Support Care Cancer (1998) 6:261-265

REVIEW ARTICLE

Gary R. Morrow
Enzo Ballatori
Susan Groshen
Ian Olver

Statistical considerations in the design, conduct and analyses of antiemetic clinical trials

An emerging consensus

Presented in part at the MASCC Consensus Conference on Antiemetic Therapy, Perugia, 28–29 April 1997

G.R. Morrow, Ph.D., M.S. (✉)
University of Rochester Cancer Center, Rochester, N.Y., USA

E. Ballatori, M.D.
University of L'Aquila, Department of Internal Medicine and Public Health, Medical Statistics Unit, L'Aquila, Italy

S. Groshen, Ph.D.
University of South California School of Medicine, Norris Comprehensive Cancer Center, Department of Preventive Medicine, Los Angeles, Calif., USA

I. Olver, M.D.
Royal Adelaide Hospital Cancer Centre, Adelaide, South Australia, Australia

Abstract Various aspects of trial design and planning for clinical testing of antiemetic therapies administered to cancer patients are considered. It is generally felt that a randomized double-blind parallel-arm design is the best. Ways of achieving adequate power of such studies are discussed briefly, as is the need for previous identification of primary and secondary end-points. Finally, summary recommendations are given.

Key words Cancer chemotherapy · Clinical trials · Statistics · Trial design

Introduction

A number of recent summaries have described desirable features of the planning, design, and analysis of antiemetic trials in patients being treated for cancer [1, 3, 5, 10, 17, 19, 20]. In phase III testing, a consensus has emerged that the optimum design is a randomized double-blind parallel arm design. The study must have adequate power to detect clinically meaningful differences. Stratification by previously identified prognostic factors could be considered. Adequately measured primary and secondary end-points need to be clearly identified in advance. These specific points are described more fully below.

Randomization

Randomization distributes known and unknown prognostic factors among the arms of a trial in an unbiased manner, so that the effects of these prognostic factors are averaged out in tests of statistical significance [21]. Randomization avoids the conscious or subconscious assignment of certain patients to particular arms of a study. Nonrandomized designs are unacceptable for phase III comparative antiemetic trials.

Double blind

Every effort should be made to double blind all comparative antiemetic trials, since both nausea and vomiting may be influenced by external stimuli and suggestion as well as the direct emetogenic effect of the chemotherapy, and be-

cause nausea is self-reported. However, it must be recognized that blinding may be difficult to achieve. Informed consent information that describes different side effects for different arms of the study may make blinding difficult. If the drugs being compared have differing side effects both patients and treating staff may be able to accurately ascertain one or more of the arms [23]. If subjective criteria are used, it is essential that maximum efforts be made to ensure that at least the patient is blinded to the treatment assignment.

Parallel design

A parallel design should be chosen for antiemetic studies. A crossover design is intuitively appealing and has been successfully used in a variety of fields. The crossover design is often advocated because it allows for patient preference, avoids interpatient variability, and requires a smaller sample size. In the context of antiemetic studies, however, the latter two assumptions are likely to be false [12]. Because the second course of treatment may be different from the first course, the within-patient variability may be no less than the between patient variability. And since, in a crossover antiemetic trial, it is advisable to test for period and carry-over effects, the planned sample size may be greater than with a parallel study.

A further challenge to study integrity in a crossover design is the high probability of a substantial patient loss between subsequent chemotherapy cycles. With a crossover design it is also not possible to evaluate the antiemetic efficacy over multiple cycles. Finally, there are also ethical and clinical concerns about changing a treatment that has been successful in the first period.

Stratification for prognostic factors

There is an emerging consensus that prognostic variables influence the outcome of antiemetic studies. Age, gender, prior long-term exposure to alcohol, and prior chemotherapy experience are generally considered to be strongly associated with nausea and vomiting. There is less consistent evidence for patient susceptibility to motion, patient anxiety, and setting of the chemotherapy [18]. There are probably further unexplored prognostic factors, such as emotional state, and as yet unexplored interactions among prognostic factors. Known and suspected factors should be prospectively assessed and evaluated as part of large antiemetic trials. Subgroup analysis based on these characteristics should be planned in order to further understand their potential role.

