PREDICTION MODELS AND DECISION SUPPORT:

CHANCES AND CHALLENGES

ISBN/EAN: 978-90-393-6314-0

University Medical Center Utrecht, The Netherlands
thesis, with a summary in Dutch

title: Prediction Models and Decision Support: Chances and Challenges
Dutch title: Klinische Voorspelmodellen en Beslissingsondersteuning: Kansen en Uitdagingen

cover & chapter page design by studio ilse van klei / www.ilsevanklei.nl
printed by drukkerij wihabo / www.wihabo.nl

We gratefully acknowledge the financial support by the Netherlands Organisation for Health Research and Development (ZonMW, project number 945–16–202).

Further financial support by the Division of Anesthesiology, Intensive Care and Emergency Medicine and the Julius Center for Health Sciences and Primary Care of the University Medical Center Utrecht.

PREDICTION MODELS AND DECISION SUPPORT:

CHANCES AND CHALLENGES

Klinische Voorspelmodellen en Beslissingsondersteuning: Kansen en Uitdagingen
(met een samenvatting in het Nederlands)

Proefschrift

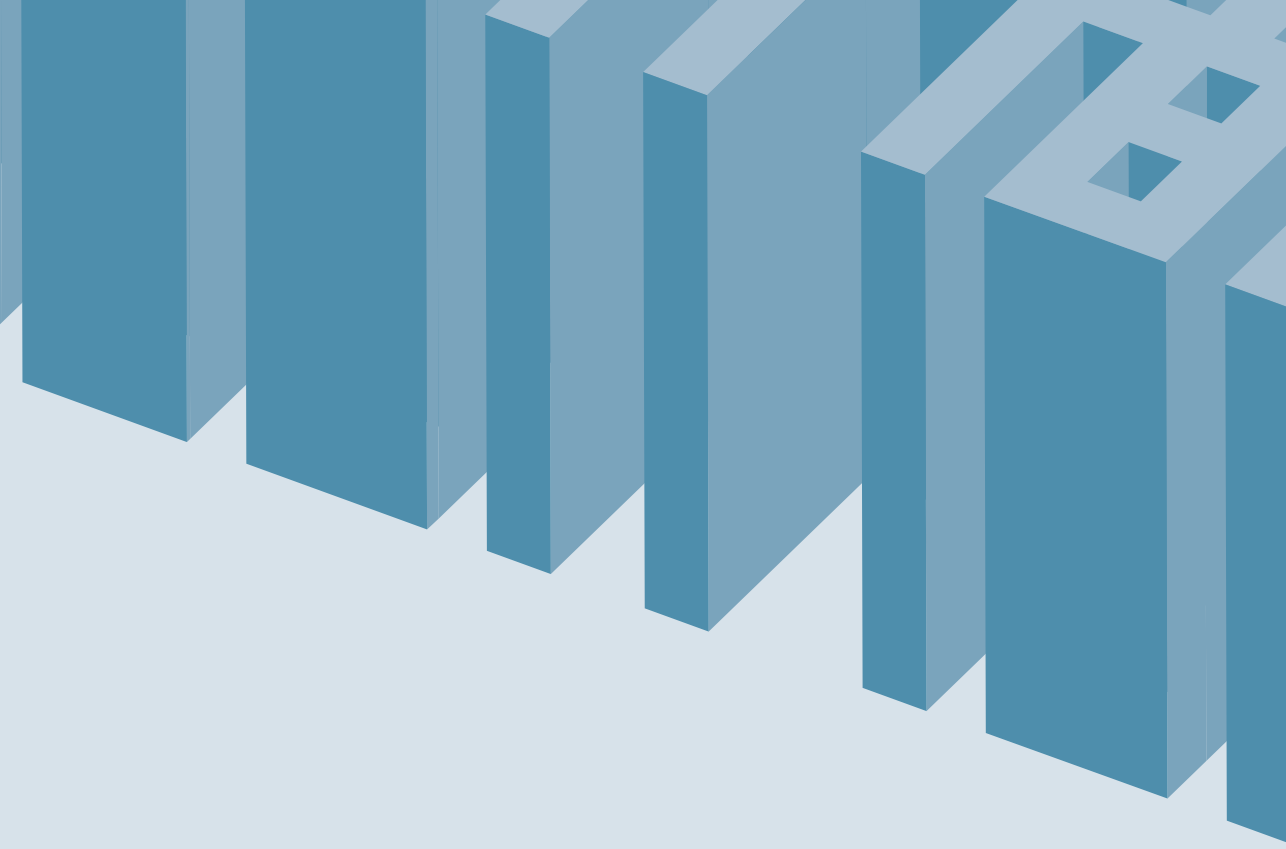
ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 12 mei 2015 des middags te 4.15 uur

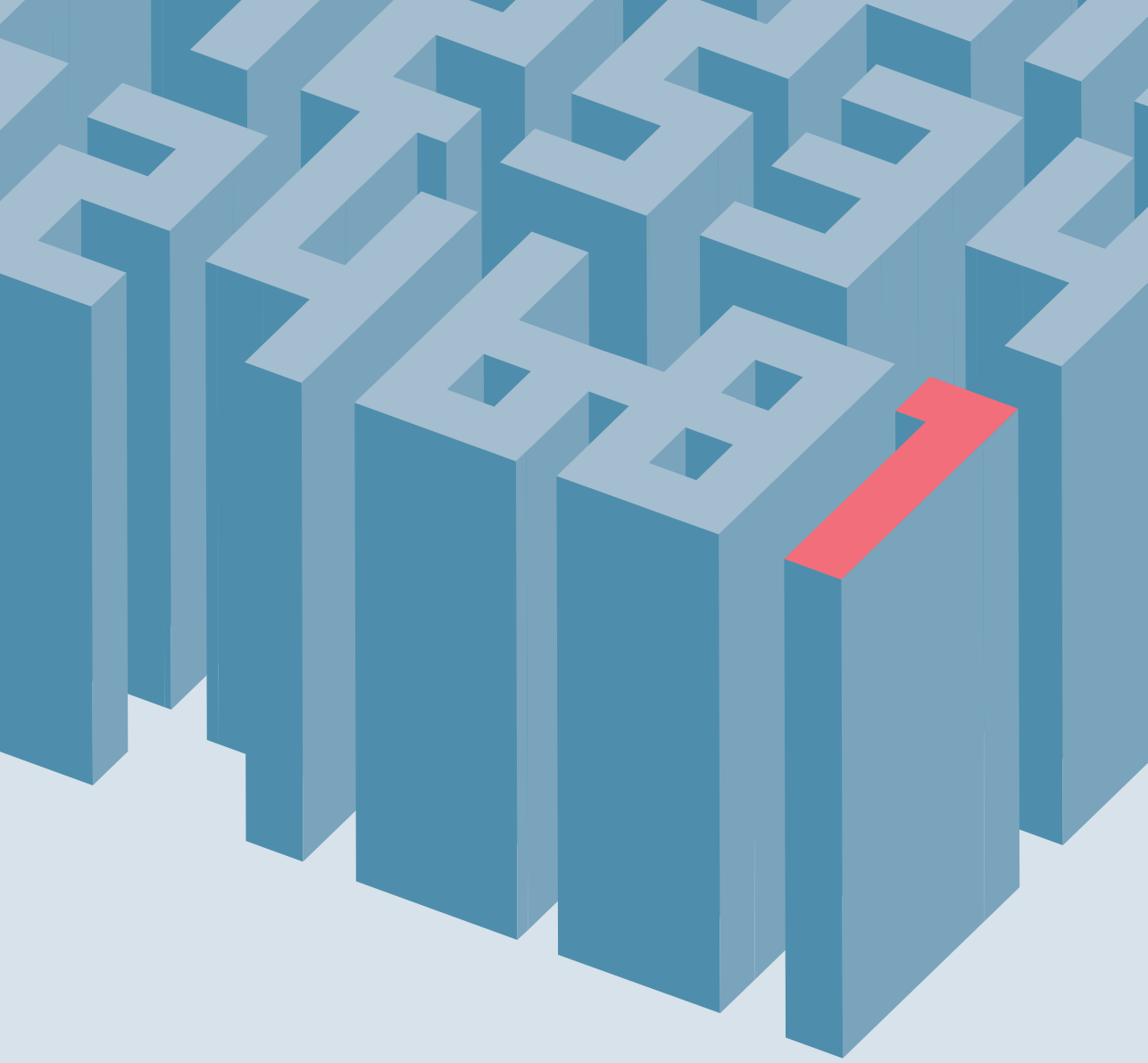
door

Teus Hubertus Kappen
geboren op 19 juli 1981 te Berlicum

Promotoren:	Prof. Dr. C.J. Kalkman
	Prof. Dr. K.G.M. Moons
	Prof. Dr. W.A. van Klei
Copromotor:	Dr. Y. Vergouwe

table of contents





1 INTRODUCTION

When visiting a doctor with signs and symptoms, people who feel ill often want to know which disease they have, i.e. their diagnosis. Second, they want to know what the course of their disease will be when they choose treatment and when they choose for no treatment – in other words, they want to know their prognosis. Diagnosis and prognosis are therefore key elements in medical decision making. A patient's diagnosis determines which interventions are potentially available to treat the disease. The patient's prognosis justifies the actual use of those interventions, weighing the possible benefits of any intervention against its possible harms. When a patient presents with signs and symptoms, choosing the most appropriate next step requires predicting the most likely diagnosis and prognosis. At the heart of medical decision making, physicians consider many combinations of patient and disease characteristics – or predictors – instead of relying on a single characteristic.[1]

prediction models

Prediction models are tools that objectively combine these multiple predictors to aid physicians in making individualized diagnostic and prognostic predictions and to facilitate risk-tailored decision making. Examples of such prediction tools are the Framingham Risk Score for cardiac events, the Ottawa Ankle Rules, and the EuroScore for cardiac surgery risk.[2–4] Multiple predictors are combined into a mathematical model that can be used to predict a diagnostic probability of a current disease state or a prognostic risk of a future disease state.[5,6] In a prediction model development study, patient data is used to select the optimal set of predictors and estimate their relative weights.[7] As the model is made to optimally fit the patient data from the development study, its ability to predict the probabilities or risks of new patients has to be verified in an external validation study.[8,9]

The ability to predict a diagnosis or prognosis in new patients is only a first step towards its actual aim. Even a good predictive performance does not guarantee that the prediction model has value to the clinical judgment of the physicians.[10–12] Whether presenting the predictions of a mathematical model to physicians and patients will improve both decision making and patient outcome may be assessed in so-called 'prediction model impact studies'.[13,14] In such impact studies, the predictions of individual patients are presented to a group of physicians, and their decision making is compared to a group of physicians to whom the predictions are not presented. Although a change in decision making guided by the predictions is required to change any subsequent health outcomes in patients, the ultimate goal of the use of prediction models is to show that their use indeed improves patient outcomes. Therefore it is quite remarkable that a lot of prediction models are being developed, much less are externally validated, and only a few are studied for their impact on decision making – let alone on patient outcome.[13,14]

clinical decision support

As a prediction model is a mathematical formula, it at least requires an interface to communicate its predictions to physicians and patients. This may vary from risk scores that are easy to remember, via paper-based scoring systems, to electronic systems also known as clinical decision support systems or tools. With the increasing availability of computers and

the emergence of electronic patient records, clinical decision support tools have become an appealing format for prediction models. The advantages of an electronic system are abundant: e.g. model calculations require little effort from physicians, additional information may be presented with the model predictions, and several decisional aspects related to these model predictions may be easily integrated into a single system.[15,16] There is now ample evidence that clinical decision support changes decision making.[15,17]

However, similar to prediction models, the claim that clinical decision support improves patient outcome remains unsubstantiated.[17–20] The gap between the effects on decision making and the effects on patient outcome has resulted in various studies that provide many recommendations to improve the uptake and the effectiveness of prediction models and clinical decision support tools.[15,16,19,21–24] Recommendations vary from governance, logistics, training and support, to the design features of the decision support tool. One of such design factors is the addition of therapeutic recommendations to the model predictions (directive implementation format) instead of presenting the model predictions alone (assistive implementation format).[11,15,25,26] Nonetheless, most of the scientific evidence for recommending a directive prediction model format is derived from systematic reviews in which studies with and without specific recommendations were compared for their rate of successful implementation. Such systematic reviews do not provide a direct comparison of an assistive and a directive format for a single prediction model or decision support tool within a single population, and may therefore be subject to bias or generalizability issues. Variations in the design of a single prediction model or decision support tool should thus be compared in a more direct manner, to improve our understanding of the gap between the effects of prediction models on decision making and patient outcome.

case example: postoperative nausea and vomiting

In this thesis the implementation of a prediction model for postoperative nausea and vomiting (PONV) is discussed as an example of a prediction model impact study on physician decision making and actual patient outcomes. PONV is considered an undesirable outcome. For patients undergoing surgery, PONV is very unpleasant. Since PONV does not cause permanent damage or death, but affects – on average – 30% of patients after general anesthesia,[27–29] it is known as the 'big little problem'.[30] Specific antiemetic drugs with a well-established efficacy can be administered during anesthesia to prevent PONV, a practice known as 'PONV prophylaxis'.[31,32] Previously developed and validated prediction models are available to assess individual PONV risks.[33–35] As predicted risk-tailored prophylaxis for PONV is recommended by several guidelines, the conditions were set to use PONV as an example to study the impact of a prediction model in clinical practice.[28,36]

objectives of this thesis

The main objective of this thesis is to study how the effects of the use of a prediction model in clinical practice are achieved. This was studied at four levels:

- The decision making of physicians: the risk-tailored administration of PONV prophylaxis
- Patient outcome: a possible decrease in the incidence of PONV
- Clinical decision support: features of the decision support tool that determine the prediction model and decision support tool effectiveness, in particular using an assistive format (without recommendation) versus a directive format (including recommendation)
- The design and analysis of impact studies

outline of this thesis

Before the prediction model for PONV – that was studied in this thesis – was implemented in the anesthesia practice of a large academic hospital in the Netherlands, the prediction model was first validated. Subsequently, the model was updated – or tailored – to the setting in which it was tested for its impact, to ensure an optimal predictive model performance in that local setting (**chapter 2**). In a cluster-randomized trial of over 12,000 patients, the prediction model for PONV was then implemented and studied for its effects on clinical practice. At the start of anesthesia, physicians who were randomized to the intervention group were 'exposed' to model-based risk predictions of PONV for their individual patients without a corresponding recommendation (assistive format). The decision making of physicians and patient outcomes in the intervention group were compared to that of the 'care-as-usual' control group – i.e. the unexposed physicians and their patients (**chapter 3**). At the end of the cluster-randomized trial, the participating physicians were interviewed and surveyed to understand how the model's risk predictions were perceived by the physicians, and how the predictions influenced their decision making (**chapter 4**). In a prospective before-after study, the prediction model was again implemented, yet this time using a directive format – i.e. the risk predictions included corresponding therapeutic recommendations (**chapter 5**). The lessons that were learned from studying the implementation of a prediction model were summarized into recommendations for the design of future impact studies on prediction models (**chapter 6**). The final chapter provides a general discussion of this thesis, put into the perspective of making decisions in complex decision situations **(chapter 7)**.

references

1. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2):e1001380. PMID:23393429.

2. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J*. 1991;121(1 Pt 2):293–298. PMID:1985385.

3. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med*. 1992;21(4):384–390. PMID:1554175.

4. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg*. 1999;16(1):9–13. PMID:10456395.

5. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338(feb23 1):b375. PMID:19237405.

6. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. (Gail M, Tsiatis A, Krickeberg K, Wong W, Sarnet J, eds.). New York, NY: Springer New York; 2009:1–507. doi:10.1007/978–0–387–77244–8.

7. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338(mar31 1):b604. PMID:19336487.

8. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–698. PMID:22397946.

9. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338(may28 1):b605. PMID:19477892.

10. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009;119(17):2408–2416. PMID:19364974.

11. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(jun04 2):b606. PMID:19502216.

12. Ferrante di Ruffano L, Hyde CJ, Mccaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *Br Med J*. 2012;344(February):1–9. PMID:22354600.

13. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209. PMID:16461965.

14. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. PMID:23393430.

15. Lobach D, Sanders GD, Bright TJ, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. *Evid Rep Technol Assess (Full Rep)*. 2012;(203):1–784. PMID:23126650.

16. Osheroff JA, Teich J, Levick D, et al. *Improving Outcomes with Clinical Decision Support:An Implementer's Guide*. 2nd ed. Chicago, IL: Healthcare Information and Management Systems Society (HIMSS); 2012.

17. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. 2011;18(3):327–334. PMID:21422100.

18. Black AD, Car J, Pagliari C, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med*. 2011;8(1):e1000387. PMID:21267058.

19. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S-A. Computerized clinical decision support for prescribing: provision does not guarantee uptake. *J Am Med Inform Assoc*. 2010;17(1):25–33. PMID:20064798.

20. McGinn CA, Grenier S, Duplantie J, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. *BMC Med*. 2011;9(1):46. PMID:21524315.

21. McGinn CA, Gagnon M-P, Shaw N, et al. Users' perspectives of key factors to implementing electronic health records in Canada: a Delphi study. *BMC Med Inform Decis Mak*. 2012;12:105. PMID:22967231.

22. Ash JS, Sittig DF, Guappone KP, et al. Recommended practices for computerized clinical decision support and knowledge management in community settings: a qualitative study. *BMC Med Inform Decis Mak*. 2012;12(1):6. PMID:22333210.

23. McCormack JL, Ash JS. Clinician perspectives on the quality of patient data used for clinical decision support: a qualitative study. *AMIA Annu Symp Proc*. 2012;2012(Table 1):1302–1309. PMID:23304409.

24. Varonen H, Kortteisto T, Kaila M. What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. *Fam Pract*. 2008;25(3):162–167. PMID:18504253.

25. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. *BMJ*. 2004;328(7435):343–345. PMID:14764503.

26. Kawamoto K, Houlihan C a, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765. PMID:15767266.

27. Fernández-Guisasola J, Gómez-Arnau JI, Cabrera Y, del Valle SG. Association between nitrous oxide and the incidence of postoperative nausea and vomiting in adults: a systematic review and meta-analysis. *Anaesthesia*. 2010;65(4):379–387. PMID:20151955.

28. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2014;118(1):85–113. PMID:24356162.

29. Kappen TH, Moons KGM, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. *Anesthesiology*. 2014;120(2):343–354. PMID:24105403.

30. Kapur PA. The big "little problem." *Anesth Analg*. 1991;73(3):243–245. PMID:1831014.

31. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350(24):2441–2451. PMID:15190136.

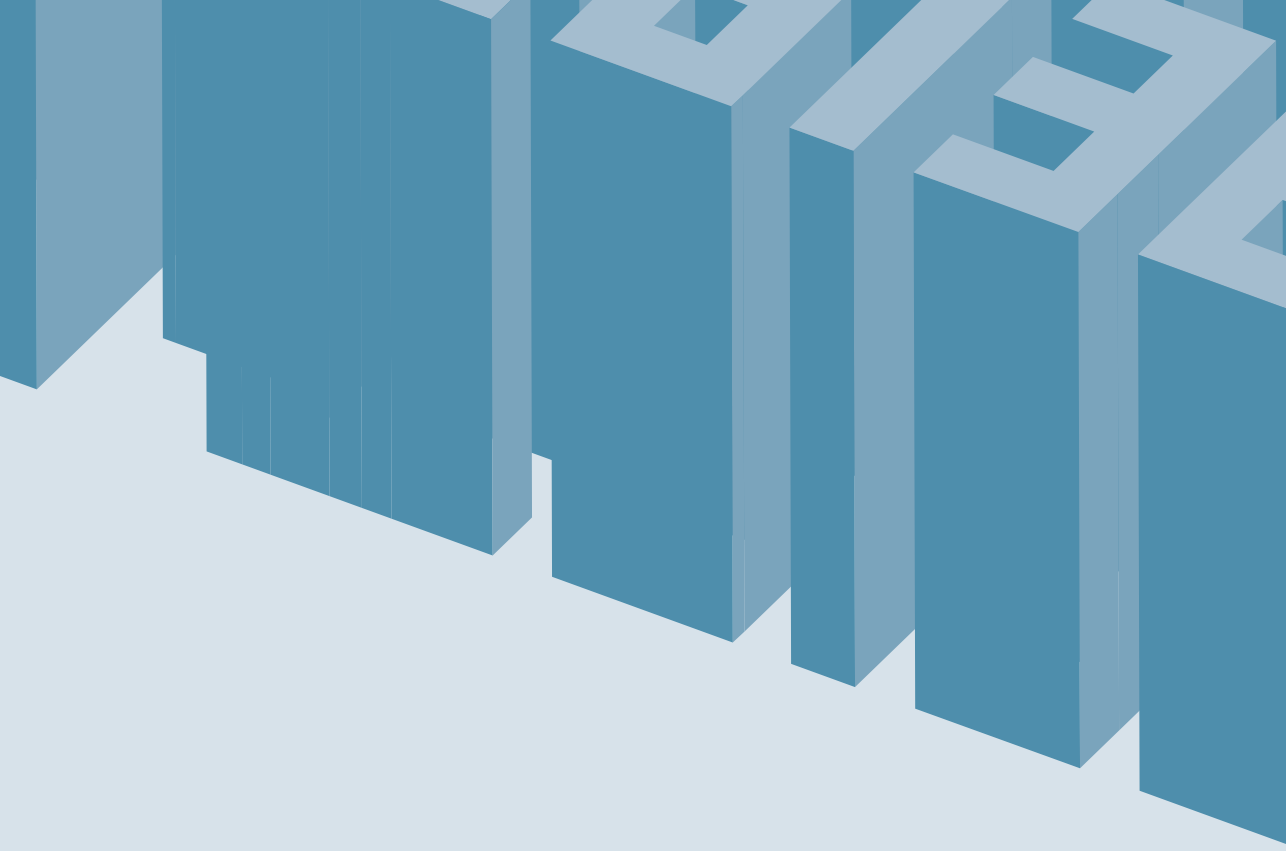
32. Carlisle JB, Stevenson CA. Drugs for preventing postoperative nausea and vomiting. *Cochrane Database Syst Rev*. 2006;3(1):CD004125. PMID:16856030.

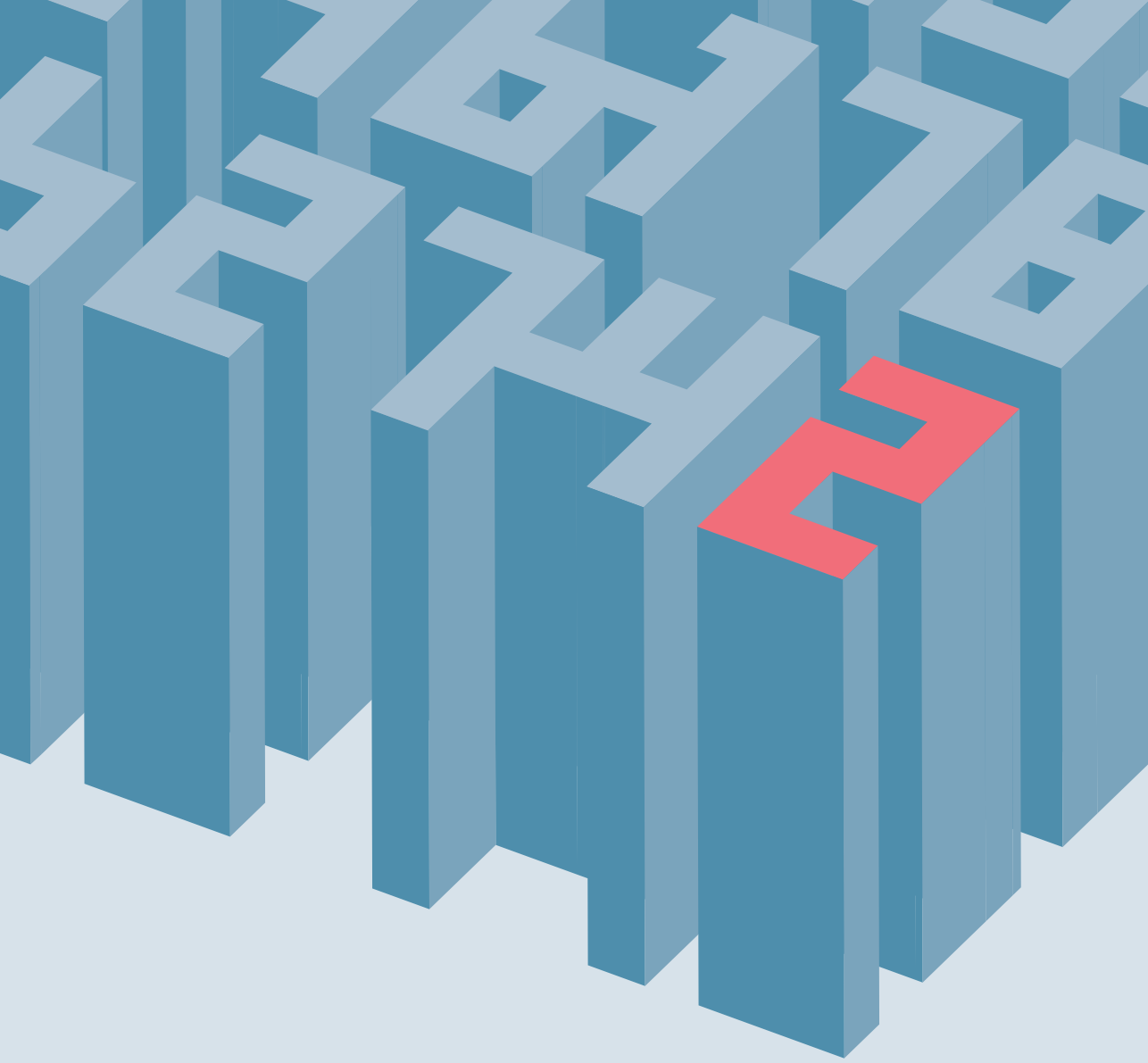
33. Koivuranta M, Läärä E, Snåre L, Alahuhta S. A survey of postoperative nausea and vomiting. *Anaesthesia*. 1997;52(5):443–449. PMID:9165963.

34. 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(3):693–700. PMID:10485781.

35. Van den Bosch JE, Moons KG, Bonsel GJ, Kalkman CJ. Does measurement of preoperative anxiety have added value for predicting postoperative nausea and vomiting? *Anesth Analg*. 2005;100(5):1525–1532. PMID:15845719.

36. Gan TJ, Meyer T a, Apfel CC, et al. Society for Ambulatory Anesthesia guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2007;105(6):1615–1628. PMID:18042859.





ADAPTATION OF CLINICAL PREDICTION MODELS FOR APPLICATION IN LOCAL SETTINGS

T.H. Kappen, Y. Vergouwe, W.A. van Klei, L. van Wolfswinkel, C.J. Kalkman, K.G.M. Moons

Medical Decision Making. 2012 May-Jun;32(3):E1–10

abstract

Background: When planning to use a validated prediction model in new patients, adequate performance is not guaranteed. For example, changes in clinical practice over time or a different case mix than the original validation population may result in inaccurate risk predictions.

Objective: To demonstrate how clinical information can direct updating a prediction model and development of a strategy for handling missing predictor values in clinical practice.

Methods: A previously derived and validated prediction model for postoperative nausea and vomiting was updated using a dataset of 1,847 patients. The update consisted of: 1) changing the definition of an existing predictor; 2) re-estimating the regression coefficient of a predictor; 3) adding a new predictor to the model. The updated model was then validated in a new series of 3,822 patients. Furthermore, we considered several imputation models to handle real-time missing values, in order to anticipate on possible missings in predictor values during actual model use.

Results: Differences in clinical practice between our local population and the original derivation population guided the update strategy of the prediction model. The predictive accuracy of the updated model was better (c-statistic 0.68; calibration slope 1.0) compared to the original model (c-statistic 0.62; calibration slope 0.57). Inclusion of logistical variables in the imputation models, besides observed patient characteristics, contributed to a strategy to deal with missing predictor values at the time of risk calculation.

Conclusions: Extensive knowledge of local, clinical processes provides crucial information to guide the process of adapting a prediction model to new clinical practices.

Prediction models allow clinicians to estimate a patient's prognosis, or a patient's probability of having a particular diagnosis.[1,2] Clinicians are increasingly using prediction models in clinical practice to guide their decision making. Prior to its use, clinicians should decide whether the model will provide accurate predictions for their patients.[1,3–5]

At least two possible scenarios should be considered. First, the local population may differ from the derivation population or the model is no longer up to date with current scientific knowledge. Using the original prediction model without an adaptation to local circumstances may render inaccurate predictions in the new patients. Second, predictor values may be missing when the model is actually being used to calculate risks (real-time missings). For example, the test or device to assess a predictor's value is not available. Simply leaving the predictor out of the model will lead to inaccurate predictions.

In recent methodological literature, strategies have been developed to update an existing prediction model to overcome differences between the derivation and local populations.[6,7] Update techniques aim to improve the performance of the original model in new patients by using additional patient information of the local population. The information that is already available in the original model is retained rather than discarded.[8] Knowledge on differences in case mix or changes in clinical practice over time may allow for specific adjustments to a subset of predictors, in addition to an overall adjustment to the original model.

Further, strategies have been reported to handle real-time missing predictor data when a prediction model is used in new patients.[9] Recently suggested methods to handle real-time missing predictor data are mostly based on statistical techniques. Clinicians – or other health care workers – may be aware of specific reasons why predictor data are missing. For example, logistics can cause missing values when a predictor's value is only available in a paper based format, while the predicted risk is automatically being calculated by an electronic decision support system. Consequently, this local, external information may be combined with statistical techniques to provide accurate estimates of missing predictor values in the local population.

This case-study demonstrates how we used insight in recent developments and local circumstances to optimize a model for predicting postoperative nausea and vomiting (PONV), before using the model in our clinical practice. Specific antiemetic drugs can be administered during anesthesia to prevent PONV, a practice known as 'PONV prophylaxis'.[10] For high risk patients, several antiemetic drugs need to be combined to effectively prevent PONV. A prediction model might enable clinicians to apply a risk-tailored approach for PONV prophylaxis and prevent unnecessary costs and possible side effects, in contrast to administering multiple drugs to all patients.[11–13]

methods

the original PONV prediction model

Previously, a prediction model for PONV occurrence within 24 hours of surgery was derived and internally validated.[14] Data of 1,389 inpatients were used, who were scheduled for elective surgery between 1997 and 1999 in the Academic Medical Center Amsterdam, The

Netherlands (hereafter referred to as derivation dataset). All patients had participated in a randomized trial on differences in PONV incidence after either intravenous or inhalational general anesthesia.[15] The model is shown in Table 2.1, left column. Initially, 7 categories were used to describe data on type of surgery: superficial, laparoscopic, upper abdominal, lower abdominal, strabismus, middle ear, and other. As the risk of PONV was significantly higher for lower abdominal surgery and middle ear surgery compared to all other categories in the derivation dataset, both were combined into high risk surgery versus all other types of surgery.

updating the PONV prediction model

clinical considerations

The decision to update the original model was based on three clinical considerations. First, we compared the predictors in the model with the latest reported evidence on important PONV predictors. Second, we considered possible changes in anesthesia practice since the model derivation. Third, we considered possible differences in case mix between the derivation dataset and our local patients.

According to recent literature, upper abdominal and laparoscopic surgery are procedures that have a risk of PONV similar to lower abdominal surgery, as all three types of procedures are forms of abdominal surgery.[16,17] Upper abdominal and laparoscopic surgery were not included in the original model, because of small numbers in the derivation dataset. Hence, we decided to change the definition of the predictor high risk surgery into 'middle ear surgery or any type of abdominal surgery'.

