

# How to study postoperative nausea and vomiting

C. C. APFEL<sup>1</sup>, N. ROEWER<sup>1</sup> and K. KORTTILA<sup>2</sup>

<sup>1</sup>Department of Anesthesiology, Julius-Maximilians-University, Wuerzburg, Germany, and <sup>2</sup>Department of Anesthesia and Critical Care, Helsinki University Central Hospital, Helsinki, Finland

Anesthesiological journals are flooded by innumerable studies of postoperative nausea and vomiting (PONV). Nevertheless, PONV remains a continuing problem with an average incidence of 20–30%. This paper should provide essential information for the design, conduct, and presentation of these studies. It should also increase comparability among future studies and help clinicians in assessing and reading the literature on PONV.

First, future studies should address new and relevant questions instead of repeatedly investigating prophylactically given antiemetics whose main results are predictable (e.g. already proven by meta-analysis). Second, group comparability should be based on well-proven risk factors and a simplified risk score for predicting PONV. Endless listings of doubtful risk factors should be avoided. Third, a realistic sample size estimation should be performed, i.e. in most cases at least 100 patients per group are necessary. Fourth, nausea, vomiting and rescue medication should be recorded and reported separately with the corre-

sponding incidences (and number of patients with these separate symptoms), and the main end-point should be PONV. The entire observation period should cover 24 h. Additional reporting of the early (0–2 h) and delayed (2–24 h) postoperative period is desirable and should consider single and cumulative incidences. Lastly, interpretation of results should take into account the study hypothesis, sources of potential bias or imprecision, and the difficulties associated with multiplicity of analysis and outcomes.

Received 3 December 2001, accepted for publication 30 April 2002

**Key words:** Postoperative nausea and vomiting.

© Acta Anaesthesiologica Scandinavica 46 (2002)

OVER the years, several thousand studies on postoperative nausea and vomiting (PONV) have been published and the number seems to increase with several hundred papers published each year. But the incidence of PONV remains fairly constant with an average of 20–30% (1). Perhaps inappropriate questions have been asked or the conduct of the studies did not allow meaningful conclusions. Further, heterogeneity between studies compromises meta-analysis, so that a recent consensus conference suggested a minimal standard for the conduct of randomized controlled trials (RCT) in the prevention or treatment of PONV (2). A **consolidated standard of reporting trials** (CONSORT) has been suggested previously, and the checklist and flowchart is strongly recommended to everybody who is involved in clinical trials, i.e. also in the field of PONV (3). More specifically, Korttila has described important issues to be considered in ‘the study of postoperative nausea and vomiting’ (4). Now, 10 years later, some new insights have evolved so that an update has become necessary. We hope that the discussed methodology,

which should not substitute but extend the CONSORT statement, may help investigators in designing, conducting and reporting studies, and may also help clinicians in assessing and reading the literature on PONV critically.

## Aims of the study

Numerous studies have been published in the past decades comparing the prophylactic use of one antiemetic with placebo, and meta-analyses have shown that the efficacy of such approaches is limited to a relative risk reduction of approximately 30–40% (5). It is therefore not surprising when a simple study whose main aim is the comparison of one antiemetic vs. placebo is rejected on the grounds of lack of novelty when a previous meta-analysis has already quantified the effect. Moreover, Aspinall and Goodman have pointed out that the denial of effective treatment is ethically not justified (6).

The question to be addressed should lead to an answer that is most likely to improve our understanding and/or the possibility of avoiding or treating PONV.

Financial support: institutional sources only.

Two examples are given:

1. The treatment of nausea and vomiting appears to be much more cost-effective compared with prophylactically given antiemetics (7). However, treatment studies are rare because informed consent must be taken from many more patients than can finally be analyzed in the study. Here, dose-response studies are needed to determine the minimal effective dose to treat established PONV.
2. In high risk patients, who can now be identified by simplified and validated risk scores, it may be ethically questionable to wait until they suffer PONV (8–11). In the light of the limited efficacy of a single antiemetic, several investigators tried to increase efficacy by combining two antiemetics that are still not statistically unequivocally proven (12). However, Scuderi et al. was able to report an almost complete elimination of PONV by applying a multimodal strategy (13). Accordingly, a systematic investigation of several antiemetic strategies using a factorial design is needed so that all possible combinations can be compared. This may indicate the best combination in terms of cost-effectiveness and risk/benefit ratio, but this requires a high number of patients to allow for interaction analyses with sufficient power.

