

STATISTICS IN MEDICINE

Per-Protocol Analyses of Pragmatic Trials

Miguel A. Hernán, M.D., Dr.P.H., and James M. Robins, M.D.

Pragmatic trials are designed to address real-world questions about options for care and thereby guide decisions by patients, clinicians, and other stakeholders. Pragmatic trials are often analyzed according to the intention-to-treat principle, which requires that patients assigned to a treatment strategy are kept in that group during the analysis, even if they deviated from their assigned treatment strategy after randomization.¹⁻³ The result of an intention-to-treat analysis is affected by the trial-specific pattern of adherence to the treatment strategies under study and therefore may not be directly relevant for guiding decisions in clinical settings with different adherence patterns. In fact, the publication and dissemination of the intention-to-treat result may change adherence in clinical settings, possibly rendering the result outdated.

Herein, we address issues that may arise from uncritical reliance on the intention-to-treat principle in pragmatic trials,^{1,4,5} outline alternative analytic approaches (known as per-protocol analyses), and provide guidance on how to choose among them. Specifically, we argue that the appropriate design, conduct, and analysis of pragmatic trials require an explicit and careful definition of the effect of interest, an a priori specification of the statistical analysis plan, and the collection of high-quality longitudinal data.

 DEMYSTIFYING INTENTION-TO-TREAT EFFECTS

An intention-to-treat analysis estimates the effect of treatment assignment in a particular trial, not the effect of treatment itself. Intention-to-treat effects are agnostic about postrandomization decisions, including treatment discontinuation and the use of concomitant therapies prohibited by the study protocol. For example, consider two pragmatic trials of a new active treatment versus standard of care. In the first trial, half the patients assigned to the active treatment actually received

it and the other half did not. In the second trial, all the patients assigned to the active treatment received it. In neither study did any patient assigned to standard of care receive active treatment. An intention-to-treat analysis may show a treatment effect in the first trial but not in the second. This could occur even if the biologic effect of active treatment were identical in the two studies. Furthermore, in a head-to-head trial of two active treatments that have differential adherence because of a mild, easily palliated side effect, an intention-to-treat analysis may misleadingly indicate a beneficial effect of the less efficacious treatment.

When there is incomplete adherence, intention-to-treat analyses may make unsafe interventions appear to be safe, may lead to a declaration of noninferiority even when one treatment is clearly superior, and may result in an effective intervention appearing to be ineffective if the poor adherence was due to misplaced concerns about effectiveness or toxicity. As a result, intention-to-treat effects may not provide clinically useful information in some situations. For example, to decide whether to use a certain contraception method, a couple would want to know the expected effectiveness when the method is used as indicated and not the estimated effectiveness in a population in which, for example, 40% of couples do not use the method properly.

Health care professionals and patients would like to have an effect measure that, unlike the intention-to-treat effect, is not influenced by the degree of adherence. In our example above, a couple would have an interest in knowing the effect of the contraception method if all trial participants had used it as indicated in the protocol. This effect, which is referred to as the per-protocol effect, is what would have been observed if all patients had adhered to the trial protocol.

It is sometimes argued that the intention-to-treat effect is the effect of interest in pragmatic

trials, because it reflects the effectiveness of treatment in clinical settings outside the trial. However, that argument fails to recognize that the degree of adherence outside the trial may change drastically after patients learn of the trial's findings. Therefore, the actual effectiveness of the treatment in the community may differ greatly from the intention-to-treat effect estimate from the trial.⁴ In contrast, the per-protocol effect from the trial is not affected by incomplete adherence and may therefore be of more interest for patients who are considering whether to use the treatment.

IMPROVING PER-PROTOCOL ANALYSES

The validity of both intention-to-treat and per-protocol effect estimates requires correct adjustment for selection bias due to differential loss to follow-up.⁶ Moreover, the validity of per-protocol effect estimates also requires correct adjustment for confounding due to incomplete adherence to the assigned treatments or use of off-protocol concomitant therapies. Because both adherence and loss to follow-up may be influenced by social and clinical factors that occur after randomization, we refer to the biases that arise because of lack of adjustment for these factors as postrandomization confounding and selection bias.

To adjust for incomplete adherence when estimating per-protocol effects, investigators often perform an analysis in which a participant's follow-up is terminated at the time, if any, at which they cease to follow the protocol; that is, data from participants are censored when the participants deviate from their assigned treatment strategy. To increase the probability that this form of per-protocol analysis yields a valid estimate of the per-protocol effect, three general rules need to be followed.