TABLE 2.1 REGRESSION COEFFICIENTS FOR PREDICTORS AND INTERCEPT OF THE ORIGINAL PREDICTION MODEL, THE UPDATED PREDICTION MODEL, AND THE NEW PREDICTION MODEL

Predictor	Original model	Updated model	Newly developed model
Age (years)	–0.022	–0.017	–0.010
Female gender	0.46	0.36	0.63
Current smoking	–0.63	–0.50	–0.068
History of PONV or motion sickness	0.76	0.60	0.85
Lower abdominal or middle ear surgery	0.61	–	–
Abdominal or middle ear surgery*	–	0.48	0.63
Isoflurane and/or nitrous oxide anesthesia†	0.72	–	–
Inhalational anesthesia‡	–	0.35	0.33
Outpatient surgery	–	–1.16	–1.14
Intercept	0.15	0.12	–0.65

* In the updated model and newly developed model this predictor replaced 'Lower abdominal or middle ear surgery' from the original model. In the updated model and the newly developed model it included lower abdominal, upper abdominal, and laparoscopic surgery in addition to middle ear surgery

† As compared to intravenous anesthesia using propofol

‡ As compared to intravenous anesthesia using propofol. In the updated model and newly developed model this predictor replaced 'Isoflurane and/or nitrous oxide anesthesia' from the original mode

PONV = postoperative nausea and vomiting

The derivation dataset included patients who had surgery in the middle to late 1990s. At that time, nitrous oxide was commonly used as an additive agent to ether derivatives for inhalational anesthesia. Both ether derivatives and nitrous oxide increase the risk of PONV. Currently, nitrous oxide is much less used in the Netherlands, in response to the discovery of several adverse effects including PONV.[18] The main ether derivative of that time, isoflurane, is increasingly being replaced by new, shorter acting inhalational agents such as sevoflurane and desflurane, which are considered to cause less PONV.[19] As predictive effects of inhalational anesthesia could be different at the present time, we decided to include the changes in anesthesia practice in the definition of the predictor inhalational anesthesia and its regression coefficient in the model.

The original prediction model was derived on surgical inpatients only. In general, PONV is less common in outpatient procedures than in inpatient procedures. Consequently, the original PONV model was likely to overestimate the risk of PONV for outpatients. The lower incidence in outpatients is most likely the result of a lower case complexity. However, case complexity is a resultant of several characteristics, which individually are not necessarily predictors of PONV. Outpatient surgery may therefore serve as a proxy variable in which the predictive effects of several of these case characteristics are combined. Since outpatient surgery is currently routine practice in many hospitals including ours, we decided to extend the model with an additional predictor for outpatient surgery.

updating dataset

Using a prospective cohort design, we consecutively included 1,847 elective surgical patients who were operated from June to December 2004 under general anesthesia at the University Medical Center Utrecht, The Netherlands (referred to as update dataset). The study included adult inpatients (*n* = 1,203) and outpatients (*n* = 644). All predictor variables of the original model and the outcome variable PONV within 24 hours were recorded similar to the data collection of the derivation dataset.

update methods

The original prediction model was updated with logistic regression analysis. For every patient in the update dataset, the linear predictor (lp_0) was calculated based on the six regression coefficients (β_i) of the original model and the corresponding predictor values (x_i) of that case ($lp_0 = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_6 x_6$) (see Table 2.1 for predictor variables). According to clinical considerations as described above, the model was updated using recently proposed methods.[6,7] According to recent literature, upper abdominal and laparoscopic surgery would have similar predictive effects of PONV to lower abdominal surgery. Consequently, we assumed that the change in definition of high risk surgery would have the same regression coefficient as compared to the old definition. In contrast, the change in definition of inhalational anesthesia was expected to lower its regression coefficient. Therefore, the model was revised for inhalational anesthesia ($x_{inhalational}$). To include outpatient surgery, the model was extended with an additional predictor (x_{outpat}). The intercept (α) and slope ($\beta_{overall}$) were estimated

to incorporate the overall effect of the new data. This resulted in the updated model with linear predictor (lp_1):

$$lp_1 = \alpha + \beta_{overall} lp_0 + \gamma_{inhalational} x_{inhalational} + \beta_{outpat} x_{outpat} \quad (1)$$

In formula (1), $\gamma_{inhalational}$ represents the deviation from the recalibrated regression coefficient for inhalational anesthesia, and β_{outpat} represents the regression coefficient for the extra predictor outpatient surgery (yes / no). All regression coefficients of the original model were multiplied by $\beta_{overall}$ to result in the updated (recalibrated) regression coefficient. In addition, $\gamma_{inhalational}$ was added to the recalibrated regression coefficient of inhalational anesthesia to result in its final regression coefficient in the updated model.

For comparison purposes, a new model was developed from the update dataset using the same predictor variables as the updated model. However, the regression coefficients for the newly developed model were estimated directly from the update dataset and did not include any information from the original dataset. Consequently, the regression coefficients of the newly developed model represented the 'closest fit' to the update dataset. The new model was used to evaluate the success of adaptation of the updated model to the local setting.

Multiple imputation was used to impute missing data within the update dataset (*aregImpute* from the *Hmisc* package of R and S-plus software). Regression coefficients of formula 1 and regression coefficients of the newly developed model were estimated in 10 completed datasets and averaged.[20–22]

validation of updated model

In the next phase, the updated model and the newly developed model were externally validated in a third cohort of 3,822 elective surgical in- and outpatients, prospectively collected between March 2006 and February 2007 at the University Medical Center Utrecht, The Netherlands (referred to as *external validation dataset*). Following multiple imputation of missing data, ROC curves were plotted to compare the predictive performance of all models within the external validation dataset. Discrimination was further assessed with the c-statistic.[23–25] Calibration within the external validation dataset was assessed with calibration intercept and slope.

handling of real-time missing predictors

When either the updated model or the newly developed model will be used in our local population, various predictor values are likely to be missing on occasion. Similar to dealing with missing values when analyzing the data of a conducted study, missing predictor values should also be dealt with during model application – i.e. in real-time. We developed an imputation strategy that can be used to estimate missing predictor values when applying the prediction model in practice. The strategy contained various imputation models.

At the start of a procedure, i.e. the moment of risk calculation, the predictors age, gender, inhalational anesthesia, and outpatient surgery are always available from the electronic patient record and will never be missing. We derived imputation models for the predictors high risk surgery, smoking status, and history of PONV or motion sickness using data from the update

dataset. Although the planned surgical procedure is always known at the start of the procedure, the fact whether it is a high risk surgery or not requires some additional recoding in order to use it as a predictor in an electronic decision support system. When the information has not yet been recoded at the moment of risk calculation – causing 'high risk surgery' to be missing – the observed values of other predictors alone may not be informative enough to impute missing values on 'high risk surgery'. Two process variables, which indicate the nature of the surgical procedure and which are always electronically available at the start of the procedure, were therefore added: surgical specialty of the treating clinician and operating room location. Surgical specialty provides valuable information whether high risk surgery is being performed. For example, gynecologists often perform abdominal surgical procedures, while ophthalmologists never perform abdominal or middle ear procedures. Consequently, a surgical specialty is considered a risk specialty if either abdominal or middle ear surgery is included in the regular scope of procedures of that specialty.

When surgical specialty is missing as well, the probability of a surgical specialty performing high risk surgery will be estimated from the particular operating room. Surgical specialties are mostly restricted to one or two specific operating rooms. Over 50,000 electronic patient records on procedures in the four years preceding the study were used to determine if an operation room should be considered to accommodate surgical specialties which perform high risk surgery. Subsequently, either the surgical specialty or the probability of a surgical specialty which performs high risk surgery was included into the imputation model for high PONV risk surgery, in addition to the available predictors. The ability to predict surgical specialty from the location of the operating room was assessed from the external validation dataset using the c-statistic, and the calibration intercept and slope.

The imputation strategy as a whole was validated in cases with completed outcome data of the external validation dataset. The three predictors, high risk surgery, smoking status, and history of PONV or motion sickness were imputed. The PONV risk was calculated with the imputed values for all three predictors. The performance of the updated model using our imputation strategy was compared to the performance of the updated model using two alternative strategies to dealing with real-time missing predictors: overall mean imputation (impute the dataset incidence for each predictor value) and ignoring the missing predictor (set predictor value to zero). Reclassification of the imputation strategy in comparison to overall mean imputation was expressed as a category-free net reclassification improvement (NRI (>0)) with bootstrapped 95% confidence intervals (*boot* package of R and S-plus software).[26,27]

results

update of the prediction model

The calibration slope ($\beta_{overall}$) of the original model was 0.79, with an intercept adjustment (α) of –0.03 (see formula 1). The regression coefficient correction for inhalational anesthesia ($\gamma_{inhalational}$) was –0.22. As a consequence, the new regression coefficient for inhalational anesthesia was 0.35 ($0.79 * 0.72 - 0.2$, Table 2.1, right column). The additional reduction of

TABLE 2.2 PATIENT AND PROCEDURAL CHARACTERISTICS OF THE UPDATE DATASET AND THE EXTERNAL VALIDATION DATASET

	Update dataset	External validation dataset
	(*n* = 1847)	**(*n* = 3822)**
Age*, years, mean (SD)	47 (16)	50 (17)
Female sex, *n* (%)	1023 (55)	2059 (54)
Current smoking, *n* (%)	595 (32)	1145 (31)
History of PONV, *n* (%)	452 (25)	737 (28)
Motion sickness, *n* (%)	182 (10)	234 (9)
Use of inhalational anesthesia, *n* (%)	292 (16)	546 (15)
Outpatient surgery, *n* (%)	1126 (61)	1971 (52)
PONV occurrence within 24h, *n* (%)	644 (35)	928 (24)

SD = standard deviation; PONV = postoperative nausea and vomiting

the predictive effect was expected, since newer inhalational agents and the less frequent use of nitrous oxide had reduced the incidence of PONV. Outpatient surgery was less frequently related to PONV than inpatient surgery, as expressed by a regression coefficient of –1.16 (β_{outpat}).

The distributions of the patient characteristics in the update dataset and the external validation dataset were similar for most variables (Table 2.2). Patients from the update dataset were slightly younger of age (mean: 47 years vs. 50 years), received more often inhalational anesthesia (61% vs. 52%), and underwent more often outpatient surgery (35% vs. 24%) compared to patients from the external validation dataset. Predictor values in the update dataset were missing in 23% of the cases for history of PONV, in 18% for motion sickness, and in 1% for current smoking. For the external validation dataset missings occurred in 30% of the cases for history of PONV, in 29% for motion sickness, in 7% for type of surgery, and in 4% for current smoking. Patients from the update dataset suffered less frequently from PONV (35% vs. 47%), which is probably the result of the higher number of outpatients in the update dataset. Outcomes were missing in 6% of the cases in the update dataset and in 25% of the cases in the external validation dataset. The larger percentage of missing outcomes in the external validation dataset was caused by a technical problem in (the logistics of) the data collection.

In the external validation dataset, discrimination was very similar for the updated model and the newly developed model, and lower for the original model (Figure 2.1). Application of the updated model in the patients of the external validation dataset resulted in a c-statistic of 0.68 (95% CI 0.66–0.70) and a calibration slope of 1.00 (95% CI 0.89–1.10) with a calibration intercept of 0.34. When applying the newly developed model in the external validation dataset, the model had a c-statistic of 0.68 (95% CI 0.67–0.70) and a calibration slope of 0.89 (95% CI 0.80–0.99) with a calibration intercept of 0.29. For comparison purposes: the c-statistic of the original model in the external validation set was 0.62 (95% CI 0.60–0.64) and the calibration slope 0.57 (95% CI 0.48–0.66).

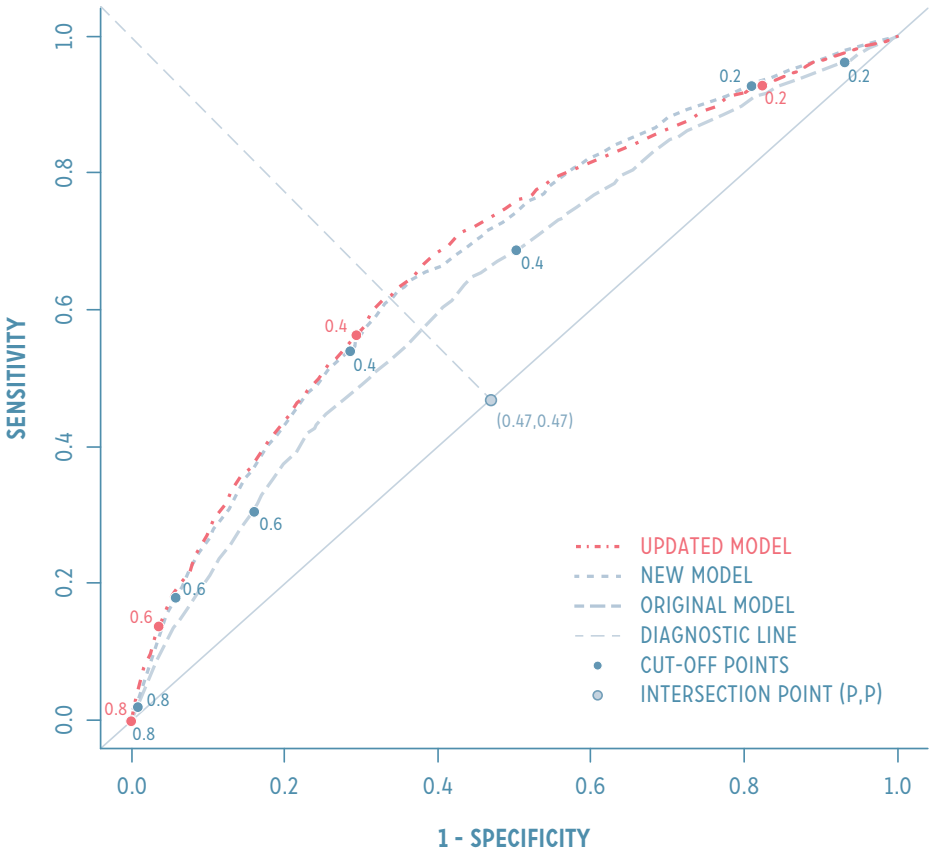


FIGURE 2.1 Receiver Operating Characteristic (ROC) curves of all three prediction models in the external validation dataset. Numbers near cut-off points indicate the predicted risks related to the cut-off points. Point (p,p) is the point with the proportion of patients who experienced PONV (p = 0.47). The point (p,p) and point (0,1) were used to construct the diagnostic line.

handling of real-time missing predictors

The imputation strategy incorporated three imputation models: one model for high risk surgery, one for smoking status, and one for history of PONV or motion sickness (Figure 2.2). First, high risk surgery was imputed, based on observed predictors and additional process variables. As a consequence, the additional information was indirectly incorporated in the imputation of missing values on smoking status and history of PONV or motion sickness. The value of high risk surgery was imputed with a logistic regression model that included age, gender, outpatient surgery, and surgical specialty as variables. When surgical specialty was unavailable, it was estimated from the location of the particular operating room, with a c-statistic of 0.96 (95% CI 0.95–0.96), a calibration slope of 1.03 (95% CI 0.97–1.10), and a calibration intercept of 0.68. Subsequently, missings on smoking status and history of PONV or motion sickness were estimated with imputation models that included the variable high risk surgery,

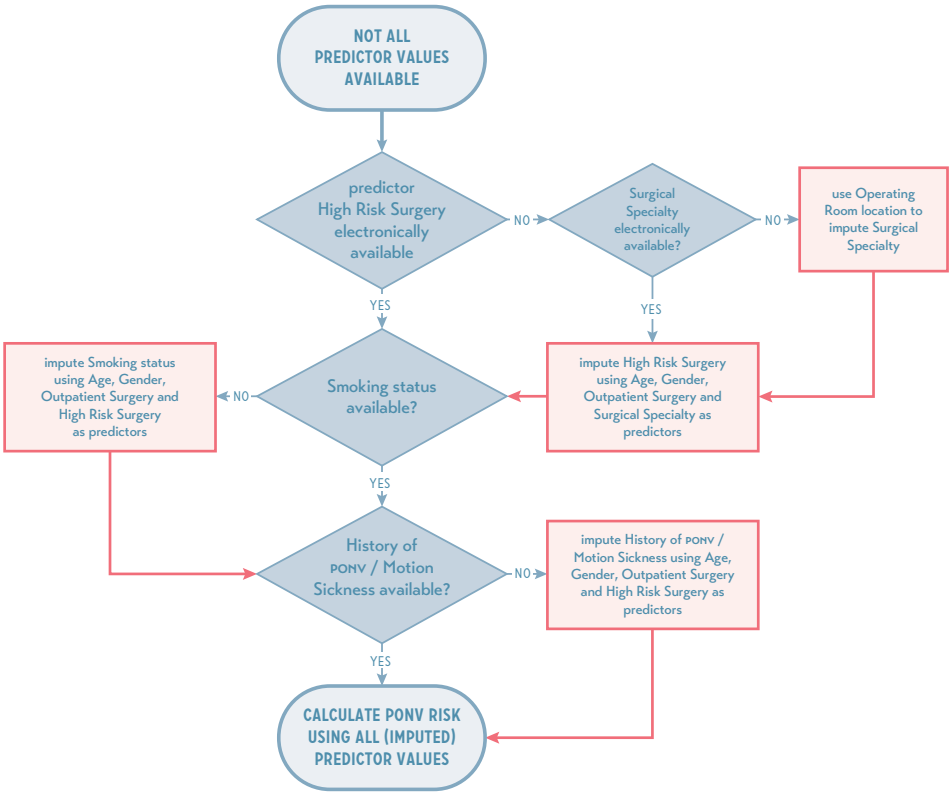


FIGURE 2.2 Flow chart of the imputation strategy for real-time missing predictors. PONV = postoperative nausea and vomiting.

in addition to age, gender, and outpatient surgery. The predictor inhalational anesthesia was not included in any of the imputation models, as a physician generally decides on the type of anesthesia after calculation of the PONV risk. Regression coefficients of the imputation models are not presented, as they reflect the local situation of our hospital. Using the imputation strategy, the updated model showed a c-statistic of 0.67 (95% CI 0.65–0.69), and a calibration slope of 1.00 (95% CI 0.86–1.13) with a calibration intercept of 0.54. For the other two strategies the c-statistic (0.65; 95% CI 0.63–0.67) and calibration slope (0.98; 95% CI 0.84–1.13) were slightly lower, with a small difference in calibration intercept (0.54 for overall mean imputation versus 0.65 for ignoring predictor). In comparison to overall mean imputation, the total NRI (>0) for the imputation strategy was 44% (95% CI 37–51%), with 33% within patients with PONV and 11% within patients without PONV. With the imputation strategy ready, the updated PONV prediction rule was further optimized for future use in our hospital.

discussion

This study demonstrates the great value of information on the local patient population and local clinical practice, at the time that a prediction model is going to be used in that practice. Additionally, potential users of clinical prediction models do not only need to consider possible differences between the derivation population and the local population in terms of case mix, but also newly published evidence that is relevant to their patient population and changes in clinical practice over time. We used local information and scientific knowledge to adapt a prediction model to our local setting, i.e. extending the definition of one of the predictors; adjusting the predictive effects; and adding a new predictor. The updated prediction model showed much better predictive performance in an external validation dataset as compared to the original model. A newly developed model, i.e. without incorporating information of the original model, showed similar discrimination in the external validation dataset as the updated model, but calibration was not as good as the updated model. Further, information on local logistics was used for the development of a real-time imputation strategy when predictor values are missing in clinical practice.

When we considered using a prediction model for PONV in our clinical practice, we encountered three situations that required adjustment of the original prediction model. First, we extended the definition of one predictor, as we expected that the scope of the predictor would be different in our local population compared to the derivation population. We left the predictive effect of the predictor unchanged. In our example, two types of abdominal procedures were known from literature to be important risk procedures, with PONV risks similar to procedures of the original index category of the predictor high risk surgery. Due to an absent association in the derivation dataset, the two procedures had not been included in the original index category. As literature considered the additional procedures to be high risk procedures, we extended the definition of high risk surgery without re-estimating its regression coefficient.

Second, the effect of a predictor was expected to be different for our local population than for the derivation population.[5–8] New inhalational anesthetics had become available since the derivation of the original prediction model. Consequently, the original estimate of the predictor effect would no longer be appropriate, resulting in incorrect risk estimates in our patients. We updated the original regression coefficient of the predictor inhalational anesthesia to include the new inhalational agents into the definition of the predictor.

Third, as our local population included an additional subset of patients at a lower risk of PONV (outpatients), a new predictor was added to the prediction model. We assumed that the new predictor did not disturb proportionality between regression coefficients of the original predictors. The additional predictor enabled us not only to use the prediction model for our inpatient procedures, but also for our outpatient procedures.

Real-time missings of predictor data are likely to be an issue during actual use of any prediction model in daily practice.[9] Relying solely on observed predictor values to estimate the missing predictor values will probably not render the best predictions for those patients.

Clinicians – or other health care workers – may be aware of variables which might serve as proxy for the missing predictor. In our study, logistics were expected to cause missing values for the predictor high risk surgery. Our strategy for real-time imputation of missing predictors prevented a drop of 0.02 in the c-statistic (0.67 vs. 0.65), as compared to overall mean imputation and ignoring the predictor. This 0.02 drop in c-statistic may be considered a relevant decrease in the accuracy of predicted risks of individual patients, since the c-statistic is relatively insensitive. The high net reclassification improvement confirms this result.[26,27] Hence we believe that developing the imputation strategy was worth the effort, as it is likely that such a strategy improves PONV risk prediction by using 'the best guess' for that patient's value using proxy surgery variables.[9]

The goal of updating the PONV prediction model was to adapt the original model to the current clinical practice of the anesthesia department in our hospital. We successfully adapted the model to our clinical practice, as was confirmed by the predictive performance of the updated model in the external validation sample. Discrimination of the updated model was similar to previous validation studies of (other) PONV prediction models (c-statistic of 0.68) and was better than discrimination of the original model (c-statistic of 0.62).[28–30] Moreover, the calibration slope of the updated model was 1.00, whereas it was only 0.57 for the original model. The importance of a better calibrated model is reflected in the use of the model in clinical practice. Using a patient's individual PONV risk as predicted by the prediction model, a clinician may decide to administer specific antiemetic drugs during the procedure which may reduce the risk of PONV. Each class of these antiemetic drugs reduces the risk by a relative 25 percent, with an additive effect for each additional class administered.[10] It is important for clinicians to accurately know their patients' individual PONV risks to weigh the benefit of administering one or more antiemetic drugs against possible side effects of the drugs. Unfortunately, the updated prediction model systematically underestimated the actual PONV risk in the external validation dataset (reflected by a calibration intercept of 0.34). An important reason for the systematic miscalibration is the difference in PONV incidence between the external validation dataset and the update dataset ($\log([0.47 / 0.53]/[0.35 / 0.65]) = 0.50$ on the log-odds scale). The calibration intercept of 0.34 is less than the 0.50 of the crude incidence difference on the log-odds scale. As the calibration intercept did not fully cover the crude incidence difference, the difference in PONV incidence was partly resolved by the predictors in the model (0.50–0.34 = 0.16 on the log-odds scale). Since it is easy to adjust for underestimation by adapting the model intercept,[31,32] the updated model is preferred over the original model.

Although it might seem intuitive that a prediction model, specifically developed in a local setting, will perform better in the local population than an update or adjustment of an existing model to that local situation, the updated model might actually be preferred over a newly developed model because of improved generalizability.[5–7] This is also reflected by our results. In new patients of the same institute (the external validation dataset) the updated model actually performed better than the newly developed model, notably in terms of calibration. A second indication of improved generalizability of the updated model over the newly developed model can be observed from the absence of a predictive effect in the new dataset for

'current smoking'. As current smoking is a consistent predictor of PONV throughout literature,[12] the absence of a predictive effect of current smoking seems implausible.

This case-study is an example of how a prediction model can be adapted to local circumstances. One limitation is that the required adaptations will not be the same for all prediction models and clinical settings. For example, differences in case mix may specifically occur between derivation and local populations, whereas all prediction models will be subject to some degree of change in clinical practice over time. Second, this study does not provide an exhaustive summary of all possible techniques to enhance updating of a prediction model or to handle missing predictor data when applying a prediction model in real-time. Alternative methods to handle real-time missing predictor data have recently been summarized.[9] Third, we updated the PONV prediction model to derive accurate predictions for our own patients. We cannot conclude from this study whether use of the updated model in clinical practice will be successful. A formal impact study should be performed to evaluate the effects of the prediction model on clinical practice and patient outcome.[5,33,34] Finally, individual patient data were required to update the prediction model and to develop the imputation strategy. Owing to a previous study within our hospital, data were already available at the time we planned to use the model. Still, the update dataset was too small to develop imputation models for high risk surgery using the proxy variables, and required an additional dataset to estimate surgical specialty.

To conclude, the intention to use a prediction model in new patients should trigger researchers and clinicians to decide whether they indeed expect an accurate performance of the prediction model in their local population. In our example, we expected problems with generalization of the original prediction model to our patients, despite previous validations of the model. An update of the model proved most valuable to improve its performance in our current population, and a strategy to handle missing predictor data enabled us to actually use the model in clinical practice, even when a predictor value would be missing. As these problems with generalization and application of clinical prediction models are likely to be common, it is conceivable that any prediction model will need to be optimized before it is actually used in a new setting. When optimization is indeed required, a sufficiently large dataset from the local population should be readily available, or will have to be collected. Clinical insights into the lack of generalization will have to be combined with statistical methods to optimize a prediction model's performance in future patients.

references

1. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209. PMID:16461965.

2. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338(feb23 1):b375. PMID:19237405.

3. Justice a C, Covinsky KE, Berlin J a. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515–524. PMID:10075620.

4. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338(may28 1):b605. PMID:19477892.

5. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(jun04 2):b606. PMID:19502216.

6. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: A study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567–2586. PMID:15287085.

7. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol*. 2008;61(1):76–86. PMID:18083464.

8. Janssen KJM, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KGM. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth*. 2009;56(3):194–201. PMID:19247740.

9. Janssen KJM, Vergouwe Y, Donders ART, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem*. 2009;55(5):994–1001. PMID:19282357.

10. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350(24):2441–2451. PMID:15190136.

11. Watcha MF, White PF. Postoperative nausea and vomiting. Its etiology, treatment, and prevention. *Anesthesiology*. 1992;77(1):162–184. PMID:1609990.

12. Gan TJ, Meyer T a, Apfel CC, et al. Society for Ambulatory Anesthesia guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2007;105(6):1615–1628. PMID:18042859.

13. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2014;118(1):85–113. PMID:24356162.

14. Van den Bosch JE, Moons KG, Bonsel GJ, Kalkman CJ. Does measurement of preoperative anxiety have added value for predicting postoperative nausea and vomiting? *Anesth Analg*. 2005;100(5):1525–1532. PMID:15845719.

15. Visser K, Hassink E a, Bonsel GJ, Moen J, Kalkman CJ. Randomized controlled trial of total intravenous anesthesia with propofol versus inhalation anesthesia with isoflurane-nitrous oxide: postoperative nausea with vomiting and economic analysis. *Anesthesiology*. 2001;95(3):616–626. PMID:11575532.

16. Ahmed AB, Hobbs GJ, Curran JP. Randomized, placebo-controlled trial of combination antiemetic prophylaxis for day-case gynaecological laparoscopic surgery. *Br J Anaesth*. 2000;85(5):678–682. PMID:11094579.

17. Tseng L-H, Liou S-C, Chang T-C, Tsai S-C, Soong Y-K, Wong S-Y. A randomized blinded study of the incidence of postoperative nausea and vomiting in women after major gynecologic laparoscopic surgery. *J Minim Invasive Gynecol*. 2006;13(5):413–417. PMID:16962524.

18. Myles PS, Leslie K, Chan MT V, et al. Avoidance of nitrous oxide for patients undergoing major surgery: a randomized controlled trial. *Anesthesiology*. 2007;107(2):221–231. PMID:17667565.

19. Philip BK, Kallar SK, Bogetz MS, Scheller MS, Wetchler B V. A multicenter comparison of maintenance and recovery with sevoflurane or isoflurane for adult ambulatory anesthesia. The Sevoflurane Multicenter Ambulatory Group. *Anesth Analg*. 1996;83(2):314–319. PMID:8694311.

20. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59(10):1087–1091. PMID:16980149.

21. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons; 1987:320. doi:10.1002/9780470316696.

22. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092–1101. PMID:16980150.

23. Cox DR. Two Further Applications of a Model for Binary Regression. *Biometrika*. 1958;45(3–4):562–565. doi:10.1093/biomet/45.3-4.562.

24. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387. PMID:8668867.

25. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Validity of prognostic models: when is a model clinically useful? *Semin Urol Oncol*. 2002;20(2):96–107. PMID:12012295.

26. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–138. PMID:20010215.

27. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11–21. PMID:21204120.

28. Eberhart LH, 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(4):480–488. PMID:10757586.

29. Apfel CC, Kranke P, Eberhart LHJ, Roos A, Roewer N. Comparison of predictive models for postoperative nausea and vomiting. *Br J Anaesth*. 2002;88(2):234–240. PMID:11883387.

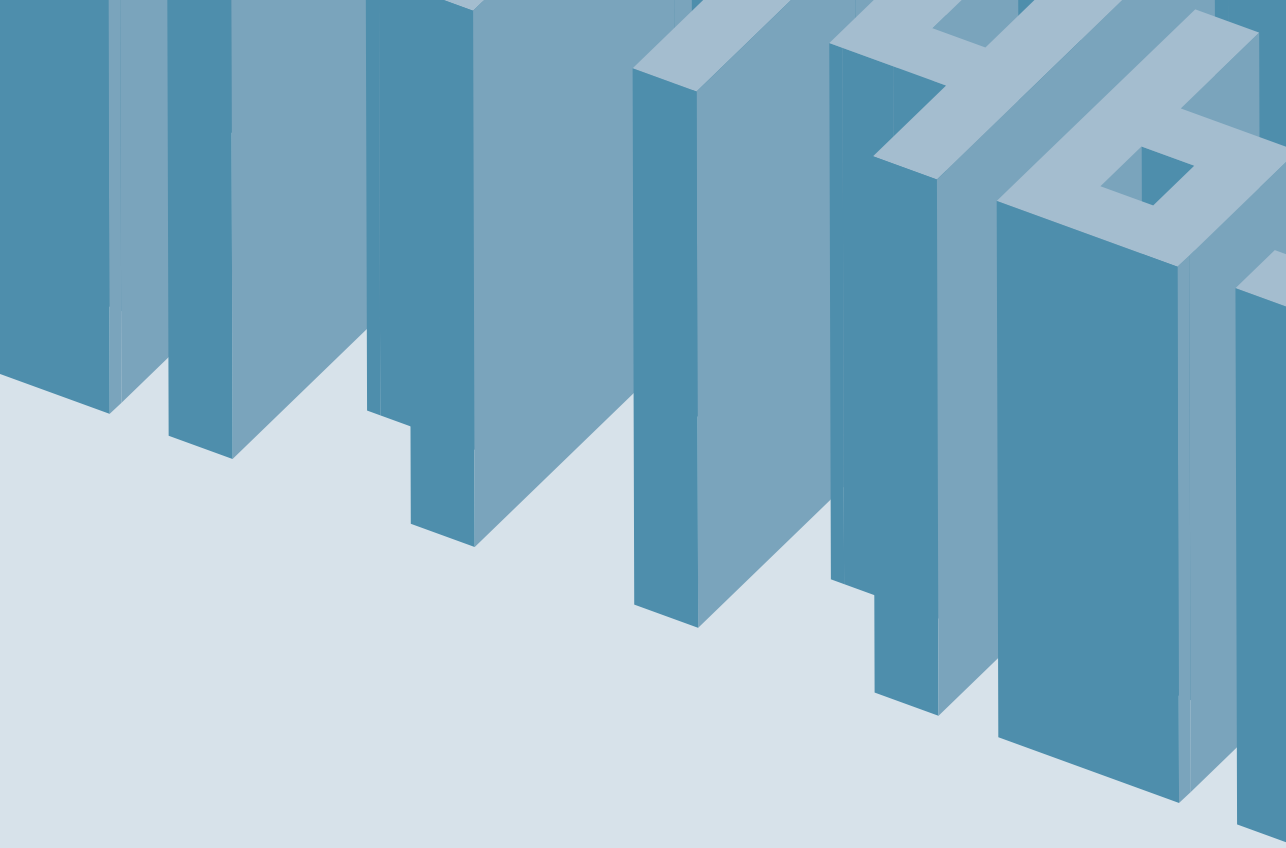
30. Van den Bosch JE, Kalkman CJ, Vergouwe Y, et al. Assessing the applicability of scoring systems for predicting postoperative nausea and vomiting. *Anaesthesia*. 2005;60(4):323–331. PMID:15766334.

