

# Making Sense of Intention-to-Treat

Kristin L. Sainani, PhD

When analyzing data from randomized clinical trials, an intention-to-treat (ITT) analysis gives the highest level of evidence. However, this type of analysis is frequently the subject of both confusion and controversy. This article will review the ITT principle, why it is important, how it is applied and misapplied, and alternative strategies and their limitations.

## CASE STUDY

The Spine Patient Outcomes Research Trial (SPORT) on disk herniation will be used as a case study throughout this article [1]. In this trial, 501 patients with disk herniation were randomized to receive surgery (lumbar discectomy) or standard nonoperative care. The ITT analysis showed no effect for surgery, but both a non-ITT reanalysis of the data and a parallel observational cohort study [2] showed a strong benefit for surgery. These conflicting results incited much discussion and debate about the ITT approach (see Sidebar for more details).

## WHAT IS ITT?

The ITT principle stipulates that all participants who are randomized must be included in the statistical analysis and analyzed according to the treatment group to which they were randomly assigned, regardless of what treatment, if any, they actually received. Thus, patients will be analyzed as a member of their initial randomization group even if they refuse or discontinue the intervention, “cross over” to another randomization group, miss follow-up measurements, or otherwise violate study protocol. In SPORT, only 60% of 245 patients randomized to surgery actually received surgery, but the ITT analysis counted all 245 as members of the surgery group. Similarly, 45% of 256 patients randomized to nonoperative care decided to undergo surgery sometime during the study, but the ITT analysis retained all 256 as members of the nonoperative care group.

## WHY ITT?

The ITT analysis may seem, on first reflection, to run counter to common sense. If the patients didn't receive surgery, it seems unfair to count them as if they did.

However, ITT has 2 strong rationales. First, it is necessary to preserve the strengths of randomization. The purpose of randomization is to ensure that potential prognostic factors are balanced between the treatment groups. This balance may be lost if some patients are excluded from the analysis or analyzed according to how they self-selected rather than how they were randomized.

For example, excluding nonadherent patients may introduce bias because patients who adhere to treatments tend to have better outcomes regardless of whether the treatment is effective. In one study [3], those who adhered to the trial drug (clofibrate) had reduced mortality; but those who adhered to the placebo pill had the same reduction in mortality. Analyzing patients according to the treatment they actually received is also likely to bias the results; for example, if patients who self-select to surgery have a stronger belief in the efficacy of surgery this may create a spurious association between surgery and positive outcomes. It may seem justifiable to exclude patients who are nonadherent for reasons outside of their control, such as those who die before receiving treatment; but this can also create bias. Take a hypothetical example presented by Montori and Guyatt [4]. In a stroke trial, patients are randomized to receive surgery or nonoperative care. Some patients in both study arms die early in the trial; because there is a lag between randomization and surgery, some patients in

**K.L.S.** Division of Epidemiology, Department of Health Research and Policy, Stanford University, HRP Redwood Building, Stanford, CA 94305. Address correspondence to: K.L.S.; e-mail: [kcobb@stanford.edu](mailto:kcobb@stanford.edu)  
Disclosure: nothing to disclose

Disclosure Key can be found on the Table of Contents and at [www.pmrjournal.org](http://www.pmrjournal.org)

**Case Study Side Bar: The Spine Patient Outcomes Research Trial (SPORT) on Disk Herniation**

**Design:** In a 2-year trial of 501 disk herniation patients, 245 were randomly assigned to receive surgery (lumbar discectomy) and 256 were randomly assigned to receive standard, nonoperative care. An additional 743 patients who refused randomization were enrolled in a parallel observational cohort study, in which patients self-selected to receive surgery ( $n = 521$ ) or nonoperative care ( $n = 222$ ). The primary outcomes were 3-month, 1-year, and 2-year changes on subjective scales of bodily pain, physical function, and disability.

**Nonadherence in the randomized trial:** By the end of the 2 years, only 140 (60%) of surgery patients randomly assigned to surgery had received surgery; and almost half (45%) of the nonoperative care group had received surgery.

