

Principles of Clinical Trial Design

By Michelle Nottage and Lillian L. Siu

A FUNDAMENTAL TENET of ethical research is the requirement that the research makes a valuable contribution to medical knowledge. Clinical trials that do not contribute because of irrelevant hypotheses or unsound methodology are not ethical. The principles that apply to clinical trial design should also be used when evaluating reports in the literature (Table 1)

RESEARCH QUESTION

All trials should address a research question of importance. However, it is easier to perform a clinical trial comparing similar strategies that may exert only a small impact on practice than to perform a trial comparing very different strategies, even though results from the latter could produce large shifts in practice. Some research questions of fundamental importance have never been satisfactorily answered because of an inability to recruit patients or physicians. The following examples help to illustrate this point:

- *Should chemotherapy for metastatic, but asymptomatic, colorectal cancer be instituted at diagnosis or when symptoms develop?* A North American intergroup study designed to answer this question closed after 4.7 years because of poor accrual (67 of a planned sample size of 144 patients) (unpublished).
- *Should we screen for prostate cancer?* A National Cancer Institute trial aiming to enroll 74,000 men began to randomize patients in November 1993.¹ This trial is still open to accrual although it was estimated that accrual would be complete in 3 years.
- *Should patients with metastatic renal cell carcinoma undergo nephrectomy?* Results of this trial were published recently after accrual of 246 patients from 80 institutions over a period of 7.3 years.²

These difficulties are in marked contrast to a group of four trials comparing single-agent cisplatin therapy with cisplatin combination chemotherapy in non-small-cell lung cancer. These trials managed to accrue more than 400

patients each in approximately 16 months, despite median survival times of between 6 and 9 months in all arms.³⁻⁶ Many oncologists would agree that the regimen selected in the control arm (single-agent cisplatin) is not one that is in wide clinical use. More importantly, the lack of quality-of-life assessments in these trials renders their question of palliation in advanced lung cancer patients unanswered.

Once established, the research question should be put into a testable form—the hypothesis. The statement of the hypothesis should be explicit as to the dependent (outcome) variable(s) and the independent variables. One usually attempts to disprove the null hypothesis, as it is easier to refute a hypothesis than to prove it is true. In single-arm studies, the null/alternate hypotheses will be framed as “the response rate of tumor x to treatment y is $\leq/\geq z\%$.”

STUDY DESIGN

Internal validity is the degree to which study results reflect the true situation. Internal validity is threatened by systematic error or bias and by poor precision. To maintain high internal validity, the investigator must appropriately select subjects, ensure valid and reliable measurements of suitable end points, and then use appropriate methods of analysis.

Formulation of the research question into a hypothesis has two effects: it requires the selection of a single end point, and it will guide choice of an appropriate study design. Improvement in survival is the ultimate goal of much oncology research, so use of survival as an end point in comparative studies is usually valid. However, large improvements in survival are rare in the treatment of advanced solid cancers. Hence, improving quality of life in patients may be a more realistic goal. Surrogate end points (such as relapse-free survival, time to progression) refer to events that are believed to be precursors of the ultimate outcome (especially survival). Thus a surrogate end point will have the advantage of requiring shorter follow-up or fewer patients. However, surrogate end points should be used with caution if their predictive ability for the true end point is not certain. Response rate is not a surrogate for survival or quality of life. Cause and effect is not established by the association of response with prognosis; response may simply be a predictive factor for patients destined to fare well regardless of therapy.⁷ A toxic treatment may yield high response rates at the expense of quality of life.

From the Princess Margaret Hospital, University Health Network, 610 University Ave, Toronto, Ontario M5G 2M9, Canada.

Address reprint requests to Lillian L. Siu, MD, FRCPC, Princess Margaret Hospital, University Health Network, 610 University Ave, Toronto, Ontario M5G 2M9, Canada; email: lillian.siu@uhn.on.ca.

© 2002 by American Society of Clinical Oncology.
0732-183X/02/2018s-42s/\$20.00

Table 1. Evaluation of Clinical Trials

-
- Does the trial address an important question?
 - Is the design of the study appropriate?
 - Are the end points appropriate and well defined?
 - What is the probability that the study is false positive or false negative?
 - Does the report reflect the results and is the study generalizable to influence oncology practice?
-

Within the selected design, various steps should be taken to minimize the possibility of an error. Two major errors can occur in hypothesis testing: rejecting the null hypothesis when it is true (false-positive, type I/alpha error) and accepting the null hypothesis when it is false (false-negative, type II/beta error). By convention, an acceptable risk of a type I error is set at 5% ($P = .05$). Therefore, if multiple comparisons of equivalent outcomes are made, some statistically significant positive results will arise by chance alone. Whether to adjust P values for multiple comparisons is controversial, as is the best method of doing so. Clinicians must be especially aware of this problem when analyzing data from subgroups, data using multiple end points, and repeated, unplanned examinations of data after different time intervals. The possibility of a chance positive result should be highlighted if one result among many is significant. An alternative method for analysis of subgroups is statistical testing for an interaction among the intervention, the subgroup characteristic, and the treatment outcome. Positive results in an unplanned subgroup analysis are best regarded as hypothesis generating. If a differential response based on background characteristics is expected, this should be accounted for in the design phase with an appropriate sample size.