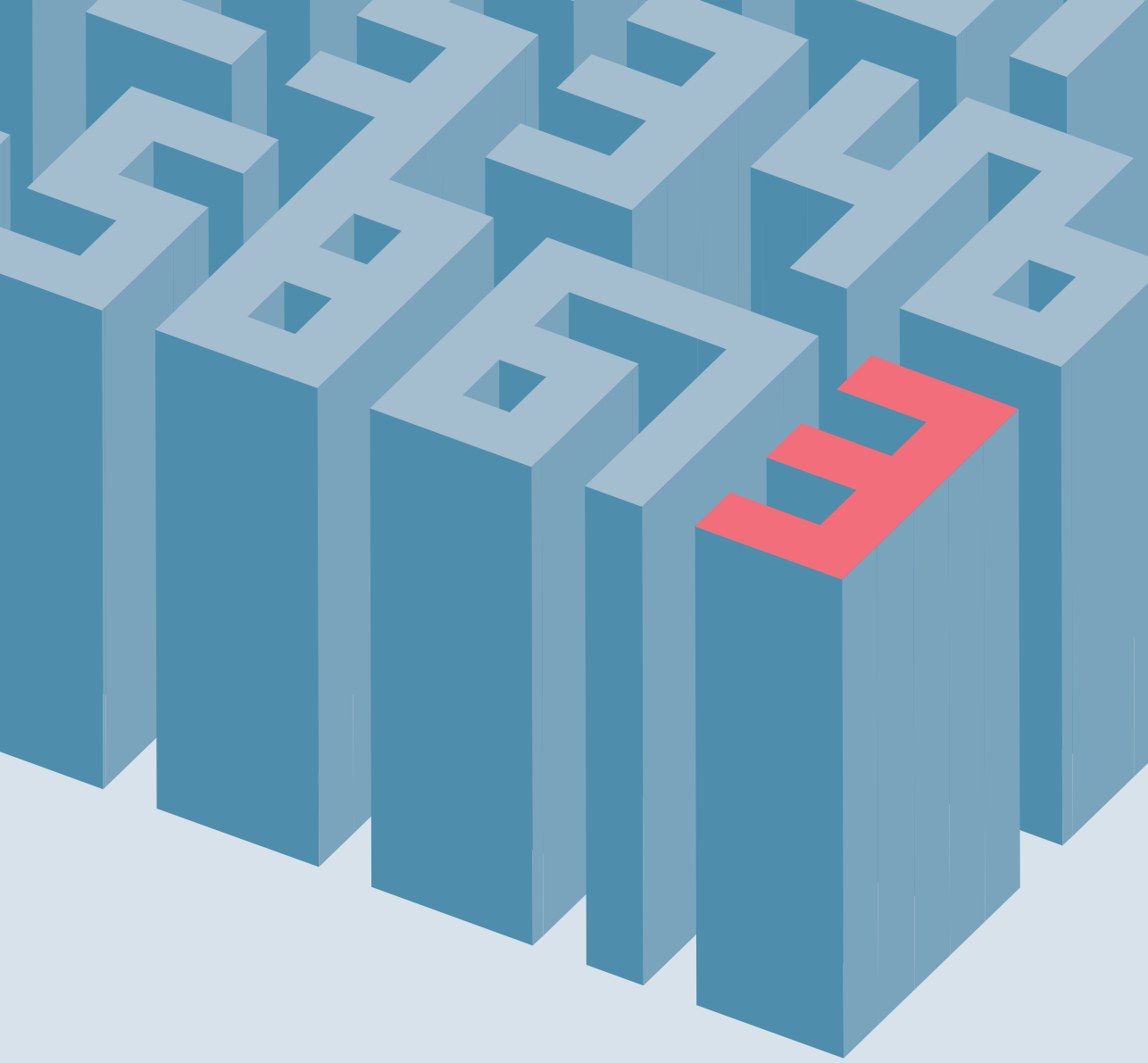
31. Ivanov J, Tu J V., Naylor CD. Ready-made, recalibrated, or Remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery. *Circulation*. 1999;99(16):2098–2104. PMID:10217648.

32. Van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401–3415. PMID:11122504.

33. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009;119(17):2408–2416. PMID:19364974.

34. Moons KGM. Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem*. 2010;56(4):537–541. PMID:20133890.





IMPACT OF RISK ASSESSMENTS ON PROPHYLACTIC ANTIEMETIC PRESCRIPTION AND THE INCIDENCE OF POSTOPERATIVE NAUSEA AND VOMITING: A CLUSTER-RANDOMIZED TRIAL

T.H. Kappen, K.G.M. Moons, L. van Wolfswinkel, C.J. Kalkman, Y. Vergouwe, W.A. van Klei

Anesthesiology. 2014 Feb;120(2):343–54

abstract

Background: Clinical prediction models have been shown to have moderate sensitivity and specificity, yet their utility will depend on implementation in clinical practice. We hypothesized that implementation of a prediction model for postoperative nausea and vomiting (PONV) would lower the PONV incidence by stimulating anesthesiologists to administer more 'risk-tailored' prophylaxis to patients.

Methods: A single-center cluster-randomized trial was performed in 12,032 elective surgical patients receiving anesthesia from 79 anesthesiologists. Anesthesiologists were randomized to either exposure or non-exposure to automated risk calculations for PONV (without patient-specific recommendations on prophylactic antiemetics). Anesthesiologists who treated less than 50 enrolled patients were excluded during the analysis to avoid too small clusters, yielding 11,613 patients and 57 anesthesiologists (intervention group: 5,471 and 31; care-as-usual group 6,142 and 26). The 24-hours incidence of PONV (primary outcome) and the number of prophylactic antiemetics administered per patient were studied for risk-dependent differences between allocation groups.

Results: There were no differences in PONV incidence between allocation groups (crude incidence intervention group 41%, care-as-usual group 43%; OR 0.97, 95% CI 0.87–1.1; risk-dependent OR 0.92, 95% CI 0.80–1.1). Nevertheless, intervention group anesthesiologists administered more prophylactic antiemetics (rate ratio 2.0, 95% CI 1.6–2.4) and more risk-tailored than care-as-usual group anesthesiologists (risk-dependent rate ratio 1.6, 95% CI 1.3–2.0).

Conclusions: Implementation of a PONV prediction model did not reduce the PONV incidence, despite increased antiemetic prescription in high-risk patients by anesthesiologists. Before implementing prediction models into clinical practice, implementation studies that include patient outcomes as an endpoint are needed.

To support physicians in their decision making, clinical guidelines increasingly include prediction models or risk scores in their recommendations.[1,2] The growing popularity of prediction models can be explained by their perceived objectivity in modeling complex interactions to predict a patient's risk, in contrast to physicians' clinical judgments, which are typically heuristic and harder to deconstruct. The effects of a prediction model on clinical processes and patient outcomes should be quantified in a so-called 'impact analysis', where the model is implemented in daily practice and compared to care-as-usual.[3–6]

Current guidelines on the management of postoperative nausea and vomiting (PONV) recommend risk-tailored, prophylactic treatment based on risk estimates from a prediction model, to prevent unnecessary costs and possible side effects, in contrast to administering multiple drugs to all patients.[7,8] Although several PONV prediction models are available,[9–12] their actual impact on clinical practice is still being questioned.[13–15] Several studies demonstrated improved guideline adherence when a PONV prediction model was implemented, although their effect on clinical practice was limited.[16–18] However, comparative randomized studies assessing the actual impact of risk-dependent prophylaxis on the incidence of PONV are rare. Without such studies one still cannot be confident that PONV prediction models will outperform clinical judgment and improve patient outcomes.

The current study was a cluster-randomized trial, in which we compared a group of physicians who were randomly 'exposed' to model-based risk estimates of PONV for their patients, to a control group of physicians who provided care-as-usual. We hypothesized that systematic implementation of a validated PONV prediction model would result in improved patient outcome by lowering the incidence of PONV, as a result of an increase in risk-tailored antiemetic prophylactic treatment by physicians. Hence, we also used the present experience to address advantages and disadvantages of such impact studies.

methods

design and participants

Between March 16, 2006 and December 21, 2007, a single-center, cluster-randomized trial was performed to study the effects of implementing a prediction model for PONV on the incidence of PONV and on the administration of antiemetic prophylaxis (clinicaltrials.gov; identifier NCT00293618). To prevent contamination and therefore dilution of the implementation effect, the study was cluster-randomized on the physician level rather than patient level.[3,19,20] Anesthesiologists of a Dutch university hospital (University Medical Center (UMC) Utrecht) were randomly assigned to exposure to predicted risks of PONV as calculated by the prediction model (intervention group) or not (care-as-usual group).

All anesthesiologists and senior residents, henceforth referred to as 'anesthesiologists', were enrolled in the study and randomized by one of the authors. Case mixes among anesthesiologists were expected to differ due to differences in their professional profile, such as differences in experience level or anesthesia sub-specialties. Therefore, at the start of the study permuted-block randomization was used (block size according to the strata; 1:1

allocation ratio, using PASS software) to stratify on anesthetic experience (senior resident, junior attending, senior attending) and anesthesia sub-specialty (no sub-specialty, cardiac anesthesia, pediatric anesthesia). Anesthesiologists could enter or leave the study due to initiation or termination of their employment. Anesthesiologists who entered after the start of the study were randomized only in a stratified way when a sufficiently large block was available, otherwise simple randomization was applied. Allocation sequences were generated by the second author (KGMM), who was not involved in patient care, and automatically assigned to the name of the anesthesiologist to ensure concealment. The generated list of the assignment of anesthesiologists was then given to the first author (THK) who enrolled and informed the anesthesiologists.

All adult patients undergoing general anesthesia for elective, non-cardiac surgery who had visited the outpatient preoperative evaluation clinic were eligible for this study. For elective surgery 98% of patients typically visit the preanesthesia evaluation clinic prior to their procedure. Exclusion criteria were pregnancy, postoperative admission to the intensive care unit, overnight ventilation at the postanesthesia care unit, and inability to communicate in Dutch or English. At the start of the study, patients undergoing intracranial surgery were no longer transferred to the intensive care unit and did no longer require postoperative mechanical ventilation, in contrast to when the study protocol was written. Therefore, the exclusion criterion for intracranial surgery in the study protocol was changed to admission to the intensive care unit.

All eligible patients from the time of study initiation were automatically included using the anesthesia information management system. During the enrollment phase of the study it became apparent that some anesthesiologists would treat no or only a few patients. Consequently, the number of patients for some anesthesiologists would be too small for an analysis with the anesthesiologists as clusters, for mixed effects regression models would not converge. Therefore, the study protocol was amended by excluding anesthesiologists from the analysis when they treated less than 50 enrolled patients during the entire study period to enable cluster-based analysis of the trial as it was originally planned.

According to Dutch law, research protocols that do not subject patients to a particular treatment or that require them to behave in a particular way, do not apply to the Medical Research Involving Human Subjects Act. As the decision support tool in our study protocol only provided evidence-based information to physicians, the institutional ethical review board waived the need for individual informed consent and approved the study protocol (Medical Ethics Review Board, UMC Utrecht, 05–288).

the prediction model

The implemented prediction model was originally developed in a population of a different university hospital in the Netherlands and had already been externally validated.[21] The model was subsequently updated and optimized for implementation in the UMC Utrecht, where the present study took place.[22] The model consisted of seven predictor variables: age; sex;

current smoking; type of surgery; inhalational anesthesia; ambulatory surgery; and history of motion sickness or PONV (for full model description see Table 3.1).

intervention

intervention group

To study the effect of systematically presenting the patient's PONV risk to the responsible anesthesiologist, we implemented the prediction model as an 'assistive' decision support tool. The decision support tool was integrated into our custom-made anesthesia information management system (AnStat software, CarePoint Nederland BV, Ede, The Netherlands), written by one of the authors (LVW). In our hospital, the individual names of each anesthesia team member are registered in the anesthesia information management system at the start of the anesthetic case. When the anesthesiologist was part of the intervention group, the decision support tool presented the patient's calculated PONV risk on the computer screen of the anesthesia information management system during the rest of the anesthetic case. Anesthesiologists then decided if, which and how many prophylactic antiemetics would be administered in view of the patient's individual risk. The presented risk was thus not accompanied with a specific therapeutic recommendation.

Anesthesiologists of the intervention group were provided with several consecutive educational sessions before patient enrollment, at the start and throughout the study

TABLE 3.1 REGRESSION COEFFICIENTS OF THE PONV PREDICTION MODEL

Predictor	Updated model
	Odds Ratio (95% CI)
Age (years)	0.98 (0.98–0.99)
Female sex	1.44 (1.14–1.82)
Current smoking	0.61 (0.48–0.77)
History of PONV / motion sickness	1.82 (1.44–2.31)
Surgery with a high PONV risk*	1.62 (1.14–2.30)
Inhalational anesthesia†	1.42 (1.12–1.79)
Outpatient surgery	0.31 (0.24–0.41)
Intercept‡	1.13 (0.73–1.74)
Model performance characteristics§	
Model discrimination as C-statistic (95% CI)	0.68 (0.66–0.70)
Calibration slope (95% CI)	1.00 (0.89–1.10)
Calibration intercept	0.34

Probability of PONV as estimated by the model = 1 / (1 + exp (–(0.12 – 0.017*age + 0.36*female sex – 0.50*current smoking + 0.60*history of PONV or motion sickness + 0.48*surgery with a high PONV risk + 0.35*inhalational anesthesia – 1.16*outpatient surgery)))

* Definition of this predictor was 'abdominal or middle ear surgery'

† As compared to intravenous anesthesia using propofol

‡ For the intercept, the second column represents the baseline odds not the odds ratio

§ Model performance was validated in a subset of patients (between March 2006 and February 2007) treated by anesthesiologists of the care-as-usual group.[22]

CI = Confidence Interval; exp = exponential; PONV = postoperative nausea and vomiting

period. These sessions aimed to inform the intervention group about the study background, how the prediction model estimated a patient's individual risk of PONV, and how the local protocol on antiemetic prophylaxis could be used according to that predicted risk. The local protocol was based on the six-factorial trial of Apfel and consisted of the dosage, timing and efficacy of prophylactic antiemetic drugs and the use of total intravenous anesthesia (see section *Methods, Outcome and Follow-up*, fifth paragraph).[23] Although the efficacy of different antiemetics strategies was discussed during the educational sessions, there was no specific recommendation made on how many antiemetic interventions should be applied at a particular predicted PONV risk.

Anesthesiologists of the intervention group were informed of the allocation status of their colleagues to promote discussion among anesthesiologists of the intervention group on how to use the model and its predictions. They were instructed to avoid discussing PONV with anesthesiologists randomized to the care-as-usual group. To enable anesthesiologists of the intervention group to reflect on their individual prophylactic management of PONV, they received individualized feedback by email after the first twelve months of study, which had been planned prior to the start of the study. This included the incidence of PONV among the patients they had treated, the overall (hospital wide) incidence of PONV and the quantity of antiemetics administered.

care-as-usual group

Anesthesiologists of the care-as-usual group were not exposed to the patient's calculated PONV risk. At the start of the study, they only were informed about the goal of the study and their randomization status. Although anesthesiologists of the care-as-usual group were not actively informed of the allocation status of their colleague physicians, additional masking was deemed impossible. As antiemetic management was not standardized in any of the two allocation groups, anesthesiologists of the care-as-usual group were simply asked to manage PONV as usual. At that time, the existing, local protocol for administration of PONV prophylaxis only included a preferable order for antiemetic drugs, their dosage and timing of administration (see section *Methods, Outcome and Follow-up*, fifth paragraph).

outcome and follow-up

As recommended in various guidelines of prediction-model impact studies,[3,19] implementation of the prediction model was studied in two steps: the effects of the prediction model on patient outcome (the incidence of PONV) as the primary outcome, and the change in physician behavior (administration of risk-dependent PONV prophylaxis) caused by the prediction model as the secondary outcome. The other secondary outcomes as stated in the clinical trial registration – i.e. cost-effectiveness and attitudes of physicians towards prediction models – were considered to be beyond the scope of this article and will be discussed in subsequent papers.

The primary outcome PONV was defined as the occurrence of at least one of the following events within the first 24 hours after surgery: an episode of nausea, an episode of vomiting, or the administration of any rescue antiemetic. For nausea, the patient was asked to rate their feeling of nausea on a three-point verbal rating scale (no / yes, a bit / yes, definitely) and for the

analysis the variable was dichotomized to any nausea (no / yes). Vomiting was defined as the expulsion of gastric contents and was recorded as a binary outcome (no / yes). Research nurses collected data on the occurrence of postoperative nausea using a validated questionnaire.[12,24]

Data were collected at the postanesthesia care unit (30 min. and 60 min. after arrival, and when leaving the unit), and after 24 hours post-surgery on the ward, or by phone when patients had already been discharged. The outcome variable for PONV was coded as missing when any of the follow-up measurements had not been completed. Although research nurses were unlikely to be aware of a patient's allocation status due to high patient volumes, active masking of allocation was deemed impossible and therefore not performed.

The change in physician behavior caused by the prediction model was defined as the difference in the rate of administration of PONV prophylaxis between study groups. The rate of administration of PONV prophylaxis was defined as the number of interventions that an anesthesiologist applied to a patient with the aim to prevent PONV, i.e. the number of prophylactic antiemetics per patient. Administration of ondansetron, droperidol, dexamethasone, or a combination as well as selecting total intravenous anesthesia instead of inhalational anesthesia were considered as prophylactic antiemetics and their administration was recorded in the anesthesia information system.[24]

Dosage of prophylactic antiemetic drugs was according to the existing local protocol: 1) ondansetron 4 mg IV, 30 min. before emergence of anesthesia; 2) droperidol 1.25 mg IV, 30 min. before emergence of anesthesia; 3) dexamethasone 4 mg IV, after induction of anesthesia. At the postanesthesia care unit and the ward the PONV protocol consisted of rescue-treatment with an antiemetic drug: either one of the above antiemetics drugs if not previously administered, or metoclopramide 20 mg IV. There was no active surveillance of adverse events during this study.

statistical analysis

Since we aimed to assess the impact of implementing a PONV prediction model on the actual patient outcome (PONV), sample size was based on an estimated PONV incidence of 30% in the care-as-usual group,[10,25] and a relative risk reduction of 25% per antiemetic.[23,26] As intervention group anesthesiologists were expected to provide more than one antiemetic to high-risk patients, the overall relative risk reduction for the intervention group was estimated at 33%, i.e. an absolute risk reduction of 10%. Detection of this 10% reduction in a randomized trial without cluster-randomization would require 295 patients per group, using a 2-sided alpha of 0.05 and power of 0.80. The sample size was adjusted for cluster-randomization using an inflation factor based on an average cluster size of 175 patients, and an intraclass correlation coefficient of 0.1, resulting in 5,430 patients for each group.[27,28] The sample size was considered sufficient for physician behavior, as less power was required to detect a difference in prescription of prophylactic antiemetics between allocation groups. Hence, around 11,000 patients were expected to be required.

Analysis was performed under the intention-to-treat principle. All statistical analyses were performed in R software (version 2.14.0). Statistical significance was defined as a two-sided alpha of 0.05. Continuous variables were visually assessed for a normal distribution using

histograms, QQ-plots. Parametric variables were expressed as means with standard deviations, nonparametric variables were expressed as medians with interquartile ranges, and discrete variables were expressed as numbers with percentages.

As this was a cluster-randomized trial, mixed effects regression analyses were used to take clustering into account: logistic regression for the incidence of PONV and Poisson regression for the number of prophylactic antiemetics (the function for generalized linear mixed effects models (*glmer*) from the *lme4* package in R software). Results of regression analyses were presented as odds ratios with 95% confidence intervals (CI) or rate ratios with 95% CI.

Fixed effects included allocation group, predicted risk of PONV, interaction between allocation group and predicted risk, and study time. The interaction term was included to quantify to what extent the difference in treatment effect (between intervention and care-as-usual) differed across predicted risks; e.g. an odds ratio below one would signify that a reduction in PONV due to the directive approach was greater in patients with higher risks. Study time was included in the model to adjust for a possible learning effect for anesthesiologists exposed to the prediction model.

In addition to the fixed effects, random effects were included for the intercept, allocation group, predicted risk, interaction between allocation group and predicted risk, and study time. A random intercept was included for anesthesiologists to account for small differences in PONV risks between their individual patient populations. Random slopes were included to account for different PONV prophylaxis strategies among anesthesiologists. A random slope was included for study time as the possible learning effect was not expected to be similar for each individual anesthesiologist. The intraclass correlation coefficient was calculated for the PONV incidence (using the *glmer* code in R) to quantify the amount of clustering at the anesthesiologists' level, which represents the resemblance of patients treated by the same anesthesiologist.[29]

Although randomization was stratified for anesthesia sub-specialty, some anesthesiologists may still have worked more with particular surgical specialties, treating patients with particular baseline risks of PONV. Due to a relatively small number of anesthesiologists (expected $n = 60$) treating a large number of patients, differences in case mix between anesthesiologists were expected to be magnified in observed baseline characteristics of patients. As the prediction model already included the most important patient characteristics, the mixed effects analyses were automatically adjusted for differences in patient characteristics by inclusion of the predicted risk as a variable. Therefore, additional analyses for both primary and secondary outcomes were performed to adjust for case mix differences by including procedure specific variables as fixed effects: surgical specialty and type of patient (ambulatory surgery yes/no). As the differences in baseline characteristics of patients were expected, these additional adjustments were amended to the study protocol before the start of the analysis.

Before multivariable modeling, all continuous variables were tested for nonlinearity using restricted cubic splines, including predicted PONV risk.[30] Missing data were multiple imputed ($n = 10$) using a regression approach in R (the function for multiple imputation (*aregImpute*) from the package *Hmisc* in R software). Imputation of missing study variables was based on predictors, outcome variables, and other perioperative data.[31-34] As PONV was coded missing

when any of the follow-up measurements was incomplete, non-missing follow-up measurements of PONV were added to the imputation process to serve as auxiliary variables to impute missing values for PONV. Subsequently, the imputed values for PONV were included into the mixed effects regression model, instead of deleted. The anesthesiologists were added as an extra variable in the imputation model to take into account the multilevel structure of the data.

Three post-hoc sensitivity analyses were performed to test the robustness of the results on three different areas of possible uncertainty. The three areas of possible uncertainty were the effects of multiple imputation of missing values; the definition of the outcome variable PONV and its consequence for the results; and the exclusion of anesthesiologists who treated less than 50 enrolled patients during the study period.

The sensitivity analysis on missing values was performed using a similar mixed effects model as the main analysis, but on complete cases only. For PONV, patients with any missing follow-up time points were discarded during the complete case analysis. For the analysis on prophylactic antiemetic interventions, all data was available and no patients were discarded.

For the sensitivity analysis on the primary outcome definition, we restricted the definition of PONV to serious nausea or vomiting at any of the follow-up time points. In contrast to the main analysis, minor nausea (the middle category of the three-point nausea scale) and the use of rescue antiemetics were not considered PONV for this sensitivity analysis.

As the exclusion of anesthesiologists with a low patient count during the study resulted in reduction in a sample size and as the discarded group of patients may be selective, a sensitivity analysis examined the impact of this exclusion. The reason to exclude anesthesiologists with a low patient count was to enable performing a cluster-based analysis, i.e. the mixed effects models. Therefore, this sensitivity analysis consisted of conventional logistic and Poisson regression analyses in all randomized patients, i.e. without the random effects and the anesthesiologists with a low patient count not excluded. Furthermore, anesthesiologists who treated no patients during the study period were still excluded from the analysis.

results

A total of 79 anesthesiologists were randomized, who together treated 12,032 enrolled patients (Figure 3.1). Of the 79 anesthesiologists, 6 were excluded as they treated no patients during the study period and 16 were excluded as they treated less than 50 patients per individual during the study period (in total 397 patients). An additional 22 patients were excluded because of a technical error in the anesthesia information management system. This resulted in 11,613 patients treated by 57 anesthesiologists (31 in the intervention group; 26 in the care-as-usual group) to be analyzed. Anesthesiologists of the intervention group treated fewer patients per anesthesiologist (median 162 patients per anesthesiologist, interquartile range 120–207) than anesthesiologists of the care-as-usual group (median 236 patients per anesthesiologist, interquartile range 128–300, Table 3.2).

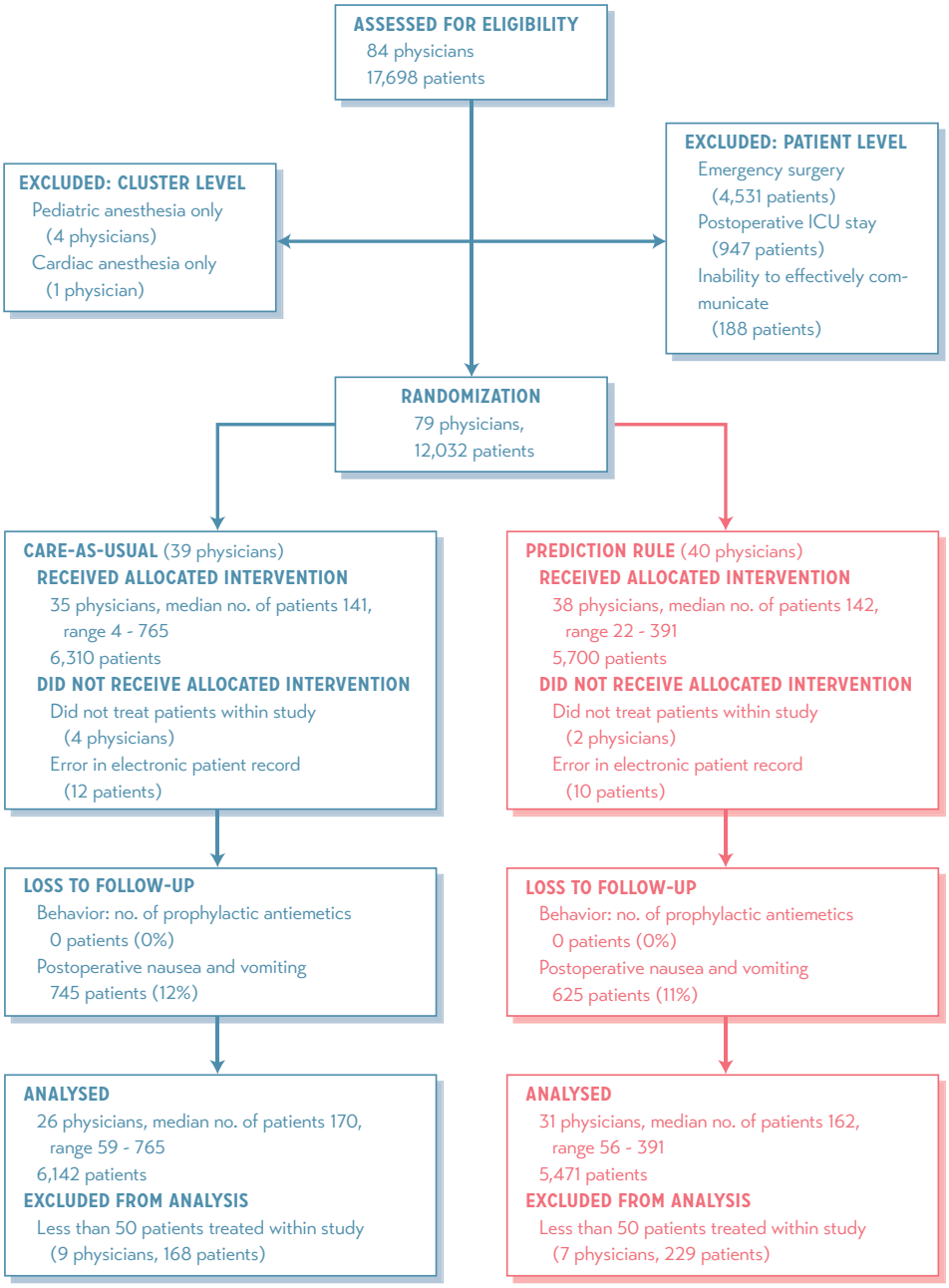


FIGURE 3.1 Flow diagram of the progress of anesthesiologists and patients through the phases of the study. Anesthesiologists represent either attending anesthesiologists or senior anesthesiology residents. ICU = Intensive Care Unit

TABLE 3.2 BASELINE CHARACTERISTICS OF ANESTHESIOLOGISTS

	Care-as-Usual Group	Intervention Group
	(*n* = 26)	**(*n* = 31)**
Number of patients treated, median (IQR)	170 (128–300)	162 (120–207)
Age*, years, mean (SD)	43 (9)	43 (10)
Female sex, *n* (%)	11 (42)	15 (48)
Entered during study, *n* (%)	5 (19)	8 (26)
Left department during study, *n* (%)	4 (15)	7 (23)
Specialty: cardiac anesthesia*, *n* (%)	2 (8)	2 (6)
pediatric anesthesia*, *n* (%)	3 (12)	4 (13)
pain medicine*, *n* (%)	1 (4)	4 (13)
Senior resident*, *n* (%)	12 (46)	11 (35)
Entered as senior resident during study, *n* (%)	4 (15)	4 (13)
Senior resident became attending anesthesiologist during study, *n* (%)	5 (19)	4 (13)

* Characteristic was documented at the moment of inclusion in the study
IQR = interquartile range; SD = standard deviation

TABLE 3.3 BASELINE CHARACTERISTICS OF PATIENTS

		Care-as-Usual Group	Intervention Group
	N*	**(*n* = 6142)**	**(*n* = 5471)**
Age, years, mean (SD)	11613	51 (17)	50 (17)
Female sex, *n* (%)	11613	3307 (54)	2988 (55)
Outpatients, *n* (%)	11613	1646 (27)	2086 (38)
ASA class, *n* (%)	11325		
1		2211 (37)	1985 (37)
2		3111 (52)	2765 (52)
3		648 (11)	581 (11)
4		15 (0)	9 (0)
Current smoking, *n* (%)	11134	1765 (30)	1598 (30)
Surgery with a high PONV risk, *n* (%)	10545	807 (15)	541 (11)
History of PONV / Motion sickness, *n* (%)	8837	1511 (33)	1397 (33)
Use of inhalational anesthesia†, *n* (%)	11613	3212 (52)	3321 (61)
Predicted risk of PONV, %, mean (SD)	11613	39 (15)	37 (15)
Apfel risk score‡, %, mean (SD)	11613	42 (19)	42 (19)
Operation duration, min, median (IQR)	11613	101 (63–167)	95 (59–151)

* N represents the total number of non-missing observations for that characteristic
† As compared to intravenous anesthesia using propofol
‡ For calculation of the Apfel risk score, the observed use of postoperative opioids was used instead of the expected
ASA = American Society of Anesthesiologists; IQR = interquartile range; min = minute;
PONV = postoperative nausea and vomiting; SD = standard deviation

Of the 11,613 patients, 5,471 (47%) were treated by intervention group anesthesiologists and 6,142 (53%) by 'care-as-usual' anesthesiologists (Table 3.3). The patients' mean ages and sex distribution were comparable. The intervention group included more outpatients (38% vs. 27%) and fewer procedures with a high PONV risk (11% vs. 15%); all other predictors were comparable between the two groups. The predicted PONV risk was slightly lower in the intervention group: 37% (SD 15%) versus 39% (SD 15%). No adverse events were reported by the study participants.

In total, 80% of all follow-up measurements were completed (intervention group 80%; care-as-usual group 79%), with 88% of all patients having at least one follow-up measurement completed (intervention group 89%; care-as-usual group 88%). As the primary outcome for a single patient was considered missing when any of the follow-up measurements were missing, 70% of patients had all follow-up measurements completed (n = 8,104), whereas the remaining 30% had their outcome variables imputed. The crude effect of prediction model implementation on patient outcome is shown in Table 3.4, columns 3 and 4. The crude incidence of PONV within the first 24 hours after surgery was 41% for patients in the intervention group, and 43% in the care-as-usual group. Intraclass correlation was low (0.020).

