RANDOMIZATION INFERENCE WITH AN INSTRUMENTAL VARIABLE

PAUL R. ROSENBAUM

Abstract. Two examples of randomization inference with an instrumental variable are presented, one concerning a randomized clinical trial with imperfect compliance, the other concerning the economic returns to education with a weak instrument. Examples are from two articles: Greevy, R., Silber, J., Cnaan, A., Rosenbaum, P. R. (2004) Randomization inference with imperfect compliance in the ACE-inhibitor after anthracycline randomized trial. *Journal of the American Statistical Association*, **99**: 7-15. and Imbens, G. and Rosenbaum, P. R. (2005) Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *Journal of the Royal Statistical Society*, A, 168, Part 1, to appear.

1. Notation and Review

1.1. **Strata or Matched Sets.** S strata defined by pretreatment covariates, $s = 1, \ldots, S$, with n_s subjects in stratum s, and $N = \sum n_s$. Write $Z_{si} = 1$ if the i^{th} subject in stratum s is treated, $Z_{si} = 0$ if control. There are $m_s = \sum_{i=1}^{n_s} Z_{si}$ treated subjects in stratum s. Matched pairs is the special case with $n_s = 2$, $m_s = 1$ for $s = 1, \ldots, S$. When $S = 1$, drop s subscript, writing n, Z_i, etc.

1.2. **Randomization.** Ω is the set of the $K = \prod_{s=1}^{S} \binom{n_s}{m_s}$ possible values $\mathbf{z}$ of the $N-$dimensional treatment assignment $\mathbf{Z} = (Z_{11}, Z_{12}, \ldots, Z_{S,n_S})^T$ with $m_s = \sum_{i=1}^{n_s} z_{si}$ for $s = 1, \ldots, S$. Randomization: $\mathbf{Z}$ picked at random from Ω, that is, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$ for each $\mathbf{z} \in \Omega$.

1.3. **Treatment Effects.** Subjects observed for K time periods, $k = 1, \ldots, K$, but data for later periods are often missing due to later entry of the subject into the trial and analysis of all data at a particular date. Under treatment, $Z_{si} = 1$, person i in stratum s would have responses y_{Tsi1}, ..., y_{TsiK}, and under control, $Z_{si} = 0$, responses y_{Csi1}, ..., y_{CsiK}; eg Neyman (1923), Rubin (1974). Observed response from this person is Y_{si1}, ..., Y_{siK} where $Y_{sik} = Z_{si}\, y_{Tsik} + (1 - Z_{si})\, y_{Csik}$ if this person is observed for k periods and is missing otherwise. In randomization inference, quantities that depend on the random assignment Z_{si} of treatments, such as Y_{sik}, are random variables, but quantities that do not depend on Z_{si}, such as y_{Tsik} or y_{Csik} are fixed aspects of the finite population of N subjects.

Date: Oct 2004.

Supported by a grant from the US National Science Foundation.

1.4. **Randomization Test of No Effect .** Null hypothesis of no treatment effect is: $H_0 : y_{Tsi1} = y_{Csi1}, \ldots, y_{TsiK} = y_{CsiK}$ for all s, i. Wei and Lachin (1984) proposed an extension of the Mann-Whitney-Wilcoxon test based on pairs of subjects, si and sj at each time k:

$$\begin{aligned} U_{sijk} &= 1 \; if \; Y_{sik} > Y_{sjk} \\ &= -1 \; if \; Y_{sik} < Y_{sjk} \\ &= 0 \; if \; Y_{sik} = Y_{sjk} \; or \; if \; either \; is \; missing, \end{aligned}$$

so that $T_{sk} = \sum_{i=1}^{I} \sum_{j=1}^{I} Z_{si} \left(1 - Z_{sj}\right) U_{sijk}$. If H_0 were true, $Y_{sik} = y_{Tsik} = y_{Csik}$ is fixed under H_0, so U_{sijk} is also fixed under H_0. The Wei and Lachin statistic is $T = \sum_s \sum_k T_{sk}$, which is the usual Mann-Whitney-Wilcoxon statistic if $S = K = 1$. T can be written as a linear rank statistic, so that the null randomization distribution of T is derived from stratified sampling of fixed scores from a finite population.