## Materials and methods

Researchers who intend to perform a clinical trial are strongly recommended to involve somebody with a sufficiently profound statistical knowledge at an very early stage. Thus the following lines can by no means comprehensively cover all statistical aspects.

### *Study design*

The highest level of evidence is achieved by studies being randomized, controlled and double-blinded. Randomization is important to decrease the risk of imbalances of the risk factors and confounding variables between the study groups (14, 15). To avoid an investigator's bias, the randomized sequence should be concealed until the interventions are assigned (16). This is not blinding, which is also important and means that the group allocation is unknown during the assessment of the outcome (17). Control groups are important to quantify the benefit of the intervention and can be either an effective treatment or placebo. Placebo is usually preferred because there is no common gold-standard for preventing PONV (7). The main advantages of using placebo instead of an active treatment may be that (a) the necessary sample size required to

detect a significant difference is much smaller and (b) the chances of demonstrating a significant difference are higher. However, denial of effective antiemetic treatment may not be justified in patients at high risk for PONV (6), so that a frequently used antiemetic, such as the current antiemetic standard of that hospital, could serve as an alternative. Double blinding of patients and the assessors is similarly important (17). In contrast to postoperative vomiting this is particularly important when assessing the subjective nature of postoperative nausea or patient satisfaction.

### *Patient selection*

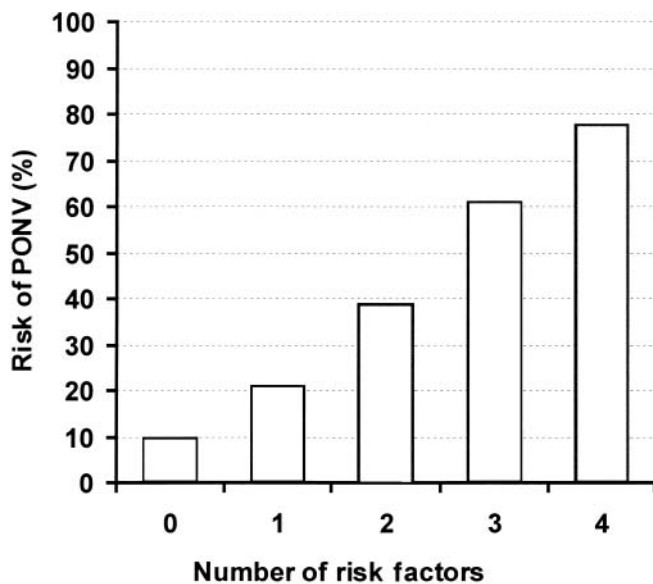
It may be wise to reduce the number of confounding variables by excluding patients who present with underlying pro-emetogenic diseases and/or who are currently on antiemetic drugs. Previously patient selection was mainly based on those types of surgery known to be associated with a high incidence of PONV. However, in the past decade multivariable analyses of large prospective trials revealed that the different incidences of PONV after most operations are mainly caused by the associated risk factors and less by the operation itself (9–11, 18–21). It was also demonstrated that a combination of the most important risk factors can be used in an operation-independent risk score to predict the probability of PONV (8, 9, 20, 22), and proved to be valid for other centers (10, 11, 22, 23). It is therefore not surprising that patients at risk for PONV can be better identified by using a simplified risk score instead of classification into, more or less voluntarily defined, high- or low-risk types of surgery (unpublished observation). Thus, instead of selecting patients undergoing just one type of surgery, we recommend the use of a validated and simplified risk score, such as that proposed by Koivuranta and colleagues (9) or that suggested by Apfel and colleagues (8), to identify patients with an increased risk (Fig. 1).

