A finite population test of the sharp null hypothesis for Compliers

Wen Wei Loh
Department of Statistics
University of Washington
wloh@u.washington.edu

Thomas S. Richardson
Department of Statistics
University of Washington
thomasr@u.washington.edu

Abstract

In a randomized experiment with non-compliance, testing whether treatment exposure X has an effect on the final response Y is often of scientific interest. We propose a finite-population permutation-based test of the null hypothesis that X has no effect on Y within Compliers. Our method builds on tests for principal stratum direct effects described in (Nolen and Hudgens, 2011).

1 INTRODUCTION

Fisher's test is well-known as a non-parametric test for independence between two binary variables. Within the context of a randomized experiment, it may also be seen as a test of the 'sharp causal null hypothesis' that every individual in a finite population has the same outcome Y regardless of which of two treatments X they receive. In the language of (Neyman, 1923) this may also be expressed by saying that every individual has the same potential outcome under both treatments. Central to the inferential approach used here is the idea that sampling variability *only* arises due to different assignments of individuals to treatment via randomization: the set of individuals in the population and their values are regarded as fixed.

Randomized experiments with 'non-compliance' arise in many situations. For example, in a randomized clinical trial there may be patients who do not take the treatment they are prescribed, possibly due to side-effects. In 'encouragement' studies, in which a randomly selected subset of subjects are given some incentive to avail themselves of a treatment, the inducement may be insufficient for some. In such studies every unit has an *assigned* treatment (Z) that was randomized, a treatment received (X) and a final outcome Y. In many such settings it is also reasonable to make the assumption that Z has no (direct) effect on Y except through X, sometimes called an 'exclusion restriction'.

We will restrict attention to binary treatment and outcomes. We will use $Z = 0$ and $Z = 1$ to indicate assignment to placebo and treatment, respectively; similarly $X = 0$ and $X = 1$ will indicate whether the subject received treatment or placebo. Finally we use $Y = 1$ and $Y = 0$ to indicate good and bad outcomes.

Using the potential outcome framework it has been argued by Angrist et al. (1996) among others, that in studies with non-compliance with a binary treatment, one should attempt to find the effect of treatment on the subset of patients who would take the treatment if asked to do so and would not if assigned not to do so. Since such patients conform with assigned treatment Sommer and Zeger (1991) term this subset of patients the 'compliers'. Imbens and Rosenbaum (2005) use randomization-based inference to obtain valid confidence intervals for the treatment effect under an additive structural model even when the instrument is 'weak'.

In this paper we consider the problem of testing the sharp null hypothesis of no effect *for compliers*. Our basic goal is to apply Fisher's exact test to the subpopulation of compliers. If somehow it were revealed to us which individuals in the population were compliers then we could simply restrict attention to this subset. Under random assignment for the whole population each member of the complier subpopulation should have the same probability to be *assigned* to treatment versus control. Since for compliers $Z = X$, under the exclusion restriction, for the complier subpopulation the null hypothesis that Z has no effect of Y is equivalent to the null hypothesis that X has no effect on Y. Still supposing that we knew which individuals were compliers we could then test this null hypothesis by performing Fisher's exact test on the (X, Y) subtable, or equivalently the (Z, Y) subtable.

However, we face the obvious difficulty that, though we

know that compliers will have $Z = X$, this condition is necessary but not sufficient. For example, an individual with $Z = X = 0$ may be a 'Complier' or they may be someone who would not have taken treatment even if (counter to fact) they had been assigned to it, in other words a 'Never Taker'.

An obvious response to this problem would be to consider all *logically* possible values for the number of compliers in any given (Z, X, Y) stratum that may contain them (i.e. for which $Z = X$), and then to carry out Fisher's test for the subtable showing X and Y for the compliers. Taking the maximum over all of the resulting p-values would then give a valid p-value for the null-hypothesis.

However, this procedure suffers from two related defects:

- it will have no power to reject the null hypothesis, since it is logically possible (though extremely unlikely) that there are no compliers in a given stratum;
- this approach ignores the information provided by strata in which $Z \neq X$, that do not contain compliers.

We will assume that there are no patients who consistently do the opposite of their assignment, i.e. who would take placebo if assigned to treatment, and would take treatment if assigned to placebo, sometimes called 'Defiers' (Chickering and Pearl, 1997). It follows from this assumption that any individuals with $(Z = 1, X = 0)$ are Never Takers. Under random assignment of assigned treatment (Z), the proportion of Never Takers in the $Z = 1$ arm should be approximately the same as in the $Z = 0$ arm. Since Never Takers will have $X = 0$ when assigned to $Z = 0$, this information then narrows the range of probable values (under the randomization distribution) for the number of Compliers in the $(Z = 0, X = 0)$ stratum. Conversely, under the assumption of no Defiers the number of individuals in the $(Z = 0, X = 1)$ stratum will narrow the range of probable values (under the randomization distribution) for the number of compliers in the $(Z = 1, X = 1)$ stratum.

More formally, our approach, following (Nolen and Hudgens, 2011), is to consider only those possible values for the number of Compliers in a given (Z, X) stratum that do not indicate large imbalance between the $Z = 1$ and $Z = 0$ arms, or in other words that do not lead us to reject the randomization null hypothesis at a pre-specified significance level γ. We then carry out Fisher's exact test in the implied (X, Y) table for Compliers. Taking the maximum over these p-values and adding γ then provides a conservative p-value.

