Published in final edited form as:

Epidemiology. 2018 November ; 29(6): 817–820. doi:10.1097/EDE.0000000000000914.

Reporting of Baseline Characteristics to Accompany Analysis by Instrumental Variables

Todd A. MacKenzie¹, A. James O'Malley¹, and Kimon Bekelis^{1,2}

¹Dartmouth College, Hanover, New Hampshire, U.S.A

²Good Samaritan Hospital, West Islip, New York, U.S.A

Abstract

A property and criticism of instrumental variable (IV) estimators are that they estimate the exposure effect in the *compliers*; individuals whose exposure is influenced by the instrument. It is conceivable that the exposure effect is different in individuals whose exposure is not influenced by the instrument. For that reason, it is useful to know who the compliers are. We present methods for reporting averages and other statistics, and comparing them between the compliers and non-compliers, which are applicable to scenarios in which the instrument and exposure are fixed in time and dichotomous. The methods are illustrated in a comparison of outcomes between teaching and nonteaching hospitals in patients undergoing brain or spinal surgery.

Keywords

Local Average Treatment Effect; Complier Average Causal Effect; Treatment Effect Heterogeneity

Introduction

Using instrumental variables to estimate treatment effects has the potential to bypass bias due to confounding (endogeneity). This method is far from a panacea as it is not possible to empirically validate an instrumental variable: It is not possible to prove or disprove that a putative instrument does indeed satisfy the definition of an instrumental variable, with the exception of randomization as designed into clinical and other trials. Despite this weakness the method has been steadily gaining in popularity as a way to resolve confounding in observational studies.

A property and criticism of instrumental variable estimators is that they provide an estimate of the treatment effect on a subset of the population, sometimes referred to as the *compliers* [1] or population *on the margin* [2]. It is plausible that the treatment effect is different in the remainder of the population, e.g. heterogeneity of the treatment effect. The concept of the

Corresponding Author: Todd.A.MacKenzie@Dartmouth.edu.

Conflicts of interest: None.

The authors have no conflicts of interest to report.

Availability of data and code for replication: All code is available as specified in the manuscript. The data is not available without permission from the Statewide Planning and Research Cooperative System (SPARCS) of the State of New York.

complier is clear when the instrumental variable is dichotomous and causal [3]. In this case it is possible to define a latent categorization known as *principal strata* [4, 5, 6, 7] consisting of three categories; compliers, *takers* and *non-takers* (also known as “always takers” and “never takers” [8, 9]).

The compliers are individuals whose treatment is determined by the instrument. The takers are those who would receive treatment regardless the value of the instrument; the non-takers are those who would not. The property that the instrumental variable estimator provides an estimate of the treatment effect on the compliers opens a criticism: The treatment effect in the takers or non-takers may be different. The effect of treatment in the non-takers is relevant because there may be scenarios in which they would take the treatment. For instance, those who are not affected by the instrument of one study (e.g. provider preference) may be affected by the instrument of another study (e.g differential geographic distance to treatments being compared, gene in another location). Due to the latency of the principal strata it isn't possible to definitively categorize each subject in a sample. For this reason, instrumental variable estimators have been criticized as applying to an unknowable strata. However, it is possible to characterize the compliers [10, 11, 12, 1]. Parametric models combined with Bayesian methods have been proposed for modeling the probability of being a complier in terms of baseline characteristics [13].

In this brief report we show nonparametric methods for presenting and comparing baseline characteristics among the compliers, takers and non-takers for the setting of a binary fixed-in-time instrument that has a causal effect on a binary fixed-in-time exposure.

Methods

Let R represent the binary instrument (e.g. randomization). Let $\{X(r)\}_{r=0,1}$ be the *potential* treatment indicators, and $X = X(R)$ the observed treatment indicator. The potential treatment indicators stratify the population into four *principal strata* [4], Compliers ($X(r) = r$), Takers ($X(r) = 1$), Non-takers ($X(r) = 0$) and Defiers $X(r) = 1 - r$. We assume that the Defier strata has probability zero, known as the *Monotonicity* assumption, which facilitates partial identification of the Taker and Non-taker strata using observations of R and X as shown in Table 1.

