

Predictive models for postoperative nausea and vomiting

Editor—We read with interest the publication by Apfel and colleagues.¹ This is, without doubt, the most important paper on

prediction of postoperative nausea and vomiting (PONV) yet published. We would like, however, to comment on several points.

First, the authors have done a very laudable search of the literature on prediction of PONV until about summer 2000, including two rather unknown models published in a German language refresher course book and a symposium handout. It might, however, be of interest that recently (and thus probably after the time when the study was designed), several new models for prediction of the incidence of PONV have been published.²⁻⁵ Also, the algorithm of Gan—cited from the symposium handout by Apfel—has recently been published in modified form in a major journal.⁶ The search for the ‘best’ predictive model is likely to continue.

Second, it was sound to test the various models in patients from a centre that was not involved in the development of the scores, probably the University of Marburg.¹ However, there is neither an indication as to which university (Würzburg or Marburg) analysed the data, nor as to who designed the study. As the first and the last author of the publication (who are from Würzburg) are connected with development of one of the scores,⁷ and are outspoken critics of another score that has been analysed,⁸ this raises the question of bias.

Third, we agree with Apfel and colleagues that an ‘ideal’ score should be simple. That is, it should have no more than four or five items. It should allow prediction without technical aids, i.e. have items each of which is loaded with an equal additive risk of e.g. 20%, as is approximately the case for both the simplified Apfel-score,⁷ and the simplified Koivuranta-score.⁹ Scores requiring an anaesthesia data management system³ or a neuronal network⁸ are in our opinion useful for research only, unless preoperative visits are computer based.

We would like to add two important points. It would be desirable to have a mnemonic attached to the ‘ideal’ score, something like **A**bstinence from nicotine, **P**rior PONV, **F**emale sex, **E**mesis while travelling, **L**onger than 1 h surgery; giving **APFEL**, although these are the items of the Koivuranta-score. The other point is that several scores include variables that preclude their use as *clinical* predictors because they require knowledge that is not available prior to anaesthesia. As an example, some scores include the use of postoperative opioids in their formulae (e.g. Ulleval-score², Apfel-score⁷). For many minor to intermediate procedures it is neither practical nor ethical to decide in advance whether or not the patient will receive opioids after surgery. These scores cannot be calculated beforehand or even during the anaesthetic. Even if used for scientific purposes only, these scores are problematic as some interventions may influence both PONV and the need for postoperative opioids. For example, in a trial comparing total intravenous anaesthesia (TIVA) with propofol and remifentanyl to balanced anaesthesia with isoflurane and fentanyl, TIVA might increase postoperative opioid use. This could lead to spuriously higher Apfel-scores in the TIVA group in spite of correct randomization or stratification.

Scores including the duration of surgery are also problematic (e.g. Junger-score,³ Koivuranta-score⁹), although it is probably easier to predict the duration of surgery than to predict postoperative opioid use. Even if the surgeon is quicker or slower than expected, scores dependent on duration of surgery at least allow prediction of PONV prior to extubation and a decision on prophylaxis at that point in time.

Fourth, for both the simplified Apfel-score⁷ and simplified Koivuranta-score,⁹ the most important information for clinical practice is missing: What was the actual incidence in patients with n risk factors ($n = 0, 1, 2, \dots$)?

In conclusion, we agree with Apfel and colleagues that scores for prediction of PONV are important both for clinical practice (to avoid use of costly or side-effect-prone anti-PONV drugs in patients who do not need them) and for group comparisons in

PONV research. At present, the simplified Koivuranta-score seems to be closest to the ‘ideal’ score, based on both ease of use and discriminating properties.^{1,2,4,10} If Koivuranta and Apfel permit, it could be connected with the mnemonic **APFEL** as outlined above.

W. H. Maleck
S. N. Piper
Ludwigshafen, Germany

Editor—We are pleased that our latest comparison of predictive models for postoperative nausea and vomiting (PONV) has produced so much interest.¹ In addition, we appreciate the comments from Drs Maleck and Piper about our paper and the opportunity to respond to them.