### *Factors to be controlled*

Controlling factors increase internal validity, i.e. the study results are most likely to apply to these specific conditions. So, it makes good sense to control all potentially confounding factors. This applies especially to the anesthetic technique and drugs, such as volatile anesthetics, propofol, but also to intraoperative hydration and oxygen concentration (24–28). However, it is worth mentioning that external validity may be limited if specific circumstances applied in the study are rarely used in daily practice.

### Group comparability

Factors that are not controlled and that may have an impact on PONV should be distributed evenly among the study groups. The six most important risk factors, which are female gender, non-smoking status, history of PONV, or motion sickness, age, duration of anesthesia and postoperative opioids, should be presented in a table according to the group allocation. In addition, other predictors may be of interest such as the type and dosage of volatile anesthetics and the types of surgery. However, an extensive list of questionable risk factors should be avoided. For example, although



Risk factors	Points
Female gender	1
Nonsmoking status	1
History of PONV and/or Motion sickness	1
Postoperative opioids	1
<b>Number of risk factors</b>	<b>4</b>

Fig. 1. Illustration of the simplified risk score by Apfel and colleagues (8). In a two center cross-validation the probability of postoperative nausea and vomiting (PONV) could be assessed by the number of the four most relevant risk factors. If none, one, two, three or four risk factors are present, the risk for PONV is approximately 10, 20, 40, 60 and 80%. Interestingly, predictive characteristics of simplified risk scores appeared to be at least as good as more complex models, while being much easier to handle (10, 11).

the body mass index and the menstrual cycle were believed to be risk factors for PONV, recent systematic reviews revealed that neither have consistently proven effects (29, 30)

In addition, the average risk of the groups according to a validated risk score should be reported, and if a simplified risk score is used the number of patients with the different risk levels should be given as well. The reason for this is that simply considering five or six separate risk factors with their 95% confidence intervals will show up in approximately 25% of cases with at least one risk factor out of that confidence interval being by chance [ $1-(0.95)^5 \approx 23\%$  and  $1-(0.95)^6 \approx 26\%$ ]. In that case, a similar average risk according to the risk score between the groups would demonstrate that this imbalance is very unlikely to have any effect on the results. In contrast, if several 'non-significant' risk factors project into the same direction, resulting in a significantly different average risk between the groups, the result should be interpreted with caution.

### Assessment of outcome variables

It is important to define precisely the meanings of nausea, retching and vomiting. Knapp and Beecher gave excellent definitions 45 years ago of nausea and emesis, and these are still useful today (31). Nausea is a subjective sensation which should be evaluated by the patient, not by the observer. The feeling is best described as the desire to vomit without the presence of expulsive muscular movements. When nausea becomes severe, the secretion of saliva is increased and is associated with vasomotor disturbances and sweating. The feature that distinguishes retching from vomiting is the production of even the smallest amount of stomach contents. When no stomach contents are expelled, the expulsive efforts are classified as retching. Retching is usually indicative of an empty stomach and is generally as unpleasant for the patient as vomiting. With both retching and vomiting, pharyngeal muscles relax, the soft palate is elevated, the diaphragm descends, and there are spasmodic contractions of respiratory chest and abdominal wall muscles. Retching and vomiting may also be grouped together under the common term 'emetic episodes' (4). However, retching and vomiting is a brainstem reflex, which is not necessarily an aggravated form of nausea because it may well occur without it. The converse is also true. As a subjective feeling, nausea must be regarded as a conscious cortical activity not necessarily affecting the brainstem. Thus, although nausea, retching and vomiting are frequently inter-linked they may well occur on their own. Therefore,

it is important that the different outcome variables, mainly nausea, vomiting and rescue treatment, should be assessed independently (in pre-school children the assessment of nausea may not be feasible).

To quantify symptoms, a visual analog scale (0–10 or 0–100), a 11-point numerical rating scale (0–10), or a verbal rating scale (none, mild, moderate, severe) may be used (32). The whole study period should cover 24h. As very recent data demonstrated that volatile anesthetics appeared to be the major cause of early (0–2h) but not delayed (2–24h) postoperative vomiting (24), additional assessment of the early and delayed postoperative period is recommended. Ideally, the exact time of every emetic episode should be recorded. This allows performing log-rank tests for survival analyses, which may provide additional time-course information on the action of antiemetics.

