

Assumptions when Analyzing Randomized Experiments with Noncompliance and Missing Outcomes

*Analisi di Esperimenti Randomizzati
in Presenza di Inadempienze e Dati Mancanti*

Fabrizia Mealli

Dipartimento di Statistica “G. Parenti”

Università di Firenze

mealli@ds.unifi.it

Donald B. Rubin

Department of Statistics

Harvard University

rubin@stat.harvard.edu

Riassunto: Gli esperimenti randomizzati sono generalmente considerati i soli validi strumenti utilizzabili per la stima di effetti causali in ambito medico e socio-economico; tuttavia essi sono di frequente caratterizzati da diverse complicazioni, che includono la non perfetta aderenza dei soggetti al protocollo (*noncompliance*) e la presenza di dati mancanti. Partendo da alcuni esempi, in parte già analizzati in letteratura, nel lavoro vengono catalogate tali complicazioni e presentate e discusse le assunzioni necessarie per condurre l'analisi. Viene mostrato come la plausibilità delle assunzioni dipenda dalle caratteristiche dello specifico esperimento (cieco, doppio-cieco, *encouragement design*, etc.) e come una appropriata analisi Bayesiana permetta di rimuovere alcune delle assunzioni e valutare la sensibilità dei risultati rispetto a violazioni delle stesse.

Keywords: Randomized Experiments, Instrumental Variables, Noncompliance, Missing Data, Intention-to-Treat Effect

1. Introduction

Estimating causal effects of interventions is often the focus of empirical studies in medicine and the social sciences. Randomized trials are the only generally accepted tools for estimating causal effects, yet they often suffer from a number of complications, including noncompliance and missing outcomes. Here we catalogue basic complications and associated assumptions under which analysis may proceed relatively directly. Starting from simple examples, we review and propose different sets of exclusion restrictions, and discuss which ones seem to be more appropriate for different settings (randomized experiments, double-blind placebo controlled trials, encouragement designs, etc.). We will show that (1) all assumptions are well beyond those in the standard Instrumental Variable approach (Angrist et al. (1996)), (2) all involve scientific assumptions that must conform to the study set-up, (3) all are conceptually simple to understand (although not necessarily to analyze correctly), (4) some can be relaxed by conducting a full Bayesian analysis to avoid full identification. We will concentrate on examples with two treatments, binary (all-or-none) compliance and a missingness mechanism that involves only the outcome variables; this allows to discuss basic ideas and could be easily extended to multiple treatments and partial compliance.

2. Noncompliance but no missing outcome

Consider a two-arm randomized experiment that compares a new treatment versus a standard one: each individual i who participates in the study can either be assigned to the new (i.e., active) treatment, $Z_i=1$, or the standard (i.e., control) one, $Z_i=0$. Following the potential outcome approach to causal inference (i.e., Rubin Causal Model, RCM, Holland, 1986), define $Y_i(z)$ to be the two potential outcomes if an individual is assigned to treatment z ($z=0,1$); if no outcome is missing, either $Y(1)$ or $Y(0)$ is known for each subject in the experiment, regardless of compliance. Noncompliance occurs if the actual treatment patients receive is different from the nominal assignment. Here we assume a simple pattern of noncompliance, namely all-or-none: soon after randomization some subjects assigned to the new treatment will not take it, but effectively take the control, whereas some of those assigned control receive the new treatment. Let $D_i(z)$ be an indicator for the treatment received (1 for new or active, 0 for standard or control) given assignment z : $D_i = D_i(Z_i)$ be the actual treatment received. The two indicators $D_i(1)$ and $D_i(0)$ describe the compliance behavior and are used to partition the population of units into four types: compliers (for whom $D_i(z)=z$, $z=0,1$), never-takers (NTs, for whom $D_i(z)=0$, $z=0,1$), always-takers (ATs, for whom $D_i(z)=1$, $z=0,1$), and defiers (for whom $D_i(z)=1-z$, $z=0,1$). The compliance status is observed only for some subjects, although by randomization it is guaranteed to have the same distribution in both treatment arms.

Intention-to-treat (ITT, as-randomized) analysis is one of several strategies designed to deal with the noncompliance problem; as it compares outcomes according to initial group assignment, ignoring compliance behavior, it remains statistically valid irrespective of compliance being perfect or not. Problems arise when interpreting the results, however, because ITT informs us about the effect of treatment assignment, but if we want to learn something about the effect of treatment received, then additional assumptions are required. Some common assumptions, which relate also to the econometric instrumental variable framework, are the exclusion restrictions for ATs and NTs: because assignment has no effect on their compliance behavior, it might plausibly have no effect on the outcome as well. These two assumptions, together with the assumption that defiers do not exist (also called monotonicity assumption, Angrist and Imbens, 1994) allow the identification of the ITT effect for compliers. Only for compliers can we hope to learn something about the effect of treatment received in this experiment, as never-takers (always-takers) are never (always) observed taking the treatment in this experiment. Note that the ITT effect for compliers is attributable to treatment received by assumption, and it is the desire to make this attribution more plausible that underlies the practice of blinding or double-blinding in medical evaluation, a practice that, for example, is typically impossible in encouragement designs. When noncompliance is present, the global ITT effect is usually regarded as being conservative, i.e. a positive ITT effect is manifested only if there is really a positive effect of treatment. It can be shown, that this intuition may sometimes be wrong; relaxing some assumptions, in particular the exclusion restriction for NTs or ATs, one can sometimes detect whether assignment itself has an effect on the outcome.