2 POTENTIAL OUTCOME FRAMEWORK

We now formalize the foregoing development. Recall the following:

- Z is the randomized treatment assignment, where 1 indicates assignment to drug;
- X is the treatment exposure subsequent to assignment, where 1 indicates drug received;
- Y is the final response, where 1 indicates a desirable outcome, such as survival.

The potential outcome X_{z_i} is the treatment X a patient *would* be exposed to if assigned $z = i$. Using these potential outcomes we may define four generic compliance 'types' t_X listed in Table 1. We denote the set of such types by $\mathbb{D}_X$.

The potential outcomes are linked to the observed outcomes by the consistency axiom (Pearl, 2010), which requires that $Z = z$ implies $X = X_z$.

Table 1: Compliance Types (t_X) Based On Potential Outcomes X_z, (Imbens and Rubin, 1997)

X_{z_0}	X_{z_1}	Compliance Type t_X	
0	0	NT	Never Taker
1	0	DE	Defier
0	1	CO	Complier
1	1	AT	Always Taker

As stated above we will assume that there are no Defiers. To simplify the discussion we will also focus much of our development on the case where there are no Always Takers, so that there are only Compliers and Never Takers. In this circumstance $Z = 0$ implies $X = 0$.

2.1 EXCLUSION RESTRICTION

Without making further assumptions, the potential outcome for a given individual $Y_{x_j z_i}$ is the subjects response Y under exposure to treatment $x = j$, and treatment assignment $z = i$. Without further assumptions there are $16 = 2^{2^2}$ possible sets of values for the variables $(Y_{x_0 z_0}, Y_{x_1 z_0}, Y_{x_0 z_1}, Y_{x_1 z_1})$. However, we will assume that there is no (individual-level) direct effect of Z on Y relative X, so that for $j, i, i' \in \{0, 1\}$, we have:

$$Y_{x_j z_i} = Y_{x_j z_{i'}} \equiv Y_{x_j}. \tag{1}$$

Assumption (1) is guaranteed to hold under double-blind placebo-controlled trials in which the active

treatment is without side-effects and unavailable to patients in the control arm. The response type t_Y then simplifies to just four types, with $\mathbb{D}_Y$ as the set of such types, shown in (Table 2).

The potential outcomes for Y are again linked to the observed outcomes via the consistency axiom, so that if $X = x$ then $Y = Y_x$.

2.2 RANDOMIZATION ASSUMPTION

We make the following assumption:

$$Z \perp\!\!\!\perp \{X_{z_0}, X_{z_1}, Y_{x_0}, Y_{x_1}\} \tag{2}$$

The assumption states that the distribution of compliance and response types (t_X, t_Y) is the same in both the $Z = 1$ and $Z = 0$ arms i.e. that Z is (jointly) independent of the potential outcomes. This will hold whenever treatment assignment Z is physically randomized. Since we are applying Fisher's Exact Test we will consider randomization schemes under which a pre-determined number of units are randomly assigned to $Z = 1$ vs. $Z = 0$, as would be the case if the units assigned to $Z = 1$ were obtained by sampling without replacement from a fixed finite population; the remainder being assigned $Z = 0$.

2.3 THE INSTRUMENTAL VARIABLE MODEL

A causal graph corresponding to the model given by (1) and (2) is shown in Figure 1. The randomization assumption (2) is indicated by the absence of edges directed into Z, while the exclusion restriction (1) corresponds to the absence of a $Z \to Y$ edge. This model is also known as the Instrumental Variable (IV) model (See e.g. Angrist et al., 1996).

2.4 INTENT-TO-TREAT (ITT) ANALYSIS

One possible approach to address the issue of partial compliance is to use Intention-To-Treat (ITT) Analysis, proposed by Lee et al. (1991). The ITT effect measures the causal effect of treatment *assignment*, rather than *exposure*:

$$\text{ITT} \equiv \text{E}[Y_{z_1} - Y_{z_0}], \tag{3}$$

where, in our finite population context, the expectation here is over the possible assignments of units to treatment or placebo. However, inference for this effect is limited to situations where the experimental environment is an accurate representation of the entire population's. Moreover, given that response types and compliance types may be associated (so that exposure to treatment is 'non-ignorable') the difference of outcome averages based on treatment assignment does not provide an unbiased or even consistent estimate of the average causal effect of treatment exposure on outcome (Angrist et al., 1996).

2.5 The Average Causal Effect of X on Y

The average causal effect (ACE) of treatment exposure X on outcome Y is defined as:

$$\text{ACE}(X \to Y) \equiv \text{E}[Y_{x_1} - Y_{x_0}]. \tag{4}$$

The ACE for the sub-population of Compliers is:

$$\text{ACE}_{CO}(X \to Y) \equiv \text{E}[Y_{x_1} - Y_{x_0} \mid t_X = CO]. \tag{5}$$

Since for Compliers $X_z = z$, it follows that $Y_z \equiv Y_{X_z} = Y_{X=z}$ so that

$$\text{ACE}_{CO}(X \to Y) = \text{ITT}_{CO} \equiv \text{E}[Y_{z_1} - Y_{z_0} \mid t_X = CO], \tag{6}$$

or in words, the Average Causal Effect of X on Y for Compliers is equal to the Intent-to-Treat effect of Z on Y for Compliers.