1.5. **Intent-to-Treat Analysis.** Inference about a typical effect ignoring doses received. Eg, under the model of an additive effect,

$$y_{Tsik} = y_{Csik} + \tau, \tag{1.1}$$

if $H_0 : \tau = \tau_0$ were true, then the adjusted responses satisfy $Y_{sik} - \tau_0 Z_{si}$ would equal y_{Csik}, so the test in §1.4 may be inverted to yield confidence intervals and Hodges-Lehmann point estimates for τ. Only the randomization distribution, $\Pr\left(\mathbf{Z} = \mathbf{z}\right) = \frac{1}{K}$, and the hypothesis being tested, $H_0 : \tau = \tau_0$ in (1.1), are used as the basis for inference. The concern is this is not the effect of taking the drug, but rather the effect of being encouraged to take it. The effect of the drug is not likely to be realized if the drug is not consumed, so (1.1) is not a plausible family of hypotheses when there is noncompliance.

2. Randomization Inference with an Instrumental Variable

2.1. **Doses.** With noncompliance, the doses actually received are outcomes of the randomly assigned encouragement to take enalapril or placebo. Under treatment, $Z_{si} = 1$, person i in stratum s would have doses of enalapril $d_{Tsi1}, \ldots, d_{TsiK}$, and under control, $Z_{si} = 0$, doses $d_{Csi1}, \ldots, d_{CsiK}$. In AAA, $d_{Csi1} = \ldots = d_{CsiK} = 0$, but this is not essential to the argument, and is not true in the second example. Observed dose from this person is $D_{si1}, \ldots, D_{siK}$ where $D_{sik} = Z_{si}\, d_{Tsik} + \left(1 - Z_{si}\right) d_{Csik}$ if this person is observed for k periods and is missing otherwise.

2.2. **Hypothesis: Effect Proportional to Change in Dose.** In IV, the family of hypotheses (1.1) is replaced by the family

$$y_{Tsik} - y_{Csik} = \beta \left(d_{Tsik} - d_{Csik}\right) \tag{2.1}$$

which asserts that the effect is proportional to the change in dose. Note that the exclusion restriction is satisfied: treatment assignment Z_{si} matters for response only indirectly by influencing dose. Notice that compliance may be severely nonrandom in ways that are directly relevant to response; e.g., $d_{Tsik} - d_{Csik}$ and y_{Csik} may be strongly correlated.

2.3. **IV.** If the hypothesis $H_0 : \beta = \beta_0$ in (2.1) were true, then:

$$\begin{aligned} Y_{sik} - \beta_0 D_{sik} &= y_{Tsik} - \beta d_{Tsik} \text{ if } Z_{si} = 1 \\ &= y_{Csik} - \beta d_{Csik} \text{ if } Z_{si} = 0 \\ &= a_{sik}, \text{ say,} \end{aligned}$$

is fixed, not varying with Z_{si}. Invert the test of §1.4 applied to the adjusted responses, $Y_{sik} - \beta_0 D_{sik}$, to build confidence intervals and Hodges-Lehmann estimates for β. As in §1.5, only the randomization distribution, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$, and the hypothesis being tested, $H_0 : \beta = \beta_0$ in (2.1), are used as the basis for inference. *Notice that the two tests agree exactly about whether the null hypothesis of no treatment effect is plausible, returning the same significance level.* However, the family of hypotheses (2.1) is a more plausible family than the family (1.1) when there is imperfect compliance, because the former says that a medication not consumed will not have its biological effects.

3. Conclusions

3.1. **Noncompliance in Randomized Trials.** In randomized trials, *IV* permits inferences in which randomization forms the "reasoned basis for inference" in Fisher's phrase, so that only the randomization distribution, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$, and the hypothesis being tested are used; however, IV permits a family of hypotheses in which medications have biological effects only if they are consumed. The *IV* test and the intent-to-treat test agree exactly about whether the null hypothesis of no effect is plausible.

3.2. **Weak Instruments.** The standard method, two-stage-least-squares (2SLS), works poorly with weak instruments. Two problems: (i) a weak instrument may or may not provide limited information, and (ii) 2SLS can exaggerate the information provided, yielding confidence intervals that cover at much less than their nominal rate. Randomization inference fixes problem (ii), thereby clarifying problem (i).

4. Bibliography

4.1. Papers From Which the Talk is Derived.

References

[1] Greevy, R., Silber, J., Cnaan, A., Rosenbaum, P. R. (2004) Randomization inference with imperfect compliance in the ACE-inhibitor after anthracycline randomized trial. *Journal of the American Statistical Association*, **99**: 7-15.