First, data from participants should not be censored when they stop treatment for clinical reasons. For example, in a trial of statin therapy and cardiovascular disease, no sensible protocol will mandate that treatment be continued when rhabdomyolysis or other substantial adverse effects occur. The treatment strategy under study in this trial is not "take statin therapy continuously during the follow-up," but rather "take statin therapy continuously during the follow-up, unless otherwise clinically indicated." In this example, trial participants who stop statin therapy after having a toxic effect are not deviating from the protocol

and their data should not be censored in a per-protocol analysis.⁷

Second, data from participants should be censored when it is no longer certain that they are receiving treatment. For example, in a trial of statin therapy, a participant's follow-up should end shortly after the participant's status with respect to statin use first becomes unknown, because without recent treatment data, it is not possible to determine whether the participant continues to adhere to the protocol. The rule applies even if the outcome in the participant is later learned through other means (e.g., by searching the National Death Index).

Third, adjustment should be made for confounding due to incomplete adherence. A naive per-protocol analysis, that is, one with no adjustment for confounding, will be valid only if adherence occurred completely at random.⁴ However, because participants who adhere to the protocol and those who do not adhere generally differ with respect to prognostic factors, a per-protocol analysis that censors patient data at the time of nonadherence must adjust for prerandomization and postrandomization prognostic factors that predict adherence. It follows that data from participants generally need to be censored shortly after the stream of data on prognostic factors that predict adherence is first interrupted, even if the treatment and outcome data from the participants are still available after that time.

There is, of course, no guarantee that adjustment for the measured prognostic factors will remove all bias from the per-protocol effect estimate. In this sense, a per-protocol analysis of a randomized trial is an observational analysis that has all the usual challenges associated with potential confounding by unmeasured factors. Historically, this has cast doubt on the validity of per-protocol analyses.⁸

Much of the trepidation about per-protocol analyses can be tracked to a landmark article from 1980; a post hoc analysis of the Coronary Drug Project (a National Heart, Lung, and Blood Institute–sponsored randomized, controlled trial that was performed between 1966 and 1975 to assess the efficacy and safety of five lipid-influencing drugs in patients who had previously had a myocardial infarction) showed lower mortality among participants who adhered to placebo than among those who did not, even after adjustment for multiple prerandomization prognostic factors.⁹ The

difference in 5-year risk of death between the participants who adhered to placebo and those who did not was approximately 10 percentage points. Because statistical adjustment could not eliminate the difference in risk of death, this finding was interpreted as evidence for “the serious difficulty, if not impossibility, of evaluating treatment efficacy in subgroups determined by [adherence] to the treatment protocol after randomization.”⁹ Indeed, this finding is frequently used as a warning for investigators who deviate from the intention-to-treat principle and attempt to perform even a nonnaive per-protocol analysis, that is, one with adjustment for confounding.

The analysis of the Coronary Drug Project, however, could benefit from statistical innovations that took place in the subsequent decades. A recent reanalysis of these data¹⁰ succeeded at largely eliminating the difference in risk of death in the placebo group by using new analytic methods that improve the adjustment for, and definition of, adherence (Fig. 1).^{7,11} The difference in

5-year mortality between the participants who adhered to placebo and those who did not was 2.5 percentage points. Specifically, the analysis was designed to be analogous to the original analysis except that it used logistic regression (rather than linear regression) for the binary mortality outcome, decoupled the definition of adherence and loss to follow-up, used standardization to adjust for prerandomization covariates, and used inverse-probability weighting to adjust for postrandomization covariates.

The use of statistical methods that were unavailable at the time of the original analysis was an important component of the reanalysis of the Coronary Drug Project. The validity of these methods is easily verifiable in the reanalysis because we expect that the outcome should not be affected by adherence to placebo. This example shows how recent additions to our toolkit offer hope for improved estimation of per-protocol effects. In particular, adjustment for postrandomization factors may prove to be critical (as described in greater detail below).