Special attention should be given to the assessment of drug-specific adverse events, e.g. headache, increased sedation, restlessness, anxiety, dysphoria, sleep disturbances, dry mouth, and voiding problems. As most of these may not be reported by the patients spontaneously it is important that they are actively asked.

In the past, patient satisfaction has been advocated as the main outcome instead of PONV, which was critically entitled a surrogate end-point (33, 34). Although not all investigators may follow this argument, assessing patients' satisfaction can provide some valid additional information. For example, Scuderi and colleagues found that prophylactically given antiemetics did not improve patient satisfaction unless given to patients at increased risk for PONV (35). The assessment of patient satisfaction is a difficult psychological task and interestingly a personal interview may lead to results at least as valid as those obtained with a standardized questionnaire (some patients may misunderstand questions in a questionnaire which can be ruled out in a personal interview) (36). However, most anesthesia-related studies published on patient satisfaction are of questionable value because only rigorous methods and reliable instruments will yield valid and clinically relevant findings (37).

### *Statistics and sample size estimation*

Every randomized controlled trial should have a realistic sample size estimation to reject or accept the null hypothesis. The null hypothesis usually assumes that there is no difference between the study groups whereas the alternative hypothesis assumes that there

is a difference. For dichotomous outcomes four parameters are important.

First, the type-I error, sometimes called  $P$  or  $\alpha$ , is the probability that we erroneously reject the null hypothesis, i.e. we erroneously assume that there is a difference, when in fact there is not. Usually, we are happy with  $\alpha=0.05$ , i.e. we accept a 5% probability that the detected difference happened by chance. Second, the type-II error, sometimes called  $\beta$ , is the probability that we erroneously accept the null hypothesis, i.e. we erroneously assume that there is not a difference, when in fact there is. Usually, we are happy with  $\beta=0.2$ , i.e. we accept a 20% probability that we fail to detect a true difference. In other words we have an 80% chance of detecting a difference (if there is one), which is often referred to as the power of the study, i.e.  $\text{power}=1-\beta$ . Third, we have to have a realistic idea of the control event rate (CER). When the CER is high more patients can potentially benefit from an intervention so that eventually less patients are needed to detect a statistically significant difference. Thus, it makes good sense to use a simplified risk score for selecting patients with a PONV-risk of more than 40, 50 or 60%. Finally, we should assume a treatment effect that is realistic and clinically relevant. For prophylactically given single antiemetics this will be a relative reduction rate in the range of 30–40%. When these four parameters are known the required sample size can be determined. There are several computer programs available for sample size estimation of which we have tested the four most widely used ('*N*', idv Data Analysis and Study Planning, Gautingen, Germany; '*nQueryAdvisor*', Statistical Solutions Ltd, Cork, Ireland; '*Sample Power*', SPSS Inc., IL, USA; and '*PASS*', NCSST, UT, USA). All of these are relatively easy to use and give similar results. Alternatively, tables may be used for approximation (Table 1).

### *Working example*

Let's assume a study is intended in the general surgical department on patients at increased risk for PONV. According to a simplified risk score only patients with at least two out of the four risk factors (female gender, non-smoking status, history of PONV or motion sickness, and postoperative opioids) will be included (8). As a rule of thumb it can be assumed that if none, one, two, three or four risk factors are present the risk for PONV is approximately 10, 20, 40, 60 and 80%, respectively (8). Including only patients with at least two risk factors will result in an average CER of approximately 50% (assuming there will be very few patients with all four risk factors, i.e. with 80%). We expect with our antiemetic approach a rela-

tive risk reduction of PONV of 40%. The absolute risk reduction would therefore be 20%, i.e. we would expect a treatment event rate of 30%. Being happy with  $\alpha=0.05$  and  $\beta=0.2$  we would need 103 patients per group, i.e. approximately 200 patients for a simple randomized controlled trial. From the table we can see that for 'realistic' estimates usually more than 100 patients per group are needed (Table 1). It should be noted that sample size estimations are frequently missing in smaller studies or that the expected treatment effect was overestimated to suit the small number of patients.