Table 2: Response Types (t_Y) Under Exclusion Restriction (1), (Heckerman and Shachter, 1995)

$Y_{x_0\cdot}$	$Y_{x_1\cdot}$	Response Type t_Y	
0	0	NR	Never Recover
1	0	HU	Hurt
0	1	HE	Helped
1	1	AR	Always Recover

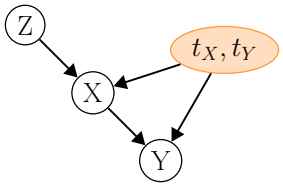

Figure 1: Graphical Representation Of The IV Model Given By Assumptions (1) And (2)

2.6 MOTIVATING EXAMPLE

We consider data from a double-blind placebo-controlled randomized trial of Cholestyramine (Efron and Feldman, 1991). Subjects were randomly assigned to two groups: one was prescribed Cholestyramine ($z = 1$), and the other given a placebo ($z = 0$).

Compliance was a continuous measure tracking the quantity of prescribed dosage consumed, over several years of treatment during the trial. The response was the average post-treatment cholesterol level, and also

a continuous variable. Both continuous measures were dichotomized in Pearl (2009), and the observed counts are shown in Table 3.

Table 3: Cholestyramine/Lipid Data

z	x	y	count	z	x	y	count
0	0	0	158	1	0	0	52
0	0	1	14	1	0	1	12
0	1	0	*0*	1	1	0	23
0	1	1	*0*	1	1	1	78

Subjects who are not assigned treatment in the control arm ($Z = 0$) could not obtain the experimental drug Cholestyramine. Thus there are two structural zeros, since $Z = 0$ implies $X = 0$. In terms of the compliance types, there are thus no Defiers and no Always Takers in the study.

Since this was a double-blind randomized control trial it may be safely assumed that Z has no effect on Y other than through X i.e. that the exclusion restriction (1) holds. Thus in this case, there are four response types t_Y, but only two compliance types t_X, which gives us eight combinations for $(t_X, t_Y) \in \{NT, CO\} \times \{HE, HU, AR, NR\}$. We will consider this simpler case during our main development, though the approach extends to the more general case in which there are also Always Takers.

3 SHARP CAUSAL NULL FOR COMPLIERS

3.1 Notation

We first introduce the notation. Let $n_{z_i x_j y_k}$ be the observable count of the number of subjects in the finite population who are assigned to treatment $z = i$, with exposure $x = j$ and outcome $y = k$. This is directly observable.

Let $n_{z_i}^{t_X, t_Y}$ be the number of subjects in the finite population with compliance type t_X, response type t_Y, and assigned to treatment $z = i$.

Similarly we will let $n_{z_i, y_k}^{t_X}$ be the number of subjects in the finite population with compliance type t_X assigned to treatment $z = i$ with outcome $y = k$.

It should be noted that the counts $n_{z_i}^{t_X, t_Y}$ and $n_{z_i, y_k}^{t_X}$ are not all known since they are not directly observable from the data.

Our interest lies in testing the individual level (or 'sharp') causal null hypothesis that there is no effect of X on Y amongst Compliers; equivalently that for Compliers:

$$Y_{x_0} = Y_{x_1}, \tag{7}$$

so that there are no Compliers who are of type 'Helped' or 'Hurt'. Note that if the individual level causal null hypothesis holds then the average causal effect of X on Y for the subpopulation of compliers (CO): $\text{ACE}_{CO}(X \to Y) = 0$.

In Table 4, the observed counts in Table 3 are decomposed both in General, and under the Sharp Null Hypothesis Equation 7.

3.2 Nuisance parameters

Under the sharp null hypothesis (7), within the Complier sub-population, no individual would have had a different outcome had they had a different exposure level X, and thus, furthermore, no individual would have had a different outcome had they been assigned to a different level of Z. Thus under the null (7), the counts

$$n_{y_0}^{CO} \equiv n_{z_0 y_0}^{CO} + n_{z_1 y_0}^{CO} \qquad n_{y_1}^{CO} \equiv n_{z_0 y_1}^{CO} + n_{z_1 y_1}^{CO}$$

would not change as we vary over all possible assignments of individuals to $Z = 0$ vs. $Z = 1$. Thus if, in addition, the number of Compliers assigned to $Z = 1$ vs. $Z = 0$ were pre-specified in advance then over repeated samplings, the four counts $n_{z_i y_j}^{CO}$ for $i, j \in \{0, 1\}$ would follow a hypergeometric distribution under the null hypothesis.

The counts n_{z_1, y_1}^{CO} and n_{z_1, y_0}^{CO} are observable from the data, as shown in the first two rows of Table 4. However, the presence of Never-Takers in the population prevents us from directly observing the other two counts: n_{z_0, y_1}^{CO} and n_{z_0, y_0}^{CO}.

