DEALING WITH MULTIVARIATE OUTCOMES IN STUDIES FOR CAUSAL EFFECTS

Donald B. Rubin
Harvard University
1 Oxford Street, 7th Floor
Cambridge, MA 02138 USA
Tel: 617-495-5496; Fax: 617-496-8057
email: rubin@stat.harvard.edu

Abstract Multivariate outcomes are common in studies for causal effects, but they are often mis-analyzed. The critical reason leading to their misuse is that even in a randomized experiment with only two treatment conditions, there are two potential outcomes associated with each measured outcome, but these can never be jointly observed. When there are two measured outcomes, one primary and another one that is intermediate in some sense, the resulting structure can therefore be very confusing, even to great statisticians such as R.A. Fisher. Two specific examples will be used to illustrate the issues using the concept of "principal stratification" (Frangakis and Rubin, 2002, *Biometrics*): the first on estimating dose-response when there is noncompliance, and the second on separating "direct" and "indirect" causal effects.

1. Introduction

Causal inference is inherently multivariate because causal effects, even just for a single experimental unit and only two exposure conditions, involves the implicit comparison of two values, the value of the outcome when the unit is exposed to the treatment condition, $Y(1)$, and the value of the outcome when the unit is exposed to the control condition, $Y(0)$. The causal effect of the treatment versus control on this unit is the comparison of the potential outcomes $Y(1)$ and $Y(0)$, for example, the difference, $Y(1) - Y(0)$. This formulation in the context of randomized experiments and randomization-based inference is due to Neyman (1923) as discussed in Rubin (1990), and was extended to nonrandomized studies in Rubin (1974) and to Bayesian inference in Rubin (1975,1978). The general perspective is now often referred to as "Rubin's Causal Model" (RCM) for a sequence of articles written in the 1970s – see Holland (1986), Imbens and Rubin (2007).

In many situations, causal inference is even more multivariate, even with just one primary outcome represented by the pair of potential outcomes $Y(1)$, $Y(0)$. Of particular interest, noncompliance with the assigned exposure condition creates another pair of potential outcomes, which are intermediate in some sense, as discussed explicitly in Angrist, Imbens, and Rubin (1996). A recent discussion in the context of a double blind randomized trial appears in Jin and Rubin (2007), where $D(1)$ is the intermediate potential outcome representing the proportion of a prescribed cholesterol-reducing drug taken when assigned to take the drug, and

d(0) is the intermediate proportion of a placebo taken when assigned to take the placebo. We discuss this example in some detail in Section 2.

A particularly challenging example of multivariate outcomes occurs when trying to describe "direct" and "indirect" causal effects, for example using surrogate markers, such as immunogenicity as measured by antibody levels in vaccine trials with human and nonhuman primates, as discussed in Rubin (2004). This type of situation is very tricky to deal with correctly, and even the great R. A. Fisher got it wrong, as discussed in detail in Rubin (2005), and overviewed here in Section 3. The framework we use to discuss the two examples is "principal stratification" due to Frangakis and Rubin (2002) and is inherently multivariate.

2.   Noncompliance and Dose-Response

For person i in a double-blind placebo-controlled randomized trial, let $Y_i(1)$ and $Y_i(0)$ be cholesterol reduction when assigned treatment and the cholesterol reduction when assigned control, respectively, and analogously, let $D_i(1)$ and $d_i(0)$ be the proportion of dose of active drug taken when assigned to take the treatment drug, and the proportion of placebo drug taken when assigned to take the control, respectively; in this study $D_i(0) = d_i(1) = 0$ by construction – that is, the amount of drug taken when assigned control and the amount of placebo taken when assigned treatment are both zero. Thus, the potential outcomes take values is a four dimensional multivariate space of $Y_i(0)$, $Y_i(1)$, $D_i(1)$, $d_i(0)$. The treatment potential outcomes, $Y_i(1)$ and $D_i(1)$, are only observed for units assigned to treatment, whereas the control potential outcomes are only observed for units assigned to control.