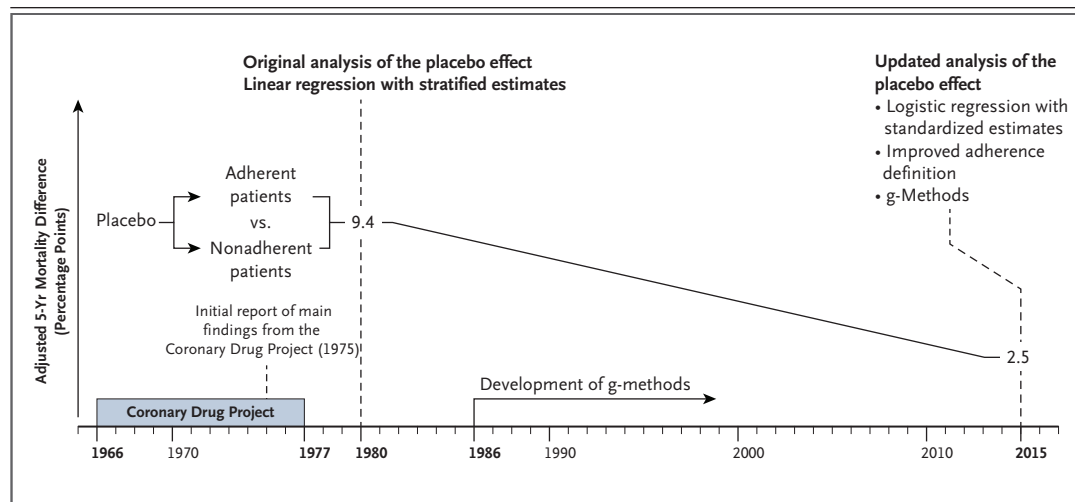


Figure 1. Adjusted Difference in 5-Year Mortality between Patients Who Adhered to Placebo and Those Who Did Not in the Coronary Drug Project Randomized Trial.

An intention-to-treat analysis of the Coronary Drug Project randomized trial (performed between 1966 and 1975) showed no difference in 5-year mortality between patients who received the active treatment (clofibrate) and those who received placebo. In 1980, a post hoc analysis showed that, in the placebo group, mortality among participants who adhered to placebo throughout the study was lower than mortality among those who did not adhere, even after adjustment for multiple baseline factors. Because placebo cannot affect mortality, the difference in mortality in the placebo group was interpreted as evidence that comparisons between patients who adhere to a regimen and those who do not are inherently biased. An updated analysis performed in 2015 used logistic regression, decoupled the definitions of adherence and loss to follow-up, used standardization to adjust for prerandomization covariates, and used g-methods to adjust for postrandomization covariates.¹⁰ In the updated analysis, the difference in 5-year mortality between the patients who adhered to placebo and those who did not was estimated to decrease from 9.4 percentage points in the 1980 analysis to 2.5 percentage points in the 2015 analysis.

ESTIMATING PER-PROTOCOL
EFFECTS FOR SUSTAINED TREATMENT
STRATEGIES

Pragmatic trials that study the long-term effects of sustained strategies in typical patients and health care settings are ideal to guide clinical decision making. However, as discussed above, they are also more vulnerable to postrandomization confounding and selection bias. Generally, the estimation of per-protocol effects for sustained treatment strategies require statistical adjustment for postrandomization prognostic factors. When these factors are not affected by prior treatment, conventional statistical methods may be appropriate. Unfortunately, conventional statistical methods cannot generally handle postrandomization prognostic factors that are affected by prior treatment.^{12,13} Appropriate adjustment for postrandomization factors affected by prior treatment requires so-called g-methods (see below), that is, methods that can adjust for confounding even when in the presence of a feedback loop between

treatment and confounders over time. g-Methods include inverse-probability weighting,¹⁴ g-estimation of structural nested models,¹⁵ the g-formula,¹⁶ and their extensions that are based on doubly robust estimation (Fig. 2).^{17,18}

g-Methods can be used to estimate per-protocol effects for sustained treatment strategies (Fig. 2), provided that data on the prerandomization and postrandomization prognostic factors that predict adherence are available. Thus, the protocols of trials need to plan for the detailed collection of data from baseline until the administrative end of the study for all participants. The Coronary Drug Project investigators refrained from adjusting for postrandomization factors in their 1980 analysis because, at that time, no g-methods were available. A number of such g-methods have been implemented with the use of standard statistical software packages; we now describe some applications to specific trials.