Differences in the occurrence of PONV were small and not statistically significant between intervention and care-as-usual group (odds ratio for allocation group 0.97, 95% CI 0.87–1.1;

TABLE 3.4 PRIMARY AND SECONDARY OUTCOMES BY DECILES OF PREDICTED RISK OF PONV

	Incidence of PONV			No. of prophylactic antiemetics*										
		Care-as-Usual Group‡ (n = 6142)	Intervention Group‡ (n = 5471)		Care-as-Usual Group¶ (n = 6142)					Intervention Group¶ (n = 5471)				
Predicted risk deciles	N†			N†,§	0	1	2	3	4	0	1	2	3	4
6–18%	790	67 (18)	82 (20)	1162	75	24	1	0	0	82	16	2	0	0
18–25%	750	102 (28)	102 (27)	1163	62	33	5	0	0	69	27	4	0	0
25–30%	795	101 (26)	113 (28)	1163	53	41	6	0	0	56	38	5	1	0
30–33%	795	123 (29)	117 (31)	1159	48	45	6	0	0	52	40	8	1	0
33–37%	823	165 (38)	133 (34)	1161	45	47	7	1	0	51	37	10	2	0
37–42%	831	199 (45)	166 (43)	1160	42	52	6	0	0	42	42	13	2	0
42–47%	815	195 (44)	163 (44)	1159	42	48	9	0	0	39	40	18	3	0
47–52%	809	251 (55)	202 (58)	1163	35	51	13	1	0	32	42	22	4	0
52–59%	835	294 (64)	216 (58)	1162	31	50	16	3	0	21	44	29	5	1
59–81%	861	337 (69)	253 (68)	1161	23	54	21	2	0	12	36	42	9	1
Overall	8104	1834 (43)	1547 (41)	11613	45	45	9	1	0	47	36	15	3	0

* The sum of the prophylactic use of ondansetron, droperidol, dexamethasone, and/or total intravenous anesthesia
† N represents the total number of non-missing observations per decile of predicted risk
‡ Data represent absolute numbers of PONV (%)
§ For the number of prophylactic antiemetics all observations were available
¶ Data represent percentages within risk decile
PONV = postoperative nausea and vomiting

TABLE 3.5 MIXED EFFECTS REGRESSION ANALYSIS ON THE INCIDENCE OF PONV

	Main analysis*,†	Sensitivity analysis		
		Complete case†,‡	Serious PONV*,§	All physicians & patients*,†,¶
	Unadjusted analysis			
Intervention group	**0.97 (0.87–1.10)**	0.93 (0.81–1.06)	0.99 (0.87–1.13)	0.98 (0.89–1.07)
Predicted risk#	**2.29 (2.07–2.54)**	2.83 (2.53–3.17)	2.27 (2.02–2.55)	2.24 (2.06–2.44)
Interaction: study group and predicted risk#	**0.92 (0.80–1.05)**	0.88 (0.74–1.04)	0.89 (0.75–1.04)	0.92 (0.82–1.03)
	Analysis adjusted for baseline differences**			
Intervention group	**0.97 (0.86–1.09)**	0.93 (0.81–1.06)	0.98 (0.86–1.13)	0.94 (0.86–1.04)
Predicted risk#	**2.37 (2.10–2.66)**	2.92 (2.57–3.31)	2.41 (2.12–2.74)	2.36 (2.15–2.59)
Interaction: intervention group and predicted risk#	**0.91 (0.80–1.05)**	0.87 (0.74–1.03)	0.90 (0.76–1.06)	0.91 (0.81–1.02)

Numbers represent odds ratios with 95% confidence intervals for the fixed effects.
All analyses included time as a covariable.
* After multiple imputation using 10 imputation datasets
† Patients with any nausea, vomiting or having received rescue antiemetics were considered cases
‡ Cases with missing variables were discarded, resulting in analysis of *n* = 8,104 patients
§ Only patients who reported definite nausea or vomiting were considered cases
¶ Unclustered logistic regression analysis in all physicians (*n* = 73) and patients (*n* = 12,010), without exclusion because of too small clusters
# Odds ratios represent predicted risks of 100% (a probability of 1)
** Baseline differences were adjusted by including surgical specialty and ambulatory surgery as covariates
PONV = postoperative nausea and vomiting

odds ratio for the interaction term of allocation group and predicted risk 0.92, 95% CI 0.80–1.1). The absence of statistical significance is reflected in Figure 3.2, as 95% confidence intervals of both groups almost fully overlap. Pre-specified adjustment for baseline characteristics did not change these results and inferences. Moreover, the sensitivity analyses on missing data, the outcome definition and exclusion of anesthesiologists and patients showed small and non-significant effects on the PONV incidence, similar to the main analyses. Numerical descriptions of both adjusted and unadjusted models, as well as the three sensitivity analyses can be found in Table 3.5.

The crude effect of the group assignment on physician behavior is shown as the percentage of patients which received a particular number of administered prophylactic antiemetics (Table 3.4, columns 5 to 15). The number and type of prophylactic antiemetics were documented for all patients.

In the main analysis, anesthesiologists of the intervention group administered more antiemetic prophylaxis than anesthesiologists of the care-as-usual group (rate ratio for allocation group 2.0, 95% CI 1.6–2.4). Moreover, when administering antiemetic prophylaxis, intervention group anesthesiologists discriminated more between patients with a high or low predicted risk than anesthesiologists of the care-as-usual group (rate ratio for the interaction

TABLE 3.6 MIXED EFFECTS REGRESSION ANALYSIS ON THE NUMBER OF PROPHYLACTIC ANTIEMETICS

		Sensitivity analysis		
	Main analysis*	Complete case†	Serious PONV‡	All physicians & patients*,§
	Unadjusted analysis			
Intervention group	**2.03 (1.64–2.50)**	2.03 (1.64–2.50)	2.03 (1.64–2.5)	1.76 (1.58–1.96)
Predicted risk¶	**2.16 (1.84–2.53)**	2.16 (1.84–2.53)	2.16 (1.84–2.53)	2.25 (2.07–2.43)
Interaction: study group and predicted risk¶	**1.63 (1.30–2.03)**	1.63 (1.30–2.03)	1.63 (1.30–2.03)	1.60 (1.43–1.80)
	Analysis adjusted for baseline differences**			
Intervention group	**1.87 (1.55–2.25)**	1.87 (1.55–2.25)	1.87 (1.55–2.25)	1.88 (1.69–2.11)
Predicted risk¶	**1.81 (1.57–2.09)**	1.81 (1.57–2.09)	1.81 (1.57–2.09)	1.71 (1.56–1.87)
Interaction: intervention group and predicted risk¶	**1.57 (1.30–1.89)**	1.57 (1.30–1.89)	1.57 (1.3–1.89)	1.63 (1.45–1.83)

Numbers represent rate ratios with 95% confidence intervals for the fixed effects.
All analyses included time as a covariable.
* After multiple imputation using 10 imputation datasets
† Identical to the primary analysis as there were no missings for the secondary outcome
‡ Identical to the primary analysis as the secondary outcome definition was not changed
§ Unclustered Poisson regression analysis in all physicians ($n = 73$) and patients ($n = 12{,}010$), without exclusion because of too small clusters
¶ Rate ratios represent predicted risks of 100% (a probability of 1)
** Baseline differences were adjusted by including surgical specialty and ambulatory surgery as covariables
PONV = postoperative nausea and vomiting

term of allocation group and predicted risk 1.6, 95% CI 1.3–2.0, Figure 3.3). Compared to anesthesiologists of the care-as-usual group, intervention group anesthesiologists administered a higher number of prophylactic antiemetics to patients with a high predicted risk, while low risk patients received a lower number of antiemetics. Both overall and risk-dependent differences in physician behavior were statistically significant (Figure 3.3, 95% confidence interval areas of both groups cross each other and only slightly overlap). Neither adjustment for baseline differences nor the three sensitivity analyses changed the results and their inferences (see Table 3.6).

discussion

Following recent guidelines on studying the impact of clinical prediction models, we performed a cluster-randomized trial on the implementation of a validated prediction model for PONV. We quantified the effects of implementing such a prediction model on both physician behavior and patient outcome. Patients did not have a substantially lower incidence of PONV when their anesthesiologists were provided with an intraoperative predicted PONV risk, despite an increased administration of risk-tailored prophylaxis by these anesthesiologists. In other words, anesthesiologists of the intervention group administered more prophylactic antiemetics to

patients at higher risk and fewer antiemetics to patients at low risk in comparison to their colleagues of the care-as-usual group. However, the tailored prescription of antiemetics in the intervention group did not result in a substantially lower incidence of PONV.

The discrepancy in results between patient outcome and physician behavior was unexpected. At the start of the study we assumed that all conditions to proceed with an impact study of our PONV prediction model had been met: the implemented prediction model had been externally validated, several clinical guidelines advocated the use of prediction models for PONV, and the efficacy of prophylactic antiemetics had been well-established by several randomized clinical trials and meta-analyses.[23,26,35–37] Since this study showed that application of the PONV prediction model did not translate into a clear benefit for patients, it is conceivable that one or more of these presumptions were wrong.

First, the predictive performance of the prediction model may actually have been insufficient to improve clinical decision making. In previous validation studies predictive performance of all existing PONV models was typically moderate (c-statistic around 0.70).[9,12] Our prediction model had comparable discrimination (c-statistic of 0.68) and it slightly underestimated the actual PONV risk.[22] With a moderate predictive performance, decisions based on the model may not have been superior to care-as-usual, i.e. clinical judgment.

Second, despite a statistically significant impact on physician behavior, the absolute impact was relatively small. In this study, four different prophylactic antiemetics were available. However, anesthesiologists of the intervention group mostly administered up to two antiemetics to high-risk patients, whereas more than two antiemetics may be indicated.[23] From the results of our study we cannot infer why anesthesiologists of the intervention group were reluctant to administer more than two antiemetics. Nonetheless, one might expect a decrease in PONV occurrence when anesthesiologists of the intervention group administered more prophylactic antiemetics to their high-risk patients.

Third, proven efficacy in randomized clinical trials is no guarantee for effectiveness in daily practice.[38–41] In previous efficacy trials prophylactic antiemetics reduced the risk of PONV by approximately 30%, with a wide variation between different studies.[26] However, study populations of these efficacy trials were often restricted to specific high-risk groups, such as women undergoing laparoscopic procedures. As the present study included a large sample of surgical patients with minimal inclusion restrictions, the actual effectiveness of 'proven' prophylactic antiemetics may indeed be lower than reported earlier.

The primary goal of the present study was to quantify the effects of implementing a clinical prediction model on patient outcome (PONV), and to interpret these effects in view of the change in physician behavior caused by the prediction model, i.e. administration of antiemetics. The absence of a direct link between increased administration of antiemetics and reduced PONV should not be interpreted as evidence that PONV prediction models – or prediction models in general – are ineffective in clinical practice. This study is only one example, and its limitations should be considered when applying the results to different settings.

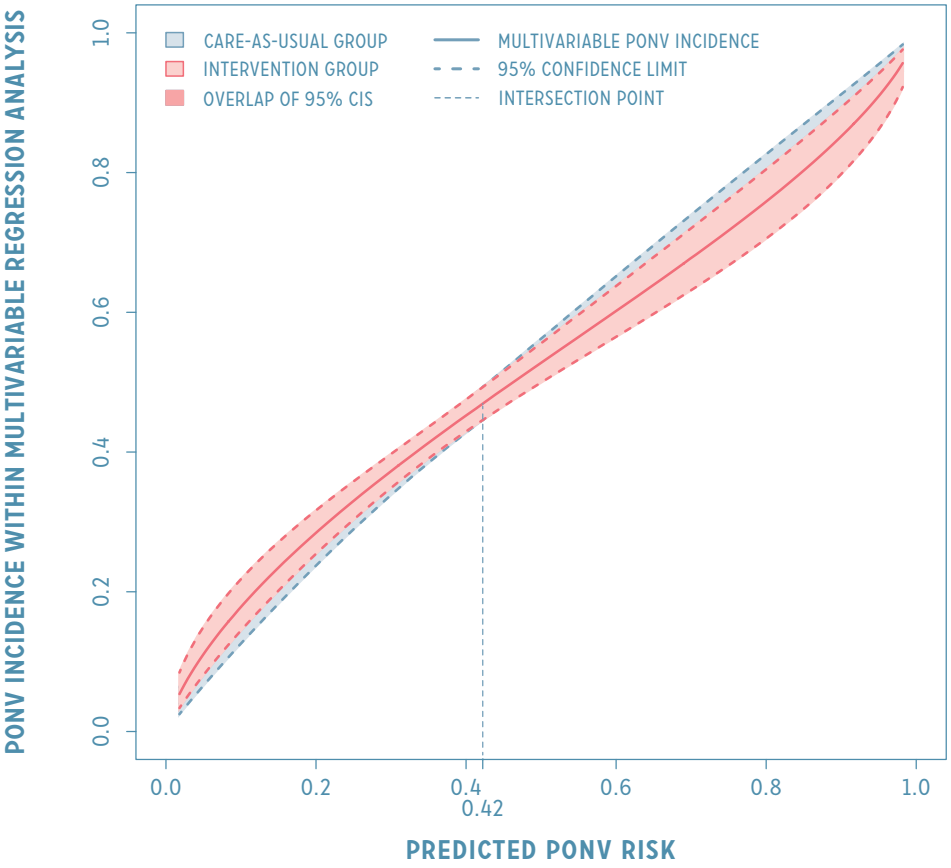


FIGURE 3.2 Graphical representation of the mixed effects regression analysis on the incidence of postoperative nausea and vomiting. Figure 3.2 may be interpreted as the occurrence of postoperative nausea and vomiting (PONV) after receiving prophylaxis, in patients with a particular predicted risk within each study group. The differences between the colored and gray areas represent the effect of implementation of the prediction model on the occurrence of PONV in patients with a particular predicted risk.

The solid lines and their 95% confidence intervals (CI) represent the fixed effects of the mixed effects regression analyses. The dotted vertical line shows the intersection point of both groups. The mixed effects models included fixed effects for the variables allocation group, predicted risk of PONV, interaction between allocation group and predicted risk, and time. Random effects were included for the intercept, predicted risk, interaction between allocation group and predicted risk, and time. The lines were calculated using the average for the variable time. The 95% CIs were calculated from the covariance matrix for the variables allocation group and its interaction term with predicted risk.

First, although we assessed a large number of outcomes, this implementation study included a single center.[3] Second, despite a large study population, differences in baseline characteristics occurred, probably due to cluster-randomization. By randomizing a relatively small sample of anesthesiologists, differences in anesthesiologists' baseline characteristics may have been magnified at the patient level, although adjustment for baseline differences in patient case mix among anesthesiologists did not change the results.

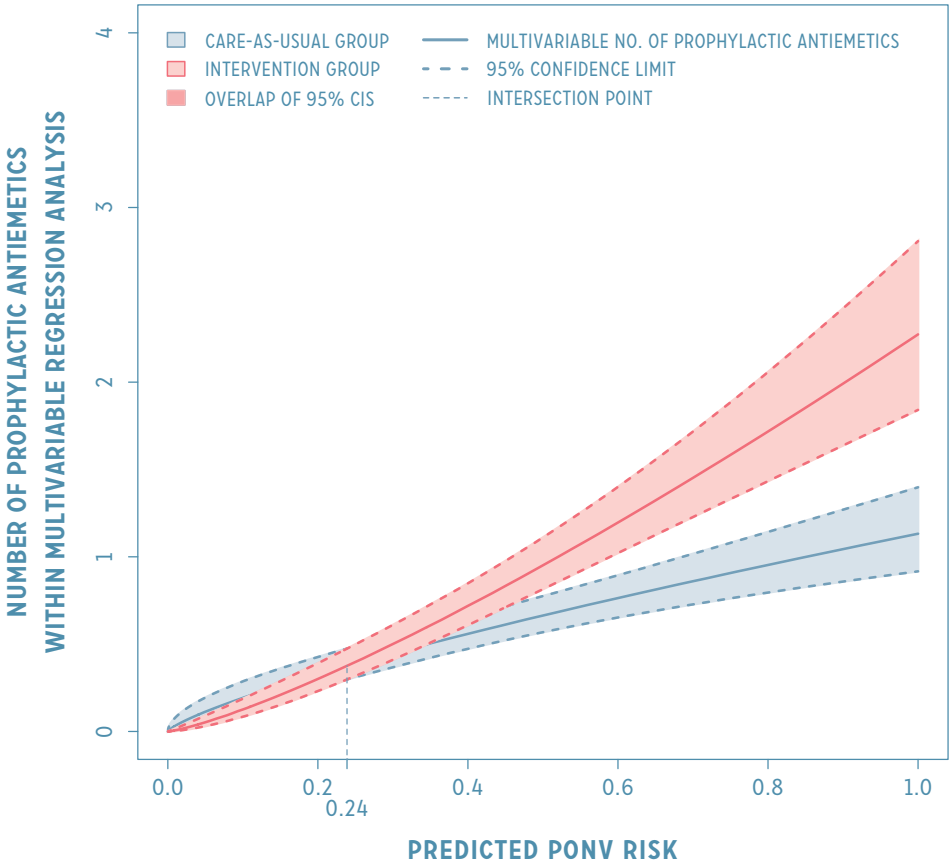


FIGURE 3.3 Graphical representation of the mixed effects regression analysis on the incidence of postoperative nausea and vomiting. Figure 3.3 may be interpreted as the number of prophylactic antiemetics a patient with a particular predicted risk of PONV would receive from any anesthesiologist within each study group. The differences between the colored and gray areas therefore represent the changes in physician behavior concerning prescription of antiemetic prophylaxis, caused by implementation of the prediction model.

The solid lines and their 95% confidence intervals (CI) represent the fixed effects of the mixed effects regression analyses. The dotted vertical line shows the intersection point of both groups. The mixed effects models included fixed effects for the variables allocation group, predicted risk of PONV, interaction between allocation group and predicted risk, and time. Random effects were included for the intercept, predicted risk, interaction between allocation group and predicted risk, and time. The lines were calculated using the average for the variable time. The 95% CIs were calculated from the covariance matrix for the variables allocation group and its interaction term with predicted risk.

Third, anesthesiologists were not naive to prophylactic PONV management. As Table 3.4 and Figure 3.2 clearly indicate, anesthesiologists of the care-as-usual group also provided prophylaxis to their patients in a risk-dependent manner, reducing the possible impact of a prediction model on the incidence of PONV. It is conceivable that anesthesiologists were already able to (heuristically) identify high-risk patients before the start of the study, as a consequence of training, clinical experience, medical literature and the pre-existent local PONV

protocol. In addition, there might have been increased knowledge and experience with the use of risk scores for PONV prophylaxis, as the implemented prediction model was updated and optimized in the same hospital and the anesthesiologists may already have been using other risk scores before the start of the study. Therefore, a Hawthorne effect cannot be ruled out as a possible explanation for the results of this study. An alternative explanation may be that some contamination may have occurred, as anesthesiologists of the intervention group may occasionally have discussed PONV and its prophylactic treatment with anesthesiologists of the care-as-usual group. However, true contamination is unlikely, as the predicted risk is not easily calculated without a proper decision support tool.

Fourth, we used an 'assistive' implementation strategy as opposed to a 'directive' approach: anesthesiologists were free to decide how to interpret the predicted risk and how to assess the need for antiemetics. Additional interventions to increase the impact on physician behavior may include a more intensive education and feedback programme.[16,42] A directive approach, which includes actionable recommendations, may have a larger impact on physician behavior.[43]

Fifth, the observed PONV incidence within 24 hours after surgery in the care-as-usual group of our study is 43%, which is high for a study population which was not selected based on a preoperative risk of PONV, such as Apfel's trial.[23] The sensitivity analysis in which we defined PONV as only serious nausea or vomiting did not change the results (Table 3.5, third column). The average predicted risks of PONV from both the implemented prediction model and Apfel's simplified risk score are very close to the observed PONV incidence in our study (Table 3.3).[10] Consequently, our study should be considered a high-risk population for PONV and results may not necessarily translate to population with a lower average PONV risk.

Sixth, the results for the PONV incidence bordered on statistical significance. The exclusion of almost a quarter of initially randomized anesthesiologists and approximately 400 patients after randomization may have prevented the results to become significant. Therefore, an unclustered sensitivity analysis in all randomized patients was performed. The sensitivity analysis produced similar results as the main analysis (Table 3.5, fourth column). Even after adding the 400 patients that were excluded, the effect of the implemented prediction model on PONV remained very small in the sensitivity analysis, and it would require a much larger set of patients to become statistically significant. The small effect size is reflected in its clinical relevance, as a crude 2% reduction in PONV is not a substantial effect from a patient perspective.

In conclusion, implementation of a previously 'validated' prediction model for PONV did not result in a clinically relevant decrease in PONV incidence, despite an increase in 'risk-tailored' application of prophylactic antiemetic strategies by anesthesiologists. Even when the use of a prediction model is consolidated in several guidelines, the discrepancy in the results of this study underscores the need to perform a formal impact analysis which includes patient outcome, before attempting to implement a prediction model into clinical practice.

references

1. Graham I, Atar D, Borch-Johnsen K, et al. European guidelines on cardiovascular disease prevention in clinical practice: executive summary: Fourth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (Constituted by r. *Eur Heart J*. 2007;28(19):2375–2414. PMID:17726041.

2. Qaseem A, Snow V, Barry P, Hornbake ER, Rodnick JE. Annals of Internal Medicine Clinical Guidelines Current Diagnosis of Venous Thromboembolism in Primary Care : A Clinical Practice Guideline from the American Academy of Family Physicians and the American College of Physicians *. *Ann Intern Med*. 2007;146(6):454–458. PMID:17261865.

3. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209. PMID:16461965.

4. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009;119(17):2408–2416. PMID:19364974.

5. Moons KGM. Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem*. 2010;56(4):537–541. PMID:20133890.

6. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–698. PMID:22397946.

7. Gan TJ, Meyer T a, Apfel CC, et al. Society for Ambulatory Anesthesia guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2007;105(6):1615–1628. PMID:18042859.

8. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2014;118(1):85–113. PMID:24356162.

9. Apfel CC, Kranke P, Eberhart LHJ, Roos A, Roewer N. Comparison of predictive models for postoperative nausea and vomiting. *Br J Anaesth*. 2002;88(2):234–240. PMID:11883387.

10. 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(3):693–700. PMID:10485781.

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(3):237–242. PMID:11861340.

12. Van den Bosch JE, Kalkman CJ, Vergouwe Y, et al. Assessing the applicability of scoring systems for predicting postoperative nausea and vomiting. *Anaesthesia*. 2005;60(4):323–331. PMID:15766334.

13. Eberhart LHJ, Morin AM. Risk scores for predicting postoperative nausea and vomiting are clinically useful tools and should be used in every patient: con--'life is really simple, but we insist on making it complicated'. *Eur J Anaesthesiol*. 2011;28(3):155–159. PMID:21192269.

14. Kranke P, Eberhart LHJ. Possibilities and limitations in the pharmacological management of postoperative nausea and vomiting. *Eur J Anaesthesiol*. 2011;28(11):758–765. PMID:21799417.

15. Pierre S. Risk scores for predicting postoperative nausea and vomiting are clinically useful tools and should be used in every patient: pro--'don't throw the baby out with the bathwater'. *Eur J Anaesthesiol*. 2011;28(3):160–163. PMID:21206276.

16. Frenzel JC, Kee SS, Ensor JE, Riedel BJ, Ruiz JR. Ongoing provision of individual clinician performance data improves practice behavior. *Anesth Analg*. 2010;111(2):515–519. PMID:20529985.

17. Kooij FO, Klok T, Hollmann MW, Kal JE. Automated reminders increase adherence to guidelines for administration of prophylaxis for postoperative nausea and vomiting. *Eur J Anaesthesiol*. 2010;27(2):187–191. PMID:19478673.

18. Kooij FO, Vos N, Siebenga P, Klok T, Hollmann MW, Kal JE. Automated reminders decrease postoperative nausea and vomiting incidence in a general surgical population. *Br J Anaesth*. 2012;108(6):961–965. PMID:22382396.

19. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(jun04 2):b606. PMID:19502216.

20. Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. *BMJ*. 1998;317(7167):1171–1172. PMID:9794847.

21. Van den Bosch JE, Moons KG, Bonsel GJ, Kalkman CJ. Does measurement of preoperative anxiety have added value for predicting postoperative nausea and vomiting? *Anesth Analg*. 2005;100(5):1525–1532. PMID:15845719.

22. Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of clinical prediction models for application in local settings. *Med Decis Making*. 2012;32(3):E1-E10. PMID:22427369.

23. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350(24):2441–2451. PMID:15190136.

24. Visser K, Hassink E a, Bonsel GJ, Moen J, Kalkman CJ. Randomized controlled trial of total intravenous anesthesia with propofol versus inhalation anesthesia with isoflurane-nitrous oxide: postoperative nausea with vomiting and economic analysis. *Anesthesiology*. 2001;95(3):616–626. PMID:11575532.

25. Gan TJ. Postoperative nausea and vomiting--can it be eliminated? *JAMA*. 2002;287(10):1233–1236. PMID:11886298.

26. Carlisle JB, Stevenson CA. Drugs for preventing postoperative nausea and vomiting. *Cochrane Database Syst Rev*. 2006;3(1):CD004125. PMID:16856030.

27. Donner A. Sample size requirements for stratified cluster randomization designs. *Stat Med*. 1992;11(6):743–750. PMID:1594813.

28. Campbell MK, Thomson S, Ramsay CR, MacLennan GS, Grimshaw JM. Sample size calculator for cluster randomized trials. *Comput Biol Med*. 2004;34(2):113–125. PMID:14972631.

29. Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. *BMC Med Res Methodol*. 2004;4.

30. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387. PMID:8668867.

31. Little RJA. Regression with missing x's: A review. *J Am Stat Assoc*. 1992;87(420):1227–1237. doi:10.2307/2290664.

32. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol*. 1995;142(12):1255–1264. PMID:7503045.

33. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59(10):1087–1091. PMID:16980149.

34. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092–1101. PMID:16980150.

35. Leslie JB, Gan TJ. Meta-analysis of the safety of 5-HT3 antagonists with dexamethasone or droperidol for prevention of PONV. *Ann Pharmacother*. 2006;40(5):856–872. PMID:16670360.

36. Neufeld SM, Newburn-Cook C V. The efficacy of 5-HT3 receptor antagonists for the prevention of postoperative nausea and vomiting after craniotomy: a meta-analysis. *J Neurosurg Anesthesiol*. 2007;19(1):10–17. PMID:17198095.

37. Karanicolas PJ, Smith SE, Kanbur B, Davies E, Guyatt GH. The impact of prophylactic dexamethasone on nausea and vomiting after laparoscopic cholecystectomy: a systematic review and meta-analysis. *Ann Surg*. 2008;248(5):751–762. PMID:18948802.

38. Bauer MS, Williford WO, Dawson EE, et al. Principles of effectiveness trials and their implementation in VA Cooperative Study #430: "Reducing the efficacy-effectiveness gap in bipolar disorder." *J Affect Disord*. 2001;67(1–3):61–78. PMID:11869753.

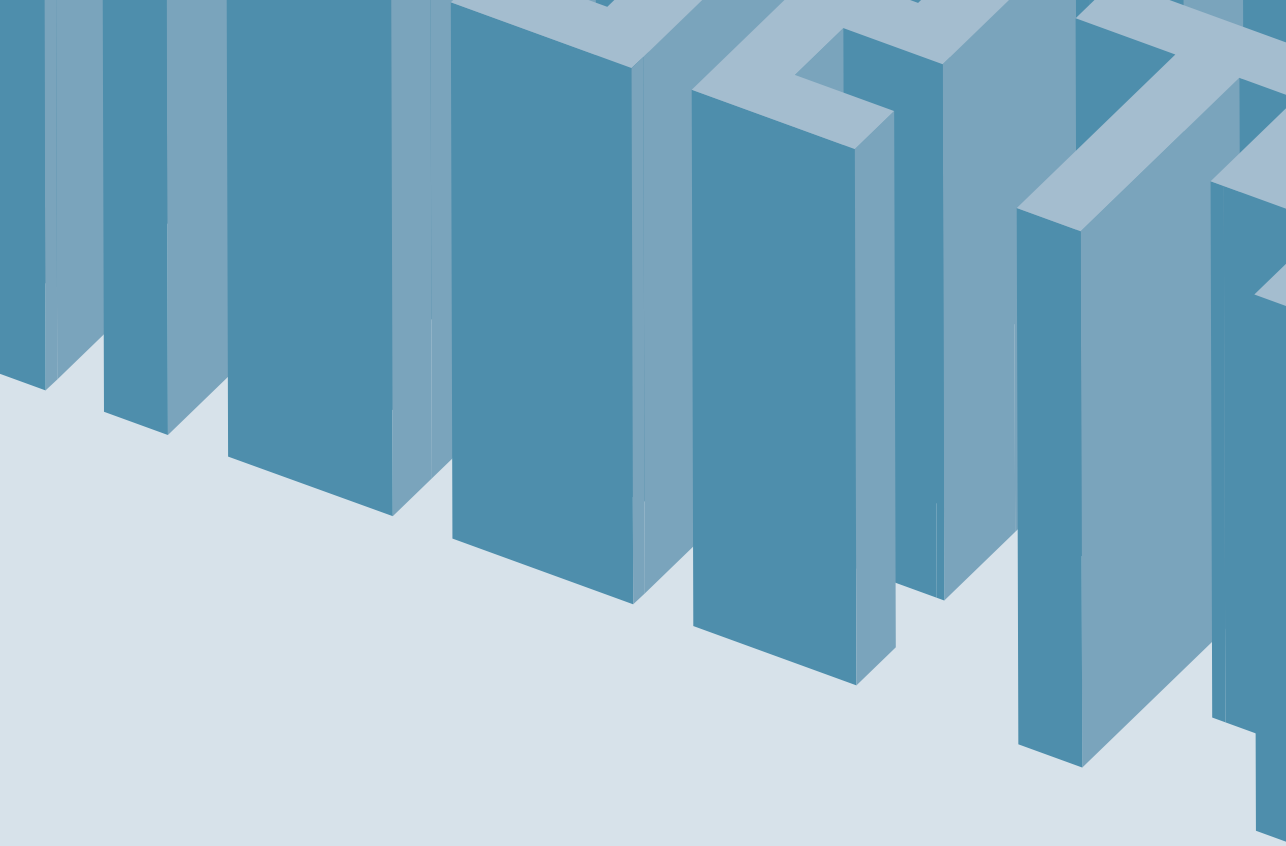
39. Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol*. 2003;3:28. PMID:14690550.