Figure 1, originally from Efron and Feldman (1991) [EF], but also reproduced in Jin and Rubin (2007), plots $Y_i(1)$ versus $D_i(1)$ in the treatment group and plots $Y_i(0)$ versus $d_i(0)$ in the control group. Figure 1 reveals a relatively clear monotonely increasing relationship between cholesterol reduction and proportion of drug dose taken, as would be expected with a drug designed to reduce cholesterol. However, there also appears to be a mild relationship between cholesterol reduction and proportion of placebo taken, which may make little sense until we realize that better placebo compliers are probably also better dietary compliers, better exercise compliers, etc., and thus the better placebo compliers are doing other things to reduce their cholesterol levels. Thus, $d_i(0)$ is an important explanatory covariate, which is only observed in the control group. The principal strata are defined by the various values of $d_i(0)$. Of importance is that, as seen in Figure 2, the distribution of drug compliance, $D_i(1)$, is substantially worse than the distribution of placebo compliance, $d_i(0)$, presumably because of negative side effects of the active drug, and so we cannot directly infer the missing $D_i(0)$ from the observed $d_i(1)$. In some sense, we want to "subtract" the dose-response curve in the placebo group from the dose-response curve in the treatment group to uncover the true biological dose-response relationship, but how to do this formally and correctly is the key issue in EF and in Jin and Rubin (2007).

Suppose that $d_i(0)$ were observed for all units and that the proportion of active drug taken, $D_i(1)$, was ignorably assigned (Rubin 1978) at some value between zero and $d_i(0)$; thus, we are assuming that the assignment mechanism for $D_i(0)$ is "latently ignorable" (Frangakis and Rubin 1999), latently because $d_i(0)$ is missing for those assigned to the treatment condition (i.e., the assignment mechanism would be ignorable if $d_i(0)$ were observed for all units. Jin and Rubin