[2] Imbens, G. and Rosenbaum, P. R. (2005) Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *Journal of the Royal Statistical Society*, A, 168, Part 1, to appear.

[3] Rosenbaum, P. R. (1996) Comment on "Identification of causal effects using instrumental variables" by Angrist, Imbens and Rubin. *JASA*, **91**, 465-468.

[4] Rosenbaum, P. R. (1999) Using combined quantile averages in matched observational studies. *App Stat* **48**, 63-78, §5.

[5] Rosenbaum, P. R. (2002) Covariance adjustment in randomized experiments and observational studies. *Stat Sci* **17**, 286-327, §6.

4.2. Examples Used in the Talk.

References

[1] Angrist, J. D. & Krueger, A. B. (1991) Does compulsory school attendance affect schooling and earnings. *Quarterly Journal of Economics*, **106**, 979-1014.
[2] Silber, J. H., Cnaan, A., Clark, B. J., et al. (2001), "Design and baseline characteristics for the ACE-inhibitor after anthracycline (AAA) study of cardiac dysfunction in pediatric onocology long-term survivors," *American Heart Journal*, 142, 577-585.
[3] Silber, J. H., Cnaan, A., Clark, B. J., et al. (2004) Enalapril to prevent cardiac function decline in long-term survivors of pediatric cancer exposed to anthracyclines. *Journal of Clinical Oncology*, 22, 820-828.

4.3. **A Few of the Many Other Papers About IV.**

References

[1] Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996) Identification of causal effects using instrumental variables (with Discussion). *JASA*, **91**, 444-455.
[2] Bound, J., Jaeger, D. A., Baker, R. M. (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *JASA*, **90**, 443-450.
[3] Maddala, G. S. & Jeong, J. (1992) On the exact small sample distribution of the instrumental variable estimator. *Econometrica*, **60**, 181-183.
[4] Nelson, C. R. & Startz, R. (1990) Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica*, **58**, 967.
[5] Sheiner, L. & Rubin, D. (1995) Intention-to-treat analysis and the goals of clinical trials. *Clin Pharm Therap* **57**, 6–15.
[6] Sommer, A. & Zeger, S. (1991) Estimating efficacy from clinical trials. *Stat Med* **10**, 45–52.
[7] Staiger, D. & Stock, J. H. (1997) Instrumental variables regression with weak instruments. *Econometrica*, **65**, 557-586.
[8] Wald, A. (1940) Fitting of lines if both variables are subject to error. *AMS* **11**, 284-300.

4.4. **Randomization Inference and Causal Effects.**

References

[1] Cox, D. R. & Reid, N. (2000) *Theory of the Design of Experiments.* New York: CRC.
[2] Fisher, R.A. (1935). *Design of Experiments.* Edinburgh: Oliver & Boyd.
[3] May, G. S., Chir, B., DeMets, D. L., et al. (1981) The randomized clinical trial: Bias in analysis. *Circulation*, 64, 669-673.
[4] Neyman, J. (1923), On the application of probability theory to agricultural experiments. Essay on principles. Section 9. (In Polish) Roczniki *Nauk Roiniczych,* Tom X, pp1-51. Reprinted in English in *Statist Sci* 1990, 5, 463-480, with discussion by T. Speed and D. Rubin.
[5] Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Ed. Psych.* **66**, 688-701.
[6] Wei, L. J. & Lachin, J. M. (1984) Two-sample asyptotically distribution free tests for incomplete multivariate observations. *JASA* **79**, 653-661.

4.5. **Inverting Randomization Tests for Additive Effects.**

References

[1] Hodges, J. & Lehmann, E. (1963) Estimates of location based on ranks. *AMS* **34** 598-611.
[2] Lehmann, E. L. (1959) Testing Statistical Hypotheses. NY: Wiley.
[3] Lehmann, E. L. (1998) *Nonparametrics.* NJ: Prentice Hall.
[4] Moses, L. E. (1965) Confidence limits from rank tests. *Technometrics* **7** 257-260.

Department of Statistics, Wharton School of the University of Pennsylvania, Philadelphia, PA 19104-6340 USA.
E-mail address: rosenbaum@stat.wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~rosenbap/index.html