Analyzing Phase III Studies in Hospice/Palliative Care. A Solution That Sits Between Intention-to-Treat and Per Protocol Analyses: The Palliative-Modified ITT Analysis

David C. Currow, BMed, MPH, FRACP, John L. Plummer, PhD, AStat, Jean S. Kutner, MD, MSPH, Greg P. Samsa, PhD, and Amy P. Abernethy, MD

Discipline, Palliative and Supportive Services (D.C.C., A.P.A.), Flinders University, Bedford Park, Adelaide, and Pain Management Unit (J.L.P.), Flinders Medical Centre, Bedford Park, Adelaide, South Australia, Australia; School of Medicine (J.S.K.), University of Colorado, Aurora, Colorado, and Department of Biostatistics and Bioinformatics (G.P.S.), and Division of Medical Oncology (A.P.A.), Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA

Abstract

Intention-to-treat (ITT) analyses are the standard way to evaluate randomized controlled trials (RCTs) to minimize Type I errors related to differential rates of noncompletion from one study arm. People in palliative care often die sooner than predicted as a direct result of disease progression, some of whom will be participating in RCTs and who will, therefore, withdraw or die after randomization for reasons unrelated to the intervention. This proportion of withdrawals is statistically negligible in other clinical disciplines, but commonplace in hospice/palliative care, creating a systematic bias away from the true effect. ITT analyses in hospice/palliative care that deem all withdrawals to be treatment failures or that impute data from deteriorating participants systematically underestimate the benefits of interventions, reducing the power of these studies. Equally unacceptable would be a per protocol analysis that excludes all withdrawals after randomization as this will underestimate toxicity. A modified analytic approach is needed on a continuum between ITT and per protocol analyses. To address data after randomization where there is a high rate of withdrawals because of death or deterioration, criteria need to include being: 1) prespecified in the original protocol; 2) clinically absolutely the result of disease progression; 3) identified by the blinded Independent Data Monitoring Committee as being unrelated to the intervention(s); and 4) accounted for in the study’s CONSORT diagram. Such data should not be included in the analysis of the primary outcome. This article aims to define a better way of balancing Type I and Type II errors in hospice/palliative care RCT analyses using the palliative-modified ITT analysis. Arguably, the palliative-modified ITT analysis should be the primary evaluation of hospice/palliative care Phase III studies but, as a minimum, should routinely be the key sensitivity analysis.

© 2012 U.S. Cancer Pain Relief Committee. Published by Elsevier Inc. All rights reserved.
Key Words
Intention-to-treat analyses, per protocol/on-treatment analyses, palliative care research, clinical trials, randomized controlled trials, phase III studies

Introduction

Randomized controlled trials (RCTs) are pivotal for objectively evaluating new interventions and require three design elements: 1) randomization; 2) inclusion of all randomized participants in the primary outcome analysis; and 3) systematic resolution or accommodation of missing data.1 This is often difficult to achieve even in well-designed and rigorously conducted studies.2,3 These principles are complemented with key aims to minimize bias (ensuring that the estimated treatment effect is as close as possible to the true effect), maximize precision (achieving the smallest possible confidence intervals), and evaluate robustness (sensitivity analyses to limitations in the data testing the assumptions underpinning the design of the study).3

In RCTs, ideally all participants would receive all allocated interventions at the times outlined in the study protocol, and all participants would contribute to the data collection as planned. As this is rarely the case, there are two data sets that are commonly analyzed in RCTs: intention-to-treat (ITT) and per protocol. Although there are a number of definitions in the literature,4 for the purposes of this article, the ones outlined in the Harmonised Tripartite Guideline—Statistical Principles for Clinical Trials will be used.3

- ITT analyses aim to include all randomized subjects in the arm to which they were allocated and account for all follow-up data, thereby creating an assessment that looks at the effectiveness of the intervention and will likely reflect its use in day-to-day practice.5 Despite this being the gold standard, ITT analyses are often poorly applied and reported.5,6 The goal of ITT analyses is to minimize Type I errors by analyzing data from all people, including those who may not receive or complete their assigned treatment while demonstrating that key clinical and prognostic factors are equivalent in all arms of the study.7 ITT is designed to minimize bias by systematically including all post-randomization events, accounting for the outcomes of all participants, and thereby creating a solid basis for statistical inference.
- Per protocol analyses explore a subgroup of the data, where participants who have had a predetermined exposure to the intervention(s) have measurable data at the primary end point and have no major protocol violations. Analyses test the efficacy of the intervention(s), thus most closely reflecting the scientific protocol underlying the study.3 Ultimately, these are the participants who have been able to comply fully with the protocol.