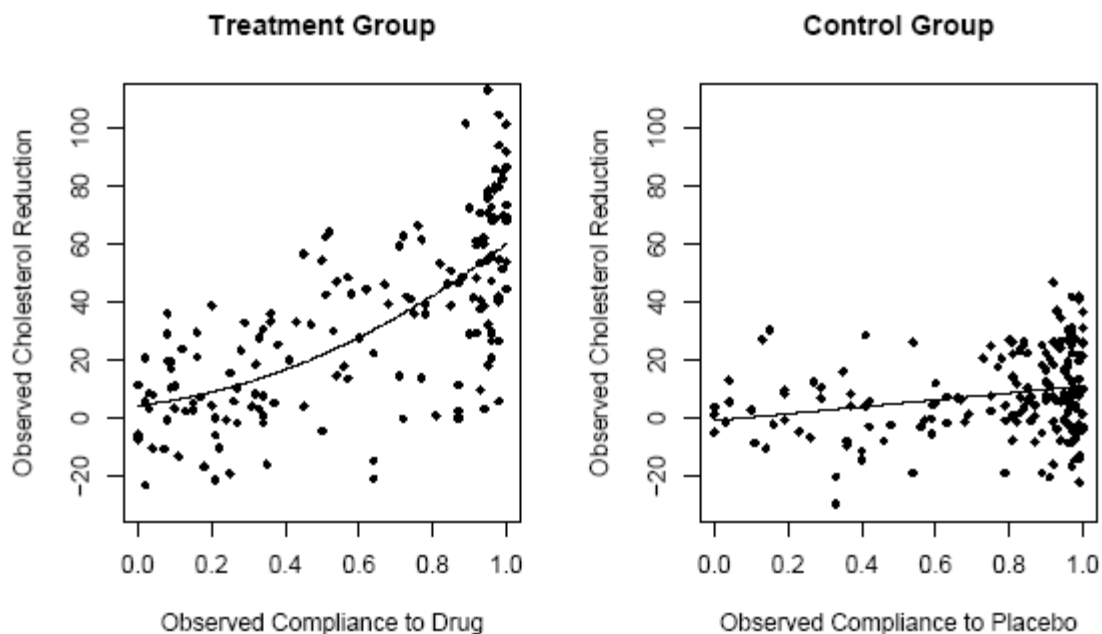**Treatment Group**   **Control Group**

FIGURE 1: Observed Dose-Response in the EF Data

(2007) conduct a Bayesian analysis under the latent ignorable and other modeling assumptions, and derive different estimated dose-response curves (i.e., $Y_i(1)$ as a function of $D_i(1)$) as functions of $d_i(0)$, where $d_i(0)$ define various principal strata.

Four estimated true dose-response curves are given in Figure 3 for four values of placebo compliance: a prefect placebo complier, $d_i(0)=1.00$; the 75th percentile placebo complier in the experiment, $d_i(0)=0.97$; the median placebo complier, $d_i(0)=0.89$; and the 25th percentile placebo complier, $d_i(0)=0.60$. Notice that dose-response is only plotted for active doses less than $d_i(0)$ because, by assumption, no active dose is assigned that is greater than $d_i(0)$. The posterior medians are represented by the bold lines and 95% posterior intervals by the light lines. Under the assumptions of Jin and Rubin's model, a very interesting conclusion arises that can be illustrated by examining the estimated cholesterol reductions at dose 0.60 for the four types of placebo compliers in the figure: The better the placebo compliance, the less the benefit from taking the active drug! Apparently, the better placebo compliers are already doing other things to reduce their cholesterol levels, whereas the worse placebo compliers are not doing so, and so they benefit more from the same dose of the active drug.

Such a conclusion would have been impossible to reach without the multivariate thinking that is forced upon us when using potential outcomes and principal stratification. That is, letting W_i indicate the condition assigned, $W_i = 1$ for treatment and $W_i = 0$ for control, the reduction from the bivariate potential outcomes $[Y_i(1),Y_i(0)]$ to the univariate observed outcome, $Y_{obs,i} = W_i Y_i(1) + (1-W_i) Y_i(0)$, and the reduction from the bivariate potential outcomes $[D_i(1), d_i(0)]$ to the univariate observed outcome, $\Delta_{obs,i} = W_i D_i(1) + (1-W_i) d_i(0)$, hide the insight revealed in Figure 3. EF (1991) effectively used the reduction to $\Delta_{obs,i}$ because they assumed $D_i(1)$ and $d_i(0)$
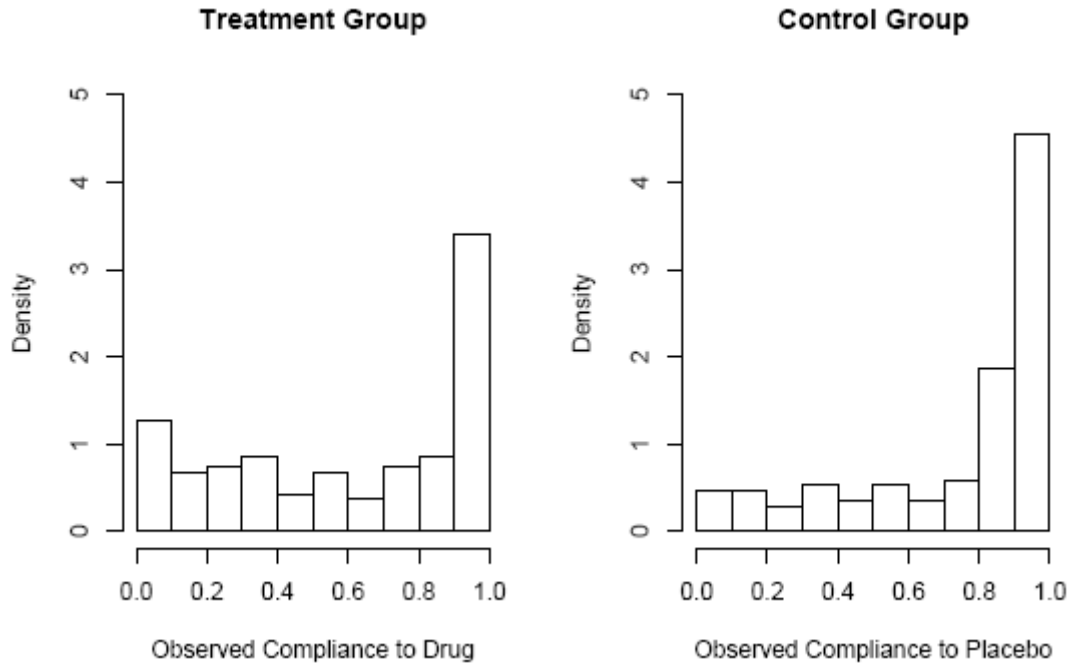
Treatment Group

