Published in final edited form as:

Semin Oncol Nurs. 2014 February ; 30(1): 74–79. doi:10.1016/j.soncn.2013.12.011.

Methodological Considerations in the Design and Implementation of Clinical Trials

Constance T. Cirrincione, MS [Senior Biostatistician],

Alliance for Clinical Trials in Oncology, Statistics and Data Center, Duke Cancer Institute, Durham, NC

Ellen M. Lavoie Smith, PhD, APN-BC, AOCN® [Assistant Professor], and

University of Michigan School of Nursing, Ann Arbor, MI

Herbert Pang, PhD [Assistant Professor]

Alliance for Clinical Trials in Oncology, Statistics and Data Center, Duke Cancer Institute, Durham, NC

Abstract

Objectives—To review study design issues related to clinical trials that have been led by oncology nurses, with special attention to those conducted within the cooperative group setting; to emphasize the importance of the statistician's role in the process of clinical trials.

Data sources—Studies available in clinicaltrials.gov using experimental designs that have been published in peer-reviewed journals; cooperative group trials are highlighted.

Conclusion—The clinical trial is a primary means to test intervention efficacy. A properly designed and powered study with clear and measurable objectives is as important as the intervention itself.

Implications for nursing practice—Collaboration among the study team members, including the statistician, is central in developing and conducting appropriately designed studies. For optimal results, collaboration is an on-going process that should begin early on.

Keywords

Clinical Trials; Study Designs; Cooperative Groups; Nurse Scientist

The ultimate goal of research, especially via clinical trials, is to improve clinical practice. Over time with careful adherence to the scientific process, an experimental intervention, possibly a drug, device or activity, becomes the new standard of care, only to be replaced subsequently by another once-experimental intervention. So the process repeats itself and science moves forward.

In all research, the investigator formulates a study question, which determines the design of the study. The investigator begins with a systematic review of the literature to identify gaps

© 2013 Elsevier Inc. All rights reserved.

Address correspondence to: Constance T. Cirrincione, MS, Duke Cancer Institute, Hock Plaza, 2424 Erwin Road, Suite 802, Durham, NC 27705. connie.cirrincione@duke.edu phone: 919-681-5404.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

in the science that justify conducting a new research study. If the experimental intervention is an agent or previously untried combination of agents, the first step beyond testing the drug in animals is to determine a tolerated dosage based upon observed side effects. This process of discovery takes place in the framework of a phase I trial. A next step is to document if the drug ‘works’ by testing it at the dosage level found tolerable in the previous (phase I) step. This is typically accomplished through a phase II study which determines whether or not the experimental intervention performs at some pre-determined criterion level that warrants additional investigation. Often phase II trials allow stopping the trial prior to full accrual if there is evidence early on that the new intervention is performing too far below or above the criterion level. In both cases, the intent of early termination is to protect patient safety, conserve resources and, if warranted, bring the experimental intervention to higher-level testing as quickly as is feasible. Prior to embarking on a full-fledged comparative trial, a pilot trial is sometimes undertaken to work out any kinks before expending resources on the larger study. The goal of the next step is to determine whether the new intervention is an improvement over the standard of care (control), which may be an intervention, no intervention at all, or placebo. This is usually accomplished in a phase III trial in which patients are randomly assigned to the standard of care or one or more experimental interventions.

There are several steps to ensure that all types of studies are conducted in a rigorous scientifically sound manner and to ensure that conclusions are appropriately drawn from the study results. This article discusses the role of the statistician in this process. Table 1 lists examples of several nurse-led published studies.[1–11] We will refer to these throughout the discussion.

DESIGNING THE STUDY

Primary outcome or endpoint

The investigator begins the research process by posing at least one question of interest. Is the purpose to describe a clinical problem? To compare the effects of different interventions? These answers relate to the *study objective* and will drive the study design. For instance, if an investigator wishes to compile and synthesize information from published studies, the task is descriptive and may require no formal comparisons using statistical testing. On the other hand, an investigator may wish to determine if an intervention, such as an agent, device, treatment or practice ‘works’, or shows efficacy. As part of the study team, the statistician ensures well-focused and precise study objective(s). Statistical input during study development is well worth the time and effort to ensure a robust study with valid conclusions.