Some participants are expected to have poor outcomes even in the absence of intervention, and those patients at the highest risk of the poorest outcomes also are least likely to get the full intervention in any study. ITT analyses were introduced to account for two scenarios in the conduct (and hence outcomes) of RCTs when data are not available for the primary end point because of dropout differentially between study arms:

- When participant data are missing because they had not received their allotted intervention (as a result, e.g., of differential access to the intervention as an inherent design function or inadvertent variation within a study). For example, there is immediate access to one intervention to which people are randomized (medical treatment of ischemic heart disease) but delayed access to another (surgical intervention for the same indication because of delays in the availability of the operating suite or processes for optimizing a patient’s readiness for major surgery). Analyzing this trial on a per protocol basis by excluding deaths in the surgical arm that occurred after randomization but before surgery was available would systematically overestimate the benefit of surgery compared with medical treatment because, in this case, the death could be attributed directly to the delays in accessing definitive treatments in some circumstances; or
• When data are missing on participants after commencing the intervention in a non-random way. For example, although there is immediate access to both interventions after randomization, one arm has a differential withdrawal rate because of toxicity from the intervention or because of a perceived lack of effect. Data will not be available after withdrawal. A per protocol or on-treatment analysis that excludes data from participants who withdrew under these circumstances will overestimate the benefits in the arm with excess withdrawals. If the participant has withdrawn but is still alive, then those data are potentially available.

The basis of all designs and analyses of RCTs is to minimize biases. There are two biases that work in opposing directions from the true effect that need to be taken into account in hospice/palliative care studies: 1) excluding all subjects because of post-randomization events (which will systematically overestimate the benefit of the intervention) and 2) considering all withdrawals to be failures or have outcomes imputed that do not take into account the deterioration in the days leading up to withdrawal or death (which will systematically underestimate the benefit of the intervention). These divergent biases also may be of different magnitudes in the same study, with the impact of this second bias of much larger magnitude in hospice/palliative care research than in any other clinical discipline. The impact of these countervailing biases is, therefore, of sufficient magnitude to require a specific solution that further refines the ITT paradigm to reach the best estimate of effect. The least biased estimate of the true effect, therefore, will lie on a continuum between ITT and per protocol analyses—the palliative-modified ITT.

The Problem

Although there are challenges in designing and conducting RCTs in hospice/palliative care, such studies are crucial, ethically defensible, and feasible if conceived with the needs of participants in mind. A characteristic of hospice/palliative care clinical trials is that a substantial proportion of participants is expected to deteriorate clinically or die during the study, independent of the study intervention. Importantly, the deterioration or death of these participants will not be individually foreseen at the commencement of the study because of limits in the ability of clinicians to prognosticate. Some of these people will be participating in RCTs at the time of clinical deterioration or death.

For most controlled clinical trials where a small proportion of people withdraw after randomization for reasons unrelated to the intervention, the effect statistically and clinically is negligible. By contrast, where this proportion is higher, the use of conventional ITT analyses systematically underestimates clinical benefit when the outcomes are analyzed. The larger the proportion of such withdrawals, the larger the underestimate of benefit will be (Table 1). With a large number of withdrawals and the assumption in ITT that all withdrawals are nonresponders or where data have been imputed from deteriorating participants, their inclusion in analyses will generate a reduction in the power of the study (Fig. 1). This is partially offset by the increase in total numbers contributing to the analysis. Consequently, a rather large proportion of participants must withdraw to cause a substantial change in the P-value (Table 1). The net result is that potentially beneficial interventions may be discarded erroneously even when a high-quality Phase III study is conducted in a hospice/palliative care population.

Interventions can cause premature mortality in hospice/palliative care research as in any area of clinical pursuit, so arbitrarily moving to a per protocol assessment as the primary analysis is not the solution. As in any setting, to use only a per protocol analysis would be unlikely to be the best estimate of net clinical benefit in day-to-day practice. Furthermore, poor compliance, use of excluded medications, loss to follow-up, and “nonignorable missing data” that should have been available for analysis diminish the value of a per protocol approach.

The aim of this article was to propose an analysis plan and its parameters that sit conceptually and statistically on a continuum between ITT and per protocol analyses.

A Worked Example

Estimates of treatment benefit become more biased as the proportion of withdrawals unrelated to the intervention rises. In palliative care
RCTs, rates of withdrawal unrelated to the intervention may be as high as 50% of participants randomized.\textsuperscript{10} When palliative interventions are tested, the anticipated clinical effect sizes are usually large. The corollary is that required sample sizes are relatively small for an adequately powered study, given the size of the clinical effect sought between groups. With the latter, the likelihood of a Type II error is high even if there are a relatively small number of withdrawals unrelated to the intervention and ITT analysis is used.

