



Published in final edited form as:

Psychiatr Ann. 2008 December ; 38(12): 772–783. doi:10.3928/00485713-20081201-10.

Intent-to-Treat vs. Non-Intent-to-Treat Analyses under Treatment Non-Adherence in Mental Health Randomized Trials

Thomas R. Ten Have¹, Sharon-Lise T. Normand², Sue M. Marcus³, C. Hendricks Brown⁴, Philip Lavori [Professor (Biostatistics) and Chair]⁵, and Naihua Duan [Professor of Biostatistics (in Psychiatry)]⁶

¹Department of Biostatistics and Epidemiology, University of Pennsylvania, 607 Blockley Hall, 423 Guardian Dr., Philadelphia, PA 19104.

²Department of Health Care Policy, Harvard Medical School; Department of Biostatistics, Harvard School of Public Health, 180 Longwood Avenue Boston, MA 02115.

³Division of Biostatistics, Division of Health Services, Department of Psychiatry, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1230, New York, NY 10029

⁴Prevention Science and Methodology Group, Departments of Epidemiology and Biostatistics, College of Public Health MDC-56, University of South Florida, 13201 Bruce B Downs Blvd Tampa, FL 33612

⁵Department of Health Research and Policy, Stanford University School of Medicine. HRP/ Redwood Bldg. 259 Campus Drive, Stanford, CA 94305-5405

⁶Departments of Biostatistics and Psychiatry, Columbia University; Director, Division of Biostatistics, New York State Psychiatric Institute, 1051 Riverside Drive, Unit 48 New York, NY 10032

Introduction

The intent-to-treat (ITT) principle has long been mandated by the Food and Drug Administration (FDA) as the primary design and analysis strategy for industry clinical trials and also has been adopted widely in government-funded randomized clinical trials (1–4). Intent-to-treat analysis aims to estimate the effect of treatment *as offered*, or *as assigned*. This analysis entails comparisons of randomized groups and include outcome data for all randomized participants regardless of their status regarding non-adherence to assigned treatment protocols and missed assessment encounters. Petkova and Teresi (5) attributed the term “intent-to-treat” to Hill (6) with a common refrain “once randomized, always analyzed.” FDA regulations emphasize this point in more formal language: “The intention-to-treat principle implies that the primary analysis should include all randomized subjects. Compliance with this principle would necessitate complete follow-up of all randomized subjects for study outcomes.” (4).

While the ITT principle has been the dominant design and analysis paradigm for clinical trials in a variety of contexts, other approaches, which we refer to as “Non-ITT analyses,” aim to estimate the effect of treatment *as delivered* or *as received* (as opposed to “*as assigned*” under the ITT approach) to account for treatment non-adherence. These Non- ITT analyses are commonly presented as secondary analyses in terms of as-treated or per-

protocol treatment effects along with ITT results (7–9). Indeed, the FDA allows for such supplementary results: “Under many circumstances, it (use of the full analysis set) may also provide estimates of treatment effects that are more likely to mirror those observed in subsequent practice.” (4). Such sentiments have been voiced not only about data analysis, but also the need to collect adherence data as outcomes in addition to clinical outcomes (10–12,7).

It is important to recognize that the ITT and Non-ITT strategies differ not only in terms of the estimation procedure, but also in terms of the underlying research goal. Given the distinction between the effect of treatment “*as assigned*” corresponding to the ITT approach and the effect of treatment “*as received*” addressed by the Non-ITT strategies, the investigator needs to choose carefully which treatment effect is the primary research goal for a specific study. The as-received treatment effect of the Non-ITT approaches attempts to measure the effect of the experimental treatment relative to the control condition when all patients adhere to the assigned treatment condition. Such an effect is usually the primary research goal for the development of new treatments. In contrast, the as-assigned treatment effect of the ITT analysis is usually more pertinent for the evaluation of the effectiveness of the treatment in terms of the public health benefits of administering the treatment in the community in light of inevitable treatment non-adherence. A treatment with a high as-received treatment effect might not yield a high as-assigned treatment effect if the adherence rate is low when the treatment is offered. Such a distinction has implications for the relationships among data-based estimates of these effects for specific studies, which are addressed below in the sections on the ITT and non-ITT strategies. Related distinctions of ITT and non-ITT treatment effects are made in terms of treatment efficacy versus effectiveness (13). In the ensuing discussion, we refer to the treatment effects of Non-ITT analyses as “as-received treatment effects” and the effects of ITT analyses as “as-assigned treatment effects.”

In addition to the above distinction between the goals of the different analytic techniques, other distinctions need to be considered in terms of the two types of deviations from the ideal study procedures: 1) missed assessment encounters due to either intermittent missed encounters or drop-out from the study; and 2) non-adherence to the randomly assigned treatment protocol. The ITT principle necessitates that all planned data collection occur for each patient regardless of her treatment adherence status (1). There are several advantages to collecting outcome data even when a patient has stopped taking the treatment. First, it facilitates the use of ITT analyses to estimate the as-assigned treatment effect. Second, it facilitates the use of causal inference methodologies to assess as received treatment effects in the presence of treatment non-adherence. These methodologies address confounding factors, both measured and unmeasured, that might affect both the adherence status and the outcome. Addressing unmeasured confounding factors is especially challenging, and usually requires strong assumptions. We discuss in this paper how these alternate assumptions can be assessed with the help of randomization and good predictors of adherence and testable modeling assumptions (14). In a clinical trial context, treatment non-adherence may take several forms depending on the type and timing of non-adherence and the treatment arm. Whatever form it takes, treatment non-adherence should be clearly defined by study investigators before the start of the study. For patients assigned to the experimental arm, different types of treatment non-adherence include declining to take the assigned, taking an alternative treatment such as the comparison treatment or a non-study treatment rather than the assigned treatment, deviating from study protocol by taking the assigned treatment but not according to the study protocol (e.g., less than the number of prescribed pills), and finally dropping out of the study completely thus ending treatment and the collection of trial outcomes. For patients assigned to the control arm, the nature of non-adherence also depends on the nature of the control condition, namely, whether an active comparison

treatment, a placebo “treatment”, or no action is specified for these patients. If the control condition specifies a comparison treatment or a placebo, non-adherence might entail not taking the assigned treatment, taking an alternative treatment, or taking the assigned treatment but not according to the study protocol. If the control condition specifies no treatment, non-adherence might entail taking the experimental treatment, or taking an alternative treatment. The form of non-adherence might need to be taken into consideration in as-treated (AT) analyses. For example, control patients who received the experimental treatment contrary to the assignment might need to be analyzed in the experimental arm.

