Biostatistics (2002), **3**, 2, *pp.* 173–177
Printed in Great Britain

Rejoinder to discussions on clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms

CONSTANTINE E. FRANGAKIS, DONALD B. RUBIN, XIAO-HUA ZHOU

We thank the editors and discussants for the opportunity to have this commentary

BERRINGTON AND COX

Berrington and Cox (B&C) raise three main points.

Stability of potential outcomes and unit of randomization

B&C express the view that, because the unit of randomization is the physician, the unit of study and comparison should also be the physician, not the patient. We certainly agree that a valid analysis must respect the unit of randomization (the physician), and to do so when faced with noncompliance at the level of a subunit (physician-patient pair) was the primary motivation for writing this paper. That is, the principal estimand in our analysis involves only the complying subunits, yet assignment is at the unit level. In consequence, a valid analysis for the complier average causal effect must have the subunit as an entity while also accounting for the randomization at the unit level.

Our approach to this particular application involves an assumption, 'stability', which concerns the potential outcomes. There is one set of potential outcomes at the subunit level when they are assigned treatment and another set of potential outcomes when they are assigned control. Stability includes the 'no interference' assumption (Cox, 1958, p. 19), which requires that whether or not a discussion takes place between a patient–physician pair is a function of whether or not that pair was assigned to discuss AD, but not of whether or not other pairs were assigned to discuss AD. Stability may be more or less plausible depending on the manner in which the design is implemented, but randomization itself plays no role in this plausibility. In our application, stability is reasonable because, for a given physician, the encouragements were sent separately for every patient's visit, and a physician's responses are expected to be specific to each patient, although of course correlated within a doctor, which is allowed by our analysis.

The issue of randomization at the unit (here, physician) level is distinct from the issue of stability at the subunit level, and implies how data must be validly analyzed, from either the randomized-based perspective (Fisher, 1925; Neyman, 1923) or the Bayesian perspective (Rubin, 1978). We fully adhere to this dictum, as our analysis accounts for the randomization at the physician level. We thank B&C for forcing clarification on this important point.

If possible deviations from stability at the subunit level are of concern, we can define potential outcomes and principal strata at a coarser level (e.g. physicians), which, as B&C also indicate, would result in multilevel compliance, although then more assumptions would be needed to draw causal inferences (e.g. see Section 5). Therefore, given the level at which stability is assumed plausible, the critical issues of drawing causal inferences under noncompliance remain the same in principle, as now discussed.

Approaches to noncompliance

We fully agree with B&C that approaches to noncompliance should be appropriate to the specific case. For example, although a common approach to noncompliance is to assume the exclusion restriction (Angrist *et al.*, 1996), in this paper we stated explicitly that we do not make this assumption that outcome depends only on treatment received (Section 3.2, paragraph 2). Our model (Section 3.3) explicitly allows that, for subunits who would receive the same treatment no matter the assigned treatment (i.e. always-takers and never-takers), assignment may have an effect on the outcome.

Also, regarding reasons for noncompliance in our application, all included patient–physician pairs were able to have discussions of AD, and, based on existing literature, we believe a main reason for low overall AD discussion (even after physician encouragement) is physician–patients' unfamiliarity with ADs. Furthermore, our framework explicitly allows compliance to be differential across subunits (physician–patient pairs), and to be differential across assignment arms within subunits, and further allows explicit modeling of measured factors for compliance (model for compliance principal strata, Section 3.3). For example, of particular relevance are our results that the younger the patient, the lower the probability of being a 'discussion-complier' versus a 'never-discussant' but that the probability of being an 'always-discussant' does not change as much, suggesting that, when the patient is younger, physician–patient pairs do not value ADs as much.

Net-treatment comparisons versus causal effects

B&C present summary statistics of proportions of AD completion stratified by encouragement and observed discussion, also presented by us in our Table 1. When one of the stratifiers is, as here, a post-randomization variable, comparisons among such proportions are called 'net-treatment' comparisons (Cochran, 1957; Rosenbaum, 1984). B&C seem to suggest that these comparisons provide evidence that the discussion has a large effect on the probability that the AD form is completed, an implied causal inference. However, because net-treatment comparisons stratify on the post-randomization treatment received, they do not generally reflect causal effects either of that variable or of randomization itself on the outcome, as has been demonstrated in practice (e.g. The Coronary Drug Project Research Group, 1980) and in theory (Rosenbaum, 1984).

Our framework of principal strata (Frangakis and Rubin, 2002) was developed to address precisely this issue: the framework defines estimands that adjust for the post-randomization treatment received, and are always causal effects, unlike net-treatment comparisons. The resulting approach based on principal strata still estimates the causal effects of the randomized treatment. A revealing analysis that respects randomization can be done, with or without exclusion assumptions, as justified in Section 3.2 of this paper, and demonstrated by our results, as well as evidenced by results in Imbens and Rubin (1997a) and Hirano *et al.* (2000).