Consider four hypothetical studies (Table 1). These parallel-arm, double-blind RCTs compare a current therapy with a new intervention. If the analysis is on an ITT basis, all people randomized will be included. \textsuperscript{1} Table 1 contrasts the four studies where increasing numbers of participants withdraw because of death or deterioration before the primary census point for reasons unrelated to the intervention and the net effect of these withdrawals.

In a hospice/palliative care trial evaluating a new approach to controlling a symptom, if a patient dies after randomization but before receiving the intervention, this does not say anything about the efficacy of the intervention but only that the patient did not survive long enough to provide any data for evaluation. Omitting these data is explicitly in keeping with the criteria laid out in the Harmonised Tripartite Guideline.\textsuperscript{3}

### Table 1

<table>
<thead>
<tr>
<th>Proportion Who Withdrew With Disease Progression Unrelated to Intervention and Omitted From Analysis\textsuperscript{a}</th>
<th>n</th>
<th>Pearson Chi-Square</th>
<th>P-value</th>
<th>Benefit</th>
<th>Participant Flow for Negative Outcomes</th>
<th>Randomization Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, 0.0%</td>
<td>106</td>
<td>3.34</td>
<td>0.068</td>
<td>Yes\textsuperscript{b}</td>
<td>Not applicable</td>
<td>26</td>
</tr>
<tr>
<td>B, 4.0%</td>
<td>102</td>
<td>3.41</td>
<td>0.065</td>
<td>No</td>
<td>Completed, no toxicity</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Toxicity\textsuperscript{c}</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Other reason\textsuperscript{c}</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Disease progression\textsuperscript{c}</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C, 7.5%</td>
<td>98</td>
<td>3.62</td>
<td>0.057</td>
<td>No</td>
<td>Completed, no toxicity</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Toxicity\textsuperscript{c}</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Other reason\textsuperscript{c}</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Disease progression\textsuperscript{c}</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>D, 21%</td>
<td>84</td>
<td>3.86</td>
<td>0.049</td>
<td>No</td>
<td>Completed, no toxicity</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Toxicity\textsuperscript{c}</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Other reason\textsuperscript{c}</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No Disease progression\textsuperscript{c}</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

\textsuperscript{ITT = intention-to-treat.}
\textsuperscript{a}Models A and D are identical to Models 1 and 3 in Table 2.
\textsuperscript{b}Benefit unchanged for all models in this table.
\textsuperscript{c}Withdraw.

A Proposed Solution

In hospice/palliative care, a logical extension of this principle sees a larger group of people have their data excluded if the person deteriorates or dies before reaching the primary endpoint of the study and that person’s deterioration or death is clearly unrelated to the intervention being evaluated. The proportion of participants withdrawing from Phase III studies for reasons unrelated to the interventions being tested is the singular problem in applying the tenets of ITT directly to hospice/palliative care studies.

In proposing a solution that omits these participants’ data, the principles underpinning the Harmonised Tripartite Guideline on the analysis of clinical trials still reflect: 1) absolute adherence to randomization as the basis for trial conduct; 2) the need to account for all
participants and their outcomes; and, through this, 3) creation of a solid base from a known and defined subgroup of people at randomization for the primary analysis.

The key requisite in considering the omission of data from participants in the primary analysis is that there is a proportionally equal withdrawal from each study arm as a result of factors related to disease progression or sudden death but not the study intervention itself. Any study where there is differentially high withdrawal in one arm must be considered to have a potential study-related effect, and a conventional ITT analysis should be undertaken as the primary analysis.

To illustrate this using the identical study from Table 1, 22 of 106 study participants withdrew from the study because of disease progression or death unrelated to treatment (Table 2). Using a conventional ITT analysis, all withdrawals are deemed nonresponders and the denominator over which responses will be calculated is unnecessarily high (106), potentially leading to systematic bias. Alternatively, participants will have their outcomes imputed, but this is likely to have a preceding period of deterioration of variable length, again unrelated to the intervention. In reality, an unknown and undefinable proportion of the 22 individuals with disease progression or “expected” death unrelated to the intervention would have responded to the intervention. Some would have derived no benefit, and a proportion would have experienced some toxicity that limited their use of the intervention had they been able to participate in the trial longer.

Omitting data from these 22 participants in the palliative-modified ITT analysis is proposed.