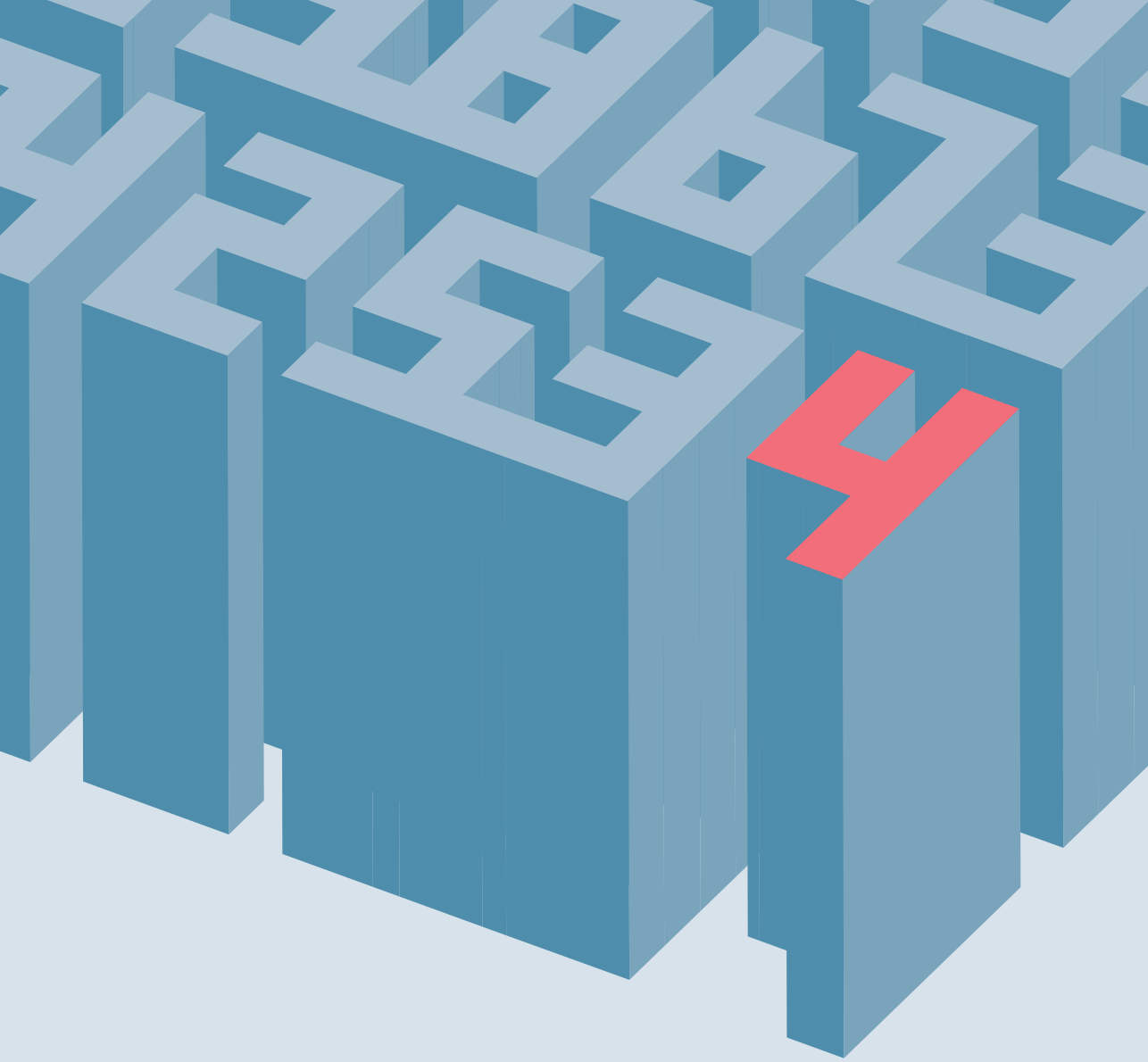
40. Hallfors D, Cho H, Sanchez V, Khatapoush S, Kim HM, Bauer D. Efficacy vs effectiveness trial results of an indicated "model" substance abuse program: implications for public health. *Am J Public Health*. 2006;96(12):2254–2259. PMID:16809591.

41. Karanicolas PJ, Montori VM, Devereaux PJ, Schünemann H, Guyatt GH. A new 'mechanistic-practical" framework for designing and interpreting randomized trials. *J Clin Epidemiol*. 2009;62(5):479–484. PMID:18468856.

42. Weingarten SR, Henning JM, Badamgarav E, et al. Interventions used in disease management programmes for patients with chronic illness-which ones work? Meta-analysis of published reports. *BMJ*. 2002;325(7370):925. PMID:12399340.

43. Kawamoto K, Houlihan C a, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765. PMID:15767266.





HOW CLINICAL PREDICTION MODELS IMPACT DECISION MAKING BY PHYSICIANS: AN EXAMPLE STUDY

T.H. Kappen, K. van Loon, M.A.M. Kappen, L. van Wolfswinkel, Y. Vergouwe, W.A. van Klei, K.G.M. Moons, C.J. Kalkman

Submitted

abstract

Background: Prediction models may facilitate risk-based management of healthcare conditions. In a large cluster-randomized trial, presenting calculated risks of postoperative nausea and vomiting (PONV) to physicians (assistive approach) increased risk-based management of PONV. This increase did not improve patient outcome – i.e. PONV incidence. This prompted us to explore how prediction tools guide the decision-making process of physicians.

Methods: Using mixed-methods we interviewed eight physicians to understand how predicted risks were perceived by the physicians, and how they influenced decision making. Subsequently, all participating physicians were surveyed for how their perceptions were influenced by the presented risks.

Results: Although the prediction tool made physicians more aware of PONV prevention, the physicians reported three barriers to use predicted risks in their decision making. PONV was not considered an outcome of utmost importance; decision making on PONV prophylaxis was mostly intuitive rather than risk-based; prediction models do not weigh benefits and risks of prophylactic drugs.

Conclusions: Combining probabilistic output of the model with their clinical experience may be difficult for physicians, especially when their decision-making process is mostly intuitive. Adding recommendations to the predicted risk (directive approach) was considered an important step to facilitate the uptake of a prediction tool.

Prediction models are introduced into medical practice to facilitate physician decision making. A good model accurately predicts a patient's risk of a current (diagnostic model) or future (prognostic model) health condition. This information may then be used by healthcare providers and patients – ideally in shared decision making – to decide on the most appropriate course of action, either treatment, preventive strategies, or a combination of these. It is often recommended that before wide spread use in daily practice, the actual impact of diagnostic or prognostic prediction models on clinical decision making and patient outcomes should be formally studied in so-called 'prediction model impact studies'.[1–4]

An important choice in such studies is the model's format or presentation: will the clinicians only be presented their patient's individual predicted outcome risk – an 'assistive' format – or will they receive a corresponding therapeutic or preventive management recommendation in addition to the risk – a 'directive' format.[1,4] Some authors suggest that a directive format will have a larger effect on decision making and is more likely to improve patient outcomes, but others fear that such an approach merely results in 'cook book medicine' and compromises clinical intuition.[1,4–7] Notwithstanding this debate, there is very limited empirical evidence on how clinicians react when predicted risks obtained from prediction models are presented to them. We recently performed a large impact study of the implementation of a prediction model in which we explicitly selected an assistive format. This impact study allowed us to explore how clinicians actually perceive the use of prediction models and use their predicted risks in their decision making.

The impact study was designed as a large cluster-randomized trial on the use of a previously developed, validated and recalibrated prediction model for postoperative nausea and vomiting (PONV).[8,9] PONV is a common problem in the first 24 hours after surgery with general anesthesia, affecting – on average – 30% of all surgical patients.[8,10,11] At the start of anesthesia, physicians who were randomized to the intervention group were 'exposed' to model-based risk estimates of PONV for their individual patients, as calculated by an on-line prediction model. When compared to the unexposed physicians – i.e. the 'care-as-usual' control group – physicians who were presented the patients' individual PONV risks indeed administered more prophylactic antiemetic drugs to their patients. However, the overall increased administration of antiemetic prophylaxis did not result in a clinically relevant decrease in either the overall or risk-dependent incidence of PONV. This discrepancy was unexpected. From the results of the trial alone we were unable to discern between possible causes for this discrepancy. One possible cause was that simply presenting risk scores only had a minimal impact on the physicians' behavior. Other possible causes were that the prediction model did not accurately target patients who will respond to the antiemetic effects of the prophylactic drugs or that the actual effectiveness of the prophylactic drugs was not as high as their studied efficacy.[12,13]

We hypothesized that physicians with a greater inclination to treat PONV prophylactically would have administered more risk-dependent prophylaxis during the original RCT and thus their patients may have had a lower incidence of PONV. Hence, we performed an additional mixed-methods study using both qualitative and quantitative methods to improve our

understanding of how predicted risks are perceived, how they influence physician decision making, and which presentation format is considered most useful. We aimed to gain insight in why clinical implementation of a validated and recalibrated predictive model did not improve patient outcomes in our example cluster-RCT. Furthermore, the results of this mixed-methods study may help to formulate general recommendations for the design and conduct of impact trials and implementation of prediction models in clinical practice.

methods

example study

The example study was a cluster-randomized trial on the clinical effects of implementing a prediction model for PONV.[8] This cluster-RCT included 57 physicians of the Anesthesiology department at the University Medical Center Utrecht who treated over 11,000 patients. After randomization of the physicians, we compared antiemetic prescription and PONV outcomes from the intervention group of physicians who were 'exposed' to their patients' PONV risks to the care-as-usual group. The model implementation strategy was assistive, i.e. the risk estimates were shown during the anesthetic case, but without any patient-specific recommendations on PONV prophylaxis.[1,3] During the cluster-RCT, the physicians of the intervention group administered more risk-dependent PONV prophylaxis to their patients than physicians of the care-as-usual group. This increase in risk-dependent PONV prophylaxis in the intervention group did not result in a lower incidence of PONV.

design of present study

The present mixed-methods study consisted of a sequential qualitative and quantitative phase, performed at the end of the cluster-RCT. We first collected qualitative data on possible facilitators and barriers for the risk-dependent prophylactic treatment of PONV through face-to-face, semi-structured interviews with physicians who participated in the cluster-RCT. The interviews were structured according to an interview guide constructed by the first and third author (THK and MAMK) and were conducted by the third author (MAMK). Interview transcripts were analyzed by the second and first author (KVL and THK), using a grounded theory approach to identify differences in physicians' conceptions and barriers on prophylactic PONV management. Second, a structured survey was conducted in all physicians who participated in the RCT to quantify whether exposure to predicted risks as calculated by the prediction model resulted in differences in conceptions and barriers between physicians of the intervention group and the care-as-usual group. Additional information on the methods can be found in the Appendix, chapter 8.

interviews

Physicians were selected for the interviews by stratified random sampling from the pool of enrolled physicians. Stratification allowed for selection of physicians in different career stages (resident or junior attending versus senior attending) and different randomization statuses

within the RCT (intervention versus care-as-usual). Two physicians per stratum was expected to provide sufficient variation in the conceptions and barriers of physicians to administer prophylactic antiemetics to their patients during general anesthesia. In total eight physicians were interviewed.

A semi-structured interview schedule was devised using both open and closed questions. Questions were formulated according to four prespecified topics: 1) perceived patient burden of PONV by the physician; 2) professional risk stratification: how does the physician identify patients at risk for PONV; 3) the decision-making process regarding preventive antiemetic strategies; 4) attitude towards the use of prediction models and decision support for prophylactic treatment of PONV. The specific phrasing and order of questions were only meant as a guide; interviewer and interviewees were encouraged to explore the different themes during the conversation. The interviews ranged from 20 to 60 minutes. Data saturation was achieved after eight interviews and additional interviews were not expected to provide new insights.[14–16] The individual responses to each question were sorted, coded and then rearranged into themes. Axial coding was used to group the themes into broader thematic topics, using Microsoft Excel. The themes and topics were subsequently integrated into central themes.

structured survey

To quantify possible differences between study groups observed during the interviews, all physicians participating in the RCT received a web-based survey. The survey addressed the same four topics used in the interviews. For each topic a specific set of questions was formulated to quantify possible differences in attitude towards that particular topic. For each question, the numerical answer of each physician was standardized to the mean score and standard deviation of all physicians, i.e. both allocations groups combined. The sum of all individual standardized scores of a physician was then used as a summary z-score for that particular topic (see Appendix Section 2). For each allocation group, the means and standard deviations of the summary z-scores per topic were calculated. Multiple testing was avoided by summing all four summary z-scores for each physician into a total summation z-score, which was tested using a two-sample Student's t-test. Results for the two additional questions for the intervention group physicians were presented as absolute frequencies with percentages. A two-sided alpha of <0.05 was considered statistically significant.

During the interview analysis the question arose whether physicians from the intervention group preferred an actionable recommendation instead of simply being presented a predicted PONV risk, and – in contrast – care-as-usual physicians would have preferred the prediction model over the recommendation. This question was addressed in the survey using a post-hoc analysis on a selection of questions from 'Attitudes towards prediction models and decision support'. The questions which addressed either the predicted risk, a recommendation or both were pooled and analyzed using a Wilcoxon rank test (see Appendix Section 2).

adjusting the cluster-randomized trial analysis

The original RCT studied the PONV incidence (primary outcome) and the number of prophylactic antiemetics administered per patient (secondary outcome) for risk-dependent differences

between allocation groups using mixed effects regression analyses.[8] We hypothesized that physicians with a greater inclination to treat PONV prophylactically would have administered more risk-dependent prophylaxis during the original RCT and thus their patients may have had a lower incidence of PONV. As the total summation z-score from the web-based survey was expected to reflect this inclination of physicians to treat PONV prophylactically, we adjusted the original RCT analysis of both outcomes for the total summation z-score (see Appendix Section 2).

results

interviews

The results of the eight interviews were rearranged into seventeen themes according to prespecified topics of the interview model (Table 4.1 and Appendix Section 1). Four overarching categories – or central themes – emerged from the 17 themes, which are described in the following sections.

TABLE 4.1 THEMES FROM THE INTERVIEW ANALYSIS

Topic 1: Perceived severity of PONV	
1A.	Knowledge of the frequency of PONV
1B.	Perception of the frequency of PONV among their patients
1C.	Feedback on PONV occurrence increases awareness
1D.	PONV is a side effect of general anesthesia
Topic 2: Professional risk stratification: how does a physician identify patients at risk for PONV?	
2A.	Variation in the use of risk factors
2B.	Risk factors may be experience-based or evidence-based
2C.	Weighing factors: no clear-cut risk appraisal
Topic 3: Making decisions on prophylactic antiemetic strategies	
3A.	Perceived benefits and harms of prophylactic antiemetic strategies
3B.	Decisions on PONV prophylaxis are made implicitly
3C.	PONV prophylaxis: only when indicated
Topic 4: Attitude towards the use of prediction models and decision support	
4A.	Intervention group: the predicted risk is higher than their own assessment
4B.	Intervention group: no conscious roadmap for using the predicted risk
4C.	Care-as-usual group: the predicted risk is mainly a reminder
4D.	Authority in making decisions is more important than decision support
4E.	Prediction models do not consider other comorbidity
4F.	The predicted risk does not reflect the underlying mechanisms of PONV
4G.	Willingness to use decision support may be dependent on familiarity

For a full description of the themes see Appendix Section 1

PONV is not a main area of attention

Awareness of PONV as a clinical problem and perceived patient burden appeared to be low. The interviewees of both groups reported that they considered PONV to be a side effect of general anesthesia (theme 1D, Table 4.1), and that it should not be 'excessively' treated because of a low burden relative to more serious complications (themes 3B and 3C, Table 4.1 and quote 1, Table 4.2). Knowledge of the different elements of a PONV prevention strategy

TABLE 4.2 QUOTES FROM THE INTERVIEW ANALYSIS

No.	Quote	Interviewee
1	*You see, you can prevent nausea for a number of people, but when opposed by one person suffering from a surgical site infection [author's note: a perceived risk of dexamethasone PONV prophylaxis], your net benefit is lost.*	junior attending care-as-usual group
2	*There is definitely added value for droperidol, I think that is a good drug....I rarely administer droperidol intraoperatively...It provides some drowsiness, but in the low dose that we give, it is hardly a problem.*	senior attending intervention group
3	*Because of the study it [author's note: awareness of PONV] has increased a little bit. However, it is still not a problem which is regularly brought up during case discussions... Perhaps it is just something that slips your mind quite easily.*	resident intervention group
4	*Let's put it this way: to systematically provide PONV prophylaxis to all patients is impossible, because of the side effects. Although, when indicated for a specific patient, I believe you can accept the side effects.*	senior attending care-as-usual group
5	*No, it is not a conscious process. It is more that you roughly know the risk factors, you have an overall view of the patient, and then you start thinking: Does this patient has a high risk or not? But in my mind, I am not actually calculating the risk.*	resident intervention group
6	*When you open the anesthesia information management system, you immediately see it. Well, it is not that I immediately am consciously aware of it, sometimes I am, but not always. It occasionally happens that at the end of the procedure you think: 'O yes, the risk was high, shall I...*	senior attending intervention group
7	*I believe I would be inclined to administer more PONV prophylaxis, because you know that the predicted risk is based on an extensive calculation using several factors.*	junior attending intervention group
8	*You do not benefit that much from a predicted risk. It is more a way to direct your attention to PONV, because to actually solve the problem itself you need more details than just a predicted risk.*	senior attending care-as-usual group
9	*On one hand it is nice to treat a patient based on a true evidence-based recommendation, on the other hand all doctors want to decide for themselves... So I have mixed feelings about this. I believe it will be a matter of getting used to it. At a certain point it will become a custom, and especially when it is has been proven to work, you would be a fool not to follow the advice.*	senior attending intervention group
10	*The recommendation is probably only aimed at PONV, so I do believe that you should continue to consider which patient you have in front of you and what drugs you are going to administer.*	resident intervention group
11	*You can only treat a patient when you know the actual cause of the problem. Without that, the predicted risk will not do you any good, you need to know which factors contributed to that particular risk.*	senior attending care-as-usual group
12	*Of course it is nice to get some advice. On the other hand, all doctors, including me, feel that they can decide for themselves... I am in charge and I am also responsible.*	senior attending intervention group

appeared to be quite fragmented. There was a large variation in reported PONV risk factors among all interviewees (themes 2A and 2B), and several interviewees from both groups were reluctant to administer either dexamethasone, droperidol or both without being able to explain their reluctance (theme 3A, quote 2). Additionally, it was reported not to be a regular topic of discussions among the medical staff of the anesthesia department (theme 1D, quote 3).

intervention group: limited effect of the implementation strategy

The strategy for implementation of the prediction model seemed to have resulted in a limited increase in knowledge and awareness regarding PONV prevention (themes 1A, 2A, 2B and 3A, Table 4.1). Compared to interviewees from the care-as-usual group, interviewees from the intervention group seemed to be more aware of the frequency of PONV (themes 1A and 1B). Interviewees from the intervention group seemed only a little more aware of the efficacy of prophylactic antiemetics (theme 3A), whereas knowledge of other aspects of PONV prevention was similar in both study groups (themes 2A and 2B, Table 4.1).

unconscious process of decision making

For most physicians the process of deciding on PONV prophylaxis is implicit. Physicians usually weighed several factors in their decision to administer prophylactic antiemetics: they make a risk assessment based on patient and procedure characteristics (themes 2A and 2B, Table 4.1 and Appendix Section 1) and subsequently weigh the risk against benefits, harms and contraindications of antiemetics (themes 3A and 3C, Table 4.1 and quote 4, Table 4.2). The physicians did not report a well-defined process of assessing and weighing all these factors (themes 2C and 3B, quote 5), so the process predominantly takes place 'unconsciously' – or intuitively.

presenting predicted PONV risks may not be sufficient to change professional decisions

Although interviewees from the intervention group reported to use the predicted risks in their decisions on PONV prophylaxis, they were unable to explicate how they used the predicted risk in their decisions on PONV prophylaxis (theme 4B, Table 4.1 and quote 6, Table 4.2). The interviewees from the intervention group expected that when a recommendation on PONV prophylaxis would be added to the predicted PONV risk, they would follow the advice (theme 4G, quote 7). In contrast, the interviewees from the care-as-usual group reported that they would consider the presented risk of PONV mainly as a reminder (theme 4C), and that a recommendation on PONV prophylaxis would not have much added value (theme 4G, quote 8).

The intuitive use of the predicted PONV risk and a stated preference for an actionable recommendation by intervention group interviewees suggested that being presented only with a predicted risk may be difficult to use in a clinical decision. Adding a risk-corresponding treatment recommendation may assist physicians in interpreting the predicted risk for a decision on PONV prophylaxis (theme 4G, quote 9).

multifaceted and individualized decision making

The physicians of both groups reported one distinct barrier for using a prediction model: they felt that a prediction model does not take into account all aspects of a specific patient. The prediction model only predicted the risk for a specific outcome and did not weigh the benefit of treatment against the expected harms and contraindications for a particular patient with specific characteristics and comorbidity (theme 4E, Table 4.1 and quote 10, Table 4.2). Furthermore, the physicians indicated that even a patient-specific risk prediction is not truly patient-specific. The predicted PONV risk is an overall risk and does not reflect the specific pathway for becoming nauseous (theme 4F, quote 11). The physicians felt that decision support should not be implemented at the cost of their authority to make case-by-case decisions. It is a physician's responsibility to make an individualized decision whether treatment of a specific condition is warranted for his or her patient – even when they decide against the recommendation (theme 4D, quote 12).

structured survey

In the original RCT there were baseline differences in the characteristics of the physicians across the two groups (Table 4.3).[8] Physicians reported similar risk thresholds to the question which was asked before the RCT: 'Above which risk threshold would you administer PONV prophylaxis to a patient?'. Of the 57 physicians analyzed in the original RCT, 53 (93%) completed the survey (intervention group 29 (94%); care-as-usual 24 (92%)). Overall, physicians of the intervention group had positive total summation z-scores, whereas physicians of the care-as-usual group had negative scores (z-scores 0.66 (SD 2.4) vs. −0.80 (SD 2.4); $p<0.05$). This difference in z-scores indicated that intervention group physicians had a more positive

TABLE 4.3 BASELINE CHARACTERISTICS OF PHYSICIANS FROM THE SURVEY

	Care-as-Usual Group	Intervention Group
	(*n* = 26)*	(*n* = 31)*
Number of patients treated, median (IQR)	170 (128–300)	162 (120–207)
Age†, years, mean (SD)	43 (9)	43 (10)
Female sex, *n* (%)	11 (42)	15 (48)
Entered during study, *n* (%)	5 (19)	8 (26)
Left department during study, *n* (%)	4 (15)	7 (23)
Specialty: cardiac anesthesia†, n. (%)	2 (8)	2 (6)
pediatric anesthesia†, *n* (%)	3 (12)	4 (13)
pain medicine†, *n* (%)	1 (4)	4 (13)
Senior resident†, *n* (%)	12 (46)	11 (35)
Entered as senior resident during study, *n* (%)	4 (15)	4 (13)
Senior resident became attending anesthesiologist during study, *n* (%)	5 (19)	4 (13)
Risk threshold for administering PONV prophylaxis†, median (IQR)	30 (30–40)	30 (20–50)

* Number of physicians that participated in the cluster-randomized trial and were eligible for the survey

† Characteristic was documented at the moment of inclusion in the study

IQR = interquartile range; SD = standard deviation

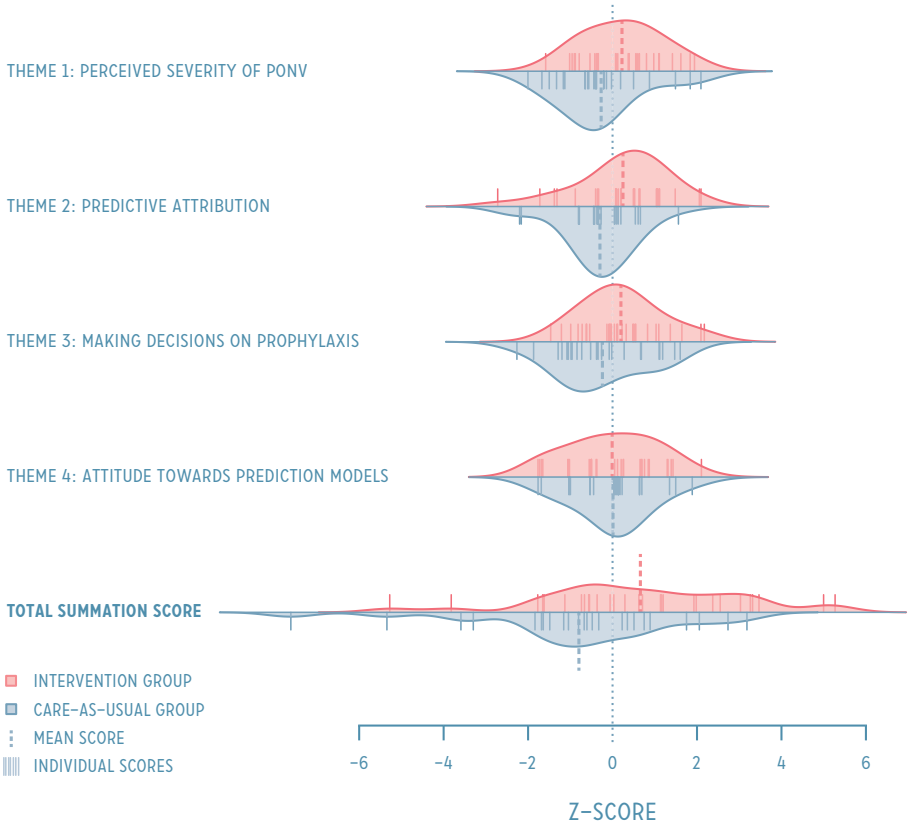


FIGURE 4.1 Density plot of the results the structured survey. The graph displays the differences in the distribution of the summary z-scores and total summation z-score between the physicians of both study groups. A positive z-score indicates that a physician had a more positive attitude towards prevention and risk-based management of postoperative nausea and vomiting (PONV) than the average attitude of the overall group of physicians.

TABLE 4.4 SURVEY SUMMARY Z-SCORES

	Care-as-Usual Group	Intervention Group
	(*n* = 24)*	**(*n* = 29)***
Theme 1: Perceived burden of PONV	−0.27 (1.0)	0.22 (0.9)
Theme 2: Professional risk stratification	−0.30 (0.8)	0.25 (1.1)
Theme 3: Making decisions on prophylaxis	−0.24 (1.1)	0.20 (0.9)
Theme 4: Attitude towards prediction models	0.01 (0.9)	−0.01 (1.1)
Total summation z-score	**−0.80 (2.4)†**	**0.66 (2.4)†**
Summation z-score for themes 1–3	−0.81 (1.9)	0.67 (1.8)

Data represent means (standard deviations) of the summary z-scores of the physicians within the same study group. A positive value for a study group indicates a more favorable attitude towards the theme as compared to the average of all physicians.

* The number of physicians who completed the survey

† $p < 0.05$

PONV = postoperative nausea and vomiting

attitude towards PONV prevention and the risk-dependent administration of PONV prophylaxis (Table 4.4, Figure 4.1). Similar to the results of the interviews, the post-hoc analysis on adding a recommendation to the predicted risk indicated that the intervention group physicians had a greater preference for the recommendation, which was statistically significant (median score intervention group 1; care-as-usual group 0; $p<0.05$).

adjusting the cluster-randomized trial analysis

When the original regression analyses were adjusted for the total summation z-score of the survey, results were similar to the original results for both the primary and secondary outcome (see Appendix Section 2, Table 8.5). There was no association between the total summation z-score itself and the incidence of PONV (odds ratio 1.0, 95% CI 1.0–1.0) or the administration of prophylactic antiemetics (rate ratio 1.0, 95% CI 1.0–1.0).

discussion

We studied how the implementation and use of an assistive prediction model allowed physicians to apply risk-tailored prevention or treatment strategies, and identified facilitators and barriers for risk-tailored management. In our example study, the implementation of the prediction model for PONV facilitated physicians in three ways: the physicians became more aware of the outcome PONV, were better informed on its risk factors, and had a more positive attitude towards preemptive management of PONV. Nevertheless, the facilitating effect of the prediction model on actual behavior – and thus patient outcomes – was modest. Physicians reported several barriers for using risk predictions from a prediction model, which may have counteracted the facilitating effect (see Box 4.1). First, the predicted outcome – i.e. PONV – was not a main area of attention for most physicians. Second, the physicians did not use a conscious process for making decisions on PONV prophylaxis. Third, the presentation of only outcome risks or probabilities – i.e. without further therapeutic management guidance – made it difficult to link the proper management to the predicted risks in individual patients. Fourth, a prediction model does not weigh the benefits and risks of prophylactic drugs with regard to the patient's comorbidity. The physicians argued that administration of PONV prophylaxis should remain an individualized decision made by a physician, not by a model. Furthermore, physicians who indicated to be more inclined to treat PONV prophylactically did not administer more prophylaxis to their patients than physicians with a lesser inclination. This suggests that the balance between the facilitating and barrier effects were in favor of the latter.

From this example study, we can learn how physicians may act upon the use of risk prediction models to guide their decision making in individual patients. The data suggest that the clinical decision-making process may be much more intuitive than physicians realize. In cognitive psychology literature, this intuitive, non-analytical process of decision making is referred to as a 'Type 1' process, as opposed to conscious and analytical reasoning, referred to as a 'Type 2' process.[17–19] Humans have a higher capacity for reasoning when they use the rapid Type

1 processes instead of the much slower Type 2 processes. The results of our study further indicate that only presenting the model's predicted risks – without corresponding treatment recommendations – will generally force doctors to use Type 2 reasoning. Physicians have to integrate their clinical experience with the probabilistic information from the prediction model, and translate this into a meaningful decision on subsequent actions and management, weighing all benefits and harms for a particular patient. This may actually be a difficult and cognitively demanding task, which makes the prediction tool seem less useful to physicians.

One might argue that this process is exactly what clinical judgment is about and that it is a physician's job to make such complicated decisions.[20–22] However, in our study the physicians also reported that the outcome PONV was not a main area of attention. This barrier (see Box 4.1) may work in two ways. First, 'attention' refers to information processes and task performance as it is used in human factors and ergonomics, and is related to workload.[23,24] Clinical decision making is an intricate task performed within a system with a high degree of competing information and a multitude of tasks.[25–27] Consequently, integrating an outcome risk into an otherwise subconscious decision process requires a lot of 'attention' in an already demanding environment. Second, PONV as an area of attention may refer to the physicians' knowledge of this particular condition or the perception of an outcome with limited patient burden .[28] In anesthesiology, PONV is considered the 'big little problem' – i.e., frequent but not life-threatening – in a network of other anticipated and possibly more severe consequences.[29,30] It is imaginable that physicians will prioritize their attention to preventable problems with a higher perceived burden than to preventable problems with a lower perceived burden.[31] In addition, when physicians believe that the predictive tool does not facilitate individualized decision making (the fourth barrier, see Box 4.1), they will be even less inclined to spend their time and attention to use the predicted outcome risks in their decisions.

The concept of how the three barriers may hinder physicians from using prediction models in their clinical decisions leads us to several recommendations for a successful implementation of a predictive tool in clinical practice and thus for the conduct of impact trials (see Box 4.1). The demand that a predictive tool poses on the attention of physicians may be reduced by adding treatment recommendations to the presented predicted risks or risk categories, as it reduces the cognitive effort required to translate the predicted risk into a clinical decision. This supports earlier recommendations that a directive prediction model implementation approach – i.e. adding an actionable recommendation to the predicted risk – may render better performance and patient outcomes than an assistive implementation approach – providing the risk without such recommendation.[1,4–6,32]

Prompted by the results of this study, we recently performed an add-on study to the cluster-RCT, in which we added therapeutic recommendations to categories of predicted risk. We indeed found a larger effect on the decision making as well as on the reduction in outcome incidence than with the assistive implementation approach of the cluster-RCT.[33] This supported our finding that adding management recommendations to predicted risks improves the uptake and impact of prediction models in practice. Secondly, automatic provision of the predicted risk that is smoothly integrated with the physician's workflow may also be key factors

BOX 4.1 SUMMARY BOX

Barriers for using risk prediction models in clinical practice

- The predicted outcome is not a main area of attention for the physicians.
- The decision making process of physicians is intuitive rather than analytical.
- The probabilistic knowledge of the outcome is difficult to use in decision making, certainly when the probability estimate it is not accompanied by corresponding (e.g. therapeutic) management recommendations (assistive prediction model).
- A prediction model does not weigh the benefits and risks of prophylactic drugs with regard to the patient's comorbidity.

Recommendations for a successful implementation of a prediction model in clinical practice

- Adding an actionable recommendation to the predicted risk (directive prediction model)
- Presentation of the predicted risk should be automated and smoothly integrated with the physician's workflow.
- The reasoning and research evidence of the underlying prediction model to show how risks are actually estimated, should also be available to physicians.
- A prediction model will be better perceived by physicians when it predicts outcomes that are relevant to them and their patients.

to reduce the required amount of attention and promote the use of the predicted outcome risks.[6,32] Third, a predictive tool may better aid physicians in making individualized decisions when the reasoning and research evidence that resides behind the tool is also presented to the physicians.[6,32] When physicians are better informed of which factors actually contribute to a predicted risk for a particular outcome, it is easier for them to understand how the predicted risk relates to their clinical experience. It was beyond the scope of this article to determine whether making physicians better informed would indeed improve their decision making or that the preference of the physicians for more information was an expression of overconfidence.[20,21,34–36] Finally, a predictive tool will likely be better perceived by physicians when it predicts outcomes that are relevant to them and their patients.