If several null-hypothesis are tested (confirmatory testing) an  $\alpha$ -correction for multiple comparisons may be considered. The most conservative approach would be the Bonferroni correction, i.e. the intended  $\alpha$ -error is divided by the number of tests (38). However, the need and the method to be used for an  $\alpha$ -correction is not undisputed, as it may be based on

wrong assumptions and therefore increases the type-II error. Accordingly, some statisticians recommend 'reporting the unadjusted *P*-values and confidence limits with a suitable note of caution with respect to interpretation' (39).

## Results

As mentioned earlier, group comparability should be demonstrated by proven risk factors and a validated risk score (Table 2). It is recommended that a simplified risk score is used (8) because there is sufficient evidence to suggest that they are at least as valid as more complex risk scores (10, 11) The assessed outcome variables should be given with absolute and relative incidences and the corresponding 95% confidence intervals. In contrast to simple *P*-values, confidence intervals convey, at the same time, the clinically relevant information of the magnitude of an effect and

Table 1

Required sample size per group for two proportions (two-sided, calculated using nQuery Advisor 4.0).

Control event rate	Relative reduction rate	Absolute reduction rate	Treatment event rate	Alpha = 0.05			Alpha = 0.01		
				beta = 0.2	beta = 0.1	beta = 0.05	beta = 0.2	beta = 0.1	beta = 0.05
40%	25%	10%	30%	376	496	609	550	695	827
	30%	12%	28%	260	342	419	380	478	569
	40%	16%	24%	145	189	231	210	263	312
	50%	20%	20%	91	119	144	131	164	194
50%	25%	12.5%	37.5%	262	345	423	383	483	574
	30%	15%	35%	183	240	293	266	334	397
	40%	20%	30%	103	134	163	149	186	221
	50%	25%	25%	66	85	103	94	117	138
60%	25%	15%	45%	186	244	299	271	341	405
	30%	18%	42%	131	171	209	190	239	283
	40%	24%	36%	75	98	118	108	135	160
	50%	30%	30%	49	63	75	70	86	102

Table 2

Essential variables to be reported in PONV studies

Risk factors/risk	Control group	Treatment group 1	Treatment group 2
Female gender, n (%)			
Age, years (95% CI)			
Non-smoking status, n (%)			
History of postoperative nausea and vomiting, n (%)			
History of motion sickness, n (%)			
Duration of anesthesia, min (95% CI)			
Type of surgery and surgical techniques, n (%)			
Postoperative opioids, n (%)			
Number of patients with 0, 1, 2, 3 or 4 risk factors according to Apfel (8), n (%)			
or 0, 1, ..., 5 risk factors according to Koivuranta (9), n (%)			
Calculated average (mean) risk for postoperative nausea and vomiting (%)			

Table 3

An example to report outcome variables for postoperative nausea and vomiting with early and delayed symptoms. (CI = confidence intervals).

Intervals	Patients with	Control group	Treatment group 1	Treatment group 2
0–24 h	Nausea, n (% under 95% CI) Vomiting, n (% under 95% CI) Nausea and/or vomiting, n (% under 95% CI) Rescue treatment, n (% under 95% CI) Optional further definitions such as a score for the severity of symptoms, 'total response' or 'complete response'			
Optional (e.g. 0–2 h)	Nausea, n (% under 95% CI) Vomiting, n (% under 95% CI) Nausea and/or vomiting, n (% under 95% CI) Rescue treatment, n (% under 95% CI) Optional further definitions such as a score for the severity of symptoms, 'total response' or 'complete response'			
and: e.g. 2–24 h	Nausea, n (% under 95% CI) Vomiting, n (% under 95% CI) Nausea and/or vomiting, n (% under 95% CI) Rescue treatment, n (% under 95% CI) Optional further definitions such as a score for the severity of symptoms, 'total response' or 'complete response'			

should therefore be preferred wherever possible (40). When different intervals are considered, separate incidences (e.g. 0–2 and 2–24 h or 0–6 and 6–24 h) as well as cumulative incidences (0–2, 0–6, 0–24 h) should be reported (Table 3). In addition, the severity of PONV may be reported although these results may not be comparable with other studies (41). A theoretical advantage is that significant differences may be demonstrated with fewer patients. Another possibility is the use of 'complete response' or 'total response'. In that case the definitions by Korttila should be used to increase the comparability between studies (4). When questionnaires (e.g. for patient satisfaction) are used these should be printed in the appendix. Finally, a statement would be desirable indicating whether individual raw data could be made available for other investigators.