Control Group

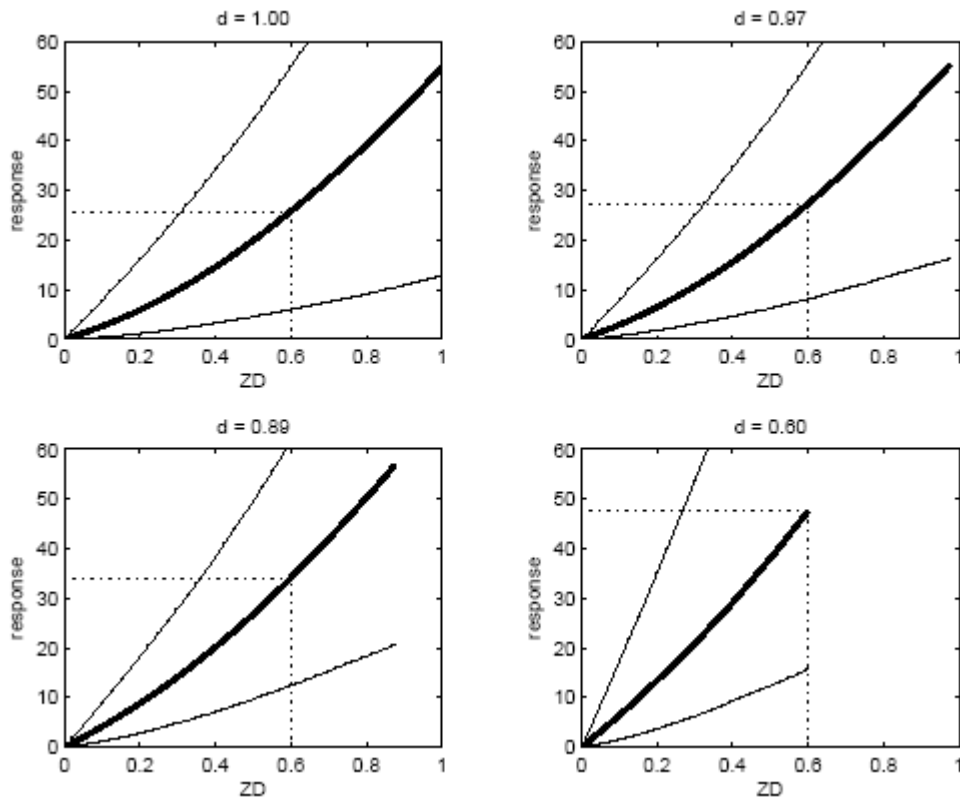FIGURE 2: Histograms of Observed Compliances in the EF Data

FIGURE 3: Estimated True Dose-Response at Four Values of $d_i(0)$:
Posterior Median and Control 95% Interval

were a monotone deterministic function of each other, and therefore assumed that $D_i(1)$ was known from $d_i(0)$, thereby reducing the four plots in Figure 3 to four points.