The next task is to identify both the relevant participant population to which the intervention applies and the evidence needed to conclude that the intervention is efficacious. While the clinical researcher knows the patient population he/she is interested in studying, the statistician is helpful in determining whether a comparison group is needed and how to take into account baseline patient characteristics that may potentially influence study outcomes. Determining if the intervention shows efficacy is closely tied to the study endpoint. The statistician translates the clinical impression of what ‘works’ into measurable terms. The study by Smith et al sought to determine whether a drug, duloxetine, reduces pain from chemo-induced neuropathy [3]. Pain level was measured before and after the intervention (duloxetine). Evidence of duloxetine efficacy was a decrease in pain after receiving study drug. Thus, the change in pain from before to after the intervention was the outcome of interest or the *primary endpoint*. The statistician ensures that the endpoint is unambiguous, objective, and measurable. How will the endpoint be objectively measured? Smith and colleagues measured pain as a score on the Brief Pain Inventory-Short Form, an objective

and validated instrument. Other studies listed in Table 1 share the strength of clearly defined study endpoints.

Statisticians also assist researchers in selecting the most appropriate measurement of the primary endpoint. In the Smith study, because the interest was in the *amount* of change in pain due to duloxetine, the endpoint was measured using a continuous scale. Had the interest been in whether or not a patient had any decrease in pain from pre-treatment, the endpoint would have been dichotomous (yes/no). This illustrates that the dichotomous measure is less precise. The precision of measurement has important implications in determining the appropriate statistical analysis to be applied, and even the required number of participants for the study. Therefore, statistical collaboration is critical during this development phase.

Intervention: Comparison with standard of care

Whether the intervention is efficacious implies that it works better than something else. Typically, the comparison is to the standard of care or control, often called usual care. Sometimes the comparison is to a placebo. Table 1 provides examples of experimental and control interventions used in several cooperative group nurse-led studies. As one example, with no standard treatment for chemotherapy-induced peripheral neuropathy, the Smith study used placebo as the control intervention. The hypothesis was that the study drug duloxetine would be associated with a larger decrease in pain than placebo.

Effect size

The investigator should quantify how large a difference between the intervention groups needs to be in order to conclude that the new intervention is effective. This difference is the *effect size*. As an example, the literature may document that a standard intervention has a 20% success rate. Investigators will consider a new intervention effective only if its success rate is at least 35%. This difference of 15% is the effect size. The amount of change considered acceptable for efficacy varies from study to study. For instance, in some cases a 10% improvement over the standard might be considered effective, while other situations might warrant a 50% improvement. Cossette and colleagues [4] conducted a randomized study to determine whether a nursing intervention focused on individual acute coronary syndromes and whether patients' perceptions of their disease and treatment would increase rehabilitation enrollment. They considered a particular intervention successful if its enrollment rate was twice that of the standard program. The enrollment rate for the current program was 15% and the new intervention would be considered effective if it enrolled at least 30% [4]. This doubling of enrollment is the effect size, the difference necessary for the experimental intervention to be considered effective.

Error probabilities

Because access to the entire target population is rare, it is important to carefully select a small number of participants from the population who are accessible. Those who participate comprise the study sample. When the sampling process is done properly, the study participants are similar to those in the target population and the study results can be generalized back to the initial target population. More specifically, suppose a study compared a new intervention with the standard one. Further, suppose results indicated that the new intervention is better than the standard. This finding discovered using a small sample may – or may not – be true for the entire target population and will not be known for certain without testing the entire population. Now suppose that data from the target population were available. Results from the study sample may agree with those of the target population (that is, be correct), or may disagree with those of the target population (that is, be incorrect). It is preferable for the probability of agreement to be high, and the probability of disagreement to be low.

Expanding upon the above, suppose that in the target population the two interventions were not different in efficacy. Based upon the study (sample) results, it may be concluded: (a) *incorrectly* that the two interventions were different or (b) correctly that they were not different. *Incorrectly* concluding that the two interventions were different (a), that is, deciding that the experimental intervention was better than the standard, when in fact it was no better, is called a 'type 1 error'. The probability of making a type 1 error is alpha (α). The probability of correctly concluding that the two interventions were not different (b), that is, deciding that the experimental intervention was no better than the standard of care, is $1-\alpha$, or confidence.

