Br. J. Cancer (1992), **66**, Suppl. XIX, S30–S34

Antiemetic study design: desirable objectives, stratifications and analyses

I.N. Olver

Royal Adelaide Hospital, North Terrace, Adelaide, South Australia, Australia 5000.

Summary Once the optimal dose and safety of an antiemetic has been established a randomised double blind parallel subjects design is recommended for phase III studies. Randomisation distributes the unknown prognostic variables so that their effects can be allowed for in tests for statistical significance. Stratification can equally distribute the known prognostic factors e.g. prior exposure to chemotherapy, strength of the emetic stimulus, age, sex and prior alcohol consumption. A cross-over design is often proposed because less patients would be required to achieve the same power as a parallel subjects study. The major problem with this design is in being able to estimate and allow for carry over effects or treatment period interactions which can interact with each other and the direct treatment effect. The study must be large enough to detect a statistically significant difference of clinical importance. Interim analyses should be preplanned and early termination should require a difference between the arms with a more significant P value than 0.05. Simple evaluation of efficacy could include quantitation of objective parameters and use of simple ordinal scales to record more subjective phenomena. In a parallel subjects design patients must assess their overall tolerance of therapy which balances the antiemetic's efficacy and toxicity.

The objective of antiemetics is to control the nausea and vomiting associated with cancer chemotherapy so that the patient tolerates the chemotherapy more easily. In assessing the efficacy of an antiemetic, the control of acute post chemotherapy emesis, delayed emesis and anticipatory emesis as well as nausea must be balanced against any adverse effects of the antiemetic.

This assessment is made more difficult by the number of prognostic factors which will impact on the major end point of controlling emesis. These relate to the characteristics of the patient population, the intensity of emesis expected from the chemotherapy delivered and the scheduling and dosage of the antiemetics (Olver *et al.*, 1986).

Firstly, an antiemetic must be given at its optimum dose and schedule. This means that new antiemetics should be formally evaluated, initially preclinically and then in phase I studies, to establish the optimum dose before randomised comparisons of efficacy with other antiemetics can occur.

The phase III study design incorporating a new antiemetic must allow for the above prognostic variables and be able to be analysed to achieve the clearest comparison of endpoints. Patient numbers should be calculated in advance to allow sufficient power to detect a clinically meaningful difference. The times for interim analyses should be predetermined.

Finally, it is an advantage if the parameters measured are as objective as possible to allow reproducibility. Only in this way can multi-institutional studies be considered. Uniform endpoints would allow more meaningful comparisons between studies.

Preclinical

Toxicological testing of new drugs is routinely performed in small and large laboratory animals but such models are also useful for gaining information on efficacy and dose response relationships. The ferret is often used because its vomiting mechanism closely parallels that of humans. For example it demonstrated the efficacy of the 5-hydroxytryptamine$_3$ receptor antagonists (Andrews *et al.*, 1987). Also the demonstration that a dose response relationship existed for metoclopramide in controlling cisplatinum induced emesis initially occurred in dogs.

We have developed a murine gastric distension model for emesis which demonstrated a dose response relationship for prochlorperazine which has also been reported clinically (Olver *et al.*, 1989). Information gained from animal models on efficacy and dosing means that clinical trials can be saved for only the most promising drugs and fewer dose levels should be required in phase I studies to identify the optimum dosing schedule.

Phase I studies

Adequate phase I evaluation of antiemetics is essential from two viewpoints. Firstly, an antiemetic must not have any severe toxicities since this would abrogate any antiemetic benefits and this must be determined before efficacy testing. Secondly, the resources expended on studies of efficacy will be wasted if the optimal dose of the drug has not been identified.

The newer antiemetics such as the 5-hydroxytryptamine$_3$ receptor antagonists have all been tested preclinically and then formal dose finding studies performed (Kris *et al.*, 1988). Older antiemetics though, such as metoclopramide and prochlorperazine have only recently been found to be more effective at higher than conventionally used doses yet had been used in clinical trials of efficacy at low doses for years (Gralla *et al.*, 1981; Carr *et al.*, 1987).