## Discussion

At this stage a very conscientious approach is required and a clear distinction should be made between a priori and a posteriori hypothesis. When an a priori hypothesis can be confirmed, i.e. the null hypothesis can be rejected because the probability  $P < \alpha$ , the conclusion is usually straightforward. But when the null hypothesis cannot be rejected, the question arises, whether the study had sufficient power to detect an expected and/or clinically relevant effect, i.e. a separate power analyses may be needed. When a

statistically significant difference is observed in a retrospective analysis of the data ( $P < \alpha$ ), this should be interpreted with caution, i.e. it is no proof of an effect but can be helpful to create an a posteriori hypothesis needing confirmation in an additional study. The issue is even more difficult when testing an a posteriori hypothesis failing statistical significance.

For illustration, let's assume that the earlier-mentioned study with 103 patients per group was performed. The incidence of PONV was 54% (52/103) in the control group and only 33% (34/103) in the treatment group. Applying a chi-square test (with continuity correction) confirms the a priori hypothesis of the reduction of PONV ( $P = 0.016$ ). Analysing the incidence of vomiting may show up with 33% (34/103) in the control group and 21% (22/103) in the treatment group. Again, this would be a reduction by approximately a third. But this time it failed statistical significance ( $P = 0.085$ ). The interpretation that 'the treatment reduced PONV but not vomiting' would be incorrect. While the statement about PONV is correct the statement on vomiting is not. The a priori defined outcome was PONV and the sample size estimation was assuming an incidence of ~50%. As the incidence of the secondary outcome is significantly lower, more patients were needed to detect the same relative reduction in vomiting, i.e. the power of the study to detect such a difference was not sufficient (only 42% in this example). It should be mentioned that this misinterpretation may not have occurred if confidence in-

Table 4

Checklist of questions for authors and referees for randomized controlled trials on postoperative nausea and vomiting.

	Reported on page #	Question	Referee rating
Title	1	Does it describe the content and/or the results of the study appropriately?	
Abstract/Summary	2	Does the abstract sufficiently describe the intention of the study, methods and results?	
Introduction	3	Are the conclusions founded by the data? Are the assumptions well proven? If not, is it clearly indicated to be a question of debate? Is the current literature considered? Is the question clinically relevant or just a repetition in a different setting?	
Methods			
Participants		Are eligibility criteria described and reasonable?	
Objective		Have the hypotheses been set a priori?	
Outcome		Are primary and secondary outcome variables described and have they been set a priori?	
Intervention		Is the intervention reasonable and appropriately described?	
Blinding, etc.		Is the method of randomization, masking (treatment allocation) and blinding (during assessment) described?	
Statistics		Is the sample size estimation based on realistic assumptions and sufficiently described?	
Results			
Recruitment		Are the number of patients randomly assigned, receiving intended treatment, completing the protocol, and analyzed for the primary outcome clearly described (e.g. flow-chart)?	
Baseline data		Are important clinical characteristics, proven risk factor and a risk score for PONV, listed for each group?	
Outcome data		Are the incidences for nausea, vomiting, and PONV reported separately and with confidence intervals?	
Ancillary results		Are further analyses indicated as prespecified or exploratory?	
Adverse events		Was the method for assessment and the reporting of adverse events appropriate?	
Discussion		Does the interpretation of the results consider sources of potential bias and difficulties associated with multiplicity of analyses and outcomes? Is the internal and external validity of trial findings considered? Are the limitations of the data and the study appropriately considered? Are the results appropriately discussed in the context of current evidence?	

tervals of the relative risks were used in the first place because it had been clear that the magnitude of the effect is similar and that only the lower incidences of vomiting contributed to a wider confidence interval.

In conclusion, most problems can be avoided by considering the CONSORT statement and the specific issues discussed, as outlined in the checklist for authors and referees (Table 4).