Inverse-probability weighting works by estimating a sequential propensity score that is then used to reweight participants over time according to

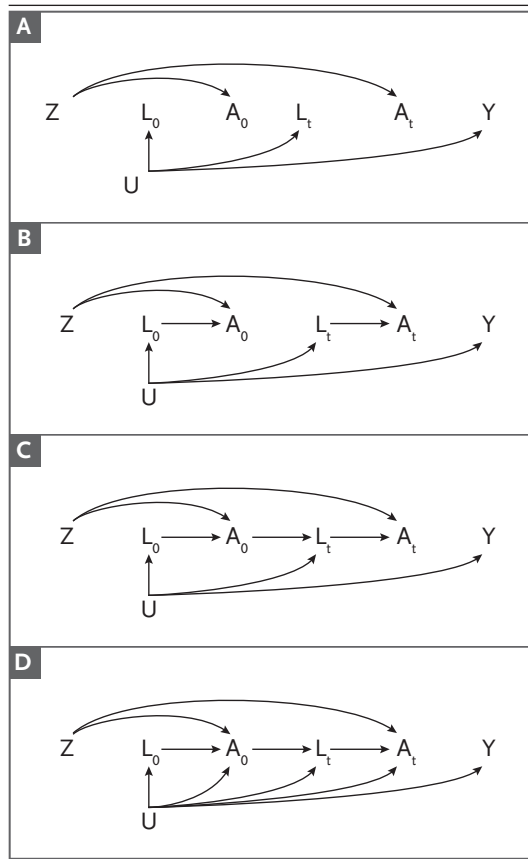


Figure 2. Four Possible Scenarios in a Randomized Trial of Sustained Treatment Strategies.

Shown are four possible scenarios in a randomized trial of sustained treatment strategies, according to the determinants of adherence to the treatment strategies.

Z is an indicator for randomization group, Y is the outcome of interest, L_0 represents prognostic factors measured at baseline (time 0), L_t represents prognostic factors measured after baseline (at time t), A_0 is an indicator for adherence to the protocol at baseline, A_t is an indicator for adherence to the protocol at time t, and U represents unmeasured baseline or postbaseline factors. To avoid clutter, we show only the first two time points, assume no losses to follow-up, and assume a null comparative effect of the treatment strategies. In scenario 1 (Panel A), adherence occurs at random (no arrows from any variable to A). The per-protocol analysis does not have to adjust for any factors. In scenario 2 (Panel B), adherence depends only on the measured factors (arrows from L to A). The per-protocol analysis has to adjust for these factors. In scenario 3 (Panel C), adherence depends only on the measured factors and the measured factors affect future adherence (arrows from L to A and from A to L). The per-protocol analysis has to adjust for these factors; the adjustment for the postbaseline factors requires g-methods. In scenario 4 (Panel D), adherence depends on both measured and unmeasured factors (arrows from L and U to A). A valid per-protocol analysis should use a form of instrumental variable estimation based on g-estimation of a structural nested model, which relies on strong assumptions.

their time-varying prognostic factors. In the resulting weighted population, treatment is mathematically independent of the measured prognostic factors, which eliminates postrandomization confounding and selection bias due to those measured factors. (Note that the method cannot eliminate bias due to factors that remain unmeasured.) Inverse-probability weighting was applied to estimate the per-protocol effect in a large Women Health's Initiative randomized trial of estrogen plus progestin hormone therapy.¹⁹ The estimated hazard ratios for breast cancer with hormone therapy versus placebo were 1.25 (95% confidence interval [CI], 1.01 to 1.54) in an intention-to-treat analysis and 1.68 (95% CI, 1.24 to 2.28) in a per-protocol analysis (the unweighted estimate was 1.54). From a patient's viewpoint, learning that the risk of breast cancer is 68% higher with hormone therapy, when taken as indicated, than with no hormone therapy, is more relevant than learning that the risk of breast cancer is 25% higher with hormone therapy in a group of women, many of whom stopped hormone therapy altogether during the follow-up.