In terms of timing, treatment non-adherence as defined by the study investigators may occur intermittently or continue until the end of study follow-up. In any case, it is important that the schedule for outcome data collection continue regardless of the type or timing of the treatment non-adherence. Finally, in one of the examples studied in this paper, treatment adherence was not defined with respect to the experimental treatment but instead with respect to physicians following guidelines for treating depressed patients. Here adherence was measured in both the treatment and control groups.

Treatment non-adherence in its different forms may be significantly more common for randomized trials in psychiatry compared to other areas of medicine that are more acutely associated with mortality. Ten Have et al. (12) reviewed a number of trials involving treatments of depression for which treatment adherence rates are found to be as low as 30%. Schulberg et al. (15) reported on low treatment adherence for an efficacy study of guideline-level treatment of depression in primary care. Despite intensive efforts to maintain high levels of patient-level adherence to the study treatments, nortriptyline or interpersonal psychotherapy, only approximately 30% of the intervention patients completed a full course of therapy. This is consistent with high rates of discontinuation of treatment with antidepressants in routine care (16) or open-label studies (17). A variety of factors have been identified as influencing adherence in behavioral trials, including psychiatric-related personal difficulties interfering with adherence and the widespread off-label use of psychiatric medication. Corrigan and Salzer (18) indicated that these factors impact treatment preferences, which in turn influence adherence to treatments in psychiatric trials even among patients who consent to participating in them. Fogg and Gross (19) contrasted similar problems with non-adherence rates in prevention studies, where interventions seem to be less imperative due to the absence of disease, to surgery trials, for which interventions are strictly controlled and in which participants may be highly motivated to adhere in light of severe conditions related to mortality. We show in this paper how standard approaches to accommodating non-adherence perform differently for behavioral outcomes and interventions than for non-behavioral medical interventions and outcomes. Hence, it is imperative that clinical trial investigators in behavioral contexts elevate the attention paid to adherence to assess its impact on outcome in clinical trials (7,10).

In the subsequent sections, we discuss several different types of analytic approaches to estimating treatment effects on outcome under the randomized clinical trial framework. The first to be addressed is the analytic approach under the ITT principle. Next, we discuss various Non-ITT analyses, including the as-treated, per-protocol, and instrumental variable analyses. Finally, we present four examples in the mental health literature that highlight the differences among these approaches. The formulae for calculating the different treatment effects under these approaches are presented in Table 1. These formulae are presented for two types of designs: 1) the control group does not have access to the experimental treatment and so cannot be measured for adherence, which is the case for three of the four example studies presented below; and 2) adherence is measured in the control group. An example of this second case occurs when adherence is defined with respect to patients’

physicians following treatment guidelines as in the fourth example below, where the intervention focused on increasing such adherence.

Intent-to-Treat Analyses

The ITT analysis aims to test and estimate the as-assigned treatment effect in the study sample. The validity of this analysis for the study sample follows from the protection against unmeasured confounding by the randomization of treatment assignment without the need to adjust for non-adherence. The resulting inference for the study sample can inform policy about the effectiveness of implementing the intervention at the population level if treatment non-adherence pattern is similar between the study sample and the target population. However, such generalizability might not hold for a target population with a different non-adherence pattern from the study sample. For example, ITT-based inference based on a sample of physically healthy patients highly adherent to their psychiatric care may not be reflective of a population receiving multiple treatments for medical comorbidities who may be less likely to adhere to psychiatric treatments.

In distinguishing between patterns of adherence at the study and population levels, a number of authors emphasize that treatment adherence in a randomized trial may be influenced by factors other than personal characteristics such as those relating to study design issues (20). For instance, patient or clinician non-adherence may occur by design due to extended enrollment periods and shortened study follow-up periods. Moreover, intensive efforts to maintain high levels of adherence to randomized treatments in some randomized trials make it difficult to extrapolate ITT estimates of as-assigned treatment effects to the community of practitioners without similar resources to sustain adherence among their patients (15). In these cases, the ITT estimates may not reflect accurately the results of implementing the corresponding interventions in practice.

Furthermore, the ITT approach does not necessarily provide a valid test and estimate of the as-received treatment effect (5,7,11), especially when treatment non-adherence rate is high. Hence, in the presence of treatment non-adherence, the common assertion that the ITT approach under-estimates the true treatment effect only applies if the goal is evaluating the as-received treatment effect but not necessarily when the focus is on the as-assigned treatment effect, as discussed above. In contrast, the Non-ITT methods discussed next in the context of estimating the as-received treatment effect may be biased for both the as-received and as-assigned treatment effects.

Non-ITT Analyses

A number of Non-ITT approaches that aim to estimate the as-received treatment effect through adjustments for non-adherence have been used in the medical literature in general and mental health research literature in particular. These methods are vulnerable to selection bias due to confounders, both measured and unmeasured, that might affect both the adherence status and outcome. Such selection bias may be classified into two categories, overt and hidden bias (21). Overt bias is attributable to observed confounders, and therefore can be explicitly adjusted for with statistical methods such as covariate adjustment or propensity scores analysis (e.g., Marcus paper in this volume). Such adjustments are made with the Non-ITT approaches. In contrast, hidden bias arises from unobserved confounders, and therefore cannot be explained entirely by covariate or propensity score adjustments of the Non-ITT approaches. Nonetheless, we consider below the instrumental variable approach as one Non-ITT method that attempts to account for hidden bias under several assumptions.

