, LSMM) methods. The moment based methods or semi-parametric method are in general less efficient than likelihood based methods. However, likelihood methods are more vulnerable to incorrect modelling assumptions, in which case moment based methods are more robust. In empirical data, although several IV methods can be applicable in the same combination of IV, exposure, and outcome, considering different methods' assumptions, target parameters being estimated are different, so the interpretations of exposure effects appear different [45]. Therefore, choosing an appropriate IV method requires attention [76].

In order to obtain ACE or LATE or ATT, along with basic assumptions, extra assumptions such as homogeneous exposure effect or monotonicity in case of heterogeneous exposure effect or no effect modification by IV, respectively should be fulfilled. These different assumptions result in the estimation of different causal effects, and hence, researchers should be aware for interpretations of the IV estimates [14].

In randomized trials, the IV of treatment assignment satisfies the assumptions by design, but in observational studies, this is not the case. In the latter situation, subject matter knowledge and theoretical motivations (why is an IV related to treatment and unrelated to patients' characteristics and outcome?) should be given especially regarding the second and third condition underlying the IV method. If the IV is weakly related to exposure and correlated with unmeasured variables, IV methods may yield biased results [100]. In addition, the main critique of any IV analysis is that the IV may affect the outcome through some pathway other than through the exposure of interest [32]. This condition cannot be verified empirically.

From a methodological perspective, the IV method is a powerful

statistical tool, given that a valid IV is present and IV analysis correctly applied. In that case, it can provide a valid estimate in the presence of measured and unmeasured confounding. However, if there is strong confounding effect, it is difficult to find an appropriate IV [13].

A limitation of our study is that we restricted ourselves to IV methods that are commonly used in epidemiologic research. We did not discuss nonparametric and Bayesian IV methods. We refer to the literature for examples of the methods [12,38,86,101-104]. Because of limited space, we did not describe mathematical models with detailed derivation of IV estimators for all methods.

In conclusion, IV analysis is potentially powerful methods to control for confounding (both measured and unmeasured). Some estimation methods (e.g., 2SLS, 2SRI) can be applied in many situations, whereas others (e.g., RE, BPM, 2SLR) can only be applied in a limited number of situations. Irrespective of the methods that are used in a particular study, in order to provide valid interpretation of the exposure effect on the outcome, researchers should be aware of the underlying methodology of the estimation method as well as key assumptions of the IV.

**Running Head:** Methods for IV estimation

## FUNDING

The PROTECT project is supported by the Innovative Medicine Initiative Joint Undertaking ([www.imi.europa.eu](http://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.

**Conflicts Of Interest:** "Olaf Klungel had received unrestricted funding for pharmaco epidemiological research from the Dutch private-public funded Top Institute Pharma (TI Pharma Grant T6.101 Mondriaan)."

## References

1. Pratt N, Roughead EE, Ryan P, Salter A (2010) Antipsychotics and the risk of death in the elderly: an instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 19: 699-707.
2. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S (2006) Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 17: 268-275.
3. Brookhart MA, Rassen JA, Schneeweiss S (2010) Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 19: 537-554.
4. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA (2009) Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 169: 273-284.
5. Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA (2006) Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and non-selective non-steroidal antiinflammatory drugs: an instrumental variable analysis. *Arthritis Rheum* 54: 3390-3398.
6. Groenwold RH, Hak E, Klungel OH, Hoes AW (2010) Instrumental variables in influenza vaccination studies: mission impossible? *Value Health* 13: 132-137.
7. Bennett DA (2010) An introduction to instrumental variables analysis: part 1. *Neuroepidemiology* 35: 237-240.