The parametric g-formula, another g-method, was applied to estimate the per-protocol effect in the International Network for Strategic Initiatives in Global Human Immunodeficiency Virus Trials (INSIGHT) Strategic Timing of Antiretroviral Treatment (START) randomized trial,²⁰ which compared the effect of immediate initiation of antiretroviral treatment with deferred initiation (until the participant had a confirmed drop in CD4 cell count to less than 350 cells per cubic millimeter or received an AIDS diagnosis) among participants who were positive for the human immunodeficiency virus and had a CD4 cell count higher than 500 cells per cubic millimeter. Although participants generally adhered well to the protocol, some deviations occurred. In particular, 30% of the participants who were randomly assigned to the deferred-treatment group started antiretroviral treatment when they had a CD4 cell count of 350 cells per cubic millimeter or higher. The per-protocol effect estimate was 20% larger than the intention-to-treat effect, which, in absolute terms, translates into an additional 0.7-percentage-point reduction (from 3.8% to 3.1%) in the 5-year risk of the primary composite outcome of any serious AIDS event, serious non-AIDS event, or death from any cause.

These examples illustrate how per-protocol

effect estimates can be presented along with intention-to-treat effect estimates when the latter are not sufficiently patient-centered. However, one should keep in mind that per-protocol analyses may insufficiently adjust for confounding and selection bias if important prognostic factors that affect either adherence or loss to follow-up are not correctly identified, measured, and adjusted for.

ESTIMATING PER-PROTOCOL EFFECTS FOR SIMPLER TREATMENT STRATEGIES

Some pragmatic trials compare treatment strategies that consist of a single intervention at baseline. For example, in a study designed to compare two different types of hernia operation, patients would be randomly assigned to undergo one of the two interventions immediately. In this research setting, an intention-to-treat analysis would provide valid estimates of both the intention-to-treat effect and the per-protocol effect because nearly all patients undergo the assigned intervention.

In other research settings, studies with a single intervention at baseline may have incomplete adherence, in which case an intention-to-treat analysis would not provide valid estimates of the per-protocol effect. For example, participants in Norway were randomly assigned to be invited to undergo sigmoidoscopy screening or to not be invited.²¹ An intention-to-treat analysis showed a lower incidence of colorectal cancer and lower colorectal-cancer-specific mortality in the sigmoidoscopy group than in the control group. No effect on overall mortality was detected. However, the intention-to-treat effects underestimated the per-protocol effects because only 70% of the participants in the sigmoidoscopy group underwent the procedure; the other 30% declined the invitation. Persons outside the study who are considering whether to undergo a sigmoidoscopy may be more interested in the per-protocol effect that would have been estimated with full adherence to the intervention. Because the treatment or intervention was administered only at baseline in this trial, it is sufficient to adjust for prerandomization factors, and therefore conventional adjustment methods would be adequate. In this research setting, g-methods are not needed to adjust for incomplete adherence. However, an analysis of the data from the sigmoidoscopy trial implied that some key adjustment variables were

missing. For instance, the overall mortality among participants who declined to undergo a sigmoidoscopy was greater than the mortality in the control group, even after adjustment for all available factors.²¹ Unlike in the Coronary Drug Project, the available data were insufficient to provide approximate comparability between those who did and those who did not adhere to the intervention. Specifically, no data on cigarette smoking and other lifestyle risk factors were obtained, and therefore an adjusted per-protocol analysis did not result in unbiased per-protocol effect estimates for overall mortality.

Instrumental variable estimation is an attractive alternative for estimating per-protocol effects in trials of a baseline intervention versus standard of care because this method, unlike inverse-probability weighting and the g-formula, does not require data on prognostic factors predictive of adherence. In contrast, the method requires that the indicator of randomization group is an instrumental variable. This indicator is an instrumental variable if its effect on the outcome is mediated entirely through the intervention itself — a condition known as the exclusion restriction, which is likely to be satisfied in the sigmoidoscopy trial.²¹ Under the exclusion restriction, an instrumental variable estimator will be unbiased for a per-protocol effect if there are no “defiers” in the trial (i.e., participants who would refuse the intervention if assigned to it but would undergo the intervention if assigned to standard of care). The absence of defiers, known as the monotonicity condition, is satisfied in the sigmoidoscopy trial because screening colonoscopy was not available to participants in the control group.

However, even when the conditions of exclusion restriction and monotonicity are satisfied, instrumental variable methods estimate the per-protocol effect only in the subgroup of participants who would have adhered regardless of their assignment (i.e., the “compliers” in the particular trial).^{22,23} In the sigmoidoscopy trial,²¹ the per-protocol estimate of the 10-year risk of colorectal cancer among the compliers exceeded the intention-to-treat estimate in the entire study population by approximately 0.12 percentage points. The clinical applicability of this per-protocol effect is limited if the effect in the trial compliers differs substantially from the effect in other patients. This limitation is compounded by the fact that the compliers in the trial are not identifiable and do not

necessarily have the same characteristics as the compliers in the community.