A common Non-ITT approach that adjusts for overt bias in attempting to estimate the as-received treatment effect is the AT analysis, which involves comparisons of groups defined by treatment adherence status (22). The AT approach has taken a number of different forms depending on how non-adherers are handled. The form of the AT analyses also depends on the study design in terms of whether the comparison group has access to the experimental treatment and is measured for adherence to the active treatment. Alternatively in the case of encouragement interventions targeted towards improving adherence to a proven treatment, adherence may be measured in both the treatment and comparison arms with respect to the delivery of the proven treatment. In these cases where the patients in the comparison group are measured for adherence, the AT analysis may contrast the adherers in the experimental treatment and comparison groups versus a non-treatment group that combines the non-adherers in both arms. Alternatively, in the cases where the participants in the comparison group are not measured for adherence, the AT analysis may compare the adherers in the experimental treatment arm to a non-treatment group combining the non-adherers in the experimental treatment arm with all participants in the comparison group.

Alternatively, the per-protocol (PP) analysis, which also adjusts for overt bias, focuses on the effect of adhering to the assigned treatment protocol. When the comparison group is measured for adherence, a PP analysis may compare adherers in the experimental treatment group with the adherers in the comparison arm. The exclusion of non-adherers under the PP approach distinguishes it from the AT method. In the case where the comparison group is not measured for adherence, the PP analysis may contrast the adherers in the experimental treatment arm with all participants in the comparison group, excluding the non-adherers in the experimental arm.

Under perfect treatment adherence, the as-assigned and as-received treatment effects are identical, and so the ITT, AT, and PP approaches yield identical estimates of both treatment effects. However, under treatment non-adherence, the as-assigned and as-received treatment effects are different, and so the ITT, AT, and PP approaches yield different results. None of these methods adequately tests and estimates the as-received treatment effect. Nevertheless, because of theoretical relationships among these individual effects, comparing them in terms of their corresponding estimates based on data may provide interesting insights on the relationships among treatment adherence, confounders, and outcome for specific studies. Accordingly, the AT and PP estimates are expected to exceed the ITT estimate if the inclusion of treatment non-adherers in the randomized to treatment group dilutes the treatment effect. However, the AT and PP approaches are not protected by randomization and thus are vulnerable to hidden bias. The above expected relationships among the ITT, AT, and PP estimates of treatment effect have been observed under various studies in clinical areas outside of psychiatry, including the Women's Health Initiative randomized trial of hormone replacement therapy (23), a randomized study of vitamin A on mortality (24), and a medication-based intervention on cholesterol in African Americans (12). However, these relationships do not seem to hold for the mental health studies we consider below, suggesting that the hidden bias that might affect the AT and PP approaches may differ between psychiatry and other areas of medicine.

Instrumental Variable Analyses

There are a number of different causal inference approaches that adjust for unmeasured confounders (i.e., hidden bias) when estimating the as-received treatment effect (7,14,25,26). They vary by estimation techniques, but have been shown to equal each other under certain assumptions. While these alternative assumptions allow for the relaxation of the assumptions about no hidden bias made by the AT and PP approaches, they require close examination either with testing based on observed data or with discussions of clinical

plausibility. This data-based testing of certain assumptions of the IV approach demands baseline predictors of adherence to treatment, as well as baseline covariates that modify the difference in adherence rates between the randomized to treatment and control groups (i.e., interaction between baseline covariates and randomized group assignment with adherence as the dependent variable). Before addressing these testable assumptions with an example, we present the most commonly used causal approach to estimating the as-received treatment effect, which is the instrumental variable (IV) method (9,16,27).

In addition to recent applications to randomized trials, the IV approach has been used to control for unmeasured confounding in observational studies (25). Instrumental variables are assumed to emulate randomization variables, unrelated to unmeasured confounders influencing the outcome. In the case of randomized trials, the same randomized treatment assignment variable used in defining treatment groups in the ITT analysis is instead used as the instrumental variable in IV analyses. In particular, the instrumental variable is used to obtain for each patient a predicted probability of receiving the experimental treatment. Under the assumptions of the IV approach, these predicted probabilities of receipt of treatment are unrelated to unmeasured confounders in contrast to the vulnerability of the actually observed receipt of treatment to hidden bias. Therefore, these predicted treatment probabilities replace the observed receipt of treatment or treatment adherence in the AT model to yield an estimate of the as-received treatment effect protected against hidden bias when all of the IV assumptions hold. (22,27). When these IV assumptions do not hold, the IV approach is vulnerable to hidden bias. However, several researchers have shown that this hidden bias of the IV estimate may be not very significant when there is relatively good adherence above 70% in the randomized trial context (14,28). Furthermore, Marcus and Gibbons (27) presented an approach that enhances the IV method in terms of protection against overt bias by adjusting for all observed confounders with a propensity score technique.

As proposed previously, theoretical relationships among the IV, ITT, and AT effects, may reveal informative relationships among their data-based estimates for a specific study. Accordingly, Little et al. (22) showed that when the AT estimate exceeds the ITT estimate, the IV estimate is typically in between those of the AT and ITT estimates. While the IV estimate tends to exceed the ITT estimate, the IV standard error also tends to exceed the ITT standard error, leading to similar p-values and inference under the two approaches (22). Horvitz-Lennon et al. (7) showed that an IV estimator can be more precise when both non-adherence and missing data are present, although there is much literature showing that the IV approach can lead to larger standard errors than the other approaches considered here especially under low to moderate adherence rates, and such increases in variability make it vulnerable to violations of the IV assumptions. Nonetheless, the IV approach is used to assess the magnitude of the as-received treatment effect in contrast to the ITT approach, which again focuses on the as-assigned treatment effect. The use of two models under the IV approach, one relating treatment received to outcome (the AT model), and the other relating randomized intervention assignment to treatment received, has led many researchers to refer to the IV method as a two-stage estimation procedure.