8. Bennett DA (2010) An introduction to instrumental variables--part 2: Mendelian randomisation. *Neuroepidemiology* 35: 307-310.
9. Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, et al. (2007) 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* 45: S116-122.
10. Angrist JD, Krueger AB (2001) Instrumental variables and the search for identification: From supply and demand to natural experiments. *J Econ Perspect* 15: 69-85.
11. Landrum MB, Ayanian JZ (2001) Causal effect of ambulatory specialty care on mortality following myocardial infarction: A comparison of propensity score and instrumental variable analyses. *Health Serv Outcomes Res* 2: 221-245.
12. Greenland S (2000) An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 29: 722-729.
13. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH (2006) Instrumental variables: application and limitations. *Epidemiology* 17: 260-267.
14. Hernán MA, Robins JM (2006) Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 17: 360-372.
15. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S (2009) Instrumental variables II: instrumental variable application in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 62: 1233-1241.
16. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S (2009) Instrumental variables I: instrumental variables exploit natural variation in non-experimental data to estimate causal relationships. *J Clin Epidemiol* 62: 1226-1232.
17. Clarke P, Windmeijer F (2010) Instrumental Variable Estimators for Binary Outcomes. Working Paper. Centre for Market and Public Organisation, Univ. Bristol.
18. Chen Y, Briesacher BA (2011) Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 64: 687-700.
19. Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, et al. (2011) Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *Am J Epidemiol* 173: 1392-1403.
20. Davies NM, Smith GD, Windmeijer F, Martin RM (2013) Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology* 24: 363-369.
21. Swanson SA, Hernán MA (2013) Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology* 24: 370-374.
22. Baiocchi M, Cheng J, Small DS (2014) Instrumental variable methods for causal inference. *Stat Med* 33: 2297-2340.
23. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB (2014) Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Ann Intern Med* 161: 131-138.
24. Greene WH, Zhang C (2003) *Econometric analysis*. Prentice hall, New Jersey.
25. Wald A (1940) The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics* 11: 284-300.
26. Grootendorst P (2007) A review of instrumental variables estimation in the applied health sciences. *Health Serv Outcomes Res Method* 7: 159-179.
27. McNamee R (2009) Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Stat Med* 28: 2639-2652.
28. Cameron AC (2005) *Microeconometrics: methods and applications*. Cambridge University Press.
29. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S (2010) Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 48: S114-120.
30. Angrist JD, Pischke JS (2008) *Mostly harmless econometrics: An empiricist's companion*. Princeton Univ Pr.
31. Bound J, Jaeger DA, Baker RM (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc* 90: 443-450.
32. Glymour MM (2006) Natural experiments and instrumental variable analyses in social epidemiology. In J Michael Oakes, Jay S Kaufman (ed) *Methods in social epidemiology*. San Francisco, Jossey-Bass.
33. Anderson TW, Kunitomo N, Sawa T (1982) Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator. *Econometrica* 50: 1009-1027.
34. Burgess S, Granell R, Palmer TM, Sterne JA, Didelez V (2014) Lack of identification in semiparametric instrumental variable models with binary outcomes. *Am J Epidemiol* 180: 111-119.
35. Angrist JD (2001) Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *J Bus Econ Stat* 19: 2-16.
36. Galárraga O, Sosa-Rubí SG, Salinas-Rodríguez A, Sesma-Vázquez S (2010) Health insurance for the poor: impact on catastrophic and out-of-pocket health expenditures in Mexico. *Eur J Health Econ* 11: 437-447.
37. Carrasco R (2001) Binary choice with binary endogenous regressors in panel data: Estimating the effect of fertility on female labor participation. *J Bus Econ Stat* 19: 385-394.
38. Baiocchi M, Rosenbaum P, Small D, Lorch S (2009) Near/Far Matching-A Nonparametric Instrumental Variables Technique for Binary Outcomes
39. Ionescu-Iltu R, Delaney JAC, Abrahamowicz M (2009) Bias-variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol Drug Saf*, 18: 562-571.
40. Lee LF (1979) Identification and estimation in binary choice models with limited (censored) dependent variables. *Econometrica* 47: 977-996.
41. Terza JV, Basu A, Rathouz PJ (2008) Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ* 27: 531-543.
42. Burgess S (2013) Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Stat Med* 32: 4726-4747.
43. Terza JV, Bradford WD, Dismuke CE (2008) The use of linear instrumental variables methods in health services research and health economics: a cautionary note. *Health Serv Res* 43: 1102-1102-20.
44. Cai B, Small DS, Have TR (2011) Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Stat Med* 30: 1809-1824.
45. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, et al. (2012) Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 21: 223-242.
46. Hausman JA (1978) Specification tests in econometrics. *Econometrica* 46: 1251-1271.
47. Suh HS (2009) The effect of using fibrates in conjunction with statins for the management of dyslipidemia in persons with type II diabetes mellitus. University Of Southern California
48. Wooldridge JM (2002) *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.
49. Wooldridge JM (2008) Instrumental variables estimation of the average treatment effect in correlated random coefficient models. *Adv Econom* 21: 93-117.
50. Stuart BC, Doshi JA, Terza JV (2009) Assessing the impact of drug use on hospital costs. *Health Serv Res* 44: 128-144.
51. Newey WK (1987) Efficient estimation of limited dependent variable models with endogenous explanatory variables. *J Econ* 36: 231-250.
52. O'Malley AJ, Frank RG, Normand SL (2011) Estimating cost-offsets of new medications: use of new antipsychotics and mental health costs for schizophrenia. *Stat Med* 30: 1971-1988.
53. Johnston KM, Gustafson P, Levy AR, Grootendorst P (2008) Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med* 27: 1539-1556.
54. Angrist JD (2001) Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat* 19: 2-28.
55. Henneman TA, Van Der Laan MJ, Hubbard AE (2002) Estimating causal parameters in marginal structural models with unmeasured confounders using instrumental variables. Berkeley, CA: The Berkeley Electronic Press.