GOETGHEBEUR AND VANSTEELANDT

We are in agreement with Goetghebeur and Vansteelandt's (G&V) first part of the discussion that, regardless of framework, it is harder to draw causal inferences if the stability assumption (see also related point in reply to B&C) is relaxed, mainly because more assumptions are then needed to address the induced multilevel compliance. For such cases, an approach along the lines introduced in Section 5 of the paper for multilevel compliance could be useful.

In the remaining part of their discussion, G&V claim that principal stratification makes more assumptions than needed and is hard to apply to more demanding data structures. We don't agree with either point. We first argue that making fewer explicit assumptions, as in G&V's approach, leads to implicit and less plausible assumptions. Second, we illustrate the flexibility of our approach with more demanding data structures.

Interpretation

Our framework of principal stratification is driven by the factors z that are explicitly controlled in the study—here, the treatment randomly assigned. A post-randomization variable D, then, is a function of the controllable factor z. Principal stratification is the sub-classification of people based on the values of post-randomization variables under all levels of the controllable factor, and principal effects are causal effects of z on primary outcomes Y conditional on principal strata (Frangakis and Rubin, 2002). This approach to defining causal effects adjusted for a post-randomization variable involves potential outcomes, which exist because each can be observed with a specific action: we will observe $D_i(z = 0)$ and $Y_i(z = 0)$ if we assign i to $z = 0$, and observe $D_i(z = 1)$ and $Y_i(z = 1)$ if we assign i to $z = 1$.

However, G&V's approach requires defining some outcomes ('treatment free' outcomes, Y_{ij0}, in G&V's paragraph 4 and equation(1)) that are not potential outcomes because they are not even potentially observable—they are non-existent. (In G&V's notation of Y_{ij0}, zero indicates treatment *received*, as opposed to treatment *assigned*.) For example, consider a physician-subject pair who discusses AD no matter what the assignment ($D_i(z = 0) = D_i(z = 1) = 1$, i.e. an always-discussant pair): the value of the outcome Y with no discussion is non-existent for this pair, and it is impossible in this study for the controllable factor z to force this D_i to be 0. Then, the assumption of G&V's approach is that *all* mechanisms that could force D_i to be 0 would result in the identical value of G&V's Y_{ij0}, but this is an implicit and generally implausible assumption on the principal stratum of 'always-discussants'.

If, on the other hand, we state explicitly how to control, in addition to z, a factor, say z', to try to change D_i, then the outcomes Y_i will still not be functions of D_i, but of the controllable factors (z, z') we used to try to change D_i. The resulting approach to causal inference adjusted for D would then still be based on a principal stratification, and its conclusions would depend on what assumptions are made on the relative frequency of the principal strata and the distribution of outcomes given principal strata.

Application to more demanding data: an illustration

We outlined in Section 5 how our approach can be used when treatment received has more than two levels. Here, we illustrate how principal stratification can formulate causal effects when treatment assigned has more than two levels and is measured longitudinally.

The Baltimore Needle Exchange Program (NEP) operates sites where drug addicts can visit, with confidentiality, and exchange a used needle for a clean one (Vlahov *et al.*, 1997; Strathdee *et al.*, 1999, Frangakis *et al.* 2001, Unpublished data). The NEP hopes to reduce HIV transmission, and the goal is to assess this from the data. For each drug user i, data combined from the NEP and a larger cohort study in Baltimore, are available at every semester t on the following variables: covariates measuring observed risk factors for HIV; the distance, $Z_{i,t}$, of the subject's domicile from the closest NEP site at that time; an indicator $D^{\text{obs}}_{i,t}$ for whether or not the subject actually exchanges drug needles at the NEP; and an indicator $Y^{\text{obs}}_{i,t}$ for whether or not the subject truly becomes HIV positive.

The standard way to evaluate NEPs is to compare HIV incidence between subjects who exchange and do not exchange. However, even after adjusting for the covariates, exchanging at the NEP is not controlled by the researchers, and subjects who exchange are probably at different (expectedly higher) risk for HIV, independently of exchange. Nevertheless, the researchers did control the location of the NEP

Table 1. *Structure of principal stratification with multilevel controlled factor in the example of needle exchange at a particular time (time index is omitted)*