There are several limitations to our study. First, the results do provide a plausible – but not definitive – explanation for the 'no-effect-result' of our cluster-RCT. Second, we performed the interviews and surveys at the end of the RCT. That poses the problem that the interviews are based on hindsight and that not everything may be accurately recollected by the physicians. Fortunately, at the time of the interviews and surveys the physicians were still making decisions on PONV prophylaxis in their daily practice, yet without the prediction model. Their responses to most of the questions likely still represented their actual decision making. Third, we discussed that integrating clinical experience with the probabilistic knowledge from the prediction model may be a demanding task. We cannot discern from our results whether this task is demanding because the physicians find it to be 'simply too much work', or because the physicians do not know how their clinical experience may be integrated with the predicted risk.

In summary, we studied facilitators and barriers in the use of clinical prediction and decision models using a large randomized study on the implementation of a prediction model for PONV. Conscious integration of a predicted risk into the doctor's decision-making process may be difficult when the decision-making process is mostly subconscious and implicit. In a high-workload environment, such integration may require too much attention from a physician to actually integrate clinical and probabilistic knowledge. When one wishes to implement a prediction model to improve clinical practice, an important consideration is how to accommodate the new information in the physicians' workflow and how to ease the demand of attention on the physicians. Adding actionable patient-management recommendations to the predicted risk categories and providing physicians with the reasoning behind the predicted risks are the first steps to reduce such demand and improve the uptake of clinical prediction models in practice.

references

1. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209. PMID:16461965.

2. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009;119(17):2408–2416. PMID:19364974.

3. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–698. PMID:22397946.

4. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. PMID:23393430.

5. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. *BMJ*. 2004;328(7435):343–345. PMID:14764503.

6. Lobach D, Sanders GD, Bright TJ, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. *Evid Rep Technol Assess (Full Rep)*. 2012;(203):1–784. PMID:23126650.

7. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S-A. Computerized clinical decision support for prescribing: provision does not guarantee uptake. *J Am Med Inform Assoc*. 2010;17(1):25–33. PMID:20064798.

8. Kappen TH, Moons KGM, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. *Anesthesiology*. 2014;120(2):343–354. PMID:24105403.

9. Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of clinical prediction models for application in local settings. *Med Decis Making*. 2012;32(3):E1-E10. PMID:22427369.

10. Fernández-Guisasola J, Gómez-Arnau JI, Cabrera Y, del Valle SG. Association between nitrous oxide and the incidence of postoperative nausea and vomiting in adults: a systematic review and meta-analysis. *Anaesthesia*. 2010;65(4):379–387. PMID:20151955.

11. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2014;118(1):85–113. PMID:24356162.

12. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350(24):2441–2451. PMID:15190136.

13. Carlisle JB, Stevenson CA. Drugs for preventing postoperative nausea and vomiting. *Cochrane Database Syst Rev*. 2006;3(1):CD004125. PMID:16856030.

14. Charmaz K. Chapter 5. In: Silverman D, ed. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Sage Publications; 2006:208.

15. Mason M. Sample size and saturation in PhD studies using qualitative interviews. *Forum Qual Sozialforschung/Forum: Qualitative Social Research* 2010;11(3).

16. Draucker CB, Martsolf DS, Ross R, Rusk TB. Theoretical sampling and category development in grounded theory. *Qual Health Res*. 2007;17(8):1137–1148. PMID:17928484.

17. Stanovich KE. *Who Is Rational?: Studies of Individual Differences in Reasoning*. Mahwah: Lawrence Erlbaum Associates; 1999.

18. Kahneman D, Frederick S. Representativeness revisited: Attribute substitution in intuitive judgment. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, UK: Cambridge Univ. Press; 2002:49–81.

19. Evans JSBT. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu Rev Psychol*. 2008;59:255–278. PMID:18154502.

20. Malterud K. Reflexivity and metapositions: strategies for appraisal of clinical evidence. *J Eval Clin Pract*. 2002;8(2):121–126. PMID:12180360.

21. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022–1028. PMID:19638766.

22. Tonelli MR. The challenge of evidence in clinical medicine. *J Eval Clin Pract*. 2010;16(2):384–389. PMID:20367870.

23. Wickens CD. Multiple resources and performance prediction. *Theor Issues Ergon Sci*. 2002;3(2):159–177. doi:10.1080/14639220210123806.

24. Wickens CD. Multiple resources and mental workload. *Hum Factors*. 2008;50(3):449–455. PMID:18689052.

25. Gaba DM, Lee T. Measuring the workload of the anesthesiologist. *Anesth Analg*. 1990;71(4):354–361. PMID:2400118.

26. Loeb RG. A measure of intraoperative attention to monitor displays. *Anesth Analg*. 1993;76(2):337–341. PMID:8424512.

27. Stevenson RA, Schlesinger JJ, Wallace MT. Effects of divided attention and operating room noise on perception of pulse oximeter pitch changes: a laboratory study. *Anesthesiology*. 2013;118(2):376–381. PMID:23263015.

28. Lewis KL, Thompson JM. Health care professionals' perceptions and knowledge of infection control practices in a community hospital. *Health Care Manag (Frederick)*. 2009;28(3):230–238. PMID:19668064.

29. Kapur PA. The big "little problem." *Anesth Analg*. 1991;73(3):243–245. PMID:1831014.

30. Macario a, Weinger M, Truong P, Lee M. Which clinical anesthesia outcomes are both common and important to avoid? The perspective of a panel of expert anesthesiologists. *Anesth Analg*. 1999;88(5):1085–1091. PMID:10320175.

31. Ash JS, Sittig DF, Guappone KP, et al. Recommended practices for computerized clinical decision support and knowledge management in community settings: a qualitative study. *BMC Med Inform Decis Mak*. 2012;12(1):6. PMID:22333210.

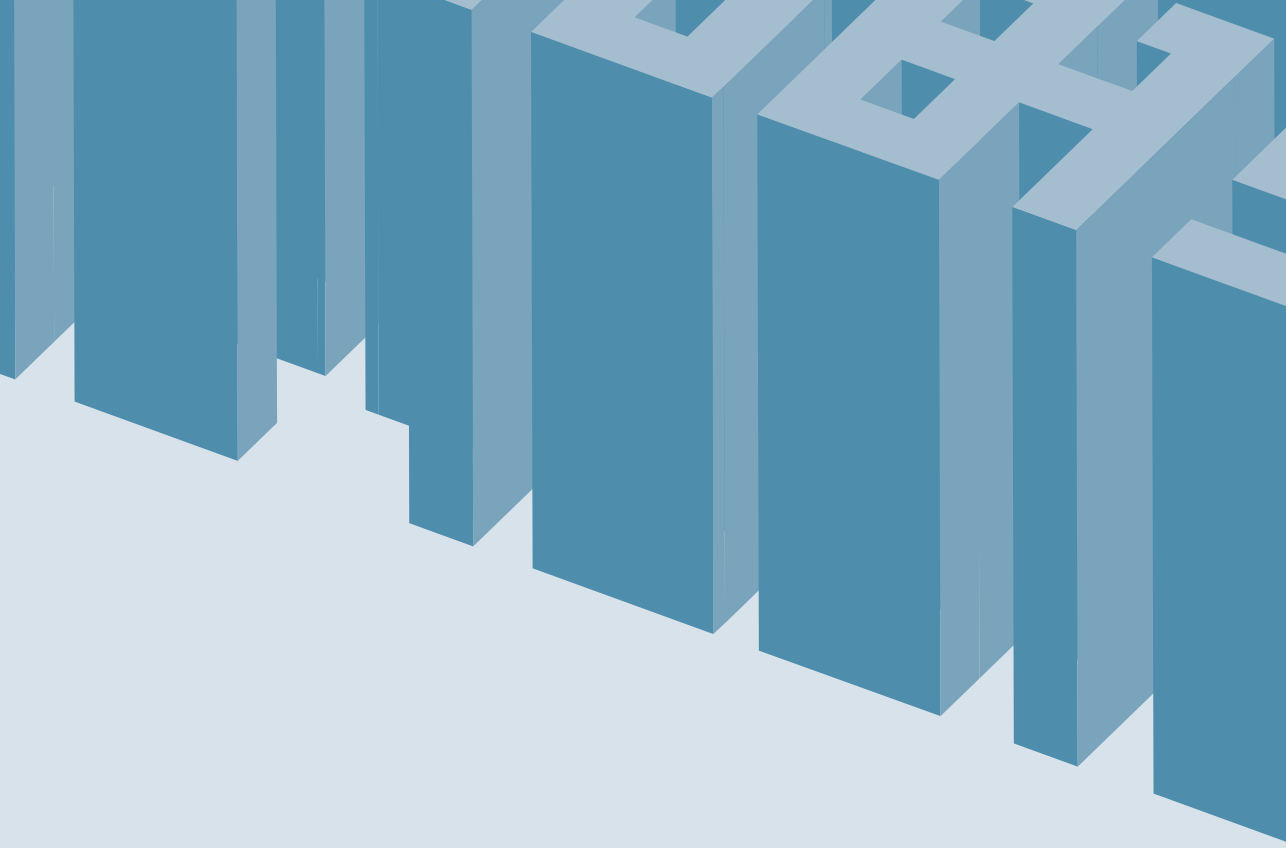
32. Kawamoto K, Houlihan C a, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765. PMID:15767266.

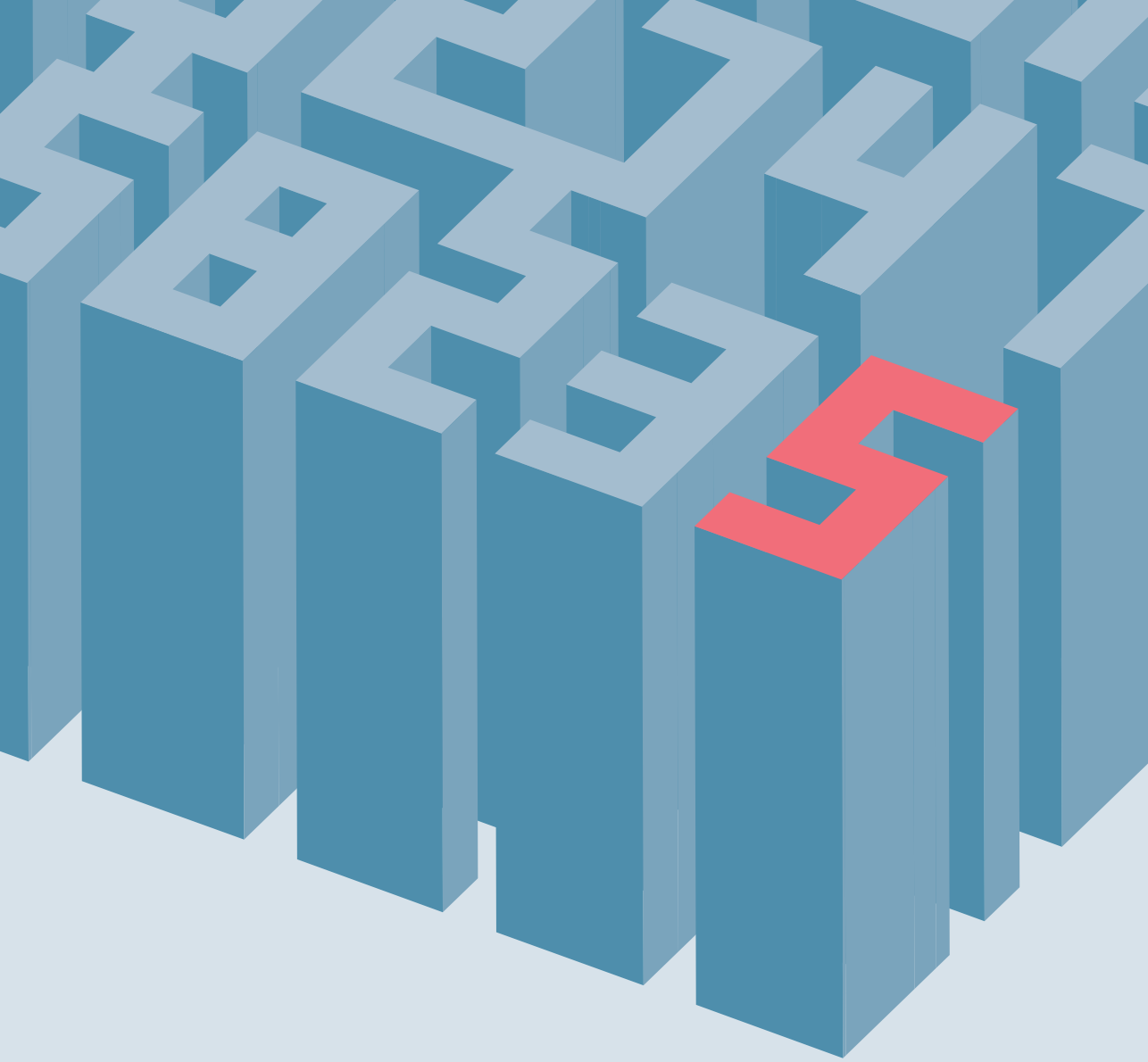
33. Kappen TH, Vergouwe Y, Wolfswinkel L Van, Kalkman CJ, Moons KGM, Klei WA Van. Impact of adding therapeutic recommendations to risk assessments from a prediction model for postoperative nausea and vomiting. *Br J Anaesth*. 2015;114(2):252–260. PMID:25274048.

34. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121(5 Suppl):S2-S23. PMID:18440350.

35. Croskerry P, Norman G. Overconfidence in clinical decision making. *Am J Med*. 2008;121(5 Suppl):S24-S29. PMID:18440351.

36. Miles RW. Fallacious reasoning and complexity as root causes of clinical inertia. *J Am Med Dir Assoc*. 2007;8(6):349–354. PMID:17619033.





IMPACT OF ADDING THERAPEUTIC RECOMMENDATIONS TO RISK ASSESSMENTS FROM A PREDICTION MODEL FOR POSTOPERATIVE NAUSEA AND VOMITING

T.H. Kappen, Y. Vergouwe, L. van Wolfswinkel, C.J. Kalkman, K.G.M. Moons, W.A. van Klei

British Journal of Anaesthesia. 2015 Feb;114(2):252–60

abstract

Background: In a large cluster-randomized trial on the impact of a prediction model, presenting the calculated risk of postoperative nausea and vomiting (PONV) on-screen (assistive approach) increased the administration of risk-dependent PONV prophylaxis by anesthesiologists. This change in therapeutic decision making did not improve patient outcome, i.e. the PONV incidence. The present study aimed to quantify the effects of adding a specific therapeutic recommendation to the predicted risk (directive approach) on PONV prophylaxis decision making and the incidence of PONV.

Methods: A prospective before-after study was conducted in 1,483 elective surgical inpatients. The before-period included care-as-usual and the after-period included the directive risk-based (intervention) strategy. Risk-dependent effects on the administered number of prophylactic antiemetics and incidence of PONV were analyzed by mixed-effects regression analysis.

Results: During the intervention period anesthesiologists administered 0.5 (95% CI 0.4–0.6) antiemetics more per antiemetic advised. This increased administration in PONV prophylaxis also resulted in a reduction in PONV incidence (OR 0.60, 95% CI 0.43–0.83), with an even greater reduction in PONV incidence in high-risk patients (OR 0.45, 95% CI 0.28–0.72).

Conclusions: Anesthesiologists administered more risk-dependent prophylactic antiemetics when a directive approach was used for risk-tailored intervention compared with care-as-usual. In contrast to the previously studied assistive approach, the increase in PONV prophylaxis now resulted in a lower PONV incidence, particularly in high-risk patients. When one aims for a truly 'PONV-free hospital', a more liberal use of prophylactic antiemetics must be accepted and lower risk thresholds should be set for the actionable recommendations.

Current guidelines on prevention of postoperative nausea and vomiting (PONV) recommend using risk-dependent strategies, where administration of antiemetic prophylaxis is based on individual risks predicted by a prediction model.[1,2] Although several prediction models have been developed,[3–6] their effect on clinical practice remains minimal – mainly because of poor implementation.[7,8] These disappointing results of risk-dependent PONV prophylaxis have resulted in an ongoing debate whether or not to shift to routinely administering multiple antiemetics to all patients, irrespective of their predicted risks.[9–11] Before switching to such a new, as yet unproven strategy of administering multiple antiemetics to every patient, the impact of risk-dependent strategies for PONV should be critically evaluated.[12–15] However, comparative studies assessing the actual impact of risk-dependent prophylaxis on the incidence of PONV are rare.[16]

In a large cluster-randomized study we previously showed that assisting anesthesiologists by only presenting the patient's calculated risk of PONV on-screen in the anesthesia information management system, but without further therapeutic directives per predicted risk, increased the number of prophylactic antiemetics administered by anesthesiologists.[17] However, this change in physician decision making did not decrease the incidence of PONV. We hypothesized that a greater impact would probably be achieved when being more directive by simply adding actionable recommendations to the presented risks.[12–14,18–20]

As a sequel to the previous cluster-randomized trial, the present before-after study aimed to quantify the effects of combining a specific therapeutic recommendation with the patient's predicted risk on both the incidence of PONV and the actual administration of risk-dependent prophylaxis.

methods

design and participants

The present study was a prospective before-after cohort study, conducted at the Anesthesiology Department of a Dutch university hospital (UMC Utrecht) in 2010. The study aimed to quantify the effects of a directive PONV prediction model approach – i.e. presenting predicted risks accompanied with non-obligatory, therapeutic recommendations – on both the incidence of PONV and the administration of antiemetic prophylaxis. Care-as-usual (see below) was studied during the before-period (January to March 2010), followed by an intervention period (April to May 2010), during which all physicians were provided with a recommendation on how many prophylactic antiemetics would be required to sufficiently lower their individual patients' PONV risks (see below).

According to Dutch law, research protocols that do not subject patients to a particular treatment or that require them to behave in a particular way, do not apply to the Medical Research Involving Human Subjects Act. As the decision support tools in our study protocol only provided evidence-based information to physicians, the institutional ethical review board waived the need for individual informed consent and approved the study protocol (Medical Ethics Review Board, UMC Utrecht, 11–553).

All adult patients undergoing general anesthesia for elective, non-ambulatory surgery who had visited the outpatient preanesthesia evaluation clinic were considered eligible for inclusion. Exclusion criteria were pregnancy, postoperative admission to the intensive care unit, overnight ventilation at the postanesthesia care unit, and inability to communicate in Dutch or English. All eligible patients from the time of study initiation were automatically included using the anesthesia information management system.

the prediction model

The implemented prediction model was originally developed in a population of a different university hospital in the Netherlands and had already been externally validated.[21,22] The model was subsequently updated for implementation at the UMC Utrecht.[23] The model consisted of seven predictor variables: age; gender; current smoking; type of surgery; inhalational anesthesia (including nitrous oxide); ambulatory surgery; and history of motion sickness or PONV (for a full model description see chapter 3, Table 3.1).

intervention

care-as-usual group

During the care-as-usual period, anesthesiologists were not exposed to any automated prognostic information by a prediction model. The prophylactic management of PONV was not standardized in any way, which was according to care-as-usual in our hospital. At that time, the existing, local protocol for administration of PONV prophylaxis only included a preferable order for antiemetic drugs, their dosage and timing of administration: 1) ondansetron 4 mg IV, 30 minutes before emergence of anesthesia; 2) droperidol 1.25 mg IV, 30 minutes before emergence of anesthesia; 3) dexamethasone 4 mg IV, after induction of anesthesia. Other prophylactic antiemetic drugs, such as NK–1 receptor antagonists, were not readily available at the anesthesia department during the study period.

At the postanesthesia care unit and the ward the PONV protocol consisted of rescue-treatment with an antiemetic drug: either one of the above antiemetic drugs if not previously administered, or metoclopramide 20 mg IV. There was no scheduled PONV prophylaxis prescribed for patients returning to the ward.

intervention group

The prediction model was implemented as a directive decision support tool in the anesthesia information management system (AnStat software, CarePoint Nederland BV, Ede, The Netherlands), a custom-made system written by one of the authors (LVW). The model presented a patient's predicted PONV risk accompanied with an advice on the number of prophylactic antiemetics to administer based on that individual's risk, i.e. a directive risk-based approach. The anesthesia information management system automatically presented this risk and the recommendation to the responsible anesthesiologist on the computer screen during each anesthetic case in each operating room.

The on-screen presentation was designed as a 'traffic light' with four colors (from green, through yellow and orange, to red). The initial color of the traffic light depended on the

patients predicted PONV risk and corresponded to the number of prophylactic antiemetics advised: from zero antiemetics (green) to three antiemetics (red). Anesthesiologists then decided whether to follow the advice and administer prophylactic antiemetics accordingly. The color – and hence the advice – did not change until an antiemetic drug was administered. The software 'adjusted' the predicted risk by a 26% relative risk reduction per antiemetic, which is the previously reported relative risk reduction.[24] When the adjusted risk fell below a threshold of 26% plus 4% for each antiemetic that already had been administered, the traffic light would turn 'green' and no further prophylactic antiemetics were advised. The 4% addition per antiemetic was aimed to ease the achievement of a 'green light' in high-risk patients, as very high-risk patients otherwise would never get a 'green light' even when treated with all available antiemetics.

Based on the calculations by the software tool, the predicted PONV risk was classified into one of four recommendation categories with an initial traffic light color: no antiemetics (green) below 26% predicted risk; one antiemetic below 41% predicted risk (yellow); two antiemetics below 62% predicted risk (orange); and three antiemetics for 62% or greater predicted risk (red). Recommendation thresholds were adapted from the 2007 Society of Ambulatory Anesthesia (SAMBA) guidelines on the management of PONV.[1] The recommendations did not specify which prophylactic antiemetics to give, but it was advised to follow the order of the existing, local protocol (see 'Care-as-usual group' section). The local protocol was easily accessible through a direct link on the main screen of the anesthesia information management system.

outcome and follow-up

The incidence of PONV was defined as the occurrence of at least one of the following events within the first 24 hours after surgery: an episode of nausea, an episode of vomiting, or the administration of any rescue antiemetic. For nausea, the patient was asked to rate their feeling of nausea during the preceding period on a three-point verbal rating scale (no / yes, a bit / yes, definitely) and for the analysis the variable was dichotomized to any nausea (no / yes). Vomiting was defined as the expulsion of gastric contents and was recorded as a binary outcome (no / yes). Research nurses and trained medical students collected data on the occurrence of postoperative nausea using a validated questionnaire.[6,25] Data were collected at the postanesthesia care unit (30 minutes and 60 minutes after arrival, and when leaving the unit), and 24 hours after surgery at the ward. The outcome variable for PONV was coded as missing when any of the follow-up measurements had not been completed.

The administration of risk-dependent PONV prophylaxis (physician behavior) was defined as the number of prophylactic antiemetics administered per patient and was recorded in the anesthesia information management system. The use of total intravenous anesthesia was not counted as a prophylactic intervention regarding the primary outcome, as it was unlikely to change during the anesthetic case. However, as inhalational anesthesia was a predictor within the prediction model, the presented recommendation did depend on the type of anesthesia used.

statistical analysis

Analysis was performed under the intention-to-treat principle. All statistical analyses were performed in R software (version 2.15.0). Statistical significance was defined as a two-sided alpha of 0.05. Continuous variables were visually assessed for a normal distribution using histograms and QQ-plots. Parametric variables were expressed as means with standard deviations, nonparametric variables were expressed as medians with interquartile ranges, and discrete variables were expressed as numbers with percentages.

The crude data on the administration and the incidence of PONV are shown into four risk categories according to the predictions and recommendations made by the decision support tool. Mixed-effects regression analyses were used for both outcomes: logistic regression for the incidence of PONV and linear regression for the number of prophylactic antiemetics per patient (*glmer*, *lme4* package, R software). A random intercept was included in the models, as the study was clustered by anesthesiologists. For both outcomes, allocation group, predicted PONV risks, and interaction between allocation group and predicted PONV risk were included as independent variables in the model. As the primary interest was a risk-dependent effect, we included the interaction term between the allocation group and predicted risk. The interaction term quantified the difference in treatment effect (between intervention and care-as-usual) across predicted risks; e.g. an odds ratio below one would signify that a reduction in PONV because of the directive approach was greater in patients with higher risks. For the PONV incidence analysis predicted PONV risks were included as a continuous variable, i.e. the PONV risk on the log odds scale. For the analysis of physician behavior, the relation between the intervention (the recommended number of antiemetics per predicted PONV risk category) and the outcome (the actual number of administered prophylactic antiemetics) was hypothetically expected to be linear (0 recommended antiemetics was likely to result in 0 administered antiemetics, 1 recommended antiemetic was likely to result in 1 administered antiemetic, etc.). Therefore, not the continuous predicted risk variable but rather the advised number of antiemetics and its interaction with allocation group were used as independent variables in a linear regression model with the administered number of antiemetics as the dependent variable.

As this was a non-randomized study, we had to adjust for potential differences between the care-as-usual and intervention group. Although inclusion of the predicted risk variable and its interaction term with allocation groups would probably adjust for most of the confounding, we a priori decided to additionally adjust for all variables from Table 5.1 that are either risk factors for PONV or may influence the decision on PONV prophylaxis (e.g. ASA class).[26] Furthermore, the anesthesiologists may have lowered their opioid prescription rates in patients with a high predicted risk to prevent PONV. As opioid usage was an intermediate variable, we performed a separate sensitivity analysis to adjust the primary outcome analysis for opioid usage.

Before multivariable modelling, all continuous variables were tested for nonlinearity using restricted cubic splines.[27] Missing data were multiply imputed (n = 10) using a regression approach in R (*areglmpute*, *Hmisc* package). Imputation of missing variables was based on predictors, outcome variables, and other perioperative data.[28–31] As PONV was coded missing

when any of the follow-up measurements was incomplete, the non-missing follow-up measurements of PONV were added to the imputation process to serve as auxiliary variables to impute missing values for PONV. Subsequently, the imputed values for PONV were included into the mixed-effects regression models, instead of deleted.[32] The anesthesiologists were added as an extra variable to the imputation model to take into account clustering in the data.

results

A total of 1,480 patients were included in the study and analyzed: 1,022 during the care-as-usual period and 458 during the intervention period. The mean predicted PONV risks were comparable between allocation periods. However, small differences in the distribution of predicted PONV risk categories existed between allocation periods. Differences in baseline of several predictor variables are likely to be related to the small differences in the distribution of the predicted PONV risk (Table 5.1).

In total, 81% of all follow-up measurements on PONV were completed (intervention group 75%; care-as-usual group 83%), with 92% of all patients having at least one follow-up measurement completed (intervention period 87%; care-as-usual period 94%). The incidence of PONV was 42%

TABLE 5.1 PATIENT CHARACTERISTICS

		Care-as-Usual Group	Intervention Group
	N*	(*n* = 1022)	(*n* = 458)
Age, years, mean (range)	1480	52 (18–94)	54 (18–88)
Female sex, *n* (%)	1480	496 (49)	193 (42)
ASA class, *n* (%)	1477		
1		337 (33)	121 (27)
2		561 (55)	272 (60)
3		116 (11)	62 (14)
4		8 (1)	0 (0)
Current smoking, *n* (%)	1453	288 (29)	109 (24)
Surgery with a high PONV risk, *n* (%)	1104	82 (12)	53 (13)
History of PONV / Motion sickness, *n* (%)	1397	212 (22)	66 (15)
Use of inhalational anesthesia†, *n* (%)	1480	472 (46)	224 (49)
Predicted risk of PONV, %, mean (SD)	1480	40 (13)	39 (12)
Predicted PONV risk in categories, *n* (%)	1480		
<26% (0 antiemetics advised)		127 (12)	60 (13)
26–41% (1 antiemetic advised)		443 (43)	217 (47)
41–62% (2 antiemetics advised)		384 (38)	161 (35)
>62% (3 antiemetics advised)		68 (7)	20 (4)
Operation duration, min, median (IQR)	1480	128 (85–188)	133 (86–206)

* N represents the total number of non-missing observations for that characteristic

† As compared to intravenous anesthesia using propofol

PONV = postoperative nausea and vomiting; IQR = interquartile range; SD = standard deviation

TABLE 5.2 INCIDENCE OF PONV PER CATEGORY OF PREDICTED RISK

			Care-as-Usual Group	Intervention Group
Predicted PONV risk in categories		N*	(*n* = 1022)	(*n* = 458)
<26%	(0 antiemetics advised)	116	19 (23)	10 (29)
26–41%	(1 antiemetic advised)	432	123 (41)	50 (38)
41–62%	(2 antiemetics advised)	354	167 (64)	48 (52)
>62%	(3 antiemetics advised)	60	41 (82)	5 (50)

Data represent absolute numbers (%) of PONV
* N represents the total number of non-missing observations for that characteristic
PONV = postoperative nausea and vomiting

during the intervention period, compared with 50% during the care-as-usual period. Table 5.2 shows the crude, risk-dependent effect of the intervention on the incidence of PONV. Confirmed by the regression analysis, there was a significant reduction in the incidence of PONV during the intervention period in comparison to the care-as-usual period (OR 0.60, 95% CI 0.43–0.83), with a greater reduction in high-risk patients (OR interaction term 0.45, 95% CI 0.28–0.72). The statistical significance of the risk-dependent reduction in PONV and the risk-dependent Numbers Needed to Treat (NNT) are reflected by Figure 5.1. Adjustment for baseline characteristics or opioid usage did not change these results and inferences. Results of the adjusted and unadjusted regression analyses can be found in Table 5.3. Differences in odds ratios for the variable predicted risk between complete case analysis and after multiple imputations can be explained by the confounder correction using all predictors from the prediction model.

TABLE 5.3 LOGISTIC REGRESSION ANALYSIS ON THE INCIDENCE OF PONV

	Unadjusted	Confounder adjusted*	Adjusted for opioid usage
		Complete case†	
Intervention period	**0.52 (0.35–0.78)**	**0.43 (0.27–0.69)**	**0.54 (0.36–0.81)**
Predicted risk‡	**3.9 (2.9–5.4)**	0.12 (0.0–7.4)	**4.0 (2.9–5.6)**
Interaction: intervention period and predicted risk‡	**0.43 (0.24–0.77)**	**0.36 (0.18–0.71)**	**0.39 (0.21–0.71)**
		Multiple imputation	
Intervention period	**0.64 (0.47–0.89)**	**0.60 (0.43–0.83)**	**0.66 (0.48–0.90)**
Predicted risk‡	**2.9 (2.2–3.8)**	0.73 (0.13–4.2)	**2.9 (2.2–3.8)**
Interaction: intervention period and predicted risk‡	**0.49 (0.31–0.77)**	**0.45 (0.28–0.72)**	**0.47 (0.30–0.73)**

Numbers represent odds ratios with 95% confidence intervals.
Bold numbers are statistically significant odds ratios.
* Adjusted for possible confounders: age (continuous, restricted cubic splines, 5 knots), gender, ASA class, current smoking, middle ear of abdominal surgery, history of PONV / motion sickness, use of inhalational anesthesia, procedure duration (continuous, restricted cubic splines, 5 knots)
† Cases with missing variables were discarded
‡ The predicted PONV risk on the log odds scale
PONV = postoperative nausea and vomiting

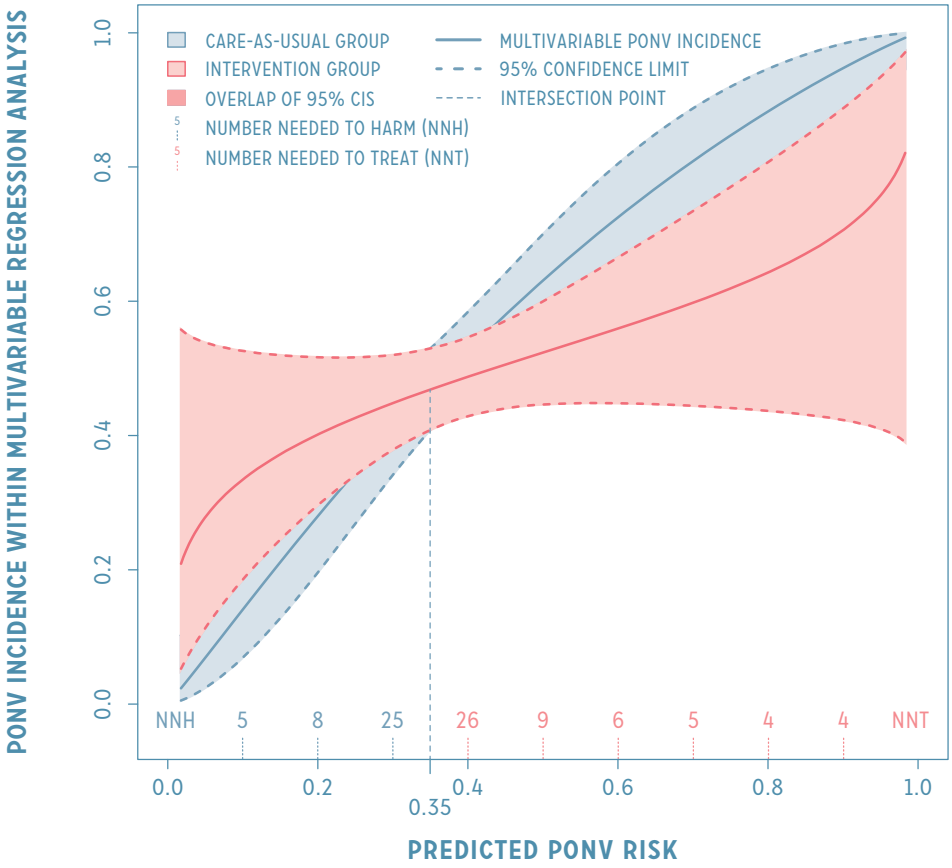


FIGURE 5.1 Graphical representation of the mixed effects regression analysis on the incidence of postoperative nausea and vomiting. Figure 5.1 may be interpreted as the occurrence of postoperative nausea and vomiting (PONV) after receiving prophylaxis, in patients with a particular predicted risk within each group. The differences between the blue and red areas represent the effect of implementation of the prediction model on the occurrence of PONV in patients with a particular predicted risk.