While the IV approach may not require the no hidden bias assumption of the AT and PP approaches, there are several tradeoffs with the IV approach involving increased variability and potential sources of bias due to factors other than unmeasured confounders not considered by the above mental health applications. Tradeoffs include increased variability of the IV treatment effect estimates and violations of assumptions that lead to bias. As described below, the increased variability of the IV estimator of as-received treatment effect in the presence of treatment non-adherence relative to the ITT, AT, and PP estimators leads to increased vulnerability of the IV approach to violations of its assumptions described

below. This increase in variability can sometimes be mitigated by including covariates, if available, that are strong predictors of treatment received and outcome in the models for outcome and treatment received. The use of predictive covariates in the model relating the randomized intervention to treatment received increases the precision of the predicted treatment probability that replaces observed treatment received in the IV model for the outcome. Including predictors of outcome in the model for outcome of course reduces the residual error at least for linear models.

One of the key assumptions for the IV approach in protecting it against hidden bias is the exclusion restriction assumption, which requires in the randomized trial context that the impact of treatment assignment is mediated entirely through the delivery or receipt of treatment such that there is no direct effect of treatment assignment independent of treatment delivered. That is, randomized assignment to the intervention does not impact the outcome through other paths other than as-received treatment. Accordingly, patients who did not receive the experimental treatment will respond similarly regardless of whether they were assigned to the experimental arm or the control arm. Likewise, patients who received the experimental treatment will also respond similarly irrespective of the arm to which they were assigned.

The exclusion restriction assumption may especially be vulnerable in unblinded studies, which arise in a number of different contexts. One such context occurs when assignment to the experimental treatment arm may enhance a participant's expectation of success, and in contrast, assignment to the control arm might dampen such an expectation. Another context involves studies that include a health professional, such as a behavioral care manager as part of the experimental intervention to enhance the delivery of the specific treatment under study. In this case, the care manager might impact patient outcomes through means other than the specific treatment, say, a pill delivered with a smile from the care manager might taste different than a pill delivered without the personal touch in the control arm, even when both patients take exactly the same pill (14,22,27,29). However, traditional medication trials, even when thoroughly blinded, might not be free from violations of this assumption. For example, among patients considered not to have received the experimental treatment, those assigned to the intervention arm might have received a larger partial dose (although considered inadequate), while those assigned to the control arm might have received a smaller partial dose, or no dose at all. To the extent that partial dose might lead to some benefits, the non-recipients in the experimental arm might have better outcomes than the non-recipients in the control arm, and consequently, treatment received might not explain the entire impact of treatment assignment.

The absence of such alternative paths is required for the standard IV procedure, but can nonetheless be assessed with IV extensions discussed subsequently in this section. These evaluations of the paths of randomized interventions apart from adherence to treatment require interactions between randomized treatment and baseline covariates on adherence to form additional instrumental variables beyond the randomized intervention (14,30).

Additional assumptions made by most causal inference approaches require that the treatment assignment of one participant does not influence the outcomes of other participants and that variations in the administration of the treatment (experimental or control) do not influence the outcome. These two assumptions are known as the "Stable Unit Treatment Assumption" (SUTVA). The assumption that the treatment assignment of one individual affects the outcomes of other individuals is not the same as the standard independence assumption made by all single endpoint, single level analyses that the outcome of one subject does not affect the outcomes of other participants. Both of the SUTVA assumptions need to be carefully considered as to whether they can be ruled out for a specific study. There are

numerous ways these assumptions might not hold, especially for behavioral interventions delivered in staff and/or patient group settings. However, medication trials are not necessarily free from these risks. For multilevel studies that randomize patients within groups (within clinics or classrooms), patients' receipt of treatment might influence other's outcomes by sharing their treatment experience, sharing their germs (for studies of contagious diseases), or even sharing their medications (which is not uncommon, e.g., in AIDS/HIV treatment programs). Furthermore in the case of provider-based administration of behavioral interventions, the resulting contextual factors that are likely to vary in the administration of the treatment and be shared by multiple patients might influence the outcomes, the care manager's smile mentioned earlier being a plausible example. Table 2 presents the example contexts of these assumptions. More research is needed on how sensitive causal inference under the IV approach is to violations of SUTVA in these contexts.

Additional assumptions are needed for interpreting the IV estimate of the as-received effect in the general population. Two alternative assumptions are often made in this case. The first alternative assumption requires that the as-received treatment effect is the same across all patients in a population (29,31). Such a no-treatment interaction assumption may not be feasible in the presence of the evidence that treatment effects often depend on personal characteristics and prognostic factors, thus leading to the push for personalized treatments (32–34). Given the likely implausibility of treatment homogeneity, an alternative assumption, known as the monotonicity assumption, has been offered, but this assumption limits causal inference to the sub-group of treatment adherent patients rather than all patients. Monotonicity is an assumption about treatment non-adherers who are not the target of inference, but nonetheless is necessary in helping to estimate the as-received treatment effect among adherent patients. More specifically, monotonicity requires that for every patient who chooses not to take the experimental treatment when randomized to it, s/he will not try to obtain the experimental treatment if randomized to the comparison group. Hence, there is a monotonic ordering of the behavior of patients with respect to potential receipt of the treatment as one moves from assignment to the comparison group to assignment to the experimental treatment group. As with the exclusion restriction, assessments of the monotonicity assumption and treatment heterogeneity are possible with the use of relationships among baseline covariates, the randomized intervention, and adherence status (14). However, these relationships are required to be strong and the corresponding models need to be specified as accurately as possible to reduce variability of the resulting IV estimates of treatment effect. Nonetheless, the possibility of data-based assessments of some of the alternative assumptions of the IV approach should be contrasted with the impossibility of assessing the no confounding of the AT and PP approaches, although the IV and PP approaches are estimating different effects of receiving treatment and abiding by the protocol in each arm, respectively.