From the last two rows of Table 4, for $k \in \{0, 1\}$, the observable counts $n_{z_0 x_0 y_k}$ are a sum of the unobservable quantities n_{z_0, y_k}^{CO} and n_{z_0, y_k}^{NT} under H_0.

The unobservable quantities n_{z_0, y_1}^{NT} and n_{z_0, y_0}^{NT} under the placebo arm ($z = 0$) may be regarded as 'nuisance parameters', since if we knew these quantities then we would know n_{z_0, y_1}^{CO} and n_{z_0, y_0}^{CO}, and could perform our hypothesis test. Thus we define the vector of nuisance parameters as:

$$\psi \equiv (\psi_1, \psi_0) = \left(n_{z_0, y_1}^{NT}, n_{z_0, y_0}^{NT}\right) \tag{8}$$

Since ψ_1 and ψ_0 are bounded above by $n_{z_0 x_0 y_1}$ and $n_{z_0 x_0 y_0}$ respectively, the space of possible values for ψ is:

$$\Psi \equiv \left\{ (\psi_1, \psi_0) : \psi_1 \in [0, n_{z_0 x_0 y_1}], \psi_0 \in [0, n_{z_0 x_0 y_0}] \right\} \tag{9}$$

3.3 Testing with nuisance parameters

Given a fixed value of ψ, we may construct Table 5, which we refer to as the 'Target Table' for our hypothesis test, with the fixed values of the four counts $n^{CO}_{z_i,y_k},\ i,k \in \{0,1\}$. Fisher's Exact Test (Fisher, 1973) may then be used to test the null hypothesis (7) *conditional on* a fixed number of Compliers being assigned to $Z=0$ and $Z=1$. Thus for a fixed value of ψ let the p-value resulting from applying Fisher's Exact Test to the Target Table 5, be $p^{CO}(\psi)$.

One possible approach to performing a hypothesis test of (7) would then be to simply compute p-values for the entire space Ψ containing every logically possible value of ψ, and then conservatively take the maximum. However, it is not hard to see that this will lead to a test with no power: for example, consider Table 7(a), which corresponds to $\psi = (\psi_1, \psi_0) = (0, 154)$. We see that the proportions of Compliers with $Y=1$ are very similar in the two Z-strata; $(78/101 = 77\% = 14/18)$. Not surprisingly the p-value from Fisher's Exact Test is 1. Thus if we were to perform a hypothesis test for every logically possible value of the nuisance parameters, clearly we would be unable to reject the null hypothesis.

However, before resigning ourselves to this procedure and concluding that there is thus no evidence against the null hypothesis, we should examine the 'Nuisance Table' shown in Table 7(b). This shows the number of individuals of each of the three types 'Complier', 'Never Taker with $Y_{x_0} = 1$' (hence either 'Always Recover' or 'Hurt') and 'Never Taker with $Y_{x_0} = 0$' (hence either 'Never Recover' or 'Helped'), that would be present in the $Z=0$ and $Z=1$ arms if $(\psi_1, \psi_0) = (0, 154)$. This shows that in order for there to be only 18 Compliers in the $Z=0$ arm, something quite extraordinary would have had to have happened: specifically, we would have quite different proportions of these three types: for example $101/165 = 61\%$ is quite different from $18/172 = 10\%$. Indeed applying Fisher's Exact Test to Table 7(b) conclusively rejects the null hypothesis of independence, giving a p-value that is essentially zero to within machine precision.

We thus propose a method that makes use of the information in the Nuisance Table, such that the set of values for the nuisance parameter is restricted to a subspace $C_\gamma \subseteq \Psi$, under which the null hypothesis of independence in the Nuisance Table is not rejected, at some pre-specified significance level γ.

3.4 Constructing C_γ

While the quantities $\psi = \left(n^{NT}_{z_0,y_1}, n^{NT}_{z_0,y_0}\right)$ under the placebo arm $(z=0)$ may not be observable, the corresponding counts $\left(n^{NT}_{z_1,y_1}, n^{NT}_{z_1,y_0}\right)$ under the treatment arm $(z=1)$ are both directly observable from the data (third and fourth row respectively in Table 4).

Under the randomization distribution, if the number of individuals assigned to $Z=0$ and $Z=1$ is fixed, then the row and column totals in the Nuisance Table should be fixed, and we may thus perform a hypothesis test.

We pre-specify a critical value of γ e.g. 0.01. For every $\psi \in \Psi$, we may then test independence between treatment assignment and compliance types using Fisher's Exact Test in Table 6, which we refer to as the Nuisance Table for ψ. We then obtain $C_\gamma \subseteq \Psi$ as the subspace containing all values of $\psi = (\psi_1, \psi_0)$ such that the test of independence in Table 6 is *not* rejected at significance level γ. Note that the size of C_γ is hence inversely proportional to γ; setting $\gamma = 0$ corresponds to the entire space Ψ. Notice that C_γ does not vary over random assignments, since the row and column sums in the Nuisance Table are fixed; however the nuisance parameter ψ does vary since it depends on how the three types that label columns in the Nuisance Table are distributed between the $Z=0$ and $Z=1$ arms.

For each fixed value of $\psi = (\psi_1, \psi_0) \in C_\gamma$, we then evaluate Fisher's Exact Test in Table 5 to find the p-values $p^{CO}(\psi)$.