## Acknowledgements

We are grateful to Dr Mark Palazzo, Charing Cross Hospital, London, UK, for his valuable help in revising the manuscript.

## References

1. Kovac AL. Prevention and treatment of postoperative nausea and vomiting. *Drugs* 2000; **59**: 213–243.
2. Apfel CC, Eberhart LHI, Kranke P, Ruesch D. Recommendations for randomized controlled trials to prevent or treat postoperative nausea and vomiting. *Anaesth Intensivmed Notfallmed Schmerzther* 2002; **43**: 69–74.
3. Moher D, Schulz KF, Altman DG, Lepage L, the CONSORT statement. Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; **357**: 1191–1194.
4. Korttila K. The study of postoperative nausea and vomiting. *Br J Anaesth* 1992; **69**: 20S–23S.
5. Tramer MR. A rational approach to the control of postoperative nausea and vomiting: evidence from systematic reviews. Part I. Efficacy and harm of antiemetic interventions, and methodological issues. *Acta Anaesthesiol Scand* 2001; **45**: 4–13.
6. Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials [see comments]. *BMJ* 1995; **311**: 844–846.
7. Tramer MR. A rational approach to the control of postoperative nausea and vomiting: evidence from systematic reviews. Part II. Recommendations for prevention and treatment, and research agenda. *Acta Anaesthesiol Scand* 2001; **45**: 14–19.
8. Apfel CC, Läärä E, Koivuranta M, Greim CA, Roewer N. A simplified risk score for predicting postoperative nausea and vomiting: Conclusions from cross-validations between two centers. *Anesthesiology* 1999; **91**: 693–700.