In Hirano et al. (2000) a medical example is given, that concerns the effects of inoculation for influenza. The study is an encouragement design, where a randomly selected group of physicians received a letter encouraging them to inoculate patients at risk for flu. A standard ITT analysis shows that encouragement decreases hospitalization rates: such a result is usually interpreted as indicating beneficial effects of the receipt of treatment (i.e., the flu shot). Relaxing exclusion restrictions (as in Imbens, Rubin, 1997) within a full Bayesian analysis allows us instead to estimate the effect of assignment (i.e., encouragement) for various subpopulations defined by compliance status: analysis suggests that encouragement has a similar beneficial effect on people who would have received the flu shot regardless of the assignment, the ATs, as on the compliers. There is thus little evidence in this experiment that the flu shot itself had beneficial effects.

3. Missing outcomes but full compliance

When there are missing outcomes, we cannot conduct a simple ITT analysis without some form of imputation of the missing values, either implicit or explicit.

Simple ITT analysis based on complete cases (evaluable population approach) is based on the Missing Completely At Random assumption, that has testable implications and is often rejected by the data. The analysis based on the worst-possible value imputation is usually not scientifically defensible, except as an extreme form of sensitivity analysis; scientifically based imputation is much preferable.

Colton et al. (2001) is an example where multiple imputation has shown its superiority over such overly conservative approaches. The study analyzes data from a multicenter randomized trial, with treatment blinding of the surgeon, evaluator and patient, aimed at showing the effectiveness and safety of Intergel, a new post surgical adhesion prevention solution, compared to an existing solution. Patients had to undergo two invasive procedures, a laparotomy and a (second-look) laparoscopy; 16 patients (6 %) did not have the second-look laparoscopy, so that their outcome was missing. The worst-possible value imputation for these patients did not make use of the substantial information available in the dataset, regarding the reasons patients did not receive a second-look: this information must be considered in any scientifically sound imputation. The multiple imputation (Rubin, 1996), applied in a completely blinded manner, was based on a matching method (similar to a “hot-deck”) and allowed the analyses to be conducted on the full ITT population. The analyses based on the multiply-imputed ITT populations support the original conclusions regarding the superiority of Intergel over the standard solution at preventing new surgical adhesions.

4. Noncompliance and missing outcomes

When both complications are present, then noncompliance and response behavior have to be jointly taken into account and modeled in some principled way.

As far as the response behavior is concerned, an appealing assumption that has been proposed (Frangakis, Rubin, 1999), and that links the compliance with the non response behavior, is Latent Ignorability: potential outcomes and potential response indicators are

assumed to be independent within each level of the compliance covariate (with the same value of the other fully observed covariates). This assumption alone does not allow full identification of the ITT effect for compliers. Additional assumptions on the outcome and on the response are required. One of these is the “compound exclusion” restriction for NTs and ATs: if assignment is assumed to have no effect on the outcome for NTs and ATs, it might plausibly have no effect on their response behavior as well.

Alternative assumptions can also be exploited, for example, assuming that the exclusion restrictions on outcome hold for NTs and ATs, but the exclusion restrictions on response hold for compliers and ATs. The rationale for this is that those who denied participation (the NTs) might be induced not to provide information on their outcome. Mealli et al. (2001) is an example of a randomized trial suffering from both complications: in the study, two methods of teaching breast-self examination (BSE) were compared to assess whether teaching programs can increase BSE practice and quality of the examination. Only 55% of the women in the treatment arm complied with their assigned treatment and only 65% provided information on the outcome. The results suggest that the usual Missing at Random assumption is less plausible here than latent ignorability, and that assignment did indeed have an effect on the response behavior of the women involved in the study.

5. Concluding remarks

The examples discussed above show that there are no universally good assumptions. Plausible assumptions are specific to each context. But explicit restrictions can often be scientifically more plausible than some implicit assumptions of standard ITT approaches.

References

- Angrist J., Imbens G.W., Rubin D.B. (1996) Identification of Causal Effects using Instrumental Variables, *JASA*, 91, 444-472.
- Colton T., Piantadosi S., Rubin D.B. (2001), Report on INTERGEL Adhesion Prevention Solution, mimeo.
- Frankgakis C.E., Rubin D.B. (1999) Addressing Complications of Intention-to-Treat Analysis in the Combined Presence of All-or-None Treatment-Noncompliance and Subsequent Missing Outcomes, *Biometrika*, 86, 365-379.
- Hirano K., Imbens G.W., Rubin D.B., Zhou X. (2000) Assessing the Effect of an Influenza Vaccine in an Encouragement Design, *Biostatistics*, 1, 69-88.
- Imbens G.W., Angrist J. (1994) Identification and Estimation of Local Average Treatment Effects, *Econometrica*, 62, 467-476.
- Imbens G. W., Rubin D.B. (1997) Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance, *Annals of Statistics*, 25, 305-327.
- Holland P (1986) Statistics and Causal Inference, *JASA*, 81, 945-970.
- Little R., Rubin D.B. (1987) *Statistical analysis with missing data*, Wiley, New York.
- Mealli F., Imbens G. W., Ferro S., Biggeri A. (2001), Analyzing a Randomized Trial on Breast Self-Examination with Noncompliance and Missing Outcomes, mimeo.
- Rubin D.B. (1996) Multiple Imputation after 18+ Years, *JASA*, 91, 473-489.