Following (Berger and Boos, 1994), the significance level for the overall procedure is then bounded above by taking the maximum and of $p^{CO}(\psi)$ over C_γ and then adding γ:

$$p^{CO}_\gamma \equiv \max\left\{p^{CO}(\psi),\ \psi \in C_\gamma\right\} + \gamma \qquad (10)$$

Following (Berger and Boos, 1994), we may show that p^{CO}_γ gives an upper bound on the probability of falsely rejecting (7) when true, and thus this is a conservative p-value. Under the sharp null hypothesis, (7):

$$\begin{aligned}
&P\left(p^{CO}_\gamma \le \alpha\right) \\
&\quad = P(p^{CO}_\gamma \le \alpha, \psi^\dagger \notin C_\gamma) + P(p^{CO}_\gamma \le \alpha, \psi^\dagger \in C_\gamma) \\
&\quad \le P(\psi^\dagger \notin C_\gamma) + P(\max_{\psi \in C_\gamma} p^{CO}(\psi) + \gamma \le \alpha, \psi^\dagger \in C_\gamma) \\
&\quad \le \gamma + \sum_{\psi'} P(\max_{\psi \in C_\gamma} p^{CO}(\psi) \le \alpha - \gamma \mid \psi^\dagger = \psi', \psi^\dagger \in C_\gamma) \\
&\qquad\qquad \times P(\psi^\dagger = \psi', \psi^\dagger \in C_\gamma) \\
&\quad \le \gamma + \sum_{\psi'} P(p^{CO}(\psi') \le \alpha - \gamma \mid \psi^\dagger = \psi', \psi^\dagger \in C_\gamma) \\
&\qquad\qquad \times P(\psi^\dagger = \psi', \psi^\dagger \in C_\gamma) \\
&\quad \le \gamma + (\alpha - \gamma) \sum_{\psi'} P(\psi' = \psi^\dagger, \psi^\dagger \in C_\gamma) \ \ \le \ \ \alpha,
\end{aligned}$$

where $\psi^\dagger$ is the (unknown) value of ψ resulting from the random assignment of Z. Here the second inequal-

ity follows since C_γ is obtained by applying Fisher's exact test to the Nuisance table at level γ; the third follows because if $\psi^\dagger \in C_\gamma$, the probability of the p-value $p^{CO}(\psi^\dagger)$ being less than $\alpha - \gamma$ is at least as large as the probability of the maximum $p^{CO}(\psi)$ obtained from all $\psi \in C_\gamma$ being less than $\alpha - \gamma$.

4 CHOLESTYRAMINE EXAMPLE

We apply the method to the data in Table 3, with γ as 0.01. Figure 2 shows a contour plot of the p-values from the first step of testing the Nuisance Table (Table 6) over the entire space Ψ. C_γ would hence be the subspace of Ψ with p-values larger than 0.01.

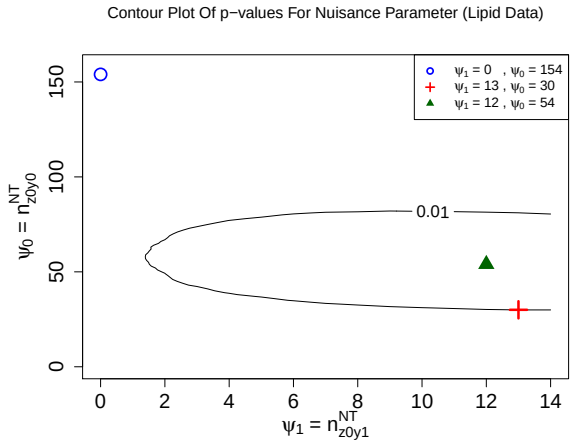

Figure 2: Contour Plot of p-values over Ψ

Given C_γ, we then proceed to test the Target Table (Table 5) for each value of $\psi \in C_\gamma$. Consider the following two examples in Table 7. In Figure 2 the blue empty circle corresponds to a nuisance parameter value of $\psi = (0, 154)$ mentioned earlier. The red filled circle corresponds to a nuisance parameter value of $\psi = (13, 30)$; see Table 7(c), (d). For this value of ψ we see that the nuisance table does not reject at $\gamma = 0.01$. In contrast to Table 7(a), when $\psi = (13, 30)$, we do reject the sharp null hypothesis under the corresponding target table Table 7(c). Thus, as in this example, our test rejects the sharp null at level α when, for every value of the nuisance parameters, either the hypothesis of independence in the Nuisance Table rejects at level γ, or the hypothesis of independence in the target table rejects at level $\alpha - \gamma$ (or both). The green triangle represents the value of ψ at which the maximum p-value is achieved.