9. Koivuranta M, Läärä E, Snare L, Alahuhta S. A survey of postoperative nausea and vomiting. *Anaesthesia* 1997; **52**: 443–449.
10. Apfel CC, Kranke P, Eberhart LHJ, Roos IA, Roewer N. A comparison of predicting models for postoperative nausea and vomiting. *Br J Anaesth* 2002; **88**: 234–240.
11. Pierre S, Benais H, Pouymayou J. Apfel's simplified score may favourably predict the risk of postoperative nausea and vomiting. *Can J Anaesth* 2002; **49**: 237–242.
12. Eberhart LH, Morin AM, Bothner U, Georgieff M. Droperidol and 5-HT<sub>3</sub>-receptor antagonists, alone or in combination, for prophylaxis of postoperative nausea and vomiting. A meta-analysis of randomised controlled trials. *Acta Anaesthesiol Scand* 2000; **44**: 1252–1257.
13. Scuderi PE, James RL, Harris L, Mims GR III Multimodal antiemetic management prevents early postoperative vomiting after outpatient laparoscopy. *Anesth Analg* 2000; **91**: 1408–1414.
14. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? *BMJ* 1999; **318**: 1209.
15. Altman DG, Bland JM. How to randomise. *BMJ* 1999; **319**: 703–704.
16. Altman DG, Schulz KF. Concealing treatment allocation in randomised trials. *BMJ* 2002; **323**: 446–447.
17. Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ* 2000; **321**: 504.
18. Palazzo M, Evans R. Logistic regression analysis of fixed patient factors for postoperative sickness: a model for risk assessment. *Br J Anaesth* 1993; **70**: 135–140.
19. Cohen MM, Duncan PG, DeBoer DP, Tweed WA. The postoperative interview: assessing risk factors for nausea and vomiting. *Anesth Analg* 1994; **78**: 7–16.
20. Apfel CC, Greim CA, Haubitz I, Goepfert C, Usadel J, Sefrin P et al. A risk score to predict the probability of postoperative vomiting in adults. *Acta Anaesthesiol Scand* 1998; **42**: 495–501.
21. Junger A, Hartmann B, Benson M, Schindler E, Dietrich G, Jost A et al. The use of an anesthesia information management system for prediction of antiemetic rescue treatment at the postanesthesia care unit. *Anesth Analg* 2001; **92**: 1203–1209.
22. Apfel CC, Greim CA, Haubitz I, Grundt D, Goepfert C, Sefrin P et al. The discriminating power of a risk score for postoperative vomiting in adults undergoing various types of surgery. *Acta Anaesthesiol Scand* 1998; **42**: 502–509.
23. Eberhart LH, Hogel J, Seeling W, Staack AM, Geldner G, Georgieff M. Evaluation of three risk scores to predict postoperative nausea and vomiting. *Acta Anaesthesiol Scand* 2000; **44**: 480–488.
24. Apfel CC, Kranke P, Papenfuß T, Rauch S, Greim CA, Roewer N. Volatile anaesthetics may be the main cause for early but not delayed postoperative nausea and vomiting. *Br J Anaesth* 2002; **88** (5): 659–668.
25. Elhakim M, el SS, Kaschef N, Essawi GH. Intravenous fluid and postoperative nausea and vomiting after day-case termination of pregnancy. *Acta Anaesthesiol Scand* 1998; **42**: 216–219.
26. Yogendran S, Asokumar B, Cheng DC, Chung F. A prospective randomized double-blinded study of the effect of intravenous fluid therapy on adverse outcomes on outpatient surgery. *Anesth Analg* 1995; **80**: 682–686.
27. Greif R, Lacyni S, Rapf B, Hickie RS, Sessler DI. Supplemental oxygen reduces the incidence of postoperative nausea and vomiting. *Anesthesiology* 1999; **91**: 1246–1252.
28. Goll V, Akca O, Greif R, Freitag H, Arkilic CF, Scheck T et al. Ondansetron is no more effective than supplemental intraoperative oxygen for prevention of postoperative nausea and vomiting. *Anesth Analg* 2001; **92**: 112–117.
29. Kranke P, Apfel CC, Papenfuss T, Rauch S, Lobmann U, Rubsam B et al. An increased body mass index is no risk factor for postoperative nausea and vomiting. A systematic review and results of original data. *Acta Anaesthesiol Scand* 2001; **45**: 160–166.
30. Eberhart LHJ, Morin AM, Georgieff M. A systematic review of studies investigating the impact of the menstrual cycle on the incidence of postoperative nausea and vomiting. *Anaesthesist* 2000; **49**: 532–535.
31. Knapp MR, Beecher HK. Postanesthetic nausea, vomiting and retching. *JAMA* 1956; **160**: 376–385.
32. Boogaerts JG, Vanacker E, Seidel L, Albert A, Bardiau FM. Assessment of postoperative nausea using a visual analogue scale. *Acta Anaesthesiol Scand* 2000; **44**: 470–474.
33. Fisher DM. Surrogate end points, are they meaningful? *Anesthesiology* 1994; **81**: 795–796.
34. Fisher DM. The 'big little problem' of postoperative nausea and vomiting: do we know the answer yet? *Anesthesiology* 1997; **87**: 1271–1273.
35. Scuderi PE, James RL, Harris L, Mims GR. Antiemetic prophylaxis does not improve outcomes after outpatient surgery when compared to symptomatic treatment. *Anesthesiology* 1999; **90**: 360–371.
36. Bauer M, Bohrer H, Aichele G, Bach A, Martin E. Measuring patient satisfaction with anaesthesia: perioperative questionnaire versus standardised face-to-face interview. *Acta Anaesthesiol Scand* 2001; **45**: 65–72.
37. Le May S, Hardy JF, Taillefer MC, Dupuis G. Patient satisfaction with anesthesia services. *Can J Anaesth* 2001; **48**: 153–161.
38. Bland JM, Altman DG. Statistics notes: Multiple significance tests: the Bonferroni method. *BMJ* 1995; **310**: 170.
39. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ* 1998; **316**: 1236–1238.
40. Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with Confidence*. 2nd edn. Oxford, UK: Oxford University Press, 2000.
41. Bellville JW, Bross IDJ, Howland WS. A method of the clinical evaluation of antiemetic agents. *Anesthesiology* 1959; **20**: 753–760.

Address:  
 Christian C. Apfel  
 Department of Anesthesiology  
 Julius-Maximilians-University  
 Josef-Schneider Str. 2  
 D-97080 Wuerzburg  
 Germany  
 e-mail: apfel@mail.uni-wuerzburg.de or apfel@ponv.de