Phase II studies

The phase II study is a single arm study which tests the antiemetic drug for efficacy, using the dose determined from the phase I dose escalation study. As larger numbers of patients are treated, more is also learned about the toxicity profile of the drug. Based upon its efficacy in phase II studies, a decision is reached on whether the drug has sufficient activity to warrant testing it in a phase III comparison with the standard antiemetic therapy.

Study design issues

For phase III comparisons of antiemetics I recommend a randomised double blind parallel subjects design.

Randomisation and stratification

Randomisation distributes the unknown prognostic variables according to a known random distribution so that their effects can be allowed for in tests for statistical significance. It also avoids any bias in patient selection (Wendle, 1979).

Stratification, prior to randomisation, is important in non-cross over studies to ensure that the known prognostic factors are equally distributed before measuring the treatment related variables (Simon, 1989). Stratification is required at the time of analysis when interactions can be identified where the effect of a treatment differs for a subset. Subset analysis of a study can only be for hypothesis generation, but performed at the conclusion of the study there is the risk that with an uneven distribution there may be insufficient patient accrual to reach meaningful conclusions.

Factors to be considered for stratification include: prior exposure to chemotherapy, strength of the emetic stimulus, age, sex and alcohol consumption. The first two of these factors are of particular importance and often studies are designed which restrict accrual to patients receiving their initial doses of highly emetogenic chemotherapy such as cisplatinum. Patients who have previously received chemotherapy and have experienced severe nausea and vomiting tend to become more refractory to antiemetics in subsequent courses. Emesis possibly becomes a conditioned response. This is well documented when anticipatory nausea and vomiting occurs after patients have received highly emetogenic agents such as cisplatinum (Moher, 1984).

In comparing antiemetic drugs it is important that the strength of the emetic stimulus be equivalent for each drug. There are some chemotherapy drugs such as cisplatinum, mechlorethamine, streptozotocin and dacarbazine which consistently cause severe emesis, while others such as 5 flourouracil or the vinca alkaloids are associated with lesser emetic responses (Tortorice & O'Connell, 1990). The dose and schedule of the drugs will also affect their emetogenicity. For example, higher doses of cisplatinum cause more severe emesis than low doses (D'Olimpio *et al.*, 1985).

Age has been reported as a prognostic factor, with younger patients not responding as well to metoclopramide as older patients (Pollera & Giarnarelli, 1989). Tetrahydrocannabinol however had its greatest efficacy reported in younger patients (Sallan, 1975). The tolerance of the antiemetic also differs with age, with younger patients experiencing more extrapyramidal reactions with metoclopramide or prochlorperazine (Goslin & Garnick, 1981).

Females have been reported as not responding as well to metoclopramide and the same may be true for the new 5 hydroxytryptamine receptor antagonists, therefore the sex ratio should be balanced in each arm of a randomised study (Pollera & Giarnarelli, 1989).

Finally, control of chemotherapy-induced emesis has been reported as easier in patients with a history of chronic high alcohol intake (D'Acquisto *et al.*, 1986).

Double blind

The use of blinding is essential to ensure the elimination of any effects due to suggestion or expectation either from the patient or the investigator. This is particularly important in antiemetic studies where many of the endpoints are subjective and unintentional bias could occur. Effective binding can be difficult if an antiemetic has easily recognisable side effects, as demonstrated in a trial of tetrahydrocannabinol (Seipp *et al.*, 1980). In a randomised comparison, blinding can be maintained by giving both drugs in a similar manner. However, the use of inert placebos as a no treatment arm for patients receiving highly emetogenic anticancer drugs raises ethical dilemmas.

Crossover vs *parallel subject designs*

The two major trial designs considered for antiemetic studies are the simple two period cross over design where each patient receives both the study drug and control in a random order over successive courses, or the parallel subjects design where the patients are randomised to receive either the study drug or the control and the two groups compared. The major advantage claimed for the cross-over trial is that less patients and cost will be required to achieve the same power as a parallel subjects study. It is likely that the variability of measurements taken in the same subject is less than the variability of measurements taken in different subjects. Also the prognostic factors are controlled since the subject provides a comparison between two treatments and can indicate which course they prefer (Bakowski, 1984). It is also intuitively appealing in a crossover trial for patients to be able to compare both treatments and state their preference.