Example Studies

We consider four example studies in the mental health literature where at least the AT and PP approaches are presented as follow-up analyses to the ITT results, occasionally accompanied with IV estimates. The comparisons of the formulae for these estimates are presented in Table 1, which may help facilitate the comparison of the corresponding example estimates for each of the studies below in Table 3. As noted at the end of the introduction, the formulae in Table 1 are presented for two types of designs, depending on whether the control group is measured for adherence: 1) adherence to the experimental treatment is not measured in the controls; or 2) controls are measured for adherence such as in the case of physician adherence to treatment guidelines. The first three example studies correspond to the first design, and the fourth example study falls under the second design.

As a first mental health example of how the AT and PP approaches relate to the ITT and IV results differently than in medical examples, Lewis et al. (8) used a randomized trial to investigate if clozapine was more effective than the other second-generation antipsychotic (SGA) drugs in treating partially resistant schizophrenia patients (Example 1). The sample consisted of 136 randomized participants aged 18–65 with DSM-IV based diagnoses of schizophrenia and a poor response to previous antipsychotic drugs. Participants were randomly allocated to clozapine or to one of the class of other SGA drugs (risperidone, olanzapine, quetiapine, amisulpride) as prescribed by the patients' respective clinicians. Adherence in both arms was quite low, as only 54% of those randomized to clozapine and 57% of the participants in the SGA arm had been taking their assigned medications at the end of one year. Because of this low adherence, the resulting standard errors for the IV estimates are at least twice as large as those for the other as-received treatment effect estimates. Focusing on one of the primary outcomes, Positive and Negative Syndrome Scale (PANSS) at 12 weeks, the estimated ITT effect of clozapine is -5.98 [SE 2.24; $p=0.004$; 95% CI= $(-10.37, -1.59)$]. The analogous AT estimate, comparing those who were taking clozapine to those who were not (non-adherers in the clozapine arm plus all participants in the SGA arm) is -1.16 [SE 2.37; $p=0.31$; 95% CI= $(-5.80, 3.49)$]. Comparing outcomes in only those patients who were adhering to their randomly assigned medications, the PP estimate of the clozapine effect on the PANSS is -5.19 [SE 2.76; $p=0.03$; 95% CI= $(-10.60, 0.22)$]. Finally, the IV estimate of the effect of receiving clozapine as opposed to the other SGAs is -13.81 [SE 5.99; $p=0.01$; 95% CI= $(-25.55, -2.07)$]. Comparing the estimates under the different approaches reveals that the AT and PP estimates are not in between the ITT and IV estimates, unlike what is expected (18). Nonetheless, the relationships among these estimates do suggest that there is a difference between those who adhere and those who do not. The most compelling difference is between the AT and PP estimates of the clozapine effect on PANSS (-1.16 vs. -5.19). Whereas the definition of the clozapine group is the same for these two estimates, the SGA comparison group is defined differently. For the AT comparison, non-adherers in the clozapine group are moved to the SGA comparison group, while for the PP comparison, non-adherers in the SGA group are removed from this comparison group. Hence, the dramatic difference between the AT and PP estimates is due to the addition of non-adherers in both randomized groups to the adherers in the SGA comparison group under the AT approach. It is apparent that non-adherers in both arms had better outcomes than did the adherers to other SGAs. Any differences between non-adherers and adherers due to unmeasured confounders will be adjusted for with the IV approach under the additional IV assumptions. The IV approach agrees more with the PP approach as both apply to the adherers, but the IV adjusts for hidden bias factors that influence adherence in contrast to the PP approach. It may be that by attenuating the treatment difference among adherers, indicating a form of negative confounding, unmeasured confounders caused the IV estimates to exceed in magnitude the corresponding PP and AT estimates. That is, the direction of the relationships between the unmeasured confounders with outcomes and adherence may be opposite from the direction of the treatment outcome relationship (i.e., positive vs. negative associations). As expected, the ITT estimate and standard error are attenuated with respect to the corresponding IV estimates, and as a result the p -values and resulting inference are similar between the two approaches.

A more extensive comparison of the ITT, AT, and IV approaches was performed by Marcus and Gibbons (27) for the analysis of a SNAP-IV impulsivity outcome at 14 months of follow-up in a randomized trial of the Multi-Modal Treatment for ADHD (MTA; Example 2). The analysis focused on a two-group comparison of the medication-only ($N=120$) versus behavior-only treatment ($N=122$) components of MTA. The non-adherence occurred in the behavioral treatment group with 33% of the group violating their randomly assigned protocol by switching to medication treatment. In contrast, all participants randomized to the

medication-only arm adhered to the protocol. Hence, the investigators treated non-adherence as participants in the behavioral arm switching to medication. The resulting AT effect size for the as-received treatment effect, comparing participants receiving only the behavioral intervention compared to those taking medication with or without the behavioral intervention is -0.20 vs. an ITT effect size of -0.27 . As with the clozapine study, the AT estimated effect size is uncharacteristically smaller than the ITT effect size. Furthermore, adjusting for potentially unmeasured confounders, the IV effect size (-0.37 with a standard deviation of 0.7) exceeds both the ITT and AT effect sizes. These unexpected relationships among the ITT, AT, and IV estimates may be due to the presence of two groups in the combined medication and behavioral component groups, as identified by Marcus and Gibbons (27) using past history of medication. For those with such a past history, the expected relationships among the ITT, AT, and IV occurred, whereas these relationships were not observed for younger patients with little history of medication use. The authors suggested that this result may have been due to a violation of the exclusion restriction made by the IV approach. Specifically, the older participants with a history of medication use may have resorted to medication because of the lack of efficacy of the behavioral component, resulting in an effect of the randomized behavioral intervention through an alternative path other than adhering to the behavioral intervention.