Although instrumental variable estimation of per-protocol effects is noncontroversial in some trials of baseline interventions, in trials of sustained interventions it requires both strong, partly unverifiable assumptions about the dose–response relation and the use of a general form of g-estimation of a structural nested model.¹⁶ The relative advantages and disadvantages of instrumental variable methods, both the classic ones and the more general versions linked to g-estimation,^{24,25} for the estimation of per-protocol effects have yet to be studied systematically in pragmatic trials. However, such further study is warranted, because instrumental variable analyses, like intention-to-treat analyses, are guaranteed to be valid under the sharp null hypothesis of no effect of either treatment or treatment assignment on the outcome of any participant, even in the absence of data on postrandomization prognostic factors, provided that selection bias due to loss to follow-up is absent.

DISCUSSION

Because pragmatic trials consume substantial societal resources, we wish to extract as much useful information as possible from them by complementing intention-to-treat effect estimates with per-protocol effect estimates. We think that obtaining clinically useful per-protocol effect estimates requires the use of statistical techniques such as instrumental variable methods (when comparing baseline interventions) or g-methods (when comparing treatment strategies that required sustained adherence).

The validity of per-protocol analyses depends not only on the choice of the appropriate method but also on an explicit definition of the per-protocol effect, an a priori specification of the statistical plan for the per-protocol analysis, and the collection of high-quality data on adherence and prognostic factors (Table 1). These three aspects need to be addressed in the study protocol to prevent suboptimal design, conduct, and analysis of pragmatic trials and to foster confidence in the study results.

An explicit definition of the per-protocol effect is key to guiding investigators in selecting the design, the data collection, and the statistical analysis plan. Specifically, the protocol needs to clarify

Table 1. Data Requirements for the Definition and Estimation of the Per-Protocol Effect in Randomized Trials with Sustained Treatment Strategies.

Available Data	Example	Identification of Adherence to the Protocol	Estimation of the Per-Protocol Effect
Randomization group assignment and outcome	A large, simple trial	Not possible	Not possible
Group assignment and outcome plus treatment received after randomization	A trial that records pill counts but not clinical information at each visit	Not possible	Not possible
Group assignment and outcome plus treatment received after randomization plus protocol-specified clinical events that either mandate or allow treatment changes	A trial that records data on protocol-specified toxic effects and contraindications	Possible	Possible if adherence is independent of prognosis
Group assignment and outcome plus treatment received after randomization plus prerandomization and postrandomization prognostic factors associated with adherence	A trial that records detailed clinical data at and after randomization	Possible	Possible; g-methods are required if prognostic factors are affected by previous treatment

that patients assigned to therapy are expected to discontinue it if they have serious adverse effects so that these discontinuations are not labeled as protocol deviations. A second benefit of an explicit definition of the per-protocol effect is the opportunity to engage patient and other stakeholders in a discussion of the treatment strategies of interest to them. When trials study complex strategies for which perfect adherence is unrealistic, one may redefine the per-protocol effect as the effect under “satisfactory adherence”²⁶ to the protocol (e.g., 80%), rather than the effect under 100% adherence to the protocol. Furthermore, when there is concern that treatment benefits may vary substantially across subgroups of patients defined by their genetics, biomarker levels, or clinical characteristics, the protocol may specify the estimation of separate intention-to-treat and per-protocol effects. A patient-centered analysis that estimates the per-protocol effects in subgroups of patients is an example of precision medicine.

An a priori specification of the statistical analysis plan for per-protocol effects is key to dispelling concerns about selective reporting of findings on the basis of “hypothesizing after results are known” (HARKing) and conducting analyses until significant values are found (p-hacking).²⁷ In the absence of such specification, investigators may understandably believe that abandoning the intention-to-treat principle leaves them vulnerable to criticism, and they choose to make the (non-

patient-centered) intention-to-treat effect their only inferential target and anchor for sample-size calculations. Because a priori full specification of the statistical analysis plan for per-protocol effects is difficult, the study protocol may include adaptive features to fine-tune per-protocol analyses without compromising the validity of the effect estimates. For example, the protocol could specify detailed rules for modeling the inverse-probability weights as a function of the postrandomization data. Furthermore, the protocol may specify a set of statistical analyses aimed at assessing the sensitivity of per-protocol effect estimates to various assumptions. As an example, the protocol may specify sensitivity analyses to quantify how the effect estimates would vary if adjustment for confounding had been imperfect (for baseline intervention studies, simple methods are available^{28,29}) and instrumental variable analyses to estimate upper and lower bounds for the effect.³⁰