The major problem is the difficulty in interpretation because factors such as carry-over effects, where the effects of the first treatment influence the response to the second. There could be treatment period interactions, where the effect of a treatment may depend on the period in which it is administered. Interpretation of the study can be difficult because these effects can't be separated and they can interact with each other and the direct treatment effect. They can reduce the power of the study therefore taking away the advantage of the cross-over design.

If one could assume that these effects were negligible compared to the direct treatment effect then a 2×2 cross over design would be appropriate. Based on previous studies we cannot definitively make this assumption. For example, anticipatory emesis is an example of a carry over effect. Testing the assumption using between subject totals and if a significant effect is found using the first period data alone to estimate the treatment effect is questionable, since the power of the preliminary test is likely to be small given the size of most study populations and the first period comparison is biased because it is conditional upon there being a significant carry over effect (Jones & Kenward, 1989).

A further disadvantage to the cross-over design is the loss of patients between period one and period two. Alternative cross-over designs e.g. increasing the number of periods studied, may require more patients to be studied and would increase the chance that patients may be lost to the study by not completing like treatments.

The difficulty of interpreting the significance of the results of a crossover study make the simpler parallel subject design more attractive. Parallel subject design studies are easier to analyse and interpret and allow easier comparisons between studies. Also patients are rarely lost to study because the initial analysis of efficacy can occur after the first treatment. It is easier to study multiple courses with a parallel design study which more closely reflects the normal treatment situation. Asking patients in each arm for a statement of overall tolerance which balances the efficacy with the toxicity of the treatment and then comparing these global assessments between arms assesses the overall group preference which is the parallel subjects design endpoint that most closely equates with the crossover study patients' expression of individual preferences.

Sample size

The sample size of an entiemetic study must be planned so that there is sufficient statistical power to detect clinically important and realistic differences. Results from previous studies can be used to estimate the response rate for the control arm. For example, if the complete response rate on the control arm is expected to be 25% and one wishes to have a 80% power to detect a 20% difference using a 2-tailed test of significance at significance level (alpha) of 0.05 a total of 196 patients will be needed, assuming that approximately equal numbers of patients will be randomised to each arm (Table I).

The timing of interim analyses should be specified in planning the study but termination of the study before the target sample size is reached should only be done if the differences are much more extreme than a P value of 0.05. For example, the probability of obtaining a statistically significant ($P = 0.05$) result by chance alone can exceed the accepted 5% and would be greater than 20% if interim analyses were performed every 6 months in a 4 year study (Fleming *et al.*, 1984).

Evaluation of efficacy

There are no standard ways for expressing the efficacy of an antiemetic but to make results easier to compare simple quantitation of parameters should be attempted. The data to be collected should reflect the control of nausea and vomiting, the side effects of the antiemetics and an overall assessment by the patients of their tolerance of the therapy (Table II).

Table I Patient numbers required in a parallel design study

Expected CR rate on control arm (%)	*CR rate on study arm (%)*	*Total patients required (%)*[a]
25	30	2578
25	35	796
25	40	330
25	45	196
25	50	132
25	55	96

[a]Total number of patients required to have an 80% power to detect the relevant differences using a 2-tailed test of significance at significance level (alph) = 0.05. Assumes approximately equal numbers of patients will be randomised to each arm. CR = Complete response.