3.   Complex Experiments: "Direct" and "Indirect" Causal Effects

Now consider the problem of adjusting for a different kind of "concomitant" variable – an outcome variable that is not the outcome of primary interest nor a measure of compliance, but may be "on the causal pathway" explaining how treatment versus control affects the primary outcome variable, Y. The following discussion is adapted from my Fisher lecture (Rubin, 2005). Fisher wrote about the problem of adjusting for such on a concomitant variable in *Design of Experiments* (DOE, Chapter IX) from the first edition in 1935 to the last (8th) edition in 1966 using the identical language:

> In agricultural experiments involving the yield following different kinds of treatments, it may be apparent that the yields of the different plots have been much disturbed by variations in the number of plants which have established themselves. If we are satisfied that this variation in plant number is not itself an effect of the treatments being investigated, or if we are willing to confine our investigation to the effects on yield, excluding such as flow directly or indirectly from effects brought about by variations in plant number, then it will appear desirable to introduce into our comparisons a correction which makes allowance, at least approximately, for the variations in yield directly due to variation in plant number itself.

He also discussed such a situation in *Statistical Methods for Research Workers* (1970, sec. 49.1), in a way that was consistent with this previous quotation.

Fisher's recommendation was to conduct an analysis of covariance (ANCOVA) of $Y_{obs,i}$ on W_i and $C_{obs,i}$, where $C_{obs,i} = W_i C_i(1) + (1-W_i) C_i(0)$. Essentially, an ANCOVA compares the average observed $Y_i(1)$ with the average observed $Y_i(0)$ for units with a common value of $C_{obs,i}$, which generally does not estimate a causal effect as we now discuss using two hypothetical examples. Suppose that Tables 1 and 2 represent very large randomized experiments of N agricultural plots; the concomitant C is the number of plants established in each plot, the primary outcome Y is the total yield in each plot, the treatment is a new fertilizer, and the control is the standard fertilizer. In each experiment, half of the units are randomly assigned to the treatment and control conditions.

The left collection of columns in Table 1 gives the potential outcomes in the first experiment. The first two rows represent those N/2 units with common values of the potential outcomes, and so randomly assigning them would result in half being assigned to control, represented by the first row, and half being assigned to treatment, represented by the second row, and analogously for the second pair of rows. The resultant observed data are represented in the right collection of columns in Table 1. Each pair of rows corresponds to a "principal stratum," where each principal stratum is defined by common values of $C_i(1)$ and $C_i(0)$, which is unaffected by treatment assignment. The left collection of columns reveal that for all units, there is a causal effect of treatment versus control on the concomitant variable, C, of size 1, but there is no treatment effect on the primary outcome variable Y for any unit. Because all units

experience a treatment versus control effect on C but no effect on Y, the answer to the question of the "direct" effect of treatment on Y, after adjusting for C, seems to be a matter of definition, where the obvious definition is zero.

Table 1. An Example With a Treatment Effect on the Concomitant, C, But No Treatment Effect on the Primary Outcome, Y.

Fraction of population	Potential outcomes				Observed data		
	C(1)	C(0)	Y(1)	Y(0)	W	C_{obs}	Y_{obs}
1/4	3	2	10	10	0	2	10
1/4	3	2	10	10	1	3	10
1/4	4	3	12	12	0	3	12
1/4	4	3	12	12	1	4	12

An examination of the right collection of columns in Table 1, however, reveals that conditioning on the observed value of the concomitant, $C_{obs,i}$, leads to a contradictory conclusion. When $C_{obs,i} = 3$, those plots that received the treatment do worse by 2 than those that received control (compare the second and third rows in the observed data). Also in this case, the regressions of $Y_{obs,i}$ on $C_{obs,i}$ in the $W_i = 0$ and $W_i = 1$ groups are linear and parallel, with constant treatment minus control difference equal to -2. So the conclusion of Fisher's recommended ANCOVA is that the treatment versus control direct effect on Y (after making allowance for any treatment effect on C) is negative! This clearly seems incorrect, as is revealed by an examination of the potential outcomes in Table 1.