Let p_C, p_T, p_N be the probabilities of sampling subjects from the Complier, Taker and Non-taker strata. From Table 1 it follows that $p_T = \Pr[X = 1 | R = 0]$ and $p_N = \Pr[X = 0 | R = 1]$, and the equation, $\Pr[X = 1 | R = 1] = p_C + p_T$ implies $p_C = \Pr[X = 1 | R = 1] - \Pr[X = 1 | R = 0]$ (or equivalently $p_C = \Pr[X = 0 | R = 0] - \Pr[X = 0 | R = 1]$). Estimators of the frequencies of the three strata p_N, p_T and p_C are found in Table 2.

Table 2 also presents estimators of the means in the Compliers, Takers and Non-takers, where $\hat{\mu}$ is the sample mean, and $\hat{\mu}_{r,x}$ is the sample mean in subjects for whom $R = r, X = x$. The formula for the Compliers follows from the identity, $\bar{X} = p_C \bar{X}_C + p_T \bar{X}_T + p_N \bar{X}_N$. Estimators for the variance in the three strata are presented next in the table, where $\hat{\sigma}^2$ is the sample variance and $\hat{\sigma}_{r,x}^2$ is the sample variance in subjects for whom $R = r, X = x$. The formula for the variance in the compliers is based on the law of total variance.

Formulas for the standard error of the mean are presented in the row labelled SEM, where σ_{rx} is the standard deviation in subjects for whom $R=r, X=x$. The appendix provides a derivation of the variance of the mean in the compliers, as well as expressions for the covariances of the three means.

The final row of Table 2 provides estimators of the cumulative distribution function for each strata, where F is the empirical distribution function (EDF) of the baseline characteristics, i.e. $\hat{F}(z) = \sum_{i=1}^n 1\{Z_i \leq z\}/n$, F_{10} is the EDF among subjects for whom $R=1, X=0$, and F_{01} is the EDF among subjects for whom $R=0, X=1$. The formula for the distribution in the compliers follows from the identity $F = p_C F_C + p_T F_T + p_N F_N$. Quantiles such as the median can be calculated for the Takers and Non-takers as usual. An estimator of the q -quantile in the compliers is $[\inf_z \{\hat{F}_C(z) \geq q\} + \sup_z \{\hat{F}_C(z) \leq q\}]/2$.

A test of equality of means (or proportions) between the three strata, $H_0: \mu_C = \mu_N = \mu_T$ can be based on the vector $\hat{\Delta} = (\hat{\mu}_C - \hat{\mu}_N, \hat{\mu}_C - \hat{\mu}_T)$ which has an expected value of (0,0) and covariance matrix equal to

$$\Sigma = \begin{pmatrix} \text{Var}[\hat{\mu}_C] - 2\text{Cov}[\hat{\mu}_C, \hat{\mu}_N] + \text{Var}[\hat{\mu}_N] & \text{Var}[\hat{\mu}_N] - \text{Cov}[\hat{\mu}_C, \hat{\mu}_N] - \text{Cov}[\hat{\mu}_C, \hat{\mu}_T] \\ \text{Var}[\hat{\mu}_C] - \text{Cov}[\hat{\mu}_C, \hat{\mu}_N] - \text{Cov}[\hat{\mu}_C, \hat{\mu}_T] & \text{Var}[\hat{\mu}_C] - 2\text{Cov}[\hat{\mu}_C, \hat{\mu}_T] + \text{Var}[\hat{\mu}_T] \end{pmatrix} \quad (1)$$

Under the null hypothesis and the assumption that $P_C > 0$ the statistic $\hat{\Delta} \Sigma^{-1} \hat{\Delta}$ has a chi-square distribution with 2 degrees of freedom in large samples.

R code for these statistics can be found at <https://github.com/toddamackenzie/BaselineCharacteristics-for-Instrumental-Variable-Analyses>.

Example

We [14] used the method of instrumental variables to estimate the effect of teaching vs nonteaching hospitals on outcomes in 186,483 patients from New York undergoing spine and cranial operations. They used a continuously valued geographic instrument; the proportion in each of the 62 counties of New York of surgical patients who received their surgery at a teaching hospital. See [15, 16] for some limitations of using geographic and preference instruments. We found that teaching hospitals increased mortality, discharge to a facility and length of stay.