The aim of stratification prior to randomization is to assure a balance of known prognostic factors among experiential groups. When the sample size is sufficiently large, imbalance among experimental groups is unlikely. If interim analyses are planned, stratification prior to randomization should be more seriously considered because of the possibility that the trial may end early, which could result in a larger imbalance of characteristics among arms. However, if the sample size is sufficiently large and analyses are planned in advance, both stratification at the time of randomization and stratification at the time of analyses may be equally effective.

Table 1 Total number of patients needed for comparison of two antiemetic regimens (2-sided, P<0.05 Chi-square test with 80% power)

% Complete control with "standard"	5% Improvement	10% Improvement	15% Improvement
40%	3146	814	372
60%	3020	752	330
80%	1890	438	176

Sample size

Sample sizes should be calculated to achieve adequate power (at least 80%; 90% is preferable) to detect differences that would be clinically meaningful. Although a large improvement may be expected (or hoped for), the study should be designed with the more modest, but still clinically relevant difference in mind. With the improved control of acute emesis with 5-HT_3 receptor antagonists an improvement of 15% or smaller may be clinically important. This will serve to increase the required sample size.

Table 1 illustrates the interrelationships among the percent complete control with a standard or comparison antiemetic agent, the percent improvement desired to be detected and the sample size needed to detect the improvement with a two-sided Chi-square test at a $P<0.05$ level and 80% power. The number of patients increases as the size of improvement to be detected decreases.

Studies designed to show equivalence of regimens should have greater power: 90% or more is recommended [6, 16]. In reporting of results, the criteria used for sample size determination should be reported, that is the desired power and what was considered a clinically important difference.

Interim analyses

The timing of interim analyses should be preplanned so that the trial is not stopped early with a spuriously positive result. The probability of achieving $P=0.05$ by chance alone where no significant result actually exists can exceed 20% if interim analyses are performed every 6 months in a 4-year study [2]. Termination of a trial prior to the target sample size requires a *P*-value smaller than 0.05 [9, 11, 24].

Intent-to-treat analysis

If not all randomized patients are adequately treated and evaluable, an intent-to-treat analysis must be performed [12, 23]. An intent-to-treat analysis requires that all patients who were assigned to a regimen be included in the denominator of the estimates of efficacy of that regimen. If an "as-treated" analysis is also performed, the results of that analysis should be compared with the results of the intent-to-treat analysis [15]. All patients who are randomized should be accounted for [14].

Selection of end-points

End-points must be clearly defined before the trial begins. Nausea and vomiting should be separately evaluated, not only because of the dependence of nausea and vomiting on different physiopathological mechanisms, but also to distinguish any different efficacy of antiemetic therapies for nausea and vomiting [7, 18, 22]. While the occurrence and severity of nausea and vomiting are of primary clinical interest, symptom duration may provide important additional information.

An analysis of the association between nausea and vomiting is recommended to evaluate whether the presence of nausea can explain or account for the relationship between the different efficacy of the antiemetic treatments and the protection from vomiting.

Issues for the design of future trials

Methodological challenges remain in exploring unresolved issues, including: treatment of patients for whom initial antiemetic therapy has failed; control of delayed emesis; prevention of emesis on repeated courses of chemotherapy; tailoring therapies to subsets of patients (giving less therapy for those at lower risk of emesis and more therapy for those at greater risk); identifying effective regimens that can be delivered on an outpatient basis; combining drugs and altering schedules to further improve the control of emesis. Future trials will involve assessment beyond the initial 24 h and include several courses. Sample size calculations will need to consider the number of patients treated and observed at later courses as well as at the first course.

Studies of delayed emesis

Delayed nausea or vomiting should be described considering both (1) the day-by-day response, for example the severity from each day measured separately, in order to evaluate the pattern of the phenomenon, and (2) a summary measurement for the whole period, for example the maximum emesis observed on day 1 or during days 2–6, or the time until emesis is first observed. Both approaches will be useful in comparisons and evaluation of potential relationship(s) between delayed emesis and potentially prognostic factors.