The variables of interest $n^{CO}_{z_0y_1}$ and $n^{CO}_{z_0y_0}$ are deterministic functions of the nuisance parameters (the last row in Table 5). The corresponding combinations of $n^{CO}_{z_0y_1}$ and $n^{CO}_{z_0y_0}$ resulting from all values of the nuisance parameter in C_γ are plotted in Figure 3. Note that while the values on the axes in Figure 2 and Figure 3 are identical, the plots are of different variables. We find

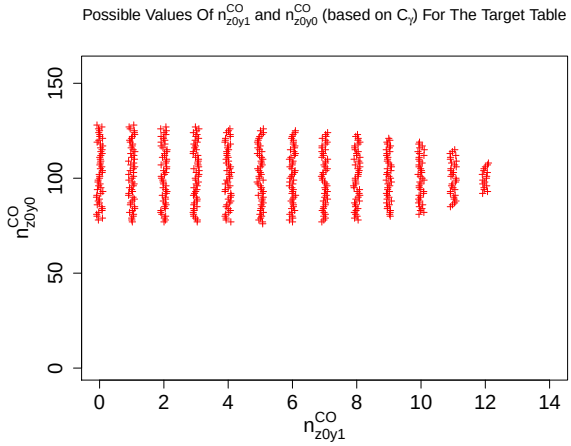

Figure 3: Different combinations of $n^{CO}_{z_0y_1}$ and $n^{CO}_{z_0y_0}$ derived from C_γ; points have been jittered.

$max\left\{p^{CO}(\psi),\ \psi \in C_\gamma\right\} = 1.5 \times 10^{-21}$. From (10), the p-value $p^{CO}_\gamma = 0.01$, which is equal to γ (to machine precision). For comparison, Fisher's Exact Test for independence of Z and Y, which tests the ITT null hypothesis that $Y_{z_0} = Y_{z_1}$ gives a p-value of 2.7×10^{-21}.

5 INCLUDING ALWAYS TAKERS

So far we have assumed that there are no Defiers *and* no Always Takers, as is appropriate when subjects assigned to the control arm ($Z = 0$) are never exposed to treatment ($X = 1$). However, it is possible to extend the method to studies where it is assumed only that monotonicity holds so that $X_{z_0} \leq X_{z_1}$. The compliance and response types corresponding to the assumption that there are no defiers are listed in Table 8. This table augments Table 4 by including the counts for the Always Takers in the relevant locations. the nuisance parameter is now a vector of length 4; the locations of the nuisance parameters are shown in Table 9. In principle we may extend the method by constructing a set C_γ for these parameters, and proceeding as before. However, the resulting procedure is computationally more challenging. For the 'Toy Data' in Table 9 our procedure does reject the sharp null hypothesis for Compliers at level 0.05.

6 SUMMARY

We have presented a method for testing the sharp null hypothesis amongst compliers in the framework of randomization-based inference.

Table 4: Compliance and Response Types under the sharp Null Hypothesis for Compliers, assuming no Defiers and no Always Takers.

z	x	y	General	$H_0 : ITT_{CO} = 0$	Lipid data
1	1	1	$n_{z_1}^{CO,AR} + n_{z_1}^{CO,HE} = n_{z_1 y_1}^{CO}$	$n_{z_1}^{CO,AR} + \quad 0 \quad = n_{z_1 y_1}^{CO}$	78
1	1	0	$n_{z_1}^{CO,NR} + n_{z_1}^{CO,HU} = n_{z_1 y_0}^{CO}$	$n_{z_1}^{CO,NR} + \quad 0 \quad = n_{z_1 y_0}^{CO}$	23
1	0	1	$n_{z_1}^{NT,AR} + n_{z_1}^{NT,HU} = n_{z_1 y_1}^{NT}$	$n_{z_1}^{NT,AR} + n_{z_1}^{NT,HU} = n_{z_1 y_1}^{NT}$	12
1	0	0	$n_{z_1}^{NT,NR} + n_{z_1}^{NT,HE} = n_{z_1 y_0}^{NT}$	$n_{z_1}^{NT,NR} + n_{z_1}^{NT,HE} = n_{z_1 y_0}^{NT}$	52
0	1	1	*0*	*0*	*0*
0	1	0	*0*	*0*	*0*
0	0	1	$\underbrace{n_{z_0}^{NT,AR} + n_{z_0}^{NT,HU}}_{= n_{z_0 y_1}^{NT}} + \underbrace{n_{z_0}^{CO,AR} + n_{z_0}^{CO,HU}}_{= n_{z_0 y_1}^{CO}}$	$\underbrace{n_{z_0}^{NT,AR} + n_{z_0}^{NT,HU}}_{= n_{z_0 y_1}^{NT}} + \underbrace{n_{z_0}^{CO,AR} + \quad 0}_{= n_{z_0 y_1}^{CO}}$	14
0	0	0	$\underbrace{n_{z_0}^{NT,NR} + n_{z_0}^{NT,HE}}_{= n_{z_0 y_0}^{NT}} + \underbrace{n_{z_0}^{CO,NR} + n_{z_0}^{CO,HE}}_{= n_{z_0 y_0}^{CO}}$	$\underbrace{n_{z_0}^{NT,NR} + n_{z_0}^{NT,HE}}_{= n_{z_0 y_0}^{NT}} + \underbrace{n_{z_0}^{CO,NR} + \quad 0}_{= n_{z_0 y_0}^{CO}}$	158

Table 5: Target Table for the sharp null hypothesis for Compliers given a fixed value of nuisance parameters $\psi = (\psi_1, \psi_0)$. Here we assume no Defiers and no Always Takers.