**Missing data in the randomized trial:** A total of 26 patients provided no follow-up data and others provided incomplete follow-up data (for example, 86 patients missed the 1-year follow-up visit).

**Results:** The following table shows the 1-year treatment effects (and 95% confidence intervals) for surgery versus nonoperative care based on: (1) an ITT analysis of the randomized trial; (2) a non-ITT analysis of the randomized trial, where patients were analyzed according to the treatment they actually received; and (3) the observational cohort study.

	Who was compared?	Bodily pain (positive effects indicate a benefit for surgery)	Physical function (positive effects indicate a benefit for surgery)	Disability (negative effects indicate a benefit for surgery)
ITT analysis (randomized trial)	Patients randomly assigned to surgery versus patients randomly assigned to nonoperative care.	2.8 (−2.3 to 7.8)	1.2 (−4.1 to 6.5)	−3.2 (−7.8 to 1.3)
Non-ITT analysis (randomized trial)	Patients who actually underwent surgery versus patients who actually received nonoperative care.	15.0 (10.9 to 19.2)*	17.5 (13.6 to 21.5)*	−15.0 (−18.3 to −11.7)*
Observational cohort study	Patients who chose to receive surgery versus patients who chose to receive nonoperative care.	10.8 (6.5 to 15.0)*	15.0 (10.9 to 19.2)*	−15.2 (−18.9 to −11.6)*

\*Statistically significant difference between surgery and nonoperative care,  $P < .05$ .

the surgery group (but not the nonoperative group) die before receiving treatment. Excluding these patients (and thus all early deaths in the surgery group) creates a spurious association between surgery and survival.

The second rationale for ITT is that it estimates the treatment effect in real-world clinical practice—where patients often don't adhere to treatments—rather than the treatment's efficacy when taken optimally. For example, in a randomized trial of a low-fat diet, only 31% of women assigned to the diet met the goal of reducing their fat intake to below 20% of total calories [5]. Restricting the analysis to just these women would overestimate the diet's benefits for most women.

## HOW IS ITT APPLIED?

The ITT principle sounds simple, but it is not straightforward to apply in practice. Many authors who claim to have used an "ITT analysis" have actually applied it inadequately [6, 7]. Authors often make mistakes in how they deal with noncompliance, false inclusions, and missing data [6, 7].

## Dealing with Noncompliance

Patients must be included in the analysis even if they never started the treatment, discontinued it, or adhered inadequately. For example, in SPORT, patients who were randomized to receive surgery, but never received it were retained in

the surgery group. Also, patients must be analyzed according to their initial randomization, even if they switched treatment arms. For example, in SPORT, nonoperative patients who had surgery were still counted as nonoperative patients.

## Dealing with False Inclusions

In some studies, patients are randomized but later discovered to be ineligible for the trial, because, for example, they don't have the disease of interest or have a contraindication to treatment that should have excluded them from the trial. A strict ITT analysis would still include these patients in the analysis ("once randomized, always analyzed"), but there is debate over whether this is necessary. According to some researchers [8], it may be acceptable to exclude these patients if: the grounds for exclusion were prespecified in the protocol, the decision to exclude is made by someone who is blinded to randomization and outcomes, and exclusions are applied equally to all study arms. If such exclusions are made, the analysis should be repeated with and without these subjects to determine the impact on results [8].

## Dealing with Missing Data

Dealing with missing data is the trickiest issue in applying ITT correctly. To perform a true ITT analysis, primary outcome data must be available for every patient in the trial, so