On the other hand, suppose that in the population the two interventions were different in efficacy. Based upon the study (sample) results, it might conclude: (c) *incorrectly* that the two interventions were not different or (d) correctly that they were different. *Incorrectly* concluding that the two interventions were not different (c), that is, deciding that the experimental intervention was no better than the standard, when in fact it was better, is called a 'type 2 error'. The probability of making a type 2 error is beta (β). The probability of (d) correctly concluding that the two interventions were different (that is, deciding that the experimental intervention was better than the standard) is $1-\beta$, or power. Further, as the probability of making a type 1 error increase, the probability of making a type 2 error decreases, and vice versa. The statistician works with the investigator to identify error rates that are reasonable for the study. As evidenced in Table 1, investigators are typically comfortable with a type 1 error rate of up to 5% to incorrectly conclude difference [2–6,8,11], and with a type 2 error rate of up to 20% to incorrectly conclude no difference [1–11].

Sometimes higher or lower error rates are appropriate [7,9]. For instance, if an experimental intervention was known to have considerably worse toxicity than the standard of care, it would be important to minimize the probability of incorrectly concluding that the experimental intervention was superior to the standard of care (making a type 1 error). Such an incorrect conclusion could result in implementing the experimental intervention when in fact it not only was not better than the standard but was also more toxic. On the other hand, suppose there were no viable options to the standard regimen for a particular illness. It would be important to minimize the probability of concluding that the experimental regimen was no better than the standard, if in fact it was (making a type 2 error), thereby bypassing a new effective treatment. In such situations, the statistician would discuss the implications of minimizing the type 1 or 2 errors in designing the study.

Often in designing a study, the effect size, α (making a type 1 error) and β (making a type 2 error) are decided upon first based upon expert clinical and statistical input. Given that information, the statistician determines the number of participants needed to answer the study question in a statistically robust manner. Refer to Table 1 for the parameters used in various published studies. All these studies are considered well-designed.

Types of Designs

The study design is the approach used to answer the study question or the study blueprint. To determine if an intervention were different from the standard, one would compare the study endpoint between patients who did versus did not receive the intervention. Thus, there would be two groups of patients. This is the case of studies cited in Table 1. To ensure that participants in the two groups had similar baseline characteristics that might influence the outcome of the study, such as, age, gender or overall functional ability, randomization is typically used. This means that all participants have an equal chance of being assigned to either group. The goal is to avoid imbalance on important characteristics between groups. The study by Smith and colleagues [3] used two groups – Group A, which received

duloxetine followed by placebo each for six weeks and Group B, which received placebo followed by duloxetine each for six weeks. This is an example of a two-group randomization.

More complicated study questions require more complex designs. An extension of the simple design above includes more than two groups, such as two or more experimental interventions with a one placebo group as the comparator. If the interventions were randomly allocated, randomization need not be limited to equal assignment. Assignment may be weighted, such as with a 2:1 weighting to intervention versus placebo. An application of unequal assignment might occur when participants want more than a 50:50 chance of being assigned to a new and popular intervention. Thus, weighted randomization can encourage enrollment. A limitation of unequally compared to equally weighted randomization, however, is the need for more participants, assuming that all other study parameters are the same.

A study may have as its primary question the comparison of two groups. A variety of factors may make it practical for all participants to receive all study interventions, but in differing sequence. Participants would be randomized to receive any study intervention first and then 'switch over' to the opposite intervention. This arrangement has been known to aid accrual, and is especially appealing to participants in the case of placebo, by ensuring that all participants receive the study intervention. This crossover design was employed by Smith et al [3] and Watkins Bruner et al [5].

Sometimes it is not practical to randomize individual participants. Instead, randomization is to a unit other than the individual, such as a hospital floor, a medical institution, or a school. This procedure might be appropriate when financial considerations or labor force availability prevent participating units, such as schools or hospitals, from implementing all study interventions. Another rationale is to avoid the possibility of those randomized to one group having undue influence on those randomized to the other group, otherwise known as contamination. As an example, breast cancer patients with symptomatic lymphedema might be randomized to either a new lymphedema intervention or standard group with minimal or no intervention. Randomization by the individual could result in both intervention and standard care individuals being in the same waiting rooms and discussing their symptoms and treatments. Non-intervention patients may then request the lymphedema intervention. Therefore, minimizing the risk of patients influencing the treatment of other patients would be aided by having all patients in the same institution receive the same intervention. This design is discussed in detail in the article by Christie and colleagues [12].