Horvitz-Lennon et al. (7) presented a mental health example (Example 3) where the relationships among the ITT, AT, and IV estimates conformed more to the expected relationships seen in medical randomized studies (22). The randomized comparison was between clozapine as the experimental treatment ($N=218$) and haloperidol as the control treatment ($N=205$) with respect to a 12 month PANSS score for hospitalized veterans with refractory schizophrenia. A significantly high proportion of patients in the clozapine arm (82%) adhered by taking clozapine with the rest non-adhering by taking haloperidol. Patients randomized to haloperidol did not have access to clozapine. The resulting ITT, AT, and IV estimates were 5.04 , 6.48 , and 5.92 . While this order is what would be expected and despite the substantially worse outcomes for non-compliers randomized to clozapine, the differences in estimates are small relative to their standard errors and thus yield similar inference.

Finally, we consider Example 4 for which the IV assumptions have been evaluated to establish its validity in comparison to the unverifiable no-unmeasured-confounding assumption of the standard approaches to adjusting for non-adherence. The assumptions are reviewed in Table 2 with respect to this last example as well as to the Marcus and Gibbons (27) and Horvitz-Lennon et al. (7) studies, Examples 2 and 3, respectively. The fourth example now under consideration differs from these two other examples in that adherence is defined differently so that it applies to both the experimental treatment and control arms rather than just to the experimental treatment arm. Specifically, the Example 4 study involved randomization of primary care practices to either usual care (UC) for depressive symptoms of their patients or a physician-level telephone encouragement (TE) intervention aimed at improving primary care physician (PCP) adherence to Agency for Healthcare and Research Quality (AHRQ)-based guidelines for treating depressive symptoms. The primary outcome, patient depressive symptoms, was measured by the Centers for Epidemiologic Studies Depression Scale (CESD). The study sample consisted of 71 clinically depressed patients who were referred to the study during a three-month period by 28 randomized PCPs in 19 primary care practices of an academic health system. The randomized encouragement intervention employed telephone communication by a behavioral health nurse with each TE patient and her or his respective PCP. The goal was to encourage PCP adherence to best practice guidelines for making treatment decisions about their patients' depression. The study investigators did not randomize individual patients, because the investigators believed that PCPs within a given practice would not be able to limit improved guideline adherence

to their TE patients and then treat their UC patients in their usual manner. The study investigators evaluated all study patients at baseline and three months of follow-up for the CESD outcome variable. The rates of binary PCP adherence to treatment guidelines, as determined by disease management chart abstractions by a study psychiatrist, were 76.5% and 40.5% for the randomized TE and UC arms, respectively. As intended, the TE intervention appeared to motivate the PCPs to adhere more to treatment guidelines. Because “treatment received” applies to the act of the physician adhering to the treatment guidelines, the control group does not represent a no-treatment received group, and so the AT method is not applicable. Therefore, only the PP result is presented here in addition to the ITT and IV results.

The ITT and PP estimates of treatment effect adjusting for baseline CESD and age (given differences in age between the randomized practices) are -3.20 [SE 1.84; $p=0.09$; 95% CI= $(-6.81, 0.41)$] and -8.64 [SE 1.71; $p<.0001$; 95% CI= $(-11.99, -5.29)$], respectively. The large magnitude of the PP estimate suggests that the patients of the PCPs who adhered to treatment guidelines in the intervention group did better than the patients treated by the adherent PCPs in the usual care group. The corresponding IV estimate of -10.0 [SE 4.16; $p=0.008$; 95% CI= $(-17.80, -1.50)$] is somewhat in agreement with the PP approach, indicating that there may not be unmeasured confounders influencing PCP adherence to the treatment guidelines.

However, such conjecture assumes that the assignment of practices to the encouragement intervention did not have an effect through a path other than PCP adherence to guidelines. For example, assignment to the encouragement intervention may have increased staff sensitivity to treating depression as well as improved patient treatment behavior, regardless of whether the PCPs in the practice actually followed guidelines. Ten Have et al. (14) assessed this assumption by extending the IV approach to estimate the effect of an alternative path as -2.40 [SE 8.55; $p=0.39$; 95% CI= $(-19.10, 14.40)$], which was not significant and also very small relative to even the ITT effect. The increased variability of this estimate relative to the variability of the ITT and PP estimates reflects the difficulty of estimating such an alternative path of the randomized intervention. The additional assumption that the treatment effect does not differ across sub-groups in the population was also evaluated. Ten Have et al. (14) showed that such treatment heterogeneity did exist, but that the corresponding treatment effect in the latent class of guideline PCP adherers in both randomization arms was still similar to those of the PP and IV estimates. Such treatment heterogeneity also showed that the monotonicity assumption did not hold, in that there was evidence of the presence of PCP's who would only follow guidelines if assigned to the usual care group (i.e., defiers with respect to the encouragement intervention). Hence, with such a sensitivity analysis, it appears that the PP and IV analyses indeed yielded robust estimates of the effect of PCP adherence to AHRQ guidelines on patient-level depression.

Discussion

Understanding treatment non-adherence and its relation to outcomes is critically important in psychiatric studies. One way of achieving this is by comparing the different analytical approaches to adjusting for non-adherence. These comparisons point to the uniqueness of mental health studies relative to more medically oriented studies. Most psychiatric randomized studies report the standard ITT results, which apply to the as-assigned treatment effects or the effectiveness of interventions in populations with similar patterns of non-adherence as those of the study samples. As-treated and per-protocol results are sometimes reported for as-received treatment effects, in attempts to adjust for treatment non-adherence in randomized studies. Recognizing the vulnerability of the AT and PP approaches to unmeasured confounding, causal approaches such as the IV method try to control for such

confounding, requiring other assumptions that are more testable than the no-confounding assumption. The testing of these alternative assumptions benefits from strong predictors of adherence, which may not be the case in mental health studies. Furthermore, adherence may not be measured well resulting in weaker results from causal approaches. Hence, as recommended in Table 3, it is imperative that adherence and its potential predictors are measured accurately, which entails better information on reasons for non-adherence and more attention to non-adherence and its predictors in the data collection stages of studies. With such prospective attention to non-adherence in psychiatric studies, a careful analysis including causal methods can help identify clinically meaningful relationships between treatment adherers and non-adherers with respect to the effect of treatment on outcome.