CO only	y_1	y_0	Row Sum
z_1	$n_{z_1 y_1}^{CO} = n_{z_1 x_1 y_1}$	$n_{z_1 y_0}^{CO} = n_{z_1 x_1 y_0}$	$n_{z_1}^{CO} = n_{z_1 x_1}$
z_0	$n_{z_0 y_1}^{CO} = n_{z_0 x_0 y_1} - \psi_1$	$n_{z_0 y_0}^{CO} = n_{z_0 x_0 y_0} - \psi_0$	$n_{z_0}^{CO}$

Table 6: Nuisance Table determining the set C_γ of values for $\psi = (\psi_1, \psi_0)$. Here we assume there are no Defiers and no Always Takers.

	CO	$NT, (AR/HU)$	$NT, (NR/HE)$	Row sum
z_1	$n_{z_1}^{CO} = n_{z_1 x_1} = n_{z_1} - n_{z_1 x_0}$	$n_{z_1 y_1}^{NT} = n_{z_1 x_0 y_1}$	$n_{z_1 y_0}^{NT} = n_{z_1 x_0 y_0}$	n_{z_1}
z_0	$n_{z_0}^{CO} = n_{z_0 x_0} - \psi_1 - \psi_0$	$n_{z_0 y_1}^{NT} = \psi_1$ $\in [0, n_{z_0 x_0 y_1}]$	$n_{z_0 y_0}^{NT} = \psi_0$ $\in [0, n_{z_0 x_0 y_0}]$	n_{z_0}

Table 7: Examples of Target and Nuisance Table for Lipid Data

(a) Target Table For Lipid Data : $\psi = (0, 154)$

CO only	y_1	y_0	Row Sum
z_1	78	23	101
z_0	14 - 0 = 14	158 - 154 = 4	18
Fisher's Exact Test p-value = 1			

(b) Nuisance Table For Lipid Data : $\psi = (0, 154)$

	CO	$NT, (AR/HU)$	$NT, (NR/HE)$	Row sum
z_1	101	12	52	165
z_0	172 - 0 - 154 = 18	0	154	172
Fisher's Exact Test p-value = $1.9 \times 10^{-29} \approx 0$				

(c) Target Table for Lipid Data : $\psi = (13, 30)$

CO only	y_1	y_0	Row Sum
z_1	78	23	101
z_0	14 - 13 = 1	158 - 30 = 128	129
Fisher's Exact Test p-value = $4.1 \times 10^{-39} \approx 0$			

(d) Nuisance Table for Lipid Data : $(\psi_1, \psi_0) = (13, 30)$

	CO	$NT, (AR/HU)$	$NT, (NR/HE)$	Row sum
z_1	101	12	52	165
z_0	172 - 13 - 30 = 129	13	30	172
Fisher's Exact Test p-value = $0.0103 > 0.01$				

Table 8: Compliance and Response Types under $H_0 : ITT_{CO} = 0$, assuming No Defiers.

z	x	y	General	$H_0 : ITT_{CO} = 0$	Toy data
1	1	1	$\underbrace{n_{z_1}^{AT,AR} + n_{z_1}^{AT,HE}}_{= n_{z_1 y_1}^{AT}} + \underbrace{n_{z_1}^{CO,AR} + n_{z_1}^{CO,HE}}_{= n_{z_1 y_1}^{CO}}$	$\underbrace{n_{z_1}^{AT,AR} + n_{z_1}^{AT,HE}}_{= n_{z_1 y_1}^{AT}} + \underbrace{n_{z_1}^{CO,AR} + 0}_{= n_{z_1 y_1}^{CO}}$	15
1	1	0	$\underbrace{n_{z_1}^{AT,NR} + n_{z_1}^{AT,HU}}_{= n_{z_1 y_0}^{AT}} + \underbrace{n_{z_1}^{CO,NR} + n_{z_1}^{CO,HU}}_{= n_{z_1 y_0}^{CO}}$	$\underbrace{n_{z_1}^{AT,NR} + n_{z_1}^{AT,HU}}_{= n_{z_1 y_0}^{AT}} + \underbrace{n_{z_1}^{CO,NR} + 0}_{= n_{z_1 y_0}^{CO}}$	5
1	0	1	$n_{z_1}^{NT,AR} + n_{z_1}^{NT,HU} = n_{z_1 y_1}^{NT}$	$n_{z_1}^{NT,AR} + n_{z_1}^{NT,HU} = n_{z_1 y_1}^{NT}$	15
1	0	0	$n_{z_1}^{NT,NR} + n_{z_1}^{NT,HE} = n_{z_1 y_0}^{NT}$	$n_{z_1}^{NT,NR} + n_{z_1}^{NT,HE} = n_{z_1 y_0}^{NT}$	5
0	1	1	$n_{z_0}^{AT,AR} + n_{z_0}^{AT,HE} = n_{z_0 y_1}^{AT}$	$n_{z_0}^{AT,AR} + n_{z_0}^{AT,HE} = n_{z_0 y_1}^{AT}$	5
0	1	0	$n_{z_0}^{AT,NR} + n_{z_0}^{AT,HU} = n_{z_0 y_0}^{AT}$	$n_{z_0}^{AT,NR} + n_{z_0}^{AT,HU} = n_{z_0 y_0}^{AT}$	15
0	0	1	$\underbrace{n_{z_0}^{NT,AR} + n_{z_0}^{NT,HU}}_{= n_{z_0 y_1}^{NT}} + \underbrace{n_{z_0}^{CO,AR} + n_{z_0}^{CO,HU}}_{= n_{z_0 y_1}^{CO}}$	$\underbrace{n_{z_0}^{NT,AR} + n_{z_0}^{NT,HU}}_{= n_{z_0 y_1}^{NT}} + \underbrace{n_{z_0}^{CO,AR} + 0}_{= n_{z_0 y_1}^{CO}}$	5
0	0	0	$\underbrace{n_{z_0}^{NT,NR} + n_{z_0}^{NT,HE}}_{= n_{z_0 y_0}^{NT}} + \underbrace{n_{z_0}^{CO,NR} + n_{z_0}^{CO,HE}}_{= n_{z_0 y_0}^{CO}}$	$\underbrace{n_{z_0}^{NT,NR} + n_{z_0}^{NT,HE}}_{= n_{z_0 y_0}^{NT}} + \underbrace{n_{z_0}^{CO,NR} + 0}_{= n_{z_0 y_0}^{CO}}$	15