56. SAS Institute Inc. (2004) SAS/ETS® 9.1. User's Guide.
57. Dowd B, Town R Does X Really Cause Y. Washington, DC: Academy Health, 2002.
58. Ahn J (2002) Beyond single equation regression analysis: Path analysis and multi-stage regression analysis. *Am J Pharm Educ* 66: 37-41.
59. Robins JM (1989) The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. Washington, DC. US Public Health Serv: 113-159
60. Robins JM (1994) Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat* 23: 2379-2412.
61. Maracy M, Dunn G (2011) Estimating dose-response effects in psychological treatment trials: the role of instrumental variables. *Stat Methods Med Res* 20: 191-215.
62. Clarke PS, Palmer TM, Windmeijer F (2011) Estimating structural mean models with multiple instrumental variables using the generalised method of moments. The Centre for Market and Public Organisation.
63. Tan Z (2010) Marginal and nested structural models using instrumental variables. *J Am Stat Assoc* 105: 157-169.
64. Vansteelandt S, Goetghebeur E (2003) Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65: 817-835.
65. Robins J, Rotnitzky A (2004) Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* 91: 763-783.
66. Clarke P, Windmeijer FAG (2009) Instrumental variable estimators for binary outcomes. Centre for Market and Public Organisation, University of Bristol.
67. Clarke PS, Windmeijer F (2010) Identification of causal effects on binary outcomes using structural mean models. *Biostatistics* 11: 756-770.
68. Didelez V, Meng S, Sheehan NA (2010) Assumptions of IV methods for observational epidemiology. *Stat Sci* 25: 22-40.
69. Goetghebeur E, Stijn V (2005) Structural mean models for compliance analysis in randomized clinical trials and the impact of errors on measures of exposure. *Stat Methods Med Res* 14: 397-415.
70. Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029-1054.
71. Foster EM (1997) Instrumental variables for logistic regression: An illustration. *Soc Sci Res*, 26: 487-504.
72. Baum CF, Schaffer ME, Stillman S (2003) Instrumental variables and GMM: Estimation and testing. *Stata J* 3: 1-31.
73. Lee L (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J Econ* 137: 489-514
74. Burgess S (2011) Statistical issues in Mendelian randomization: use of genetic instrumental variables for assessing causal associations.
75. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR (2008) Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 37: 1161-1168.
76. Yoo BK, Frick KD (2006) The instrumental variable method to study self-selection mechanism: a case of influenza vaccination. *Value Health* 9: 114-122.
77. Bhattacharya J, Goldman D, McCaffrey D (2006) Estimating probit models with self-selected treatments. *Stat Med* 25: 389-413.
78. Ramirez SP, Albert JM, Blayney MJ, Tentori F, Goodkin DA, et al. (2009) Rosiglitazone is associated with mortality in chronic hemodialysis patients. *J Am Soc Nephrol* 20: 1094-1101.
79. Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, et al. (2010) A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 63: 64-74.
80. Schmoor C, Caputo A, Schumacher M (2008) Evidence from nonrandomized studies: a case study on the estimation of causal effects. *Am J Epidemiol* 167: 1120-1129.
81. Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall/CRC.
82. Cain LE, Cole SR, Greenland S, Brown TT, Chmiel JS, et al. (2009) Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable. *Am J Epidemiol* 169: 1124-1132.
83. Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, et al. (2001) Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis. *J Clin Oncol* 19: 1064-1070.
84. Finke R, Theil H (1984) Bootstrapping for standard errors of instrumental variable estimates. *Econ Lett* 14: 297-301.
85. Ionescu-Iltu R, Abrahamowicz M, Pilote L (2012) Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 65: 155-162.
86. Balke A, Pearl J (1997) Bounds on treatment effects from studies with imperfect compliance. *J Am Stat Assoc* 92: 1171-1176.
87. Robins JM, Greenland S (1996) Identification of causal effects using instrumental variables: comment. *Journal of the American Statistical Association* 91: 456-458.
88. Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. *Econometrica* 62: 467-475.
89. Angrist J, Imbens G, Rubin DB (1996) Identification of causal effects using instrumental variables. *J Am Stat Assoc* 91: 444-472.
90. Brooks JM, Fang G (2009) Interpreting treatment-effect estimates with heterogeneity and choice: simulation model results. *Clin Ther* 31: 902-919.
91. Fang G, Brooks JM, Chrischilles EA (2012) Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 175: 60-65.
92. Didelez V, Sheehan N (2007) Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 16: 309-330.
93. Schneeweiss S, Setoguchi S, Brookhart A, Dormuth C, Wang PS (2007) Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. *Can Med Assoc J* 176: 627-632.
94. Brookhart MA, Schneeweiss S (2007) Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int J Biostat* 3: Article 14.
95. Ali MS, Uddin MJ, Groenwold RH, Pestman WR, Belitser SV, et al. (2014) Quantitative falsification of instrumental variables assumption using balance measures. *Epidemiology* 25: 770-772.
96. Stock JH, Wright JH, Yogo M (2002) A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 20: 518-529.
97. Sheehan NA, Meng S, Didelez V (2011) Mendelian randomisation: a tool for assessing causality in observational epidemiology. *Methods Mol Biol* 713: 153-166.
98. Rascati KL, Johnsrud MT, Crismon ML, Lage MJ, Barber BL (2003) Olanzapine versus risperidone in the treatment of schizophrenia: a comparison of costs among Texas Medicaid recipients. *Pharmacoeconomics* 21: 683-697.
99. Staiger D, Stock JH (1997) Instrumental variables regression with weak instruments. *Econometrica* 65: 557-586.
100. Hahn J, Hausman J (2000) A new specification test for the validity of instrumental variables. *Econometrica*, 70:163-189
101. Burgess S, Thompson SG; CRP CHD Genetics Collaboration, Burgess S, Thompson SG, Andrews G, et al. (2010) Bayesian methods for meta-analysis of causal relationships estimated using genetic instrumental variables. *Stat Med* 29: 1298-1311.
102. Mc Keigue PM, Campbell H, Wild S, Vitart V, Hayward C, et al. (2010) Bayesian methods for instrumental variable analysis with genetic instruments ('Mendelian randomization'): example with urate transporter SLC2A9 as an instrumental variable for effect of urate levels on metabolic syndrome. *Int J Epidemiol* 39: 907-918.
103. Gao C, Lahiri K (1999) A comparison of some recent Bayesian and classical procedures for simultaneous equation models with weak instruments. Department of Economics, SUNY Albany.
104. Kleibergen F, Zivot E (2003) Bayesian and classical approaches to instrumental variable regression. *J Econ* 114: 29-72.