We conjecture that this instrument is causal [3] as the local (i.e. county) preference for teaching hospitals affects a resident's choice of where they access surgery. Therefore the methods we advocate in this paper for causal instruments are applicable. To illustrate the method for characterizing the compliers we dichotomized the instrument at its median value of 92.6%. The complier average causal effect of teaching versus non-teaching hospitals on mortality was 4.9 (95% CI: 2.0, 7.8) more deaths per 1000 surgeries.

The proportion of patients who were Compliers, Takers and Non-takers for this instrument was 26.0%, 71.3% and 2.6% respectively: Only 26% of patients are influenced by living in a

county with a teaching hospital preference above or below the median value of 92.6%. Table 3 reports mean age, and the proportion of females, and race, for the Compliers, Takers and Non-takers. It also reports the difference between the Compliers and Takers, as well as the difference between the Compliers and Non-takers. There was a higher proportion of females among the Compliers than among the Takers and Non-takers. In addition the Compliers included considerably fewer whites. The median age in the Compliers was 56, 55 in the Takers and 53 in the Non-takers.

Discussion

We recommend that the statistics proposed in this report should accompany the reporting of instrumental variable analyses. An alternative is to model the strata an individual is in according to their baseline characteristics [13]. The advantage of our approach is that it is nonparametric and matches the convention of reporting descriptive statistics for baseline characteristics, such as the first table of most clinical and epidemiological studies.

If there is treatment effect heterogeneity it is possible that different instrumental variable analyses will result in substantively different estimates depending on the instrument. For instance, two popular choices for instrumental variables are formed from distances traveled and clinic (or physician) preference. The subpopulation of individuals whose treatment choice is affected by the difference in distance may be quite different from the subpopulation of individuals whose treatment choice is affected by the preference of their clinic.

Mendelian Randomization studies employ a genotype or genotypic score as an instrument for the effect of a phenotype on an outcome. In this design the subpopulation of compliers are those for whom the genotypic instrument determines the phenotype as opposed to being unaffected by the instrument. The methods used in our paper should be reported in Mendelian randomization studies to determine this subpopulation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment:

This work was partially supported by NIH Grant 1RC4MH 092717-01, and through a Patient-Centered Outcomes Research Institute (PCORI) Award ME1503-28261. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee.

Appendix A Derivations of $\text{Var}[\mu^C\{(R_i, X_i)\}_{i=1}^n]$ and the covariance of the means

The estimator $\hat{\mu}_C$ can be written

$$\hat{\mu}_C = \sum_{i=1}^n w_i Z_i \quad (2)$$

where

$$w_i = \frac{1}{\hat{p}_C} \begin{cases} \frac{1}{n} & \text{if } R_i = X_i \\ \frac{1}{n} - \frac{\hat{p}_T}{n_{01}} & \text{if } R_i = 0 \text{ and } X_i = 1 \\ \frac{1}{n} - \frac{\hat{p}_N}{n_{10}} & \text{if } R_i = 1 \text{ and } X_i = 0 \end{cases} = \frac{1}{\hat{p}_C} \begin{cases} \frac{1}{n} & \text{if } R_i = X_i \\ -\frac{1}{n} \frac{n_{1\cdot}}{n_{0\cdot}} & \text{if } R_i = 0 \text{ and } X_i = 1 \\ -\frac{1}{n} \frac{n_{0\cdot}}{n_{1\cdot}} & \text{if } R_i = 1 \text{ and } X_i = 0 \end{cases} \quad (3)$$

Therefore the variance of the estimator of the mean in the compliers is

$$\begin{aligned} \text{Var}[\hat{\mu}_C | \{(R_i, X_i)\}_{i=1}^n] &= \sum_{i=1}^n w_i^2 \text{Var}(Z_i | R_i, X_i) \quad (4) \\ &= \frac{n_{00}\sigma_{00}^2 + n_{01}\left[\frac{n_{1\cdot}}{n_{0\cdot}}\right]^2 \sigma_{01}^2 + n_{10}\left[\frac{n_{0\cdot}}{n_{1\cdot}}\right]^2 \sigma_{10}^2 + n_{11}\sigma_{11}^2}{n^2 \hat{p}_C^2} \end{aligned}$$