Table 9: Nuisance Table determining the set C_γ of values for $\psi = (\psi_{11}, \psi_{10}, \psi_{01}, \psi_{00})$, assuming No Defiers.

	CO	$NT, (AR/HU)$	$NT, (NR/HE)$	$AT, (AR/HE)$	$AT, (NR/HU)$	Row sum
z_1	$n^{CO}_{z_1} = n_{z_1x_1} - \psi_{11} - \psi_{10}$	$n^{NT}_{z_1y_1} = n_{z_1x_0y_1}$	$n^{NT}_{z_1y_0} = n_{z_1x_0y_0}$	$n^{AT}_{z_1y_1} = \psi_{11}$ $\in [0, n_{z_1x_1y_1}]$	$n^{AT}_{z_1y_0} = \psi_{10}$ $\in [0, n_{z_1x_1y_0}]$	n_{z_1}
z_0	$n^{CO}_{z_0} = n_{z_0x_0} - \psi_{01} - \psi_{00}$	$n^{NT}_{z_0y_1} = \psi_{01}$ $\in [0, n_{z_0x_0y_1}]$	$n^{NT}_{z_0y_0} = \psi_{00}$ $\in [0, n_{z_0x_0y_0}]$	$n^{AT}_{z_0y_1} = n_{z_0x_1y_1}$	$n^{AT}_{z_0y_0} = n_{z_0x_1y_0}$	n_{z_0}

Table 10: Target Table for the sharp null hypothesis for Compliers given a fixed value of nuisance parameters $\psi = (\psi_{11}, \psi_{10}, \psi_{01}, \psi_{00})$, assuming No Defiers.

CO only	y_1	y_0	Row Sum
z_1	$n^{CO}_{z_1y_1} = n_{z_1x_1y_1} - \psi_{11}$	$n^{CO}_{z_1y_0} = n_{z_1x_1y_0} - \psi_{10}$	$n^{CO}_{z_1} = n_{z_1x_1} - \psi_{11} - \psi_{10}$
z_0	$n^{CO}_{z_0y_1} = n_{z_0x_0y_1} - \psi_{01}$	$n^{CO}_{z_0y_0} = n_{z_0x_0y_0} - \psi_{00}$	$n^{CO}_{z_0} = n_{z_0x_0} - \psi_{01} - \psi_{00}$

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455.

Berger, R. and D. Boos (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association 89*(427), 1012–1016.

Chickering, D. M. and J. Pearl (1997). A clinician's tool for analyzing non-compliance. *Computing Science and Statistics 29*(2), 424–431.

Efron, B. and D. Feldman (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association 86*(413), 9–17.

Fisher, R. A. (1973). *Statistical methods and scientific inference*. New York: Hafner Press.

Heckerman, D. and R. Shachter (1995). Decision-theoretic foundations for causal reasoning. *Journal of Artificial Intelligence Research 3*, 405–430.

Imbens, G. W. and P. R. Rosenbaum (2005). Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 168*(1), 109–126.

Imbens, G. W. and D. B. Rubin (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics 25*(1), 305–327.

Lee, Y. J., J. H. Ellenberg, D. G. Hirtz, and K. B. Nelson (1991). Analysis of clinical trials by treatment actually received: is it really an option? *Statistics in Medicine 10*(10), 1595–1605.

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych X*, 1–51. In Polish, English translation by D. Dabrowska and T. Speed in *Statistical Science 5*(4), 463–472, 1990.

Nolen, T. L. and M. G. Hudgens (2011). Randomization-Based Inference within Principal Strata. *Journal of the American Statistical Association 106*(494), 581–593.

Pearl, J. (2009). *Causality: models, reasoning, and inference* (2nd ed. ed.). Cambridge, UK: Cambridge University Press.

Pearl, J. (2010). On the Consistency Rule in Causal Inference: Axiom, Definition, Assumption, or Theorem? *Epidemiology 21*(6), 872.

Sommer, A. and S. L. Zeger (1991). On estimating efficacy from clinical trials. *Statistics in Medicine 10*(1), 45–52.