The most common source of a type II error is insufficient sample size. Power ($1 - \beta$) is a theoretical property that is used to determine the sample size required for an acceptable (often 10% or 20%) risk of type II error. The other determinant of sample size is the size of the difference between treatment arms one hopes to detect. In oncology, it is rare that an absolute difference in survival exceeds 10%. It is important to note that just because a trial is not powered to detect a small difference does not mean it cannot detect this difference, only that its chance of doing so is small.

Phase I and II trials can only generate hypotheses about the value of new strategies; they cannot prove benefit. A phase I or II trial should only be undertaken if it will lay groundwork for a phase III trial or evaluate mechanisms in translational trials. Comparison of results between nonrandomized studies or historical controls is of limited value for several reasons, including selection bias and stage migration. These factors are illustrated below by the story of

high-dose chemotherapy and autologous stem-cell transplantation for metastatic breast cancer. Selection bias occurs when patients recruited for a trial differ from the source population. It often favors the adoption of a new treatment by preferential inclusion of patients with favorable prognoses (eg, young age, good performance status). When evaluating results, one should assess whether the patients in the trial are representative of the typical patient and one should ensure that all registered patients are accounted for in the analysis. Stage migration is the phenomenon in which patients with low tumor bulk (and favorable prognosis) are upstaged by improved staging techniques, with no change in the staging definitions. This leads to an apparent improvement in survival for each stage even though no actual change in outcome has occurred. Requirement for more rigorous staging for entry onto a clinical trial than in routine practice has the potential to cause stage migration. Other examples include prostate-specific antigen testing, extensive pathologic survey for axillary micrometastases in breast cancer, and positron emission tomography scanning in lung cancer. The extent to which stage migration may affect outcomes was observed in an analysis of patients referred for consideration of the National Cancer Institute of Canada Clinical Trials Group MA.5 trial, a randomized study of the addition of high-dose therapy to standard adjuvant treatment in women with very high-risk breast cancer. Patients underwent the usual staging procedures before embarking on adjuvant therapy (bone scan, abdominal ultrasound, and chest x-ray). Once referred to the trial center for consideration of study entry, consenting patients also underwent computed tomography scanning of the head, chest, abdomen, and pelvis and bilateral bone marrow biopsies. Nearly one quarter of those referred were found to have unrecognized metastatic disease as a result of the trial staging and were thus excluded from participation.⁸ Selection bias likely accounted for promising results seen in three early trials of high-dose chemotherapy.⁹⁻¹¹ These trials reported high response rates, with higher than expected complete response rates and some durable remissions. Further phase II trials ensued investigating variants of the procedure. Selection criteria included the fittest and youngest patients, with reasonable justification as these patients were to undergo an intense treatment. The median age of patients on the trials referred to above was between 37 and 41 years, clearly different from the wider breast cancer population. Many trials then further selected patients with a favorable prognosis by transplanting only those patients who responded to induction therapy. Re-analysis of over 1,500 patients enrolled onto chemotherapy trials for metastatic breast cancer at the M.D. Anderson Cancer Center revealed that meeting eligibility criteria for high-dose ther-

apy trials predicted for improved response and survival. At 5 years, 21% of high-dose candidates remained alive compared with 6% of those who were not candidates.¹² Although randomized trials were initiated, the widespread uptake of this procedure on the basis of the early results slowed accrual. In the United States, a randomized trial required the recruitment of the cooperative groups North Central Cancer Treatment Group, Eastern Cooperative Oncology Group, and Southwest Oncology Group and modification of the study design to allow a lower target before completing accrual over 7 years between 1990 and 1997.¹³ Results from this and other randomized trials did not reveal any advantage for high-dose therapy.