In studies with two primary questions, the statistician can provide insight into whether the questions can be addressed concurrently. For instance, one question might compare two treatment interventions, experimental or standard, while another question might compare the length of the intervention, either longer or standard duration. A factorial design is an efficient configuration. One factor is intervention (experimental vs. standard) and the other is length (longer vs. standard). This gives four possibilities – experimental treatment given for longer duration, experimental treatment for standard duration, standard treatment for longer duration, and standard treatment for standard duration. This design requires fewer participants than would four separate groups and answers two questions concurrently. Thus, involving a statistician early in study development can save precious participant resources.

Studies interested in whether and how the study endpoint changes over time, as with quality of life, participant attitudes, laboratory values, and physical exam features use a longitudinal design which allows testing a time effect. In its simplest form, the interest is limited to the time factor. In a study with two or more groups, the design accommodates both the group

factor and the longitudinal time factor. Such a mixed factor design allows testing both differences over time regardless of intervention group, and differences between groups regardless of time [7]. Given this design, more complex statistical modeling can also test whether there is an interaction between time and group, in other words, if changes over time are different in one group compared to the other.

CONDUCTING THE STUDY

Uniformity is the hallmark for obtaining accurate data which in turn allows accurate conclusions. To this end, the intervention plan must be clearly described in a formal protocol document that allows uniform administration. Another consideration is who will measure and report data, and where data collection will occur. The goal is objective standardized measurement. The participant's intervention experience must be defined, and the data collected and coded consistently. Participant compliance should be monitored closely. Study results can be seriously impaired if participants undergo interventions at non-scheduled intervals or non-protocol doses, or participate in concurrent non-intervention activities or medications that compromise the reported study outcomes. When intervention occurs in an unsupervised setting, such as the participant's home, various techniques may be employed to encourage adherence. These techniques might involve phone reminders, or patient medication calendars or diaries to document adherence. Such techniques attempt to ensure that the intervention has been administered completely and uniformly.

Application to Cooperative Groups

The Alliance for Clinical Trials in Oncology is one of several cooperative groups. It boasts participation by more than 10,000 cancer specialists throughout the United States. In this setting, the importance of standardization and unambiguity is obvious. Alliance clinicians recruit patients to participate in clinical trials guided by a protocol document. This document must be written clearly to ensure consistency in interpretation. Medical and health care personnel must adhere to the protocol. Data must be collected consistently in terms of content, frequency, and timing as well as units of measurement. More than ever, the importance of communication is obvious, as investigators, practitioners, and statisticians may be at different sites.

CONCLUSION

The above sections detail various aspects of the research process. To achieve high-quality effective research, collaboration among *all* study team members is key. Members participate with equal voice and input according to their specialty. Early collaboration ensures that the team is aware of the broad aspects of the study and ensures that various viewpoints are addressed early on. Including the statistician early in study team discussions, even prior to planning, enables the statistician to assist in focusing the objectives, design, and data capture. Nor does the team collaboration end with designing the study. Continued communication is necessary in accrual monitoring, data accuracy, analysis, interpretation, and reporting results.

Acknowledgments

Dr. Smith's role as Co-Chair of the Alliance Symptom Intervention Committee was supported by the Alliance Chairman's Grant. The duloxetine study was supported by in part by grant CA31946 from the NCI Division of Cancer Prevention, the Alliance Chairman's Grant, and Elli Lilly (provided drug and placebo), and in part by NCI grant U10-CA033601 for the Alliance Statistics and Data Center.