A dependence of delayed emesis on acute emesis may contribute to differences in observed efficacy between acute and delayed side effects. The association between emesis during the first 24 h and the delayed emesis should be summarized by the type of antiemetic and the type of chemotherapy treatment given the patient [13]. Comparative trials should be specifically designed to evaluate delayed emesis. The treatment of acute emesis should be administered in a standard manner. One approach is to randomize patients prior to any treatment and perform an intent-to-treat analysis. This allows evaluation both of patients who experience little or no acute emesis prior to the experience of delayed and of those patients who experience severe emesis during the acute phase.

In studies where patients are randomized after the acute phase, stratification should be based on a patient's response during the acute phase. It is important to record information on those patients who drop out because of acute nausea/emesis. In the situation that all the patients receive the same prophylaxis against delayed emesis, some information regarding the different efficacy of the various regimens against delayed emesis can also be obtained.

Studies of multiple cycles of chemotherapy

The analysis of multiple cycles of chemotherapy is complicated, because of the dependence of each cycle on previous cycles. Furthermore, there maybe a potential overlapping of acute and delayed emesis (and, depending on time between cycles, a potential overlap between delayed side effects of one cycle and anticipatory side effects of the next [17]). Statistical models need to describe nausea/vomiting over multiple cycles, allowing for the effect of drop-outs on the observed treatment efficacy as well as a potential effect on prognostic factors [4]. In summarizing the results of a trial of multiple cycles, the following should be reported: the number of drop-outs at each cycle, the reasons for their exclusion, and a summary of the primary prognostic factors.

Alternate designs

Future studies will involve the comparison of combinations of two or more drugs, different schedules or routes of administration, or different doses. For these, factorial designs [8] will allow us to ask two or more questions simultaneously, and have two important features: fewer patients are required than in the sequence of separate trials, each asking a single question; and interactions (synergy or antagonism) between treatment factors can be identified. If the goal is to select the best regimen (i.e. the best combi-

nation of treatment factors), then methods for statistical selection can be employed [4] in the setting of the factorial design. With selection methods, the regimen that achieves the best observed response is selected for use or for further study, or the two (or three) regimens with the best observed responses are selected for further study. The number of patients required is determined by probability requirement for correct selection. Methods of statistical selection have the advantage of generally requiring fewer patients than formal hypothesis-testing procedures. However, the individual questions (of scheduling, of dosing, etc.) are not addressed specifically and the regimen selected is not "proven" to be best, as is the case in formal hypothesis testing. Two-stage and multiple-stage selection methods are also available, in which a number of regimens are eliminated at each stage, allowing more patients to be assigned to those regimens which appear more effective in the earlier stages. These can also be used in early phase testing of drugs, to identify the most biologically active doses for further testing.

Recommendations

Articles such as those briefly reviewed and discussions at the consensus conference lead us to suggest the following recommendations for the conduct of clinical trials evaluating antiemetic therapy.

1. A randomized, parallel arm double-blind study is the preferred design to compare the efficacy of two or more antiemetic therapies.
2. There are factors that may influence the likelihood of emesis following chemotherapy that must be part of the study design.

a. The prognostic factors of age, gender, long-term exposure to alcohol and prior chemotherapy experience have consistently been found to influence nausea and vomiting.
b. Other potential prognostic factors should be investigat ed prospectively as part of large antiemetic trials.

3. Complete response for vomiting and nausea should be the primary end-points of antiemetic trials, and these should be evaluated separately. The association between nausea and vomiting and its relationship with treatment should be analyzed.
4. Sample sizes should be calculated to achieve adequate power (at least 80%; 90% is preferable) to detect clinically meaningful differences.
5. Interim analyses should be preplanned so that a trial is not stopped early with a spuriously positive result.
6. An intent-to-treat analysis should be performed, and all randomized patients should be accounted for.
7. For studies of delayed emesis, the criterion used to define delayed emesis should be stated and the relationship of delayed nausea/vomiting and acute nausea/emesis should be allowed for in the analysis.
8. Clinical trials should be designed to measure both nausea and vomiting over multiple cycles of chemotherapy, and the analysis should take into consideration that cycles are not independent. When reporting results the number of patients who drop out at each cycle for each arm should be specified and the reasons for drop-out should be summarized.