The authors are correct in pointing out that several additional predictive models have been published since our last systematic search. The Norwegian reference from Tropé and Raeder² should have been detectable at the end of 2000 when we performed our search and we can only speculate that there was some delay in entering this paper on Medline. However, using the Ulleval ranks results in an area under the receiver operating characteristic curve (AUC) of only 0.60 (95% confidence intervals 0.57–0.63) which is even smaller (0.58 (0.55–0.60)) when data are reduced to a binary model, i.e. when patients with more than seven points are considered to be at increased risk (Fig. 1). Perhaps its discriminating power is considerably better in the Ulleval population, but unfortunately no AUC was calculated for comparison so no conclusion can be drawn from that paper. Including this score in our paper would, therefore, have had no impact on our conclusions. The same point applies if we had performed our electronic database research in 2001, i.e. after the manuscript was accepted, since those scores were either not dealing with models for PONV (Junger and colleagues refer to rescue treatment in the PACU³), or a risk calculation was not possible from the published abstracts because the model from our coauthor Eberhart⁴ used artificial neural networks. In addition, the coefficients of the model from Yamaguchi and colleagues⁵ were not given. It may also be of interest that the original score from

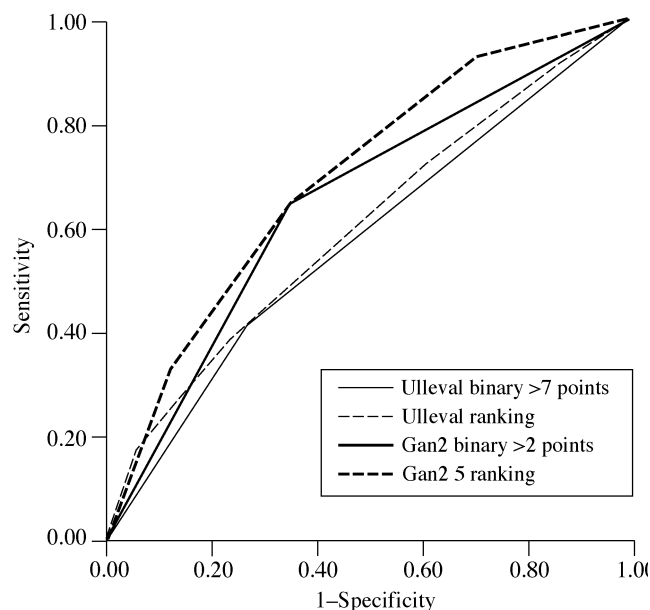


Fig 1 Receiver operating characteristic curves of both types of Ulleval score and the second score from Gan.

Gan consisted of 14 predictors while his very recently published score⁶ is basically a mixture of the risk factors from our simplified score⁷ plus some surgical risk factors. The latter score results in an AUC of 0.70 (0.67–0.72) when the five possible ranks are considered and of 0.65 (0.62–0.68) when a binary model is applied (where patients with more than two risk factors are classified as high risk) (Fig. 1). This is very much the same as the AUC of the simplified score, so that including the type of surgery did not lead to a better prediction. This corroborates our previously published results from a computer simulation which demonstrated that the limited strength of predictors (odds ratio in the range of 2 to 3) will make it very unlikely that additional predictors (such as the type of surgery) can lead to a clinically better prediction if four or five predictors are already included in the model.¹¹ Thus, despite agreeing with Maleck and Piper that the search for new predictive models should continue, it seems that the two simplified scores are currently the most useful predictive models for clinical practice.^{7,9}

The data were collected at the University of Ulm before Dr Eberhart went to the University of Marburg. As explained in the paper, none of the analysed risk scores was derived from these data to allow a fair comparison between the different models and to avoid any potential bias. However, if Maleck and Piper are concerned that there may be some bias in the analysis because we have been involved in the development of one of the simplified scores and reported favourable results in our current validation, we can only argue that the two other independent validations from Eberhart and colleagues¹⁰ (Ulm, Germany) and from Pierre and colleagues¹² (Toulouse, France) came to similar conclusions. Perhaps the best way to rule out any remaining doubts, would be a validation of risk scores based on data from their own centre. Since such validations may be complicated we would be delighted to offer Maleck and Piper, or any other colleagues, our statistical support.

We agree with the authors that an 'ideal' risk score has to be simple; the appealing (and flattering) mnemonic as suggested by Maleck and Piper may help clinicians in daily practice. However, while Apfel may be easy to remember for some colleagues because it is the German word for 'apple', we wonder whether a mnemonic taken from the English language may better serve its intention.

The concern about using unforeseeable predictors is a frequently asked question which usually relates to the use of postoperative opioids, but rarely to the duration of anaesthesia. As pointed out before, the use of postoperative opioids depends largely on institutional attitudes and the extent of surgery.^{7,13} We have now simulated what would happen if in 10% of 1566 patients, judgement as to whether they will receive postoperative opioids or not will be wrong (i.e. 108 patients will be wrongly assumed not to receive opioids and 49 patients the other way round). Interestingly, the discriminating power of the simplified score was hardly affected, as expressed by a decrease in the AUC from 0.684 to 0.668, i.e. 0.016. Thus, although the inclusion of an incomplete predictor decreases the discriminating power, it has no clinically relevant implication in this case. One reason for the robustness of the score is that the impact of one factor on the overall risk is limited. More detailed information is available on the Internet at www.ponv.de.