To further limit comparison of nonrandomized groups, patients participating in clinical trials seem to have improved survival when compared with nonstudy patients, even when the intervention is shown to not be of value and disease extent is taken into account.^{14,15} Very recent data using nearly 19,000 patients treated at the M.D. Anderson Cancer Center revealed that patients with metastatic disease had better survival if treated on a clinical trial protocol of any sort (phase I to phase III) compared with patients not treated on a trial protocol. The survival benefit was not seen for curative therapies in patients with locoregional disease.¹⁶

PHASE I

There are several objectives in performing a phase I (or safety and dose-finding trial). The primary outcome is usually the determination of a dose for future study (the maximum-tolerated dose or the recommended phase II dose). This is done by the continuous assessment of toxicity of the study agent(s). During the phase I trial, it is common to find out more about the new drug with pharmacologic evaluation (this may be required if the agent has never been tested in human subjects) and biologic correlation. Evidence of efficacy is not the major goal of such trials, but this information is usually collected. Eligible patients are usually those who have no other reasonable therapeutic option.

PHASE II

The primary goal of phase II trials is to establish sufficient efficacy to warrant further testing. The usual end point of phase II trials in oncology is tumor response rate. It should be remembered that although response is seen as an objective measure, it is subject to measurement error, and criteria for response vary. Moreover, it is not a measure of patient benefit. Further biologic correlative studies are often performed in a phase II setting, particularly in the case of newer, molecular-targeted agents. The other major goal of phase II studies is more accurate definition of toxicity

profile than is possible in the phase I setting. Selection criteria may be more restrictive in early phase trials compared with phase III trials. It is common to limit the amount of pretreatment patients may have received so as not to obscure any activity by inclusion of patients very unlikely to respond to any form of therapy.

PHASE III

Practice-changing results will most commonly derive from a randomized, blinded, and controlled trial. Randomization is essential to ensure uniformity between the treatment groups, such that the only difference is the treatment assignment. This will minimize both known and unknown sources of bias. Blinding, although often not feasible in oncology trials, has the effect of minimizing bias in the ascertainment of outcomes. The aim of phase III trials is to establish patient benefit. As such, the two most important end points are survival and quality of life. Survival is clearly an objective measure, but it has the disadvantages of requiring prolonged follow-up and large numbers of patients if deaths are few. Most physicians are familiar with the concept of high false-positive rates occurring in situations where there is a low pretest probability of an outcome. Screening studies for colorectal cancer using fecal occult blood illustrate this point. In the general population, the prevalence of colorectal cancer is low. Thus, in one study, more than half of the positive tests were false positives.¹⁷ This principle also applies to cancer trials, as marked improvements in survival are rare in the management of common metastatic tumors in adults. Hence, a single positive clinical trial has a definite possibility of being falsely positive, and it generally requires confirmation.

Quality-of-life end points are often seen as being subjective, but in fact validated instruments now exist. Improvement of quality of life is a realistic and achievable end point. Important elements in the appropriate use of quality-of-life end points include a patient-based tool, definition of a single measure as the primary end point, generation of a hypothesis about a clinically important change in the primary measure, and blind assessment where feasible. Phase III trials should be designed so that treating physicians will be able to generalize results to their practice. Thus, selection criteria should be as simple and inclusive as possible.

TRIAL CONDUCT

Failure to conduct a study in a rigorous fashion will threaten the internal validity of the results, both by allowing bias and imprecision. Strict attention should be paid to the method of enrolling and staging patients, so as not to introduce selection bias. All entered patients should be included in the reporting. Follow-up methods should be ar-

ranged so that data are not skewed by more complete or accurate follow-up in one group. A detailed description of the intervention is needed, so that it can be reproduced by others. If the trial is randomized or blinded, the investigator should ensure that appropriate methods have been selected and are adequately concealed from physicians and patients. Tools used to measure outcomes should be the most precise available. Toxicity data are vital in oncology trials and should be both precise and detailed. Means to maintain and assess compliance should be included in the study design, especially with oral therapy, or if prolonged follow-up is required.

Strict attention to the design and conduct of a trial should ensure that it is ethical in that it will contribute to medical knowledge. One also needs to obtain the consent of study participants. The three essential elements of consent are disclosure of information, understanding of the study nature, and voluntary participation. The informed consent document aids the process of informed consent by making these elements explicit and providing a record of the discussion. According to National Institute of Health guidelines, the following basic elements should be included in a consent document: description of the study, description of risks and benefits, presentation of alternative management strategies, confirmation of patient confidentiality and freedom to withdraw, mention of costs and compensation, and signatures.

EXTERNAL VALIDITY

External validity is the property of being able to generalize results from a clinical trial to one's own practice. Trials that have numerous selection criteria are

less likely to be applicable to the general oncology population. This may be particularly the case for performance status and age. It is well recognized that the age distribution of cancer patients and cancer patients in clinical trials are different, with older patients less likely to be represented in clinical trials. One must also be able to incorporate new results into the body of evidence that already exists. A single positive trial at odds with previously published results may simply reflect a false-positive result. Physicians should recognize that there is a bias in favor of publication of positive results and against negative results, so that the published literature may not accurately reflect the true situation. A recent presentation demonstrated that the probability of publication at 3 years after presentation at an American Society of Clinical Oncology meeting was 52% for positive trials and 39% for negative trials. The principle reason behind failure to publish was related to investigator preference.¹⁸

In conclusion, the conduction of clinical trials remains the most useful method of addressing important questions in oncologic practice. Patients and resources are precious, and as such, clinical trials utilizing them in the research setting must be methodologically and ethically sound. Thoughtful formulation of the research question, selection of an appropriate trial design and study end point, assurance of data quality, valid statistical evaluations and analyses, and judicious reporting will enable dissemination of valuable information to the oncology literature.