**Table 1.** *Strategies for Dealing with Missing Data on the Primary Outcome*

Strategy	Description	Advantages	Disadvantages
1. Optimal study design and execution	Design and execute studies to minimize missing data.	The best way to deal with missing data is to avoid it.	It is often not possible to avoid all missing data.
2. Statistical models that handle incomplete follow-up data	Certain statistical models can accommodate partial missing data; examples: longitudinal mixed models, Cox regression.	Avoids complete omission of participants who are missing some follow-up data.	Cannot include patients with no follow-up data; and estimates are still based on incomplete data.
3. Imputation	"Fills in" missing data points; examples: last observation carried forward, multiple imputation.	Allows the inclusion of all patients in the final analysis.	Uses values that may or may not be accurate.
4. Extreme case analysis	Assumes all missing subjects had a bad outcome, all missing subjects had a good outcome; or all missing subjects in one group had a good outcome and those in the other group had a poor outcome.	Gives a sense of the extent to which missing data could change the effect estimates.	The answers may be too extreme to be of much value, especially if the number of subjects with missing data is large.
5. Sensitivity analysis	Applies a number of different statistical strategies for dealing with missing data and compares the results.	The best statistical strategy for dealing with missing data. Gives a sense of the extent to which missing data is affecting the results.	Time consuming; may give a wide range of possible answers.

that every patient can be included in the final analysis. In reality, most trials have some dropouts and losses to follow-up. For example, in SPORT, 13 patients from the surgery group and 16 from the nonoperative care group provided no outcome data, and thus could not be directly included in the analysis; other patients provided only partial follow-up data.

Commonly, authors simply exclude patients with missing data [6, 7], but this is not optimal and can add bias. For example, if the nonoperative patients with poor outcomes are more likely to be lost to follow-up than surgery patients with poor outcomes, this will bias the results in favor of surgery. Thus, authors should use at least one of several statistical strategies for dealing with missing data, described in the following section and in Table 1.

Though many statistical models drop patients with missing data, certain models can incorporate patients with incomplete data. For example, longitudinal mixed models (used in SPORT) accommodate patients who missed some follow-up measurements and Cox regression models retain patients in the analysis for as much time as they were followed. These strategies will not help when a patient has provided no follow-up data, however.

Statistical imputation can be used to "fill in" both partial and complete missing data. The simplest imputation method is "last observation carried forward," where the last recorded value (sometimes the baseline value) is carried forward to all subsequent time points. So, if a patient reported a bodily pain score of 27 at the 3-month follow-up but missed the 6-month follow-up, a bodily pain score of 27 would be imputed for the 6-month time point. Other imputation methods (such as "multiple imputation") use more sophisticated statistical modeling to predict the value of missing outcomes based on a patient's available data.

Finally, extreme case analysis recalculates effects under extreme assumptions about the missing data—for example, that all the missing subjects had positive outcomes, all had negative outcomes, or that all missing subjects in one group had positive outcomes and in the other had negative outcomes. This gives a sense of the largest possible impact that the missing data could have on effect estimates.

Ideally, authors should use several different approaches for handling missing data and compare the results. This type of "sensitivity analysis" can help readers gauge the extent to which missing data is likely to have impacted the conclusions. In SPORT, the authors tried several different analytic strategies for filling in the missing data, and found that these alternate approaches had little impact on the effect estimates.

Of course, the best way to deal with noncompliance, false inclusions, and missing data is to design a study that minimizes all 3. An ITT analysis cannot fix a poorly designed or executed study.

## ALTERNATIVES TO ITT AND THEIR LIMITATIONS

Authors often present alternatives to ITT analysis, including: modified ITT, per-protocol analysis, and as-treated analysis. These approaches can contribute important information, but they provide a weaker level of evidence than an ITT analysis.

### Modified ITT

In recent years, the "modified ITT" analysis has become increasingly popular in the literature as an alternative to a strict ITT analysis [9]. This design allows some exclusions from the ITT population (such as patients who were deemed

ineligible after randomization or certain patients who never started treatment) if these exclusions can be justified as unlikely to bias the results. Unfortunately, modified ITT has not been clearly defined and there is a lack of consistent guidelines for its application [9]. Thus, it is subjective and opens the door for authors to manipulate or “tidy up” their data and thus introduce bias. It is imperative that authors clearly describe the modifications used and indicate whether these modifications were specified in the original protocol or were applied to the data post-hoc. Readers should exercise caution when interpreting the results of a modified ITT analysis and evaluate the likelihood that the modifications could have introduced bias.

### Per-protocol Analysis

A per-protocol analysis excludes all patients who have violated protocol, including anyone who did not adhere to treatment, switched groups, or missed measurements. This approach may give insight into the efficacy of treatment under optimal conditions. However, because it loses the balances of randomization, the results are more akin to those of an observational study than a randomized trial, and should be interpreted with caution.