References

1. Barton DL, Liu H, Dakhil SR, et al. Wisconsin Ginseng (*Panax quinquefolius*) to improve cancer-related fatigue: A randomized, double-blind trial, N07C2. *J Natl Cancer Inst.* 2013; 105(16):1230–8. [PubMed: 23853057]
2. Kelly KP, Wells DK, Chen L, et al. Caregiving demands and well-being in parents of children treated with outpatient or inpatient methotrexate infusion: A report from the Children's Oncology Group. *J Pediatr Hematol Oncol.* 2013 Apr 11. Epub.
3. Smith EM, Pang H, Cirrincione C, et al. Alliance for Clinical Trials in Oncology. Effect of duloxetine on pain, function, and quality of life among patients with chemotherapy-induced painful peripheral neuropathy: A randomized clinical trial. *JAMA.* 2013; 309(13):1359–67. [PubMed: 23549581]
4. Cossette S, Frasure-Smith N, Dupuis J, Juneau M, Guertin MC. Randomized controlled trial of tailored nursing interventions to improve cardiac rehabilitation enrollment. *Nurs Res.* 2012; 61(2): 111–20. [PubMed: 22282154]
5. Watkins Bruner D, James JL, Bryan CJ, et al. Randomized, double-blinded, placebo-controlled crossover trial of treating erectile dysfunction with sildenafil after radiotherapy and short-term androgen deprivation therapy: results of RTOG 0215. *J Sex Med.* 2011; 8(4):1228–38. [PubMed: 21235716]
6. Barton DL, Atherton PJ, Bauer BA, et al. The use of *Valeriana officinalis* (Valerian) in improving sleep in patients who are undergoing treatment for cancer: A phase III randomized, placebo-controlled, double-blind study (NCCTG Trial, N01C5). *J Support Oncol.* 2011; 9:24–31. [PubMed: 21399726]
7. Winters-Stone KM, Dobek J, Nail L, et al. Strength training stops bone loss and builds muscle in postmenopausal breast cancer survivors: a randomized, controlled trial. *Breast Cancer Res Treat.* 2011; 127(2):447–56. [PubMed: 21424279]
8. Barton DL, Wos EJ, Qin R, et al. A double-blind, placebo-controlled trial of a topical treatment for chemotherapy-induced peripheral neuropathy: NCCTG trial N06CA. *Support Care Cancer.* 2011; 19(6):833–41. [PubMed: 20496177]
9. Bakitas M, Lyons KD, Hegel MT, et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomized controlled trial. *JAMA.* 2009; 302(7):741–9. [PubMed: 19690306]
10. Ott CD, Twiss JJ, Waltman NL, Gross GJ, Lindsey AM. Challenges of recruitment of breast cancer survivors to a randomized clinical trial for osteoporosis prevention. *Cancer Nurs.* 2006; 29(1):21–31. quiz 32–3. [PubMed: 16557117]
11. Mock V, Frangakis C, Davidson NE, et al. Exercise manages fatigue during breast cancer treatment: a randomized controlled trial. *Psychooncology.* 2005 Jun; 14(6):464–77. [PubMed: 15484202]
12. Christie J, O'Halloran P, Stevenson M. Planning a cluster randomized controlled trial: methodological issues. *Nurs Res.* 2009 Mar-Apr;58(2):128–34. [PubMed: 19289934]

Table 1

Selected studies led by oncology nurses

Publication/Reference	Primary Endpoint	Intervention groups *	Design	Design parameters
Barton et al (1) #	Fatigue score	E=American ginseng C=Placebo	2-arm randomized	Alpha=Unspecified Power=90% N=300
Kelly et al (2) #	Caregiving demand score, well-being score	E=MTX infusion as outpatient C=MTX infusion as inpatient	2-arm randomized	Alpha=0.05 Power=80% N=140
Smith et al (3) #	Pain score	E=Duloxetine → placebo C=Placebo → duloxetine	2-arm randomized crossover	Alpha=0.05 Power=90% N=232
Cossette et al (4)	Enrollment in rehab program	E=Nurse-patient meeting before discharge C=Standard of care	2-arm randomized	Alpha=0.05 Power=80% N=242
Watkins Bruner et al (5) #	Erectile function score	E=Sildenafil C=Placebo	2-arm randomized crossover	Alpha=0.05 Power=90% N=332
Barton et al (6) #	Sleep quality score	E=Valerian C=Placebo	2-arm randomized	Alpha=0.05 Power=94% N=200
Winters-Stone et al (7)	Bone mineral density, bone-free lean mass, fat mass	E=Moderate-intensity resistance + impact training C=Low-intensity, non-weight-bearing stretching	2-arm randomized with mixed factors	Alpha=0.01 Power=81–99% N=106
Barton et al (8) #	Sensory score	E=BAK C=Placebo	2-arm randomized	Alpha=0.05 Power=80% N=128
Bakitas et al (9)	Quality of life, patient-reported quality of life, symptom intensity, and resource use	E=Nurse-led focused palliative care C=Standard oncology care	2-arm randomized	Alpha=0.01 Power=80% N=400
Ott et al (10)	Bone mineral density	E=Risedronate plus calcium with vitamin D + progressive strength & weight training C=Risedronate plus calcium with vitamin D alone	2-arm randomized	Alpha=Unspecified Power=85% N=218
Mock et al (11)	Fatigue level	E=Formal prescribed home-based exercise program C=Encouragement to exercise	2-arm randomized	Alpha=0.05 Power=80% N=120

* E=Experimental intervention

C=Control/standard of care/usual care

Cooperative group study.