References

1. Aapro M (1993) Methodological issues in antiemetic studies. Invest New Drugs 11:243–253
2. Armitage PW, McPherson K, Rowe BC (1969) Repeated significance tests on accumulating data. Journal of the Royal Statistical Society A 132:235–244
3. Ballatori E, Roila F, Del Favero A (1996) Methodology of antiemetic trials. In: Tonato M (ed) Antiemetics in the supportive care in cancer. Springer, Berlin Heidelberg New York, pp 35–47
4. Bechhofer RE, Santner TJ, Goldsman DM (1995) Design and analysis of experiments for statistical selection, screening, and multiple comparisons. Wiley, New York
5. Bergmann JF (1995) Méthodologie de l'évaluation des antiémétiques. Bull Cancer 82:1062–1066
6. Blackwelder WC (1982) Proving the null hypothesis in clinical trials. Control Clin Trials 3:345–353
7. Bonneterre J, Hecquet B, Adenis A, Fournier C, Pion JM, Demaille A (1991) How do patients and physicians decide which antiemetic is the best in a cross-over study? Proc ASCO 10:323
8. Byar DP, Piantadosi S (1985) Factorial designs for randomized clinical trials. Cancer Treat Rep 69:1055–1063
9. Geller NL (1987) Planned interim analysis and its role in cancer clinical trials. J Clin Oncol 5:1857–1490
10. Gralla RJ, Clark RA, Kris MG, Tyson LB (1991) Methodology in anti-emetic trials. Eur J Cancer 27 [Suppl 1]:S5–S8
11. Green SJ, Fleming TR, O'Fallon JR (1987) Policies for study monitoring and interim reporting of results. J Clin Oncol 5:477–1484
12. Groshen S (1992) Antiemetic study design: a discussion of Dr. Olver's paper. Br J Cancer 66 [Suppl XIX]:S35–S37
13. Italian Group for Antiemetic Research (1997) Delayed emesis induced by moderate emetogenic chemotherapy: do we need to treat all patients? Ann Oncol (in press)
14. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB (1991) Analysis of clinical trials by treatment actually received: is it really an option? Stat Med 10:1595–1605
15. Lewis JA, Machin D (1993) Intention to treat – who should use ITT? Br J Cancer 68:647–650
16. Makuch R, Simon R (1978) Sample size requirements for evaluating a conservative therapy. Cancer Treat Rep 62:1037–1040
17. Morrow GR (1992) Methodology and assessment in clinical anti-emetic research: a meta-analysis of outcome parameters. Br J Cancer 66 [Suppl XIX]:S38–S41

18. Morrow GR, Roscoe JA (1997) Anticipatory nausea and vomiting: models, mechanisms and management. In: Dicato M (ed) Medical Management of Cancer-Treatment Induced Emesis. Martin Dunitz, London, pp 149-166
19. Olver IN (1992) Antiemetic study design: desirable objectives, stratifications and analyses. Br J Cancer 66 [Suppl XIX]:S30–S34
20. Olver IN (1996) Antiemetic study methodology: recommendations for future studies. Oncology 53 [Suppl 1]:96–101
21. Olver IN, Simon RM, Aisner J (1986) Antiemetic studies: a methodological discussion. Cancer Treat Rep 70:555–563
22. Olver IN, Matthews JP, Bishop JF, Smith RA (1994) The roles of patient and observer assessments in anti-emetic trials. Eur J Cancer [A] 30:1223–1227
23. Seipp CA, Chang AE, Shiling DJ, Rosenberg SA (1980) In search of an effective antiemetic: a nursing staff participates in marijuana research. Cancer Nurs 21:271–276
24. Zelen M (1987) Early stopping, interim analyses, and monitoring committees: what are the tradeoffs? J Clin Oncol 5:1314–1315