Table II Antiemetic assessment

Nausea
- Severity [4 point scale (none, mild, moderate, severe]
- Duration [hours]
- Record anticipatory and post treatment

Vomiting
- Number of episodes
- Record pretreatment then hourly

Other Parameters
- Record severity [4 point scale]
- Duration

- Anxiety
- Sedation
- Specific toxic effects
  - Amnesia (benzodiazepines)
  - Dystonic reactions (phenothiazines)
  - Dysphoric reactions (cannabinoids)

Patients' overall assessment
- 4 point scale [very well to very poorly]
- Record reason for choice

The number of vomiting episodes can be counted over a period of time. The occurrence of anticipatory vomiting should be recorded and the time interval over which the vomiting is assessed should be extended to include delayed emesis in cases such as following cisplatinum chemotherapy, where this is likely to occur (Navari, 1989).

For subjective parameters such as nausea or toxicities such as sedation I favour using simple ordinal scales. Large numbers of points or a continuous visual analogue scale do not necessarily increase the quantitative accuracy of the assessment since patients may only be able to discriminate between broad grades of a subjective sensation e.g. none, mild, moderate, severe. Whether scales should be odd or even numbered to avoid the possibility of selecting a neutral middle category is problematical (Presser & Schuman, 1980).

In a parallel subjects design study a measure of overall tolerance is essential to allow a patient to balance the toxicity of the antiemetic with its efficacy. This could again be recorded on a simple ordinal scale and the reason for the patient's decision recorded.

Patient vs *observer assessments*

Both patients and observers should assess treatment outcomes in the ideal study design. Observers may be better at recording objective parameters such as the number of vomiting episodes, particularly if the patient is sedated by the treatment. Subjective sensations can only be reported by the patient either directly or by telling the observer. There is often a good correlation between patient and observer assessments and any differences which do occur can often be attributed to the rigour of recording e.g. nurses recording hourly while patients fill out assessment forms at 24 h (Fetting, 1982).

Reporting results

In reporting the results the original sample size and power estimations should be recorded. If the trial is stopped before the accrual target is reached the reasons should be given. Reasons for exclusions from analysis should also be provided particularly in a cross-over study where patients may be lost between courses one and two. Complete response rates should be recorded but I favour just reporting the raw data for lesser responses to aid interpretation of the results. The use of confidence levels rather than P values in reporting results is more informative because it makes it clearer whether results are negative or indeterminate.

Conclusions

After adequate preclinical assessment, formal phase I studies are required to establish the optimal dose for studies of efficacy and to ensure that the drug has no serious toxicities. The suggested optimal study design for phase III comparisons with established antiemetics is a randomised double blind parallel subjects design. Major prognostic factors which must be balanced in the two arms include prior exposure to chemotherapy, strength of the emetic stimulus, age, sex and a history of heavy alcohol intake. Finally, observers and patients should record their results in a manner which would allow comparisons with other studies and simple quantitation of objective parameters and ordinal scales for subjective parameters are suggested.

I would like to acknowledge the contribution of Dr Jane Matthews, Head of the Statistical Centre, Peter MacCallum Cancer Institute, for her critical review of this manuscript.

References

ANDREWS, P.L.R., BAILEY, H.E., HAWTHORN, J., STABLES, R. & TYERS, M.B. (1987). GR38032F, a novel 5-HT_3 receptor antagonist, can abolish emesis induced by cyclophosphamide and radiation in the ferret. *Br. J. Pharmacol.*, **91** (Suppl.), 417.

BAKOWSKI, M.T. (1984). Advances in antiemetic therapy. *Cancer Treat. Rev.*, **11**, 237.

CARR, B.I., BLAYNEY, D.W., GOLDBERG, D.A., BRALY, P., METTER, G.E. & DOROSHOW, J.H. (1987). High doses of prochlorperazine for cisplatin-induced emesis A prospective, random, dose-response study. *Cancer*, **60**, 2165.

D'ACQUISTO, R.W., TYSON, L.B., GRALLA, R.J. & 4 others (1986). The influence of a chronic high alcohol intake on chemotherapy-induced nausea and vomiting. *Proc. Am. Soc. Clin. Oncol.*, **5**, 257.

D'OLIMPIO, J.T., CAMACHO, F., CHANDRA, P. & 4 others (1985). Antiemetic efficacy of high dose dexamethasone vs placebo in patients receiving cisplatin based chemotherapy: a randomized double blind controlled clinical trial. *J. Clin. Oncol.*, **3**, 1133.