The collection of high-quality data before and after randomization is essential for most per-protocol analyses. Unless the trial compares baseline interventions with all-or-nothing adherence, such as in the sigmoidoscopy trial²¹ described above, it will be necessary to collect and analyze postrandomization data on adherence to the assigned treatment strategies and on the evolution of prognostic factors. Thus, embracing sound per-protocol analyses necessitates an appropriate framework for the design, conduct, and reporting³¹ of randomized trials. Of course, as for any observational

estimates, the validity of most per-protocol analyses depends on assumptions about confounding and selection bias that cannot be empirically verified.

In summary, the protocols of pragmatic trials would benefit from explicit definition of the per-protocol effect, including plans to measure adherence and postrandomization variables, and specifications of the statistical analysis plan. The added complexity is unavoidable if we wish to prioritize the patient-centeredness of the research and get the most out of pragmatic trials.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

This article was updated on October 5, 2017, at NEJM.org.

From the Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health (M.A.H., J.M.R.), and the Harvard–MIT Division of Health Sciences and Technology (M.A.H.), Boston. Address reprint requests to Dr. Hernán at the Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave., Boston, MA 02115, or at miguel_hernan@post.harvard.edu.

1. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995;57:6-15.
2. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res* 2011;2:109-12.
3. Piantadosi S. *Clinical trials: a methodologic perspective*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2005.
4. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012;9:48-55.
5. Goetghebeur E, Loeys T. Beyond intention to treat. *Epidemiol Rev* 2002;24:85-90.
6. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012;367:1355-60.
7. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000;56:779-88.
8. Dunn G, Goetghebeur E. Analysing compliance in clinical trials. *Stat Methods Med Res* 2005;14:325-6.
9. The Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *N Engl J Med* 1980;303:1038-41.
10. Murray EJ, Hernán MA. Adherence adjustment in the Coronary Drug Project: a call for better per-protocol effect estimates in randomized trials. *Clin Trials* 2016;13:372-8.
11. Robins JM. Correction for non-compliance in equivalence trials. *Stat Med* 1998;17:269-302.
12. Hernán MA, Robins JM. *Causal inference*. Boca Raton, FL: Chapman & Hall/CRC (in press).
13. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period — application to the healthy worker survivor effect. *Math Model* 1986;7:1393-512.
14. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550-60.
15. Hernán MA, Cole SR, Margolick J, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf* 2005;14:477-91.
16. Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Moltenberghs G, eds. *Longitudinal data analysis*. New York: Chapman and Hall/CRC Press, 2008:553-99.
17. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* 2005;61:962-73.
18. Van der Laan M, Rose S. *Targeted learning: causal inference for observational and experimental data*. New York: Springer, 2011.
19. Toh S, Hernández-Díaz S, Logan R, Robins JM, Hernán MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology* 2010;21:528-39.
20. Lodi S, Sharma S, Lundgren JD, et al. The per-protocol effect of immediate vs. deferred ART initiation in the START randomized trial. *AIDS* 2016;30:2659-663.
21. Holme Ø, Løberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA* 2014;312:606-15.
22. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;29:722-9.
23. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006;17:360-72.
24. Ten Have TR, Normand SL, Marcus SM, Brown CH, Lavori P, Duan N. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatr Ann* 2008;38:772-83.
25. Cole SR, Chu H. Effect of acyclovir on herpetic ocular recurrence using a structural nested model. *Contemp Clin Trials* 2005;26:300-10.
26. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. *Annu Rev Pharmacol Toxicol* 2012;52:275-301.
27. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640-8.
28. Rosenbaum PR. *Observational studies*. 2nd ed. New York: Springer, 2002.
29. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology* 2016;27:368-77.
30. Swanson SA, Holme Ø, Løberg M, et al. Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening. *Trials* 2015;16:541.
31. Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. *Trials* 2012;13:84.

DOI: 10.1056/NEJMsm1605385

Copyright © 2017 Massachusetts Medical Society.