REFERENCES

- Gohagan JK, Prorok PC, Kramer BS, et al: Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial of the National Cancer Institute. *J Urol* 152:1905-1909, 1994
- Flanigan RC, Salmon SE, Blumenstein BA, et al: Nephrectomy followed by interferon alfa-2b compared with interferon Alfa-2b alone for metastatic renal-cell cancer. *N Engl J Med* 345:1655-1659, 2001
- Sandler A, Nemunaitis J, Denham C, et al: Phase III trial of gemcitabine plus cisplatin versus cisplatin alone in patients with locally advanced or metastatic non-small-cell lung cancer. *J Clin Oncol* 18:122-130, 2000
- Gatzemeier U, von Pawel J, Gottfried M, et al: Phase III comparative study of high-dose cisplatin versus a combination of paclitaxel and cisplatin in patients with advanced non-small-cell lung cancer. *J Clin Oncol* 18:3390-3399, 2000
- von Pawel J, von Roemeling R, Gatzemeier U, et al: Tirapazamine plus cisplatin versus cisplatin in advanced non-small-cell lung cancer: A report of the International CATAPULT I Study Group. *J Clin Oncol* 18:1351-1359, 2000
- Wozniak A, Crowley J, Balcerzak S, et al: Randomized trial comparing cisplatin with cisplatin plus vinorelbine in the treatment of advanced non-small-cell lung cancer: A Southwest Oncology Group study. *J Clin Oncol* 16:2459-2465, 1998
- Anderson JR, Cain KC, Gelber RD: Analysis of survival by tumor response. *J Clin Oncol* 1:710-719, 1983
- Crump M, Goss PE, Prince M, et al: Outcome of extensive evaluation before adjuvant therapy in women with breast cancer and 10 or more positive axillary lymph nodes. *J Clin Oncol* 14:66-69, 1996
- Antman K, Ayash L, Wheeler C, et al: A phase II study of high-dose cyclophosphamide, thiotepa, and carboplatin with autologous marrow support in women with measurable advanced breast cancer responding to standard-dose therapy. *J Clin Oncol* 10:102-110, 1992
- Kennedy M, Beveridge R, Rowley S, et al: High-dose chemotherapy with reinfusion of purged autologous bone marrow following dose-intense induction as initial therapy for metastatic breast cancer. *J Natl Cancer Inst* 83:920-926, 1991
- Peters W, Shpall E, Jones R, et al: High-dose combination alkylating agents with bone marrow support as initial treatment for metastatic breast cancer. *J Clin Oncol* 6:1368-1376, 1988

12. Rahman ZU, Frye DK, Buzdar AU, et al: Impact of selection process on response rate and long-term survival of potential high-dose chemotherapy candidates treated with standard-dose doxorubicin-containing chemotherapy in patients with metastatic breast cancer. *J Clin Oncol* 15:3171-3177, 1997
13. Stadtmauer E, O'Neill A, Goldstein L, et al: Conventional-dose chemotherapy compared with high-dose chemotherapy plus autologous hematopoietic stem-cell transplantation for metastatic breast cancer: Philadelphia Bone Marrow Transplant Group. *N Engl J Med* 342:1069-1076, 2000
14. Davis S, Wright P, Schulman SF, et al: Participants in prospective, randomized clinical trials for resected non-small cell lung cancer have improved survival compared with nonparticipants in such trials. *Cancer* 56:1710-1718, 1985
15. Antman K, Amato D, Wood W, et al: Selection bias in clinical trials. *J Clin Oncol* 3:1142-1147, 1985
16. Elting L, Cooksley C: Survival benefits of clinical trials compared to conventional therapy for cancer. *Proc Am Soc Clin Oncol* 21:248a, 2002 (abstr 988)
17. Hardcastle JD, Chamberlain JO, Robinson MH, et al: Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 348:1472-1477, 1996
18. Krzyzanowska MK, Pintilie M, Tannock I: Burying of unwanted results: A survey of more than 500 large randomized clinical trials presented at ASCO meetings to determine the probability and causes of failure to publish. *Proc Am Soc Clin Oncol* 21:244a, 2002 (abstr 973)