FETTING, J.H., GROCHOW, L.B., FOLSTEIN, M.F., ETTINGER, D.S. & COLVIN, M. (1982). The course of nausea and vomiting after high-dose cyclophosphamide. *Cancer Treat. Rep.*, **66**, 1487.

FLEMING, T.R., GREEN, S.J. & HARRINGTON, D.P. (1984). Considerations for monitoring and evaluating treatment effects in clinical trials. *Controlled Clinical Trials*, **5**, 55.

JONES, B. & KENWARD, M.G. (1989). The 2 × 2 cross-over trial with continuous data. In *Design and Analysis of Cross-over Trials*. Jones, B. & Kenward, M.G. (eds) p. 16 Chapman and Hall: London.

GOSLIN, R.H. & GARNICK, M.B. (1981). Metoclopramide as an antiemetic in patients receiving cisplatin. In *The Treatment of Nausea and Vomiting Induced by Cancer Chemotherapy*, Poster D. (ed) p. 159 Masson: New York.

GRALLA, R.J., ITRI, L.M., PISKO, S.E. & 6 others (1981). Antiemetic efficacy of high-dose metoclopramide: randomized trials with placebo and prochlorperazine in patients with chemotherapy-induced nausea and vomiting. *N. Engl. J. Med.*, **305**, 905.

KRIS, M.G., GRALLA, R.J., CLARK, R.A. & TYSON, L.B. (1988). Dose ranging evaluation of the serotonin antagonist GR-C507/75(GR38032F) when used as an antiemetic in patients receiving anticancer chemotherapy. *J. Clin. Oncol.*, **6**, 659.

MOHER, D., ARTHUR, A.Z. & PETER, J.L. (1984). Anticipatory nauzea and/or vomiting in chemotherapy patients. *Cancer Treat. Rev.*, **11**, 257.

NAVARI, R.M. (1989). Comparison of intermittent versus continuous infusion metoclopramide in control of acute nausea induced by cisplatin chemotherapy. *J. Clin. Oncol.*, **7**, 943.

OLVER, I.N., SIMON, R.M. AISNER, J. (1986). Antiemetic studies: A methodological discussion. *Cancer Treat Rep.*, **70**, 555.

OLVER, I.N., ROOS, I.A.G., THOMAS, K. & HILLCOAT, B.L. (1989). Development of a murine gastric distension model for testing the emetic potential of new drugs and efficacy of antiemetics. *Chem. Biol. Interactions*, **69**, 353.

POLLERA, C.F. & GIARNELLI, D. (1989). Prognostic factors influencing cisplatin-induced emesis. Definition and validation of a predictive logistic model. *Cancer*, **64**, 1117.

PRESSER, S. & SCHUMAN, H. (1980). The measurement of a middle position in attitude surveys. *Public Opinion Q.*, **44**, 70.

SALLAN, S.E., ZINBERG, N.E. & FREI, E.III. (1975). Antiemetic effect of delta-9 tetrahydrocannabinol in patients receiving cancer chemotherapy. *N. Engl. J. Med.*, **293**, 795.

SEIPP, C.A., CHANG, A.E., SHILLING, D.J. & ROSENBERG, S.A. (1980). In search of an effective antiemetic: a nursing staff participates in marijuana research. *Cancer Nurs.*, **3**, 271.

SIMON, R.M. (1989). Design and conduct of clinical trials. In *Cancer Principles and Practice of Oncology*, De Vita, V.T., Hellman, S. & Rosenberg, S.A. (eds) p. 405. Lippincott: Philadelphia.

TORTORICE, P.V. & O'CONNEL, M. (1990). Management of chemotherapy-induced nausea and vomiting. *Pharmacotherapy*, **10**, 129.

WENDLE, H.A. (1979). Randomization in clinical trials. *Science*, **199**, 368.