The incidences for the five classes of the Apfel score were 14/73 (19%), 46/308 (15%), 192/547 (35%), 224/445 (50%) and 124/193 (64%). The incidences for the six classes of the Koivuranta score were 6/44 (14%), 41/273 (15%), 157/515 (30%), 226/478 (47%), 140/214 (65%), and 30/42 (71%). However, for clinical purposes, the absolute incidences may be less relevant than recognizing patients at low, medium, or high risk for PONV, as previously suggested.¹⁴

Finally, we would like to note that these scores are more a tool for clinical assessment than a guideline for the use of antiemetics. The patient risk (i.e. probability) of PONV is only one, albeit important, aspect but an antiemetic strategy should always take into account other important considerations. For example, if a female non-smoker undergoing a major gynaecological procedure has a PONV risk of 50% but does not want prophylactic antiemetics, we consider a 'wait and see' strategy is appropriate, and rescue treatment can be offered if needed. In contrast, a patient with a wired jaw must not vomit¹⁵ and we would consider a multimodal antiemetic strategy similar to that reported by Scuderi,¹⁶ even if the risk for PONV is low.

In conclusion, our latest validation of risk models has demonstrated that simplified risk scores appear to be the best tools to estimate the risk of PONV.

C. C. Apfel
N. Roewer
Würzburg, Germany

- 1 Apfel CC, Kranke P, Eberhart LHJ, Roos A, Roewer N. Comparison of predictive models for postoperative nausea and vomiting. *Br J Anaesth* 2002; **88**: 234–40
- 2 TROPÉ A, Raeder JC. Kan postoperativ kvalme eller oppkast forutsies? *Tidsskr Nor Laegeforen* 2000; **120**: 2423–6 [Note: in figures 2 and 3, figures and legends are incorrectly matched]
- 3 Junger A, Hartmann B, Benson M, 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**: 1203–9
- 4 Eberhart LHJ, Träger M, Nawroth A, Malaka R, Seeling W. Using an artificial neuronal network to predict postoperative nausea and vomiting. A comparison with two conventional risk scores based on logistic regression analysis. *Anesthesiology* 2001; **95**: A1104 [Abstract; published electronically only, downloaded from www.lwwonline.com April 3, 2002]
- 5 Yamaguchi H, Satsumae T, Sakaguchi M, Yamashita S, Yamamoto S. A trial of prediction of postoperative recovery status, pain, and nausea and vomiting in surgical patients using accumulated database. *Anesth Analg* 2001; **92**: S113
- 6 Gan TJ. Postoperative nausea and vomiting—can it be eliminated? *JAMA* 2002; **287**: 1233–6
- 7 Apfel CC, Läärä E, Koivuranta M, Greim CA, Roewer N. A simplified risk score for predicting postoperative nausea and vomiting. *Anesthesiology* 1999; **91**: 693–700
- 8 Apfel CC, Kranke P, Roewer N, Palazzo M, Koivuranta M. Bias in a further model for predicting PONV may not advance current knowledge. *Anesthesiology* 2000; **92**: 1489–91
- 9 Koivuranta M, Läärä E, Snare L, Alahuhta S. A survey of postoperative nausea and vomiting. *Anaesthesia* 1997; **52**: 443–9
- 10 Eberhart LHJ, Högel 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**: 480–8
- 11 Apfel CC, Kranke P, Greim CA, Roewer N. What can be expected from risk scores for predicting postoperative nausea and vomiting? *Br J Anaesth* 2001; **86**: 822–7
- 12 Pierre S, Benais H, Pouymayou J. Apfel's simplified score may favourably predict the risk of postoperative nausea and vomiting: [La cotation d'Apfel simplifiée peut prédire favorablement le risque de nausées et de vomissements postopératoires]. *Can J Anaesth* 2002; **49**: 237–42
- 13 Chung F, Ritchie E, Su J. Postoperative pain in ambulatory surgery. *Anesth Analg* 1997; **85**: 808–16
- 14 Watcha MF. The cost-effective management of postoperative nausea and vomiting [editorial; comment]. *Anesthesiology* 2000; **92**: 931–3
- 15 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

- 16** Scuderi PE, James RL, Harris L, Mims GR, III. Multimodal antiemetic management prevents early postoperative vomiting after outpatient laparoscopy. *Anesth Analg* 2000; **91**: 1408–14