The example in Table 2 illustrates a flaw in Fisher's proposed solution even when there does appear to be a well-defined "direct" treatment versus control effect on Y after controlling for C, at least for some units. The example is similar to the one in Table 1 except that first, there is a constant treatment effect on Y of size 1 for all units, and second, for one-third of the units, represented by the middle two rows, there is no treatment effect on the concomitant, C. For the other units, the treatment effect on the concomitant is 1. For the principal stratum where there is no treatment effect on the concomitant, the answer to the question about the direct effect of treatment seems to be clear: It is size 1. Yet here too, Fisher's advice yields a different and incorrect answer, despite the parallel linear regression lines in the $W_i = 0$ and $W_i = 1$ groups. The ANCOVA of $Y_{obs,i}$ on W_i and $C_{obs,i}$ implies that the "direct" causal effect of treatment on the primary outcome, after accounting for the effect of treatment on the concomitant, is of size -1, which is the average $Y_{obs,i}$ for the treated plots with $C_{obs,i} = 3$ (the average $Y_{obs,i}$ in rows 2 and 4, i.e., 12) minus the average $Y_{obs,i}$ for the control plots with $C_{obs,i} = 3$ (the average $Y_{obs,i}$ in rows 3 and 5, i.e., 13).

Table 2. An Example With a Constant Treatment Effect on the Outcome, Y, and a "Direct"
Effect for Units With No Treatment Effect on the Concomitant, C.

Fraction of population	Potential outcomes				Observed data		
	$C(1)$	$C(0)$	$Y(1)$	$Y(0)$	W	C_{obs}	Y_{obs}
1/6	3	2	11	10	0	2	10
1/6	3	2	11	10	1	3	11
1/6	3	3	13	12	0	3	12
1/6	4	3	13	12	1	3	13
1/6	4	3	15	14	0	3	14
1/6	4	3	15	14	1	4	15

4.   Discussion

Causal inference is inherently multivariate because underlying each measured outcome variable, there are at least two potential outcomes, only one of which is revealed by the assignment mechanism.   With more than one measured outcome, as when we have noncompliance or are interested in "direct" and "indirect" effects, there are at least four underlying potential outcomes, and understanding their relationships requires clear multivariate thinking, which, in my mind, only recently has been adequately clarified through the use of principal stratification (Frangakis and Rubin, 2002).

References

Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996).   Identification of causal effects using instrumental variables. *Journal of the American Statistical Association,* 91, 444-472.

Efron, B. and Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association*, 86, 9-17.

Fisher, R.A. (1935).   *Design of Experiments*. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1966).   *Design of Experiments, 8th ed.* Edinburgh: Oliver & Boyd.

Fisher, R.A. (1970).   *Statistical Methods for Research Workers* (14th ed.). Edinburgh: Oliver & Boyd.

Frangakis, C.E. and 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, 2, 365-379.

Frangakis, C.E. and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics*, 58, 21-29.

Holland, P.M. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-960.

Imbens, G. and Rubin, D.B. (2007). *Causal Inference in Statistics, and in the Social and Biomedical Sciences*. Cambridge University Press: New York [to appear].

Jin, H. and Rubin, D.B. (2007) "Principal Stratification for Causal Inference with Extended Partial Compliance: Application to Efron-Feldman Data." To appear in The Journal of the American Statistical Association.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: essay on principles, section 9. Translated in *Statistical Science*, 5, 465-480, 1990.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.

Rubin, D.B. (1975). Bayesian inference for causality: the importance of randomization. *The Proceedings of the Social Statistics Section of the American Statistical Association*, 233-239.

Rubin, D.B. (1978). Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, 6, 34-58.

Rubin, D.B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5, 472-480.

Rubin, D.B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, 31, 161-170; (with discussion and reply) 195-198.

Rubin, D.B. (2005). Causal inference using potential outcomes: design, modeling, decisions. 2004 Fisher Lecture. *Journal of the American Statistical Association*, 100, 322-331.