References

1. Lavori PW. Clinical trials in psychiatry: should protocol deviation censor patient data? *Neuropsychopharmacology* 1992;6:39–48. [PubMed: 1571068]
2. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomized controlled trials. *British Medical Journal* 1999;319:670–674. [PubMed: 10480822]
3. Lewis JA, Machin D. Intention-to-treat: who should use ITT? *British Journal of Cancer* 1993;68:647–650. [PubMed: 8398686]
4. Food and Drug Administration. International conference on harmonization: Guidelines on general considerations for clinical trials. *Federal Register* 1997;62(242):66113–66119.
5. Petkova E, Teresi J. Some Statistical Issues in the Analyses of Data from Longitudinal Studies of Elderly Chronic Care Populations. *Psychosomatic Medicine* 2002;64:531–547. [PubMed: 12021427]
6. Hill, AB. Principles of medical statistics. 7th ed.. London: The Lancet; 1961.
7. Horvitz-Lennon M, O'Malley AJ, Frank RG, Normand S-LT. Improving traditional intention-to-treat analyses: A new approach. *Psychological Medicine* 2005;35:961–970. [PubMed: 16045063]
8. Lewis SW, Barnes TR, Davies LM, Murray RM, Dunn G, Hayhurst KP, Markwick A, Lloyd H, Jones PB. Randomized Controlled Trial of Effect of Prescription of Clozapine Versus Other Second-Generation Antipsychotic Drugs in Resistant Schizophrenia. *Schizophrenia Bulletin* 2006;32:715–723. [PubMed: 16540702]
9. Oslin DW, Lynch KG, Pettinati HM, Kampman KM, Gariti P, Gelfand L, Ten Have T, Wortman S, Dundon W, Dackis C, Volpicelli JR, O'Brien CB. A Placebo-Controlled Randomized Clinical Trial of Naltrexone in the Context of Different Levels of Psychosocial Intervention. *Alcoholism: Clinical and Experimental Research* 2008;32:1299–1308.
10. Prinz RJ, Smith EP, Dumas JE, Laughlin JE, White DW, Barron R. Recruitment and retention of participants in prevention trials involving family-based interventions. *American Journal of Preventive Medicine* 2001;20:31–37. [PubMed: 11146258]
11. Gross D, Fogg L. A critical analysis of the intent to treat principle in prevention analysis. *The Journal of Primary Prevention* 2004;25:475–489.
12. Ten Have TR, Coyne JC, Salzer M, Katz IR. Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial. *General Hospital Psychiatry* 2003;25:115–123. [PubMed: 12676425]
13. Bellamy S, Lin J, Ten Have T. An introduction to causal modeling in clinical trials. *Clinical Trials: Journal of the Society for Clinical Trials* 2007;4:58–73.
14. Ten Have TR, Elliott M, Joffe M, Zanutto E, Datto C. Causal models for randomized physician encouragement trials in treating primary care depression. *Journal of the American Statistical Association* 2004;99:8–16.
15. Schulberg HC, Block MR, Madonia MJ, Scott CP, Rodriguez E, Imber SD, Perel J, Lave J, Houck PR, Coulehan JL. Treating major depression in primary care practice. Eight-month clinical outcomes. *Arch Gen Psychiatry* 1996;53:913–919. [PubMed: 8857868]
16. Katon W, Vonkorff M, Lin E, Bush T, Ormel J. Adequacy and duration of antidepressant treatment in primary care. *Medical Care* 1992;30:67–76. [PubMed: 1729588]

17. Stevens T, Katona C, Manela M, Watkin V, Livingston G. Drug treatment of older people with affective disorders in the community. *Int'l J Ger Psychiatry* 1999;6:467–472.
18. Corrigan PW, Salzer MS. The conflict between random assignment and treatment preference: implications for internal validity. *Evaluation and Program Planning* 2003;26:109–121.
19. Fogg L, Gross D. Threats to validity in randomized clinical trials. *Research in Nursing and Health* 2000;23:79–87. [PubMed: 10686575]
20. Stein REK, Bauman LJ, Ireys HT. Who enrolls in prevention trials? Discordance in perception of risk by professional and participants. *American Journal of Community Psychology* 1991;19:603–617. [PubMed: 1755438]
21. Rosenbaum, PR. *Observational Studies*. 2nd Edition. New York: Springer Verlag; 2002.
22. Little R, Long Q, Lin X. A Comparison of Methods for Estimating the Causal Effect of a Treatment in Randomized Clinical Trials Subject to Noncompliance. *Biometrics*. in press.
23. Heiss G, Wallace R, Anderson G, Aragaki A, Beresford S, Brzyski R, Chlebowski R, Gass M, LaCroix A, Manson J, Prentice RL, Rossouw J, Stefanick M. WHI Investigators. Health Risks and Benefits 3 Years After Stopping Randomized Treatment With Estrogen and Progestin. *JAMA* 2008;299:1036–1045. [PubMed: 18319414]
24. Sommer A, Zeger S. On estimating efficacy from clinical trials. *Statistics in Medicine* 1991;10:45–52. [PubMed: 2006355]
25. Bao Y, Duan N, Fox SA. Is some provider advice on smoking cessation better than no advice? An instrumental variable analysis of the 2001 National Health Interview Survey. *Health Services Research* 2006;41:2114–2135. [PubMed: 17116112]
26. O'Malley AJ, Normand S-LT. Likelihood methods for treatment-noncompliance and subsequent nonresponse in randomized trials. *Biometrics* 2005;61:325–334. [PubMed: 16011678]
27. Marcus S, Gibbons R. Estimating the Efficacy of Receiving Treatment in Randomized Clinical Trials with Noncompliance. *Journal Health Services and Outcomes Research Methodology* 2001;2:247–258.
28. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instrumental variable and endogenous explanatory variable is weak. *Journal of the American Statistical Association* 1995;90:433–450.
29. Brookhart AM, Schneeweiss S. Preference-Based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results. *The International Journal of Biostatistics* 2007;3 Article 14.
30. Dunn G, Bentall R. Modeling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine* 2007;26:4719–4745. [PubMed: 17476649]
31. Hernan MA, Robins JM. Instruments for Causal Inference An Epidemiologist's Dream? *Epidemiology* 2006;17:360–372. [PubMed: 16755261]
32. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and Moderators of Treatment Effects in Randomized Clinical. *Archives of General Psychiatry* 2002;59:877–883. [PubMed: 12365874]
33. Genomics and Personalized Medicine Act of 2006. 109th Cong., 2nd sess. (3 August 2006). S. 3822
34. Overview of NIMH Priorities
http://www.nimh.nih.gov/about/strategic-planningreports/index.shtml#priorities_overview
 “Conducting clinical trials that will provide treatment options to deliver more effective personalized care across diverse populations and settings.”