Fig. 1. Participant flow through a study and data potentially included in subsequent differing analyses. According to the Harmonised Tripartite Guideline—Statistical Principles for Clinical Trials, failure to take at least one dose of trial medication when knowledge of the assigned treatment arm cannot influence this decision, or where there are no data after randomization, may allow omission of data from these participants in a fully compliant ITT analysis. Although withdrawal from the study because of toxicity and lack of effect may be unequal, for a palliative-modified ITT analysis to be used, the withdrawal due to “deterioration or death explicitly unrelated to any study intervention” should be of equal size. Any deviation from this, a standard ITT analysis should be used. ITT = intention-to-treat.
In absolute keeping with the intent of the principles of the Harmonised Tripartite Guideline for the proposed solution, omission of participants from the primary analysis would need to comply with strict criteria that require that the withdrawal was:

- prespecified in the original protocol before the study commenced;
- the result of known clinically assessed disease progression;
- independently assessed at the time of withdrawal to be absolutely unrelated to the intervention(s) being evaluated by a blinded Independent Data Monitoring Committee (IDMC);
- a random event equally distributed across all arms in the study reflecting consistency in the likelihood of such events; and,
- transparent, by ensuring that all participants were accounted for in the study’s CONSORT flow diagram demonstrating that all groups were treated and behaved identically.

The independent blinded assessment by the IDMC sits broadly under the mandate to “assess at intervals the progress of a clinical trial” outlined in the Harmonised Tripartite Guideline and has a related precedent in the literature where an IDMC is asked to rule on the application of pre-randomization criteria as applied to individual participants. If there is any question that the withdrawal could be the result of any of the intervention(s) being evaluated, then the participant’s data must be included in the final analysis.

Table 2, therefore, contrasts ITT and per protocol (denominator = 77) analyses with the palliative-modified ITT analysis, which lies on a continuum between the other two approaches. All 22 meet the five criteria outlined previously, and in the last model, the denominator has been adjusted to 84 by omitting the 22 participants who withdrew (11 withdrawals from each group) because of disease progression or death. This change in the analysis shifts the interpretation of results statistically and, arguably, clinically. Comparison of these tables highlights an approach that sits philosophically between conventional efficacy (per protocol) and effectiveness (ITT) analyses, potentially providing clinical decision makers with a more realistic estimate of day-to-day benefit in hospice/palliative care practice.
Why not consider these withdrawals simply a problem of missing data? There are different types of missing data often divided categorically into “ignorable” and “nonignorable.” Ignorable missing data are data that are “missing completely at random” (i.e., for reasons unrelated to any characteristics or responses of the subjects) or where the probability of being missing depends on the values of variables that have actually been measured (e.g., subjects with a particular diagnosis are less likely to provide data, but their diagnosis is still known). If the probability of the outcome variable not being recorded is related to the (unknown) value of the outcome variable, this would be a case of “nonignorable” missing data. In the case being outlined for hospice/palliative care research, as the missing data are judged by the IDMC to be independent of the intervention, it becomes a consideration of “ignorable” missing data.

Clinically, in hospice/palliative care, most people decline over a period of time without dramatic changes in their overall condition, leading eventually to their death. Therefore, any measures used to impute participants’ outcomes are likely to be affected by deterioration in subjects’ conditions that lead to their ultimate withdrawal from the study. Although sudden changes in a person’s clinical condition including sudden death may occur, the scenario more frequently encountered is of progressive deterioration over several days or weeks preceding such withdrawal. Coding this when it happens after a period of deterioration as “missing data” and using mathematical imputation or last observation carried forward will not adequately address participants’ deterioration in the days leading up to withdrawal or death. The data with death or clinical deterioration are data that do not exist in contrast to studies where data exist but are simply not observed (and, in the latter case, where imputation becomes a reasonable process to put in place). Incorporating data from the phase where deterioration is unrelated to the intervention will underestimate any benefit of the intervention being assessed. To take this to its logical conclusion, if the deterioration is unequivocally unrelated to the intervention, data should only be included from such participants as a sensitivity analysis, given the difficulty determining the time point at which this deterioration independent of the intervention commences. Randomly missing data in the remaining cohort will need to be dealt with in accordance with the original study plan. Sample size calculations will need to be done on the cohort who are evaluable after withdrawal of those who have deteriorated.