### As-treated Analysis

Rather than excluding nonadherent patients, an as-treated approach analyzes them according to the treatment they actually received. The aim is to estimate the efficacy of a treatment when actually taken. An as-treated analysis was performed as a secondary analysis in SPORT. The surgery group comprised the 140 patients randomly assigned to surgery who actually received it and the 107 patients randomly assigned to nonoperative care that received surgery (adjusted for the time of surgery); the nonoperative group comprised the remaining patients. As with the per-protocol analysis, the as-treated analysis loses the balances of randomization and thus is more akin to an observational study. In SPORT, patients who crossed over from surgery to nonoperative care differed in several respects (including disease severity and income) from patients who were adherent to surgery; and patients who crossed over from nonoperative care to surgery differed from those who were adherent to nonoperative care (including in age, income, and disease severity). Thus, the as-treated groups were imbalanced in several important prognostic factors that may have influenced outcomes. Though it is possible to adjust for these differences in the statistical analysis (as was done in SPORT), statistical adjustment cannot completely remove confounding by these variables and it can only account for variables that were measured. The limitations of statistical adjustment will be discussed in a future column.

### Interpreting Results from an ITT Analysis

The SPORT trial generated much controversy because the ITT analysis gave different results than the non-ITT analysis (and parallel observational study). This is a common outcome in randomized trials: the ITT analysis gives a null result, but the non-ITT analysis finds a significant effect.

There are 2 possible reasons for this type of split result: bias introduced by the non-ITT analysis may be creating a spurious association or statistical power may be reduced in the ITT analysis, making it more difficult to detect a significant difference. It is impossible to differentiate between these 2 alternatives, and both factors may be at work.

In SPORT, the results of the as-treated analysis were almost identical to the results of the parallel SPORT observational cohort. This may seem to strengthen the case that surgery is effective, but, in fact, both designs are observational in nature and thus subject to the same biases.

In SPORT, the ITT analysis was almost certainly underpowered, as the researchers failed to factor in nonadherence when calculating sample size needs for the trial. Because the surgery group is “contaminated” with participants who did not receive surgery and the nonoperative group is contaminated with participants who received surgery, this makes the groups appear more similar and makes it harder to detect a difference between them (if one exists). Thus, the ITT analysis may have had only a slim chance of detecting a real difference between the groups.

Where does that leave us? As the SPORT authors state in their write-up, the randomized trial provides inadequate evidence for making conclusions. Thus, there is no randomized trial-level evidence that surgery is effective (or ineffective). On the other hand, the observational data (from both the as-treated analysis and the cohort study) show a strong benefit for surgery. As with any observational study, this is suggestive of an effect but does not provide the highest level of evidence. It is perhaps a frustrating conclusion, but what SPORT really adds to the literature is 2 high-quality observational datasets that show surgery is effective.

### CONCLUSIONS

ITT analysis guards against bias, thus giving the highest level of evidence for clinical research. However, ITT analysis is often applied incorrectly, so readers should critically evaluate its application, paying close attention to the handling of noncompliance, false inclusions, and missing data. Ideally, randomized trials should be designed and executed such as to minimize these problems. Alternatives to ITT analysis can provide additional information, but may introduce biases and thus should be interpreted with caution.

### REFERENCES

1. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes

- Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441-2450.
2. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgery vs non-operative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) Observational Cohort. *JAMA* 2006;296:2451-2459.
  3. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *Engl J Med* 1980;303:1038-1041.
  4. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ* 2001;165:1339-1341.
  5. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer: The Women's Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA* 2006;295:629-642.
  6. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomized controlled trials. *BMJ* 1999;319:670-674.
  7. Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350-356.
  8. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;325:652-654.
  9. Iosief A, Alessandro M, Carlo R. Modified intention to treat: frequency, definition and implication for clinical trials. 15th Cochrane Colloquium, Sao Paulo, 23-27 October 2007. Available at: <http://www.imbi.unifreiburg.de/OJS/cca/index.php/cca/article/view/5036>. Accessed December 21, 2009.