The solid lines and their 95% confidence intervals (CI) represent the fixed effects of the mixed effects regression analyses. The dotted vertical line shows the intersection point of both groups. The mixed effects models included fixed effects for the variables allocation group, predicted risk of PONV, interaction between allocation group and predicted risk, and time. Random effects were included for the intercept, predicted risk, interaction between allocation group and predicted risk, and time. The lines were calculated using the average for the variable time. The 95% CIs were calculated from the covariance matrix for the variables allocation group and its interaction term with predicted risk. The results are considered statistically significant when the 95% CIs of one study group do not overlap with the solid line of the other study group.

For each risk increment of 10%, the Number Needed to Treat (NNT) for PONV within 24 hours after surgery was calculated by taking the multiplicative inverse of the absolute risk reduction between the intervention group and the care-as-usual group. When the absolute risk reduction was negative, the Number Needed to Harm (NNH) was calculated instead.

TABLE 5.4 NUMBER OF PROPHYLACTIC ANTIEMETICS ADMINISTERED PER CATEGORY OF PREDICTED RISK

			Care-as-Usual Group (*n* = 1022)				Intervention Group (*n* = 458)			
Predicted PONV risk in categories		N*	0	1	2	3	0	1	2	3
<26%	(0 antiemetics advised)	188	**117 (92)**	10 (8)	0 (0)	0 (0)	**53 (88)**	7 (12)	0 (0)	0 (0)
26–41%	(1 antiemetic advised)	661	362 (82)	**69 (16)**	11 (2)	1 (0)	66 (30)	**141 (65)**	10 (5)	0 (0)
41–62%	(2 antiemetics advised)	546	258 (67)	98 (26)	**21 (5)**	7 (2)	35 (22)	20 (12)	**103 (64)**	3 (2)
>62%	(3 antiemetics advised)	88	31 (46)	28 (41)	9 (13)	**0 (0)**	3 (15)	3 (15)	7 (35)	**7 (35)**

Data represent absolute numbers (%) of patients who received either zero, one, two, or three of the available prophylactic antiemetics: ondansetron, droperidol, and/or dexamethasone. Bold numbers represent cells where anesthesiologists administered the exact number of administered prophylactic antiemetics as recommended for that risk category (i.e. the compliance).

* N represents the total number of non-missing observations for that characteristic

PONV = postoperative nausea and vomiting

TABLE 5.5 LINEAR REGRESSION ANALYSIS ON THE NUMBER OF PROPHYLACTIC ANTIEMETICS

	Unadjusted	Confounder adjusted*	Confounder adjusted*
	Complete case†		Multiple imputation
Intervention period	0.04 (−0.09–0.17)	0.06 (−0.10–0.22)	0.04 (−0.09–0.17)
Recommendation categories‡	**0.18 (0.14–0.23)**	0.07 (−0.05–0.20)	0.04 (−0.05–0.14)
Interaction: intervention period and recommendation categories‡	**0.48 (0.39–0.56)**	**0.49 (0.39–0.58)**	**0.50 (0.41–0.58)**

Numbers represent regression coefficients with 95% confidence intervals.

Bold numbers are statistically significant regression coefficients.

* Adjusted for possible confounders: age (continuous, restricted cubic splines, 5 knots), gender, ASA class, current smoking, middle ear of abdominal surgery, history of PONV / motion sickness, use of inhalational anesthesia, procedure duration (continuous, restricted cubic splines, 5 knots)

† Cases with missing variables were discarded

‡ Regression coefficients represent an increase in the number of administered prophylactic antiemetics per advised prophylactic antiemetics – i.e. per risk category

PONV = postoperative nausea and vomiting

The number and type of prophylactic antiemetics were documented for all patients. During the intervention period anesthesiologists complied with the recommendation of the clinical decision support tool and administered the recommended number of prophylactic antiemetics in 66% of patients. Although no actual recommendations were given during the care-as-usual period, the fictional compliance (i.e. the prescription behavior that would be recommended, if the decision rule had been active) was 20%, resulting in an absolute increase of compliance of 46%. During the intervention period 76% of the prophylactic antiemetics were administered at the appropriate time during the anesthetic case, compared with 72% during the care-as-usual period. The timing of the prophylactic antiemetics was unrelated to the occurrence of PONV, either at the postanesthesia care unit or at the ward.

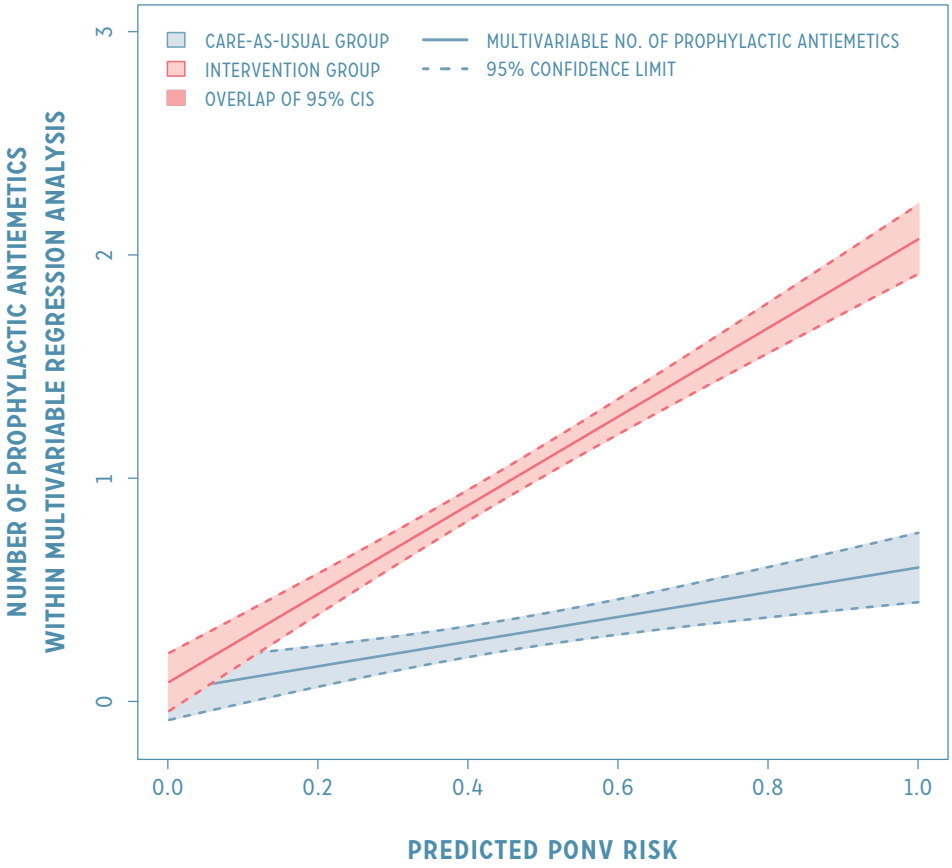


FIGURE 5.2 Graphical representation of the mixed effects regression analysis on the administration of prophylactic antiemetics by anesthesiologists. Figure 5.2 may be interpreted as the number of prophylactic antiemetics a patient with a particular predicted risk of PONV would receive from any anesthesiologist within each group. The differences between the blue and red areas therefore represent the changes in physician behavior concerning prescription of antiemetic prophylaxis, caused by implementation of the prediction model.

The solid lines and their 95% confidence intervals (CI) represent the fixed effects of the mixed effects regression analyses. The dotted vertical line shows the intersection point of both groups. The mixed effects models included fixed effects for the variables allocation group, predicted risk of PONV, interaction between allocation group and predicted risk, and time. Random effects were included for the intercept, predicted risk, interaction between allocation group and predicted risk, and time. The lines were calculated using the average for the variable time. The 95% CIs were calculated from the covariance matrix for the variables allocation group and its interaction term with predicted risk. The results are considered statistically significant when the 95% CIs of one study group do not overlap with the solid line of the other study group.

The crude, risk-dependent effects of the intervention on administration of prophylactic antiemetics and the risk-dependent compliance are given in Table 5.4. The increase in administration of prophylactic antiemetics was confirmed in the results of the linear regression analysis after multiple imputation and confounder adjustment. Anesthesiologists administered more antiemetic prophylaxis in a risk-dependent manner during the intervention period. In

this period, for each additional antiemetic advised the anesthesiologists actually administered 0.49 (95% CI 0.41–0.58) additional antiemetics (Table 5.5, regression coefficient for the interaction term, far right column). The statistical significance of the increased administration of risk-dependent PONV prophylaxis is reflected by Figure 5.2. These results were not different from the models without adjustment for confounding, as is given in Table 5.5.

discussion

We studied the effects of the implementation of risk-dependent PONV prophylaxis based on the predicted PONV risks, calculated by a prediction model. The model provided automated decision support by presenting predicted risks directly accompanied with treatment recommendations to anesthesiologists in the operating room, i.e. a directive approach was used.[12–14,20] This directive approach clearly increased administration of risk-dependent antiemetic prophylaxis to patients and also reduced the incidence of PONV within 24 hours after surgery, particularly in patients with higher risks.

The results of the present study are in contrast with the results of our previous study. The previous study – a large cluster-randomized trial – tested an assistive approach for model implementation, i.e. presenting only the risk of PONV without a therapeutic recommendation. This assistive strategy had little effect on the PONV incidence, whereas the directive strategy of the present study significantly reduced (OR 0.60, 95% CI 0.43–0.83) the incidence of PONV within 24 hours after surgery.[17] The difference in results between the two studies suggests that the impact on clinical practice may be larger when a prediction model is accompanied with an actionable recommendation to aid physicians in their decision making.

The impact on physician behavior in the present study is similar to other impact studies of PONV decision support.[7,8] Kooij et al. implemented a directive decision support tool with an absolute increase in compliance of 40%, which is comparable with the results of our present (directive) study (46%). Using an assistive approach Frenzel et al. achieved an absolute increase in compliance of 5%, with a single feedback report, which is similar to our previous (assistive) study.[17] Effects on the incidence of PONV (absolute risk reduction of 8%) for our directive approach were within the lower range of results from other studies that reported overall absolute risk reductions ranging from 8 to 35%.[16,33–37] Unfortunately, the merit of such a comparison is limited, attributable to differences in actual administration of PONV prophylaxis, study design and study analysis. Most of the other studies did not randomize, did not adjust for confounding, or did not have a proper control group, which makes it challenging to compare their results to our study.

Regardless of a significant decrease in PONV incidence, the actual impact of the directive approach on PONV occurrence seems at best moderate and does not come close to its desired potential impact – a 'PONV-free hospital'.[38] However, differences between actual and potential

impact do not imply that we should discard risk-dependent strategies for administration of PONV prophylaxis. Several interactions between clinicians and the decision support tool need to be considered for our study, before coming to a conclusion.[12]

First, the predictive performance of the prediction model may have been insufficient to improve clinical decision making. The predictive performance of our prediction model was comparable with other PONV prediction models (c-statistic around 0.70).[4,6,23] With a moderate predictive performance, decisions based on the model may not have been superior to care-as-usual, i.e. clinical judgment.

Second, the interface of the decision support tool may have affected the compliance to the recommendations. For example, desensitization may have occurred, as the color would often not change during the surgical part of the case, as prophylactic antiemetics are supposed to be administered either at the start or towards the end of the anesthetic case.

Third, despite a large increase in risk-dependent PONV prophylaxis, physicians did not fully adhere to the presented risks and therapeutic recommendations. For example, patients in the highest risk category where three prophylactic antiemetics advised, received on average two prophylactic antiemetics. Several barriers to use prediction models and decision support have been identified in the literature and may account for the incomplete compliance by the anesthesiologists.[39–42]

Fourth, the increased administration of prophylactic antiemetics may have been an effect of an overall increase in attention to PONV (a Hawthorne effect), instead of an effect of the intervention. It is the actual goal of decision support to increase the attention for a particular patient problem, hence decision support systems are sometimes referred to as 'reminder systems'.[43] In the case of a prediction model as a form of decision support, the goal is to actually improve risk-dependent decision making. As there was a risk-dependent effect on both primary and secondary outcome in our study, it is likely that the anesthesiologists indeed used the information presented by the decision support tool, instead of Hawthorne effect.

The results of our study indirectly support the suggested general multimodal prevention strategy from the recent update of the SAMBA consensus guidelines for management of PONV.[2] In a general multimodal prevention strategy prophylactic antiemetics are routinely administered to all patients, which implies that more prophylactic antiemetics will be administered to patients with low to moderate risks of PONV. The risk thresholds and treatment recommendations as used in our decision support tool were developed from clinical considerations, as literature on specific treatment recommendations for our prediction model was not available at the time of the study's initiation.[1] From the NNT in Figure 5.1 we may observe that our prevention strategy was mainly effective in patients with high risks of PONV, as the NNT dropped below ten for patients with a predicted risk percent of 50% or greater. Would one adopt a more liberal approach to administration of prophylactic antiemetics, our decision support tool would recommend more antiemetics to patients with lower risks. This may improve NNT for patients with predicted risks below 50%, resulting in a PONV incidence that may be closer to a 'PONV-free hospital' than the resultant PONV incidence in our study.

In recent literature there is an ongoing debate whether more liberal use of prophylactic antiemetics should be achieved by either a risk-dependent strategy or routine administration of several prophylactic antiemetics to every patient (a general multimodal approach).[2,9–11,36] At this moment, neither points of view have a substantive evidence base. On the one hand there is a debate on the clinical utility of prediction models, as their implementation rate remains very low. Moreover, a general multimodal approach may facilitate implementation as PONV prevention becomes a professional routine and there is no need any more for a reminder. Decision support tools may then be reserved for more complicated clinical decisions, which require interpretation of more sophisticated – and non-routine – information. On the other hand there is a debate on the side-effects of antiemetics and a reluctance to administer polypharmacy with several, potentially unnecessary drugs, and a prediction model may assist in achieving an optimal risk-to-benefit ratio. This study adds some support for both points of view. It is one of the first prospective comparative studies that demonstrate that risk-dependent prophylaxis actually decreases the incidence of PONV. At the same time, anesthesiologists were only partially compliant with the presented recommendations and the absolute decrease in PONV incidence was limited. If the anesthesia community is willing to embrace large-scale multimodal prevention of PONV – as suggested by the recently updated SAMBA guidelines on PONV management – this may enable us to evaluate whether a 'PONV-free hospital' is actually achievable with an acceptable rate of side-effects.

We conclude that risk-dependent PONV prophylaxis is not only efficacious in clinical trials, but also effective in clinical practice when a real-time, computer-based prediction model is used in combination with risk-based recommendations on PONV prophylaxis. We also conclude that implementation of a risk prediction model combined with treatment recommendations per predicted risk, yields better effects on clinical decision making and patient outcomes than the use of a prediction model without addition of such treatment recommendations. As the resultant PONV incidence remained high, a more liberal use of prophylactic antiemetics and lowering of the risk thresholds for the actionable recommendations may be needed to achieve a truly 'PONV-free hospital'.

references

1. Gan TJ, Meyer T a, Apfel CC, et al. Society for Ambulatory Anesthesia guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2007;105(6):1615–1628. PMID:18042859.

2. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2014;118(1):85–113. PMID:24356162.

3. 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(3):693–700. PMID:10485781.

4. Apfel CC, Kranke P, Eberhart LHJ, Roos A, Roewer N. Comparison of predictive models for postoperative nausea and vomiting. *Br J Anaesth*. 2002;88(2):234–240. PMID:11883387.

5. 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(3):237–242. PMID:11861340.

6. Van den Bosch JE, Kalkman CJ, Vergouwe Y, et al. Assessing the applicability of scoring systems for predicting postoperative nausea and vomiting. *Anaesthesia*. 2005;60(4):323–331. PMID:15766334.

7. Frenzel JC, Kee SS, Ensor JE, Riedel BJ, Ruiz JR. Ongoing provision of individual clinician performance data improves practice behavior. *Anesth Analg*. 2010;111(2):515–519. PMID:20529985.

8. Kooij FO, Klok T, Hollmann MW, Kal JE. Automated reminders increase adherence to guidelines for administration of prophylaxis for postoperative nausea and vomiting. *Eur J Anaesthesiol*. 2010;27(2):187–191. PMID:19478673.

9. Eberhart LHJ, Morin AM. Risk scores for predicting postoperative nausea and vomiting are clinically useful tools and should be used in every patient: con--'life is really simple, but we insist on making it complicated'. *Eur J Anaesthesiol*. 2011;28(3):155–159. PMID:21192269.

10. Kranke P, Eberhart LHJ. Possibilities and limitations in the pharmacological management of postoperative nausea and vomiting. *Eur J Anaesthesiol*. 2011;28(11):758–765. PMID:21799417.

11. Pierre S. Risk scores for predicting postoperative nausea and vomiting are clinically useful tools and should be used in every patient: pro--'don't throw the baby out with the bathwater'. *Eur J Anaesthesiol*. 2011;28(3):160–163. PMID:21206276.

12. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209. PMID:16461965.

13. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(jun04 2):b606. PMID:19502216.

14. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–698. PMID:22397946.

15. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. PMID:23393430.

16. Kooij FO, Vos N, Siebenga P, Klok T, Hollmann MW, Kal JE. Automated reminders decrease postoperative nausea and vomiting incidence in a general surgical population. *Br J Anaesth*. 2012;108(6):961–965. PMID:22382396.

17. Kappen TH, Moons KGM, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. *Anesthesiology*. 2014;120(2):343–354. PMID:24105403.

18. Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ*. 1998;317(7162):858–861. PMID:9748183.

19. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. *BMJ*. 2004;328(7435):343–345. PMID:14764503.

20. Kawamoto K, Houlihan C a, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765. PMID:15767266.

21. Van den Bosch JE, Moons KG, Bonsel GJ, Kalkman CJ. Does measurement of preoperative anxiety have added value for predicting postoperative nausea and vomiting? *Anesth Analg*. 2005;100(5):1525–1532. PMID:15845719.

22. Van den Bosch JE. Prediction of postoperative nausea and vomiting. [PhD Thesis].2006:87–115.

23. Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of clinical prediction models for application in local settings. *Med Decis Making*. 2012;32(3):E1-E10. PMID:22427369.

24. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350(24):2441–2451. PMID:15190136.

25. Visser K, Hassink E a, Bonsel GJ, Moen J, Kalkman CJ. Randomized controlled trial of total intravenous anesthesia with propofol versus inhalation anesthesia with isoflurane-nitrous oxide: postoperative nausea with vomiting and economic analysis. *Anesthesiology*. 2001;95(3):616–626. PMID:11575532.

26. Groenwold RHH, Klungel OH, Grobbee DE, Hoes AW. Selection of confounding variables should not be based on observed associations with exposure. *Eur J Epidemiol*. 2011;26(8):589–593. PMID:21796419.

27. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387. PMID:8668867.

28. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*. 1999;18(6):681–694. PMID:10204197.

29. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59(10):1087–1091. PMID:16980149.

30. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367(14):1355–1360. PMID:23034025.

31. Groenwold RHH, Moons KGM, Vandenbroucke JP. Randomized trials with missing outcome data: how to analyze and what to report. *CMAJ*. 2014. PMID:24778353.

32. Von Hippel PT. Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociol Methodol*. 2007;37(1):83–117. doi:10.1111/j.1467–9531.2007.00180.x.

33. Pierre S, Corno G, Benais H, Apfel CC. A risk score-dependent antiemetic approach effectively reduces postoperative nausea and vomiting--a continuous quality improvement initiative. *Can J Anaesth*. 2004;51(4):320–325. PMID:15064260.

34. Rüsch D, Eberhart L, Biedler A, Dethling J, Apfel CC. Prospective application of a simplified risk score to prevent postoperative nausea and vomiting. *Can J Anaesth*. 2005;52(5):478–484. PMID:15872125.

35. Sigaut S, Merckx P, Peuch C, Necib S, Pingeon F, Mantz J. Does an educational strategy based on systematic preoperative assessment of simplified Apfel's score decrease postoperative nausea and vomiting? *Ann Fr Anesth Reanim*. 2010;29(11):765–769. PMID:20950990.

36. Kumar A, Brampton W, Watson S, Reid VL, Neilly D. Postoperative nausea and vomiting: simple risk scoring does work. *Eur J Anaesthesiol*. 2012;29(1):57–59; author reply 59–60. PMID:21826006.

37. Biedler A, Wermelt J, Kunitz O, et al. A risk adapted approach reduces the overall institutional incidence of postoperative nausea and vomiting. *Can J Anaesth*. 2004;51(1):13–19. PMID:14709454.

38. Kranke P, Schnabel A, Eberhart LH, Muellenbach R, Morin AM, Roewer N. Providing effective implementation of antiemetic strategies: the postoperative nausea and vomiting-free hospital is a laudable and realistic goal. *Eur J Anaesthesiol*. 2011;28(4):308–309. PMID:21513079.

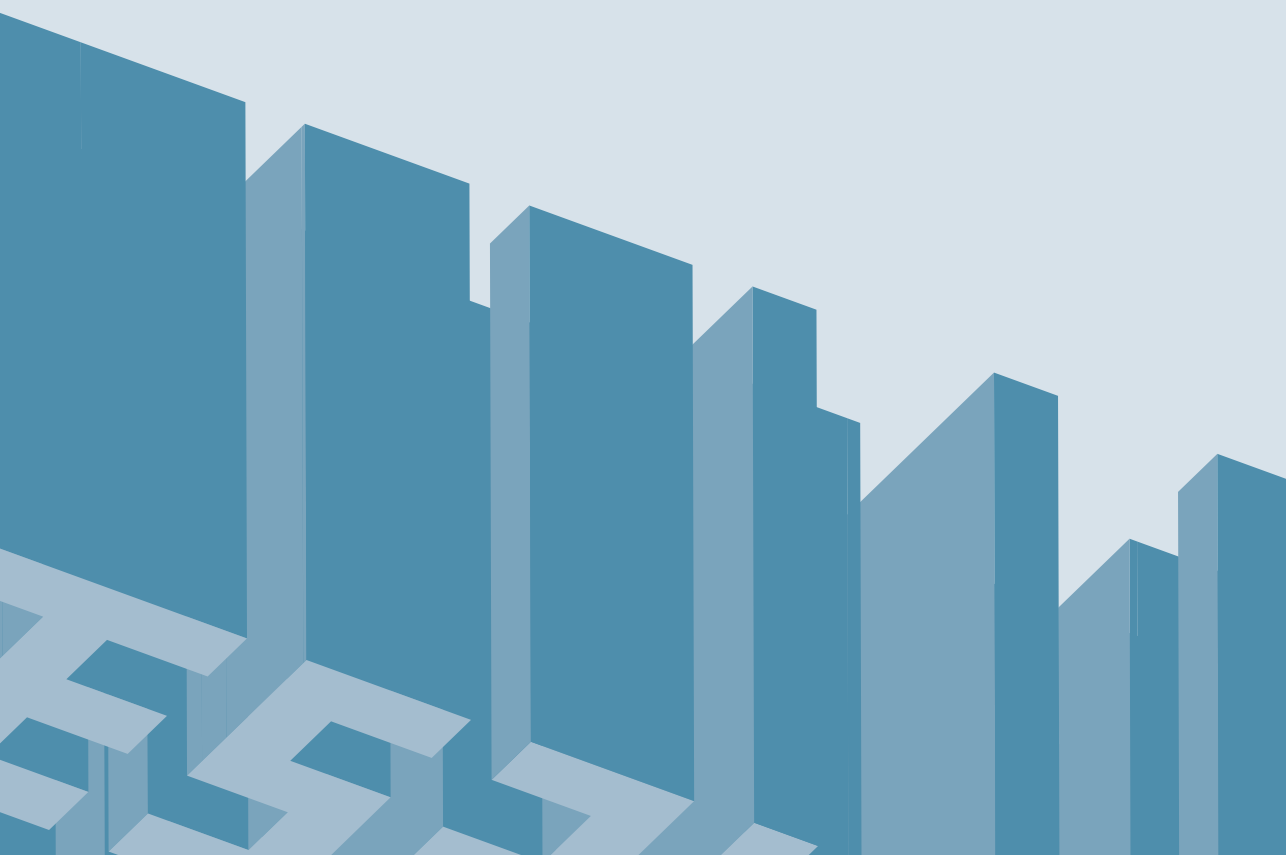
39. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282(15):1458–1465. PMID:10535437.

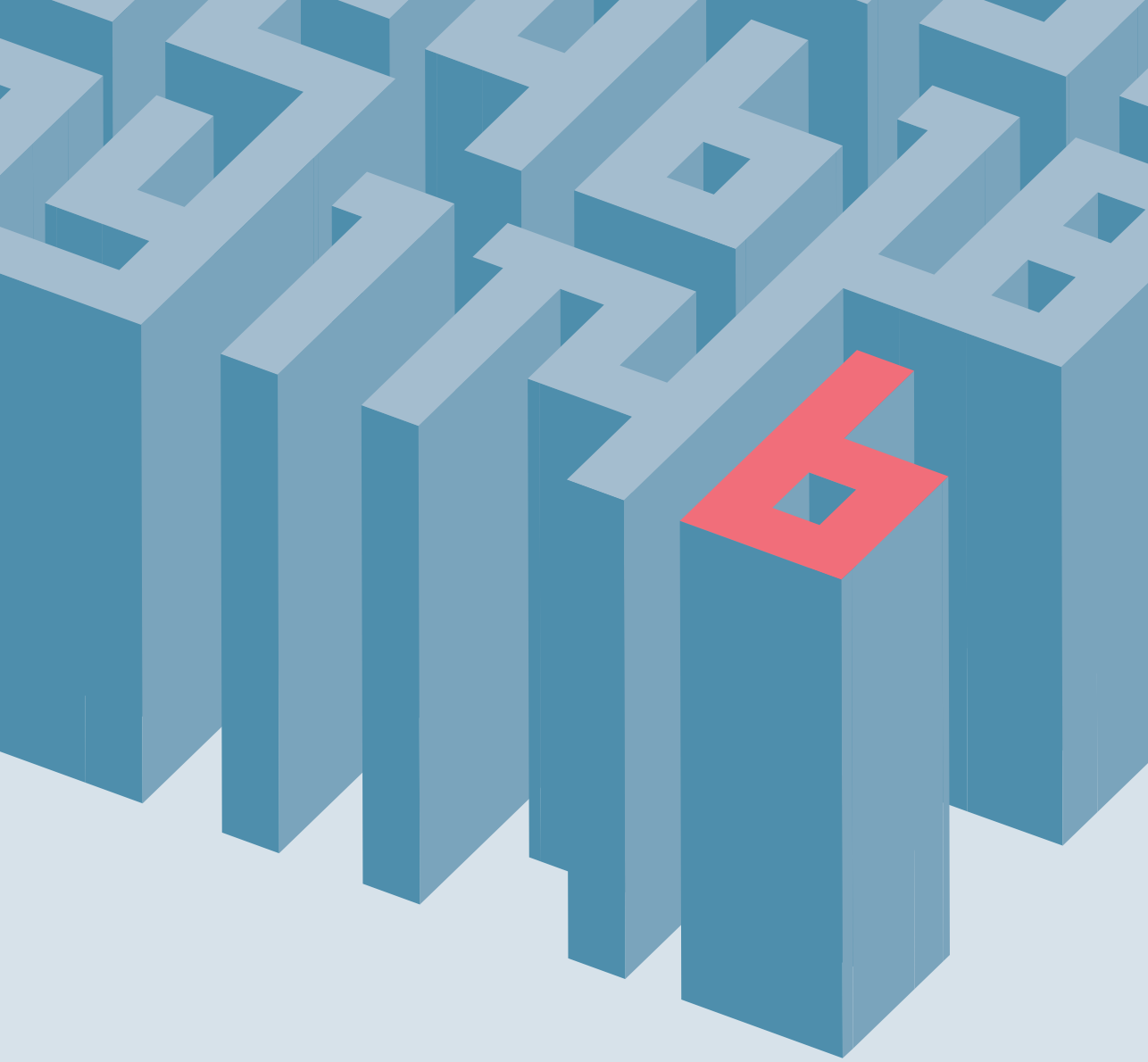
40. Grol R. Improving the Quality of Medical Care: Building Bridges Among Professional Pride, Payer Profit, and Patient Satisfaction. *JAMA J Am Med Assoc*. 2001;286(20):2578–2585. PMID:11722272.

41. Grimshaw JM, Shirran L, Thomas R, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care*. 2001;39(8 Suppl 2):II2-II45. PMID:11583120.

42. Swennen MHJ, van der Heijden GJMG, Blijham GH, Kalkman CJ. Career stage and work setting create different barriers for evidence-based medicine. *J Eval Clin Pract*. 2011;17(4):775–785. PMID:20438602.

43. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. 2011;18(3):327–334. PMID:21422100.





GENERAL DISCUSSION I: EVALUATING THE IMPACT OF THE USE OF PREDICTION MODELS IN CLINICAL PRACTICE: CHALLENGES AND RECOMMENDATIONS

T.H. Kappen, W.A. van Klei, L. van Wolfswinkel, C.J. Kalkman, Y. Vergouwe, K.G.M. Moons

Prediction models are abundant in the medical literature.[1–3] They are used to predict an individual's probability of a current (diagnostic model) or a future (prognostic model) disease, complication or other health state. Numerous clinical guidelines advocate using prediction models[4–6] and increasing online availability allows patients to calculate their own risks.[7,8] Prediction models are used to predict an individual's probability of a current (diagnostic model) or of a future (prognostic model) disease, complication or other health state. Healthcare providers and patients use such diagnostic and prognostic predictions to decide on the most appropriate course of action – ideally in shared decision making, including subsequent diagnostic testing, treatment actions, life style changes or a combination. A prediction model is developed from a particular data set to which it typically is optimally fitted.[9] A model's ability to accurately predict in other individuals should thus always be verified in so-called (external) validation studies.[10–12] Poor validation results can often be improved by simply tailoring or optimizing the model to the clinical setting at hand.[13,14] However, it is important to realize that good predictive performance is no guarantee that clinically using a model will improve health outcomes of individuals.[12,15,16] Model predictions first have to affect decision making.