Conditional on R and X the covariance of estimators $\hat{\mu}_T$ and $\hat{\mu}_N$ is

$$\text{Cov}[\hat{\mu}_T, \hat{\mu}_N | \{(R_i, X_i)\}_{i=1}^n] = \text{Cov}[\hat{\mu}_{01\cdot}, \hat{\mu}_{10\cdot} | \{(R_i, X_i)\}_{i=1}^n] = 0 \quad (5)$$

The covariance of $\hat{\mu}_C$ and $\hat{\mu}_N$ is

$$\begin{aligned}
Cov[\hat{\mu}_C, \hat{\mu}_N | \{(R_i, X_i)\}_{i=1}^n] &= Cov \left[\sum_{i=1}^n w_i Z_i, \frac{\sum_{R_i=1, X_i=0} z_i}{n_{10}} \middle| \{(R_i, X_i)\}_{i=1}^n \right] \quad (6) \\
\sum_{R_i=1, X_i=0} w_i \frac{1}{n_{10}} Var[Z_i | R_i=1, X_i=0] &= n_{10} \left(-\frac{1}{n} \frac{n_{0.}}{n_{1.}} \right) \frac{1}{n_{10}} \sigma_{10}^2 \\
&= -\frac{1}{n} \frac{n_{0.}}{n_{1.}} \sigma_{10}^2.
\end{aligned}$$

Using a similar derivation, the covariance of $\hat{\mu}_C$ and $\hat{\mu}_T$ is

$$Cov[\hat{\mu}_C, \hat{\mu}_T | \{(R_i, X_i)\}_{i=1}^n] = -\frac{1}{n} \frac{n_{1.}}{n_{0.}} \sigma_{01}^2. \quad (7)$$

References

- [1]. Baiocchi Michael, Cheng Jing, and Small Dylan S. “Instrumental variable methods for causal inference”. In: *Statistics in medicine* 33.13 (2014), pp. 2297–2340. [PubMed: 24599889]
- [2]. Card David. “The causal effect of education on earnings”. In: *Handbook of labor economics* 3 (1999), pp. 1801–1863.
- [3]. Swanson Sonja A and Hernán Miguel A. “The challenging interpretation of instrumental variable estimates under monotonicity”. In: *International journal of epidemiology* (2017).
- [4]. Frangakis Constantine E and Rubin Donald B. “Principal stratification in causal inference”. In: *Biometrics* 58.1 (2002), pp. 21–29. [PubMed: 11890317]
- [5]. Barnard John et al. “Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City”. In: *Journal of the American Statistical Association* 98.462 (2003), pp. 299–323.
- [6]. Angrist Joshua D, Imbens Guido W, and Rubin Donald B. “Identification of causal effects using instrumental variables”. In: *Journal of the American statistical Association* 91.434 (1996), pp. 444–455.
- [7]. Imbens Guido W and Rubin Donald B. “Bayesian inference for causal effects in randomized experiments with noncompliance”. In: *The Annals of Statistics* (1997), pp. 305–327.
- [8]. Bartolucci Francesco and Farcomeni Alessio. “Causal inference in paired two-arm experimental studies under noncompliance with application to prognosis of myocardial infarction”. In: *Statistics in medicine* 32.25 (2013), pp. 4348–4366. [PubMed: 23754710]
- [9]. King Gary et al. “Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme”. In: *The lancet* 373.9673 (2009), pp. 1447–1454.
- [10]. Imbens Guido W and Rubin Donald B. “Estimating outcome distributions for compliers in instrumental variables models”. In: *The Review of Economic Studies* 64.4 (1997), pp. 555–574.
- [11]. Angrist Joshua David, Pischke Jörn-Steffen, and Pischke Jörn-Steffen. *Mostly harmless econometrics: an empiricists companion*. Cram101 Publishing, 2013.

- [12]. Swanson Sonja A and Hernán Miguel A “Commentary: how to report instrumental variable analyses (suggestions welcome)”. In: *Epidemiology* 24.3 (2013), pp. 370–374. [PubMed: 23549180]
- [13]. Hirano Keisuke et al. “Assessing the effect of an influenza vaccine in an encouragement design”. In: *Biostatistics* 1.1 (2000), pp. 69–88. [PubMed: 12933526]
- [14]. Bekelis Kimon et al. “Association of hospital teaching status with neurosurgical outcomes: an instrumental variable analysis”. In: *World neurosurgery* 110 (2018), e689–e698. [PubMed: 29174238]
- [15]. Ertefaie Ashkan, Small Dylan S, and Rosenbaum Paul R. “Quantitative evaluation of the trade-off of strengthened instruments and sample size in observational studies”. In: *Journal of the American Statistical Association* (2018), pp. 1–13. [PubMed: 30034060]
- [16]. Hernán Miguel A and Robins James M. “Instruments for causal inference: an epidemiologist’s dream?” In: *Epidemiology* (2006), pp. 360–372. [PubMed: 16755261]