Previous proposals have grappled with similar problems. Shih and Quan proposed an approach to deal with the case of dropout for treatment-related reasons such as “adverse reaction” and “lack of improvement.” Patients who dropped out for “non-outcome-related reasons” were excluded from the analysis as an analytical expediency. This contrasts with the approach proposed in the current palliative-modified ITT analysis in that we purposefully exclude only dropouts for “non-outcome-related reasons” to reduce bias. Yusuf et al. have proposed criteria for subgroup analysis that include ensuring that the direction of benefit is the same for each subgroup as the overall (adequately powered) outcomes of the study. There does need to be an ITT analysis to complement the more limited analysis of the hospice/palliative care population who have genuinely evaluable end points related to actual exposure to the treatment in a trial arm and the ability to reach a census point that has the shortest possible time to be able to measure effect.

Discussion

Hospice/palliative care studies involve a heterogeneous population where: timing of referral in relation to the disease trajectory varies widely between and within services; patients experience a range of life-limiting illnesses often with multiple, complex comorbid illnesses; and the diagnosis alone poorly predicts resource utilization and prognosis. Death or clinical deterioration is anticipated for a subgroup of the whole study population but cannot be predicted for each individual participant at enrollment. This predictable deterioration in the cohort (but not in individuals) is a key reason for randomizing palliative care studies; without randomization, there is a risk of underestimating the benefit of an intervention because participants are generally getting sicker irrespective of the intervention.

It is plausible clinically to exclude from study analyses only participants whose withdrawal definitely was unrelated to the intervention.
creating a very conservative framework for this proposed analysis plan. Functional status such as the Australian Modified Karnofsky Performance Scale\textsuperscript{15} or the Eastern Cooperative Oncology Group Performance Scale\textsuperscript{16} need to be routinely collected in all hospice/palliative care Phase III studies to help inform the IDMC that there was unequivocal evidence of disease progression. With longitudinal data collected to describe global clinical deterioration in the participating population, study withdrawal can more accurately be attributed to the disease rather than the intervention in the subgroup who withdraw before the primary census point.

Issues of the influence of study withdrawal on study findings have previously been explored in the setting where the intervention may be related to deaths of participants, but the scenario under consideration here is only where withdrawal or death are categorically not related to the intervention.\textsuperscript{17} Importantly, Rubin makes the point that evaluation of an intervention is limited by withdrawals, and models need to account for this withdrawal.

Excluding any subjects from analysis on the basis of post-randomization events will be unpalatable to many scientists, as the subjects remaining in the analysis are only a subgroup of all those initially randomized. Because of this, the proposed approach does not claim that it will yield an unbiased estimate of treatment effect. It does, however, remove a known source of potentially large bias so that the resulting estimate is likely to be nearer to the true treatment effect than if the missing outcomes had been imputed mathematically or attributed simply to treatment failure.

Systematically underestimating the benefits of an intervention is unethical, especially given the time and energy that people at the end of life invest by participating in clinical trials. This potentially large source of bias in emerging evidence is scientifically unsound and poses a serious patient safety concern, given the current focus on refining the evidence base for providing better quality care at the end of life. The proposed approach allows a better balance between Type I and Type II errors in clinical populations with high levels of expected withdrawal unrelated to the interventions being evaluated where such withdrawal cannot be predicted on an individual participant basis.

**Operationalizing the Solution**

It could be argued that, given the magnitude of withdrawals unrelated to the intervention in hospice/palliative care Phase III studies, this approach should be built into every study protocol. As an absolute minimum, this approach should be a sensitivity analysis that includes a comparison of all available primary and secondary outcomes of the groups who withdrew from each arm with the remaining participants. Although comparisons of the groups who withdraw cannot categorically eliminate bias associated with withdrawal, such an analysis can indicate when withdrawal does not appear to be random. Practically, clinical researchers must build the costs of an additional IDMC review into study budgets; in the long run, studies should be less expensive because sample size goals can be smaller. The palliative-modified ITT analysis allows a systematically less biased approach to evaluating the effects of interventions being evaluated in Phase III hospice/palliative care trials.

**Disclosures and Acknowledgments**

The authors have no personal or financial conflicts of interest to disclose in relation to this work. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Dr. Abernethy has research funding from the U.S. National Institutes of Health, U.S. Agency for Healthcare Research and Quality, Robert Wood Johnson Foundation, Pfizer, Eli Lilly, Bristol Meyers Squibb, Helsinn Therapeutics, Amgen, Kanglaite, Alexion, Biovex, DARA Therapeutics, Novartis, and Mi-Co; these funds are all distributed to Duke University Medical Center to support research. In the last two years, she has had nominal consulting agreements (less than $10,000) with Helsinn Therapeutics, Amgen, and Novartis.

**References**