Table 1

Formulae for calculating without baseline covariates the ITT, AT, PP, and IV treatment effect estimates under two different clinical trial designs: 1) when adherence to the experimental treatment is not measured in the control group; and 2) when adherence is measured in the control group, e.g., adherence to treatment guidelines. We first define the character symbols used for the formulae.

Design	Estimated Treatment Effect			
	Intention to Treat	As Treated	Per Protocol	Instrumental Variable
Adherence not measured in controls	$A1 - A0 = (A11 + A10) - A0$	$A11 - (A10 + A0)$	$A11 - A0$	$ITT/P11 = (A1 - A0)/P11$
Adherence measured in controls	$A1 - A0 = (A11 + A10) - (A01 + A00)$	$(A11 + A01) - (A10 + A00)$	$A11 - A00$	$ITT/(P11 - P01) = (A1 - A0)/(P11 - P01)$

A1 = observed average for randomized to treatment group

A0 = observed average for randomized to control group

A11 = observed average for participants who receive treatment in the randomized to treatment group

A10 = observed average for participants who do not receive treatment in the randomized to treatment group

A00 = observed average for participants who do not receive the treatment in the randomized to control group

A01 = observed average for participants who do receive the treatment in the randomized to control group (e.g., take the medication in the usual care group)

P11 = Proportion receiving treatment in the randomized to treatment group

P01 = Proportion receiving treatment in the randomized to control group

Table 2

Summary of Assumptions for the IV treatment estimates for three of the four mental health studies (Marcus and Gibbons (25); Horvitz-Lennon et al. (10); 4 = Ten Have et al. (11)) discussed in the Example Section.

Assumptions	Application to Studies
Independence Assumption for the Instrumental Variable	Randomization seemed to be effective in balancing observed covariates in all three studies, except for age in the last study, so independence is supported in the first two studies, and less so in the third study.
Relationship between IV and treatment received.	In the first two studies above, when the experimental treatment was assigned, adherence ranged between 67 and 87% with none of the control group having access to the experimental treatment. In the third study, 75% of the telephone encouragement group had medication prescribed whereas only 41% in the usual care group had medication prescribed
Exclusion Restriction	Only in the third study was there empirical evidence that randomized assignment to the experimental telephone encouragement intervention did not have an alternative path to the outcome apart from through adherence to guidelines.
Monotonicity	Monotonicity for the first two studies was satisfied by definition with the control groups not having access to the respective experimental treatments. For the third study, an empirical assessment showed that the monotonicity assumption may have not held with evidence for physicians who would defy their assigned treatment by following treatment guidelines only if in the usual care group.
SUTVA	The first two studies would appear to satisfy this assumption that patient's treatment assignment does not influence the outcomes of other patients as patients wouldn't necessarily share information. In the third study, where PCPs were randomized, it may be that those in the same practices may have shared information about their assignments and therefore influenced the outcomes of the patients of the other PCPS. Another aspect of this assumption is that variations of the administration of the experimental treatments delivery of medications in different ways (pick-up versus drop-off at home) may have influenced the outcome. This may have been possible in the first two studies, which involved medication as the assigned treatment

Table 3

Summary of ITT, AT, PP, and IV treatment estimates (as applicable) for the four mental health studies discussed in the Example Section.

Study	Treatments		Outcome (1) – (2)	Estimated Treatment Effect (Stand. Error)			
	(1)	(2)		Intention to Treat	As Treated	Per Protocol	Instrumental Variable
1	Clozapine	SGA	PANSS	-5.98 (2.24)	-1.16 (2.37)	-5.19 (2.76)	*
2	Behavioral	Medicine	Impulsivity	-0.27 (*)	-0.20 (*)	*	-0.37 (*)
3	Clozapine	Halperiodol	PANNS	5.04 (*)	6.48 (*)	*	5.92 (*)
4	Telephone Encouragement	Usual Care	CESD	-3.20 (1.84)	*	-8.64 (1.71)	-10.0 (4.16)

1 = Lewis et al. (19); 2 = Marcus and Gibbons (25); 3=Horvitz-Lennon et al. (10); 4 = Ten Have et al. (11).

* = not available or applicable

Table 4

Key Recommendations

- Treat information on treatment non-adherence and missing data as separate, collecting as much information as possible on both.
- Assess the plausibility of the assumptions made with method used. For example, is there reason to believe there is treatment heterogeneity.