(a)						(b)				
Principal stratum C means	what the exchange behaviour D will be as function of distance z:					Fraction of people who get HIV, given C and as function of distance z:				
	$D(z = 6)$	$D(z = 5)$	$D(z = 4)$	. . .	$D(z = 1)$	$\lambda(z = 6)$	$\lambda(z = 5)$	$\lambda(z = 4)$	. . .	$\lambda(z = 1)$
$C = 6+$	1	1	1	. . .	1	λ_{6+}	$\longrightarrow$	$\longrightarrow$	$\longrightarrow$	$\longrightarrow\!\mid$
$C = 5$	0	1	1	. . .	1	λ_5	$\lambda_5 * R_5$	$\longrightarrow$	$\longrightarrow$	$\longrightarrow\!\mid$
$C = 4$	0	0	1	. . .	1	λ_4	$\longrightarrow$	$\longrightarrow\!\mid$ $\lambda_4 * R_4$	$\longrightarrow$	$\longrightarrow\!\mid$
. . .	. . .	. . .	. . .	. . .	. . .	. . .				
$C = 1$	0	0	0	. . .	1	λ_1	$\longrightarrow$	$\longrightarrow$	$\longrightarrow$	$\longrightarrow\!\mid$ $\lambda_1 * R_1$
$C = 0$	0	0	0	. . .	0	λ_0	$\longrightarrow$	$\longrightarrow$	$\longrightarrow$	$\longrightarrow\!\mid$

sites, in a way that can be assumed random within areas of high past observed risk as captured by the covariates. Then, the key points are that we have strong preliminary evidence that proximity to the NEP sites encourages exchange, and so we can use distance as the controlled encouragement factor in order to estimate the effect of exchange on HIV incidence (Frangakis *et al.* 2001, Unpublished data). (Although use of distance with principal stratification has similarities with use of distance as instrumental variable in different settings (e.g. McClellan *et al.*, 1994), such standard methodology of instrumental variables is not applicable when there are either non-constant treatment effects across principal strata or additional post-treatment complications, as with censored outcomes (Frangakis and Rubin, 1999, 2002).)

The exchange behavior of a person at each time, as a function of distance, defines the principal strata. It is plausible to assume that if drug user i does not exchange when the NEP site is at distance z, i would not exchange at longer distances, i.e. $D_{i,t}(z') \leqslant D_{i,t}(z)$ if $z' > z$. Then, i's principal stratum is the distance threshold above which i would not visit the NEP to exchange. Table 1(a) displays these principal strata for six levels of distance. Moreover, here, it is reasonable to assume the exclusion restriction that, for subject i, placing the NEP at distances z versus z' will affect $Y_{i,t}$ (HIV status in that semester) only if z versus z' affects i's $D_{i,t}$ (exchanging at the NEP). For the two principal strata whose exchange is the same for all distances z, there is no causal effect of distance on HIV. For each of the other principal strata, we can summarize the causal effect of distance on HIV attributable to NEP, by the ratio of two fractions: the fraction of people who get HIV if assigned at the closest versus if assigned at the longest distance. Table 1(b) shows the relation of these causal effects, $R_{1,t}, \ldots, R_{5,t}$, to the fractions of incidence of HIV given principal strata and as functions of distance.

When the sample size is large enough, the relative frequency of the principal strata and the causal effects are estimable, as functions of the covariates and time, with no further assumptions. Of course, appropriate parametric assumptions become increasingly helpful with smaller samples. Importantly, because the principal strata are collections of subjects with common propensity to exchange needles, estimation of their relative frequency and of the principal strata-specific causal effects as function of the observed covariates provides valuable information for where the effort of placing NEPs should concentrate in the future. Moreover, the property of principal strata as person-specific characteristics unaffected by the controlled factor (here distance) provides a natural and principled way to address additional complicating factors, such as loss to follow-up, that *are* affected by the controlled factor.

We thank Berrington, Cox, Goetghebeur, and Vansteelandt for their valuable comments, which we believe have reinforced the extra clarity that our approach of principal stratification brings to issues of causal inference.

REFERENCES

COCHRAN, W. G. (1957). Analysis of covariance: its nature and uses. *Biometrics* **13**, 261–281.

COX, D. R. (1958). *Planning of Experiments*. New York: Wiley.

FISHER, R. A. (1925). *Statistical Methods for Research Workers*, 1st edn. Edinburgh: Oliver and Boyd.

FRANGAKIS, C. E. AND RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

MCCLELLAN, M., MCNEIL, B. J. AND NEWHOUSE, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *Journal of the American Medical Association* **272**, 859–866.

NEYMAN, J. (1923). On the application of probability theory to agricultural experiments: essay on principles. Translated in *Statistical Sciences* **5**, 465–480. 1990

ROSENBAUM, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society* A **147**, 656–666.

STRATHDEE, S. A., CELENTANO, D. D., SHAH, N., LYLES, C., STAMBOLIS, V. A., MACALINO, G., NELSON, K. AND VLAHOV, D. Needle-exchange attendance and health care utilization promote entry into detoxification. *Journal of Urban Health* **76**, 448–460.

THE CORONARY DRUG PROJECT RESEARCH GROUP (1980). Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *New England Journal of Medicine* **303**, 1038–1041.

VLAHOV, D., JUNGE, B. BROOKMEYER, R. *et al.* (1997). Reduction in high-risk drug use behaviors among participants in the Baltimore Needle Exchange Program. *Journal of Acquired Immune Deficiency Syndrome* **16**, 400–406.

[*Received 3 December, 2001; revised 10 December, 2001*]