Discussion of Dr Olver's paper

Smyth: The last two presentations have addressed very important issues for us all. Can we concentrate initially on whether crossover or parallel studies are the most worthwhile? A lot of studies have been done on a crossover basis to take into account the influence of prognostic factors, and perhaps on the misguided assumption that this is conservative in terms of patient numbers. Dr Olver certainly made a strong plea for parallel studies.

Gralla: I think the crossover design is unnecessary. What do we learn from cross-over trials that we cannot learn from single treatment?

Smyth: Patient preference.

Gralla: I'll accept that, but apart from patient preference do we gain any objective data on anti-emetic efficacy?

Secondly, there are only two outcomes from a crossover study: either A and B are equivalent or one is better. If a patient has a complete response to drug A, is it useful, or ethical, to swop them to drug B? We are already doing the best we can for that patient. Leaving them on drug A will also allow you to monitor its continued efficacy.

Is the patient his/her own control? I think not, because the effect of the first cycle of treatment affects subsequent courses.

We must also take care in expression of results. When patients are lost to a study it is important that we express response rates as a percentage of the original denominator. Patient preference for example—if a patient fails treatment the first time and is not crossed over they are not asked their preference. This could bias results. Maybe overall satisfaction with treatment is a better measure for preference.

Smyth: If the point of a crossover design is to allow for differences in prognostic factors then you could still say that it is valid to cross a patient over. If a patient does well on drug A you still have an interest in whether they will do as well on drug B.

Gralla: That is actually a different question—it doesn't answer whether drug A is better than drug B. It may be more valid to know whether drug A continues to give you that same complete response.

Morrow: We have seen that there is no significant difference in overall results whether the study measured emesis, nausea, both or patient preference. Any of these outcomes are useful measurements. Patient preference has a lot of publicity, but it is not more or less sensitive than measuring any other outcome.

Smyth: That surprises me. If we are evaluating the treatment of a tumour then there are obviously ethical questions about crossing over drugs when there is a complete response. But, if we are evaluating something that has an effect on the patients subjective perception of the treatment then surely the patients' preference is a valid end point?

Morrow: I am just making the point that the impact of an anti-emetic will be measurable, and the magnitude of that impact, or the effect size, will be the same whether you measure change in nausea, change in vomiting, changes in both or patient preference.

Smyth: Surely if we have two arms of a trial that both produce a comparable response and 55% of patients prefer drug A, then that says that drug A has some advantages.

Soukop: Patient preference actually incorporates a number of other things besides control of vomiting. I think patient preference has led us to an understanding of some facets of treatment. For example the mode of administration—when we gave chlorpromazine intramuscularly, the injection caused a certain amount of distress. This is a factor that might not have been considered, but clearly it did affect the patients and it matters more to patients when the control of nausea and vomiting is good.

Gralla: But we do not need a crossover design to establish these points.

De Muller: In a double-blind crossover study, mode of administration is not a factor because it is the same for both arms of the trial. Patient appreciation might be a relevant factor, but I think one could measure that type of information with an overall question in a parallel study.

As long as prognostic factors are evenly balanced I don't think there is any question that the number of patients required makes a parallel study easier.

Buxton: If we are going to use a preference question why is it just yes/no? Can we not have a more sensitive rating for example—strong preference, weak preference. Also shouldn't we ask why a patient prefers a certain drug?

Morrow: We should really be collecting data to answer two questions. First, for how many people will there be an effect, this requires non-parametric tests. Secondly what is the magnitude of that effect? One could use for example a four point scale. Analyses here require parametric statistics. Both factors are needed.

Can I ask what does one do when the odd patient performs so differently from the rest of the population that they are more than three standard deviations away from the mean?

Groshen: Yes this is a problem. In such situations your standard deviation is going to be larger than the mean.

Gralla: What about a rank order test? If you have a large enough sample size these two odd patients would still stand out, but the overall trend could be assessed.

Groshen: Yes, if you are using non-parametric statistics, but if you are committed to a parametric analysis that would not work.

Morrow: Related to this is when we wind up with non-equivalent variances in our two groups. What do you suggest?