Table 1:

Identification of the principal strata according to the observations of the instrument, R , and exposure, X

Observed		Latent Categories:
R	X	<i>Principal Strata</i>
0	0	Compliers or Non-takers
0	1	Takers
1	0	Non-takers
1	1	Compliers or Takers

Table 2:

Estimators for baseline characteristics of the Compliers, Takers and Non-takers

Statistic	Compliers	Takers	Non-takers	Combined
Frequency	$\hat{p}_C = 1 - \hat{p}_N - \hat{p}_T$	$\hat{p}_T = \frac{n_{01}}{n_{00} + n_{01}}$	$\hat{p}_N = \frac{n_{10}}{n_{10} + n_{11}}$	1
Mean	$\hat{\mu}_C = \{\hat{\mu} - \hat{p}_T \hat{\mu}_T - \hat{p}_N \hat{\mu}_N\} / \hat{p}_C$	$\hat{\mu}_T = \hat{\mu}_{01}$	$\hat{\mu}_N = \hat{\mu}_{10}$	$\hat{\mu}$
Variance	$\hat{p}_C^{-1} \left\{ \hat{\sigma}^2 - \left\{ \hat{p}_N \hat{\sigma}_N^2 + \hat{p}_T \hat{\sigma}_T^2 \right\} \right.$ $\left. \left\{ \hat{p}_C (\hat{\mu}_C - \hat{\mu})^2 + \hat{p}_N (\hat{\mu}_N - \hat{\mu})^2 + \right. \right.$ $\left. \left. \hat{p}_T (\hat{\mu}_T - \hat{\mu})^2 \right\} \right.$	$\hat{\sigma}_N^2 = \hat{\sigma}_{10}^2$	$\hat{\sigma}_T^2 = \hat{\sigma}_{01}^2$	$\hat{\sigma}^2$
Standard Error of the Mean	$n^{-1} \hat{p}_C^{-1} \left\{ n_{00} \hat{\sigma}_{00}^2 + n_{01} \left[\frac{n_{1.}}{n_{0.}} \right]^2 \hat{\sigma}_{01}^2 + \right.$ $\left. n_{10} \left[\frac{n_{0.}}{n_{1.}} \right]^2 \hat{\sigma}_{10}^2 + n_{11} \hat{\sigma}_{11}^2 \right\}^{1/2}$	$\hat{\sigma}_{01} / \sqrt{n_{01}}$	$\hat{\sigma}_{10} / \sqrt{n_{10}}$	$\hat{\sigma} / \sqrt{n}$
Cumulative Distribution Function	$\{\hat{F} - \hat{p}_T \hat{F}_T - \hat{p}_N \hat{F}_N\} / \hat{p}_C$	$\hat{F}_T = \hat{F}_{01}$	$\hat{F}_N = \hat{F}_{10}$	$\hat{F}$

Table 3:

Characteristics and comparisons of the Compliers, Takers and Non-takers using means and proportions (standard error in brackets).

Variable	Compliers	Takers	Non-takers	Difference	
				Compliers vs Takers	Compliers vs Non-takers
Age (mean)	55.5 (0.1)	55.1 (0.1)	53.9 (0.3)	0.39 (0.16)	1.60 (0.31)
Female (%)	53.7 (0.5)	49.7 (0.2)	48.6 (1.0)	4.04 (0.53)	58.84 (5.15)
White (%)	44.7 (0.4)	80.2 (0.2)	57.0 (1.0)	-35.48 (0.45)	-12.31 (1.08)
Black (%)	19.7 (0.3)	5.9 (0.1)	23.5 (0.8)	13.83 (0.30)	-3.83 (0.90)
Hispanic (%)	13.3 (0.2)	5.3 (0.1)	11.5 (0.6)	7.99 (0.27)	1.86 (0.69)