Such effects on decision actions and individual outcomes can directly be quantified in so-called 'prediction model impact studies', where a prediction model is used in practice and the subsequent actions and outcomes are compared to care-as-usual.[1,3,12] Randomized study designs provide the most robust comparisons, especially when randomized at cluster level, e.g. physicians or hospitals.[3,12,17] An important aspect to consider when planning such an impact study is how the model's predictions will be presented to physicians and/or individuals. Predictions may simply be presented as a numerical probability without corresponding decision recommendation – an assistive prediction tool – or as a decision recommendation which may or may not include the numerical probability – a more directive prediction tool.[1,12] An assistive prediction tool leaves physicians more room to combine the predictions with their clinical judgment, whereas a directive prediction tool is said to have greater impact.[18–20]

Nonetheless, the vast majority of prediction models used in practice are not externally validated or optimized to the setting at hand, let alone that a proper impact study has been performed.[3,10,21,22] Recent recommendations, including by these authors, have stressed the need for prediction model validations, adjustments and impact assessments.[1,3,12] Although there is ample evidence that using prediction models change the behavior of physicians, whether they improve the quality of their decision making and subsequent patient outcomes remains highly inconsistent.[20,23–25] The efforts required to perform a large-scale cluster-randomized impact study also raises the question whether all – properly developed, validated and setting-optimized – prediction models should indeed be subjected to an impact study, and if so how this should be done.

We recently performed a series of validation and impact studies on a specific prediction model for postoperative nausea and vomiting (PONV)[26,27] that was previously developed and externally validated (Table 6.1, left column).[28,29] During these consecutive studies we encountered several problems with the study design, the statistical analysis and the interpretation of the results. We here chronologically summarize the challenges we encountered and provide

TABLE 6.1 PREDICTORS AND REGRESSION COEFFICIENTS OF THE ORIGINAL AND UPDATED MODEL

Predictor	Regression coefficients		Predictor
Original model		Updated model	
Age (years)	−0.022	−0.017	Age (years)
Female gender	0.46	0.36	Female gender
Current smoking	−0.63	−0.50	Current smoking
History of PONV or motion sickness	0.76	0.60	History of PONV or motion sickness
Lower abdominal or middle ear surgery	0.61	0.48	Abdominal or middle ear surgery*
Isoflurane and/or nitrous oxide anesthesia†	0.72	0.35	Inhalational anesthesia†
N/A	–	−1.16	Outpatient surgery‡
Intercept	0.15	0.12	Intercept

* In the updated model the predictor included lower abdominal, upper abdominal, and laparoscopic surgery in addition to middle ear surgery

† As compared to intravenous anesthesia using propofol

‡ Predictor not included in the original model

PONV = postoperative nausea and vomiting

recommendations (Box 6.1) to address them, to help decide whether and how to design, analyze and interpret a prediction model impact study.

1. tailoring the model to a local setting

Before the prediction model is implemented in a new local setting, either for clinical use or an model impact study, it is recommended that a model's predictive performance is first verified in that setting, and updated if needed.[10–12] Incorrect predictions from predictive performance likely leads to improper treatment decisions that are unlikely to improve patient outcome. Local practice, medical care and patient population may not be similar to the practice and population from which the prediction model was derived. Tailoring or updating the prediction model in that setting will overcome such differences and optimize its predictive performance.[12,14,29] It may simply suffice to adjust the baseline risk or hazard (e.g. the intercept). However, larger differences between populations or changes in clinical practice may require that the weights of predictors are adjusted or even that predictors are added to the model. Before we implemented and studied the impact of the previously developed and externally validated PONV prediction model, we first validated and updated the model in our local setting to overcome several of such differences (Table 6.1).[29] The original model showed a very poor predictive performance: c-statistic 0.62 (95% confidence interval (CI) 0.60–0.64) and calibration slope 0.57 (95% CI 0.48–0.66). After updating, these values became 0.68 (0.66–0.70) and 1.00 (0.89–1.10) respectively.

dealing with missing predictor values

Furthermore, to study the impact of a prediction model one needs to apply the (updated) model in each study participant to estimate the outcome probability. However, at the moment of probability calculation one or more predictor values may be missing (unobserved) in an individual such that the outcome probability cannot be estimated. A real-time strategy to

BOX 6.1 LESSONS LEARNT ON PREDICTION MODEL IMPLEMENTATION

I. Tailor the prediction model to the local setting

Verify the predictive performance of the prediction model in the local setting

A prerequisite for a good prediction model when implemented in practice, is that its predictions are accurate. Local practice, medical care and patient population may not be similar to the practice and population from which the prediction model was derived. Before a prediction model is used in clinical practice its predictive performance in the local setting at hand should be validated.

Update the prediction model to optimize the predictive performance in the local setting

When the prediction model does not sufficiently perform in the local setting the model should be updated. Simple adjustments to the prediction model, e.g. of the baseline risk/hazard or predictor weights, may overcome the problem of poor predictive accuracy.

Real-time strategy to handle missing predictor variables at the time of risk calculation

When estimating the model's probability, one or more predictor values may be missing. Based on the other predictors of the model, additional patient information, and information about the local clinical process and logistics the missing value may be estimated using a real-time imputation strategy. A multivariable imputation strategy is preferred over simply omitting the predictor variable or using an overall mean value.

II. Design a tailored format for the prediction tool that maximizes the ease of use

Add a decision recommendation to the predicted probabilities

A directive prediction tool provides a recommendation on further diagnostic, treatment, or preventive interventions that corresponds with the predicted probability. A directive prediction tool may be easier to use by physicians in their decision making, than when an assistive prediction tool provides only the predicted probabilities without recommendations.

Automatic calculation and presentation of the model's probability within the physician's workflow

Minimizing manual predictor value entry and integrating the estimation of the model's probability in the electronic patient record, facilitates the ease of use of a prediction model by care providers. In addition, probabilities need to be presented at the time of actual decision making. For example, physicians are more likely to use a predicted probability that automatically pops up in the electronic patient records, than a probability for which they have to manually enter patient data in a web-based prediction tool.

Provide the reasoning and/or research evidence behind the predicted probability

The reasoning and research evidence of the underlying prediction model, e.g. indicating which predictors are included in the model and their relative weights, should be explained to the users. This enhances face value, acceptation and belief in the model and thus the willingness to use the model's probabilities to guide their decision making.

III. Tailor the prediction tool to its causal pathway

Implementing a prediction tool is a complex intervention that has multiple, interacting components

A prediction tool affects patient outcome through a causal chain that consists of the model's predictions, the physicians who make the decisions, and the subsequent interventions that are applied. The components of the causal chain do not simply 'add up' to an effect on patient outcome, but interact with each other. The prediction tool may thus have unintended downstream effects on patient outcome.

Consider to study the effects on both physician behavior and patient outcome

The downstream effects of a prediction tool are not always predictable. Changes in process or behavior may not be sufficient to improve patient outcome. Nonetheless, studying the effects of a prediction tool on patient outcome typically requires more time and money.

The predicted probabilities can have unintended negative effects on decision making

The predicted probabilities – and recommendations – of the prediction tool are designed to improve the decision making of physicians. However, the probabilities may also be a distraction or a nuisance to physicians, which may have a negative effect on their decisions and patient outcome.

The predicted probabilities may be related to the pathophysiologic mechanisms of the applied interventions

The predicted probabilities are typically calculated from several patient and disease characteristics. The same patient and disease characteristics may be related to the pathophysiological mechanism of the possible interventions. However, when this is an inversed relation, a patient with a high predicted probability of a health condition will not respond well to the intervention.

Unintended effects should be considered during the design of the prediction tool

Understanding the causal chain of the prediction tool may help anticipate on unintended effects of the prediction tool. This may help to tailor the format of the prediction tool to the local clinical setting and improve the effectiveness of the prediction model as a complex intervention.

IV. Design and analysis of an impact study

Compare the use of a prediction model to care-as-usual

Physicians are not naive in patient selection and making interventional decisions. A prediction model should be compared to the usual care of physicians to demonstrate its added value to clinical decision making.

A cluster-randomized trial may be the optimal design

A cluster-randomized trial may be the optimal design to quantify the effects of a prediction model on physician behavior and patient outcome. Randomization of practices or practitioners aims to prevent contamination between study groups. Nonetheless, a cluster-randomized study does not always provide a conclusive answer (see the text of this manuscript). The time and costs to perform a cluster-randomized study should be weighed against its expected informational value.

Baseline measurements before the start of the cluster-randomized trial

It is not always possible to prevent contamination between study groups in a cluster-randomized impact study. Baseline measurements for both the intervention and care-as-usual group may provide additional information on contamination and Hawthorne effects. In addition, it allows to more closely observe the individual change in behavior of physicians that are exposed to the risk predictions.

Analyze the results of the study conditional on the predicted risk

A risk-dependent change in behavior is expected, thus both physician behavior and patient outcome should be studied for risk-dependent differences between study groups. In other words, the predicted risk is an effect modifier and should be handled as such in the statistical analysis, e.g. through stratification or include it as an interaction term with 'study group' in a regression model.

All predictor variables should be available for care-as-usual patients

A risk-dependent analysis of the results of the study requires that predicted risk can also be calculated for care-as-usual patients. When the prediction model contains only signs and symptoms as predictor variables, this will generally not pose a problem. However, more costly or more invasive predictor variables should also be collected for care-as-usual patients.

V. Understanding the results of an impact study

Additional in-depth information may improve understanding of the results

Alternative sources of information may help to understand which factor of the causal chain of a prediction tool is the weakest link and how that factor may be changed to improve the intervention. Such additional information may be acquired through both qualitative and quantitative methods.

impute the missing value is preferred over omitting the predictor from the model or imputing an overall mean value of that predictor.[29,30] Missing predictor values may better be estimated using other observed information on that patient, including the other predictors of the model. Other – less obvious – sources of information can also be considered.[29] In our example, the moment of prediction and thus the first use of the model is at the beginning of the surgical procedure. Type of surgery, one of the predictors (Table 6.1), was sometimes only available as free text. We estimated this predictor value based on the surgical specialty that was performing the procedure.

II. the format of the prediction tool

assistive versus directive prediction tools

Current literature suggests that a directive prediction tool – i.e. with risk-corresponding decision recommendations – may have greater impact on physician decision making than an assistive format – i.e. without such recommendations.[1,12,18–20] However, the scientific evidence for recommending a directive format is mainly derived from reviews in which studies with and without specific recommendations are compared for their rate of successful implementation. These reviews do not provide a direct comparison of an assistive and a directive format for

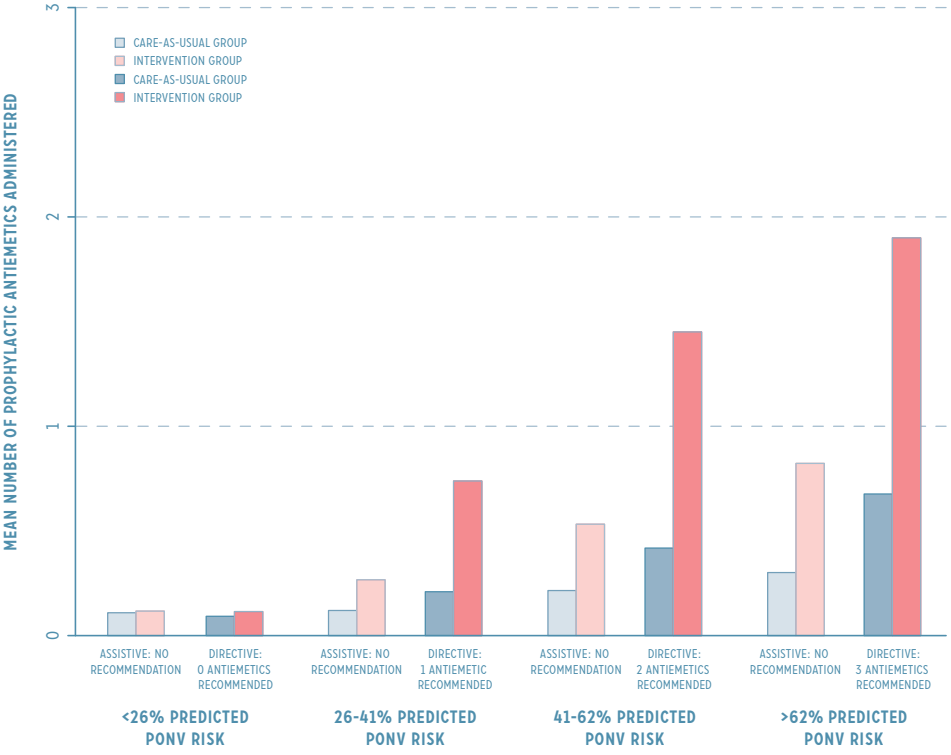


FIGURE 6.1 Comparison of physicians' administration of antiemetic prophylaxis

a single prediction model within a single population. We studied the impact of the use of an assistive versus a directive prediction tool.

To observe the impact of the assistive PONV prediction tool in a cluster-randomized trial, we randomized 57 physicians of the Anesthesia Department at the University Medical Center Utrecht and evaluated outcomes in the 12,000 surgical patients that they treated.[26] When a physician registered his/her name at the start of the procedure in the intraoperative electronic patient record, it was verified whether the physician was randomized to the intervention or care-as-usual group. For the intervention group physicians, the predicted PONV probabilities of their patients was presented on-screen during the entire procedure, but without corresponding recommendations on the number of prophylactic antiemetics to administer. We found that in the intervention group more prophylactic antiemetics were administered: overall rate ratio 2.0 (95% CI, 1.6–2.4) (Figure 6.1).

Unexpectedly, this increase in risk-dependent prophylaxis was not accompanied by a decreased PONV incidence: overall odds ratio, 0.97 (95% CI, 0.87–1.1) (Figure 6.2). Hence, the prediction tool changed physician behavior, but this did not translate into improved patient outcomes. Prompted by this result, we implemented the prediction model as a directive prediction tool and added a recommendation on the number of prophylactic antiemetic drugs, given the patient probability.[27] This study was a before-after study in the same setting,

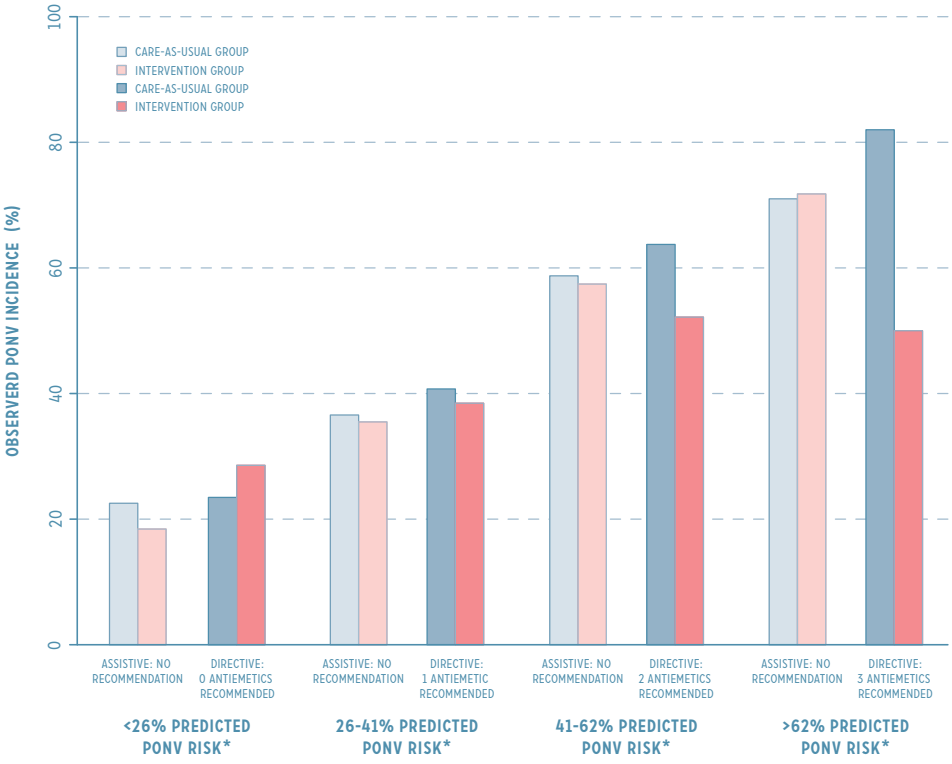


FIGURE 6.2 Comparison of the incidence of postoperative nausea and vomiting (PONV)

after the cluster-randomized trial had ended. In contrast to the assistive prediction tool in the first impact study, the directive prediction tool now also substantially improved actual patient outcome (Figure 6.2): overall odds ratio, after adjustment for baseline differences: 0.60 (95% CI, 0.43–0.83).

The results of this comparison shows that a directive prediction tool changes physician decisions more than an assistive prediction tool, with a subsequent measurable improvement in patient outcome. To our knowledge, this is one of the first comparisons that demonstrates in a single population of physicians and patients that a directive prediction tool has a greater impact on clinical practice than an assistive prediction tool.

the physician's perspective

To understand how physicians use predicted probabilities in their decision making, we subsequently performed interviews and a structured survey among the participating physicians in the intervention groups] We found (Box 6.2) that using a prediction tool may require substantial cognitive effort from physicians. Apparently it is not trivial to convert predicted probabilities from a prediction model into a specific therapeutic decision. The physicians reported that adding decision recommendations to predicted risk categories decreases the required cognitive effort and increases the ease of use of the prediction tool, which was supported by the before-after study's results.[31–33]

Specific design features can improve a prediction tool's ease of use. Automatic provision of the predicted probabilities and smooth integration with the physician's workflow are also key factors to improve usability, which may increase the impact on patient outcomes.[20,19] Furthermore, a prediction tool may better aid physicians in decision making when the reasoning and research evidence of the underlying prediction model is also provided.[20,19] When physicians are well informed of which predictors actually contribute to a predicted probability for a particular outcome, it enhances understanding of the predicted probabilities

BOX 6.2 SUMMARY OF RESULTS FROM THE INTERVIEWS AND SURVEYS*

Barriers for using risk prediction models in clinical practice

- The predicted outcome is not a main area of attention for the physicians.
- The decision making process of physicians is intuitive rather than analytical.
- The probabilistic knowledge of the outcome is difficult to use in decision making, certainly when the probability estimate it is not accompanied by corresponding (e.g. therapeutic) management recommendations (assistive prediction model).
- A prediction model does not weigh the benefits and risks of prophylactic drugs with regard to the patient's comorbidity.

PONV = Postoperative Nausea and Vomiting

* After the cluster-randomized trial of the assistive implementation of the PONV prediction model, we interviewed and surveyed the physicians that participated in the trial to explore how the predicted probabilities influenced physician decision making, and which presentation format was considered most useful.

and adherence to the corresponding management recommendations. From user feedback we learned that the tool's format should thus be designed to maximally facilitate physicians in using the predicted probabilities and recommendations for their decisions.[20,24,34,35]

III. the causal pathway of the prediction model

the importance of studying patient outcome

Although a prediction model simply presents an individual's predicted probability, when used in a decision support tool it should influence decision making sufficiently to improve the individual's health outcome. The 'causal pathway' through which a prediction model can affect patient outcomes consists of the accuracy and presentation of its predictions, the interpretation by physicians, as well as the interventions or management they apply based on the presented model information. This suggests that, provided the efficacy of the intervention is known from previous trials, one can study to what extent the use of a prediction model changes treatment decisions, and quantify the impact of the use of the model on patient outcomes from the decision changes and the efficacy of the intervention.

Some, including ourselves, previously proposed that a prediction model may not have to be studied for its effects on patient outcome when the efficacy of the intervention has been established.[12,16,36] However, the downstream effects of using a prediction model in clinical practice are not always predictable. In our example, the efficacies of the various prophylactic antiemetics were well established by several randomized trials and meta-analyses.[37,38] Nevertheless, despite an increased administration of prophylactic antiemetics in the cluster-randomized trial of the assistive model, there was no change in the PONV incidence (Figures 6.1 and 6.2). This discrepancy in results may indicate that either the predictive performance of the model was insufficient, the impact on physician decision making was still too small, or the antiemetic drugs were not as effective as they were efficacious in previous randomized trials and meta-analyses, or a combination. Given the design of the study and of the index intervention – i.e. a combination of the predicted probability, physician interpretation, and treatment effect – the individual role of each in explaining the discrepant results cannot be disentangled. Clearly, introducing clinical use of a prediction model with subsequent management actions is a complex intervention.[15,39] As the multiple components interact, the effect of a prediction tool on patient outcome is not simply the sum of the consecutive components.[40] Changes in process or behavior may thus not always be sufficient to improve outcome and it may be necessary to formally study the effects on patient outcome. The challenge is to determine when this is necessary. When one better understands which factors determine physician decision making, one could identify specific situations where there is a substantial risk of unintended effects and thus a need to study patient outcome.

unintended effects of the prediction tool

The intended effect of the prediction tool is to help physicians select the right intervention for the right patient. But physicians will regularly use other information in their decisions than only the information presented by the prediction tool (Box 6.2). As this other information is not

captured by the predictors, the predicted probabilities may have an unexpected interaction with such information. In our example, several pathophysiological mechanisms contribute to the predicted outcome of PONV and the predictors are related to these mechanisms. For example, nausea and vomiting may be caused by abdominal distension which is reflected by the predictor 'abdominal surgery' (Table 6.1), whereas anesthetics and opioids may invoke nausea through the chemoreceptor trigger zone, which monitors the presence of systemic toxins. Nonetheless, most of the antiemetic drugs act on the chemoreceptor trigger zone and may not be as effective for abdominal causes of nausea and vomiting. A patient with certain risk factors may not respond to an intervention that does not act on one of the causative pathophysiological mechanisms.

The consequence of the interaction between the model-predicted probabilities and the intervention, depends on how physicians respond to the predicted probabilities.[15] When physicians are aware of the pathophysiology underlying both the predicted risk and the expected result of the intervention, they are more likely to select that intervention (Box 6.2). In contrast, when physicians do not acknowledge the pathophysiology underlying the predicted probabilities, they may simply select any intervention or no intervention at all. In such cases a prediction tool may do more harm than good. It is even conceivable that without exposure to the prediction tool, physicians might have obtained a better outcome by simply basing their management on patient factors and pathophysiologic reasoning. Awareness of such phenomena may help to improve the presentation and format of the prediction model.[15,40,41]

IV. design and analysis of the actual impact study

study design

The most important feature of a prediction model impact study is that the exposure to the studied prediction tool (intervention group), either assistive or directive, is compared to a group that is not exposed to the model and its predictions (care-as-usual group). It is important to note that in the intervention group a learning curve may occur: after repeated exposure to the predictions in a variety of patients, physicians will become better at estimating the risk in subsequent similar patients – even without actual use of the model.[12,16] In a typical, non-clustered, randomized trial this learning curve may very likely dilute the effectiveness and impact of the model use: patients are randomized to either the intervention or care-as-usual, yet their physicians participate in both groups.[17] The effects of a learning curve may be minimized, though not completely prevented, by randomization at a cluster level, e.g. physicians or hospitals.[42] Then, each physician participates in either the intervention or care-as-usual group. By definition, only physicians of the intervention group may – unintendedly – experience a learning curve, which does not necessarily lead to a dilution of the contrast between the two groups.

The major drawback of randomization at the cluster-level is that the study requires a larger sample size and will thus more time and resources. Less costly alternatives are available, such as a non-randomized before-after study, where one obviously needs to compare and adjust for baseline differences between the two groups.[12,16] However, before deciding to conduct a prospective impact study, randomized or not, one might first conduct a decision analytic

modelling approach.[3,12,43] Before studying the impact of a prediction model in a longitudinal impact study, one may first study how a prediction model changes treatment decisions in a cross-sectional design – i.e. without studying patient outcome.[12,16] Provided that the efficacy of the treatments is known, one can actually quantify the impact of using the model on patient outcomes. Although these alternatives are more prone to bias, a negative result may indicate that a cluster-randomized trial is not yet warranted.

Once a cluster-randomized trial design has been chosen to study the impact of a prediction model, one might consider to do pre-trial observations of the potential users and their decision making behavior.[44,45] Observing the (care-as-usual) behavior of all physicians before the start of the trial has the advantage that one is able to quantify how physician decision making changes within each study group, which may help to interpret the study's results. In our example, physicians of the care-as-usual group also provided risk-dependent PONV prophylaxis to their patients without explicitly using a prediction model (Figure 6.1). Although it may simply represent their clinical expertise, after the study one cannot distinguish this from Hawthorne effects or contamination between study groups.[20,45,46] Such pre-trial observations would also have enabled us to verify whether physicians who are more inclined to treat PONV, were indeed equally distributed between the study groups.

risk-dependent analysis of the results

The aim of prediction models is to encourage risk-tailored management of individual patients. Consequently, the effects on physician decision making and patient outcome are expected to be risk dependent. High risk patients will receive more 'intense' interventions (a higher number of drugs or a higher dose) than low risk patients. This risk-dependent effect is expected to be larger in the intervention than in the care-as-usual group. The statistical analysis should thus study the results for risk-dependent differences between study groups. The predicted probability should be handled as an effect modifier, e.g. including it as an interaction term with 'study group' in a regression model. A prerequisite for such analysis is that the model's predicted probabilities can be estimated for the patients in both study groups. This requires that the patient values of the model predictors should also be observed in the control group. In our example, this was easy since only patient and procedure characteristics comprised the model predictors. However, when a prediction model includes laboratory values, genetic tests or results from invasive diagnostic procedures as predictors, obtaining this information from the patients in the control group is more cumbersome and subject to ethical constraints.

v. interpreting the results of the impact study

As implementing a prediction tool – presenting predicted probabilities in combination with subsequent management actions – is a complex intervention, it may prove difficult to interpret the results of an impact study.[15,39,40,47] When there is a large effect on decision making and patient outcome, we may simply conclude that the prediction tool is effective and should be implemented on a large scale.[48] However, more subtle effects or conflicting effects between decision making and patient outcome – such as in our cluster-randomized study (Figures 6.1 and 6.2) – are more challenging to understand.[39,41] Additional information about

the processes of decision making – or other components of the causal pathway described above – is then essential to understand the intricate downstream effects of prediction models.[15,39,40] For example, physicians reported during the interviews that one needs to know the cause of PONV to treat it correctly (Box 6.2). This made us realize that the prediction model predicts an overall probability of PONV and does not reflect a specific pathophysiologic mechanism that causes PONV. As we discussed in a previous section, we learned that the prediction model may thus have an unintended interaction with the applied interventions.

concluding remarks

Evaluating the clinical impact of using a prediction model in a large-scale comparative study requires a phased approach tailored to the specific decision and the local setting in which the model will be used. The format of the prediction tool – e.g. assistive or directive – should be carefully designed to minimize any unintended effects. Knowledge of the current behavior of the intended users, and their perspectives on risk prediction models in general, will be extremely helpful to determine how to best present the model, to design the study and interpret its results.

Currently, the number of published clinical prediction models is overwhelming.[2,3,49] It is simply impossible to study all these prediction models in prospective impact studies. There are just not enough human beings to do so: 'the world is not enough'.[50] Prospective comparative impact studies, randomized or not, are costly and require much effort, especially when patient long term effects can only be observed after substantial follow-up time.[12,16] We recommend not to envision any impact study before the developed model is at least once validated in individuals not used to develop the model (external validation).[1,3,10–12] When externally validated, subsequent tailoring to the setting in which the model impact is to be studied, as explained above, is highly recommended. We believe it is only cost-effective to perform a prospective comparative impact study when there is a reasonable chance to find a relevant positive effect on decision making and patient outcome. It may prove to be more cost-efficient when one first studies how a prediction model changes decision making by physicians in clinical practice.[12,16] This may be done by either studying the effects of implementing the prediction model on physician's behavior or on other process variables in a non-randomized study, or estimating the potential effects of model prediction and subsequent treatment effects using a decision analytical study. The results of such studies may help to decide whether or not to proceed to a comparative, prospective impact study. It can separate the wheat from the chaff. When the results of such studies and analysis indicate that a models is potentially effective, the results may also be used to plan the subsequent impact study. The results may help to optimize the design of the impact study – e.g., which subgroups will most likely to benefit from the model's use – and to improve the study's analysis – e.g., which additional information needs to be documented.

references

1. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209. PMID:16461965.

2. Bouwmeester W, Zuithoff NP a, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(5):1–12. PMID:22629234.

3. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. PMID:23393430.

4. Rabar S, Lau R, O'Flynn N, Li L, Barry P. Risk assessment of fragility fractures: summary of NICE guidance. *BMJ*. 2012;345:e3698. PMID:22875946.

5. Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(25 Suppl 2):S49-S73. PMID:24222018.

6. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2014;118(1):85–113. PMID:24356162.

7. Freedman AN, Slattery ML, Ballard-Barbash R, et al. Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol*. 2009;27(5):686–693. PMID:19114701.

8. Wishart GC, Azzato EM, Greenberg DC, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res*. 2010;12(1):R1. PMID:20053270.

9. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338(mar31 1):b604. PMID:19336487.

10. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338(may28 1):b605. PMID:19477892.

11. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: A review. *J Clin Epidemiol*. 2008;61(11):1085–1094. PMID:19208371.

12. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–698. PMID:22397946.

13. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: A study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567–2586. PMID:15287085.

14. Janssen KJM, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KGM. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth*. 2009;56(3):194–201. PMID:19247740.

15. Ferrante di Ruffano L, Hyde CJ, Mccaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *Br Med J*. 2012;344(February):1–9. PMID:22354600.

16. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(jun04 2):b606. PMID:19502216.

17. Hendriksen JMT, Geersing GJ, Moons KGM, de Groot J a H. Diagnostic and prognostic prediction models. *J Thromb Haemost*. 2013;11 Suppl 1:129–141. PMID:23809117.

18. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. *BMJ*. 2004;328(7435):343–345. PMID:14764503.

19. Kawamoto K, Houlihan C a, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765. PMID:15767266.

20. Lobach D, Sanders GD, Bright TJ, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. *Evid Rep Technol Assess (Full Rep)*. 2012;(203):1–784. PMID:23126650.

21. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009;119(17):2408–2416. PMID:19364974.

22. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ, Group G. Strengthening the reporting of Genetic Risk Prediction Studies: the GRIPS Statement. *PLoS Med*. 2011;8(3):e1000420. PMID:21423587.

23. Black AD, Car J, Pagliari C, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med*. 2011;8(1):e1000387. PMID:21267058.

24. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. 2011;18(3):327–334. PMID:21422100.

25. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med*. 2012;157(1):29–43. PMID:22751758.

26. Kappen TH, Moons KGM, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. *Anesthesiology*. 2014;120(2):343–354. PMID:24105403.

27. Kappen TH, Vergouwe Y, Wolfswinkel L Van, Kalkman CJ, Moons KGM, Klei WA Van. Impact of adding therapeutic recommendations to risk assessments from a prediction model for postoperative nausea and vomiting. *Br J Anaesth*. 2015;114(2):252–260. PMID:25274048.

28. Van den Bosch JE, Moons KG, Bonsel GJ, Kalkman CJ. Does measurement of preoperative anxiety have added value for predicting postoperative nausea and vomiting? *Anesth Analg*. 2005;100(5):1525–1532. PMID:15845719.

29. Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of clinical prediction models for application in local settings. *Med Decis Making*. 2012;32(3):E1-E10. PMID:22427369.

30. Janssen KJM, Vergouwe Y, Donders ART, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem*. 2009;55(5):994–1001. PMID:19282357.

31. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q*. 1989;13(3):319–340. PMID:4679168.

32. Kortteisto T, Komulainen J, Mäkelä M, Kunnamo I, Kaila M. Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care. *BMC Health Serv Res*. 2012;12:349. PMID:23039113.

33. McGinn CA, Gagnon M-P, Shaw N, et al. Users' perspectives of key factors to implementing electronic health records in Canada: a Delphi study. *BMC Med Inform Decis Mak*. 2012;12:105. PMID:22967231.

34. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S-A. Computerized clinical decision support for prescribing: provision does not guarantee uptake. *J Am Med Inform Assoc*. 2010;17(1):25–33. PMID:20064798.

35. Ash JS, Sittig DF, Guappone KP, et al. Recommended practices for computerized clinical decision support and knowledge management in community settings: a qualitative study. *BMC Med Inform Decis Mak*. 2012;12(1):6. PMID:22333210.

36. Lilford RJ, Chilton PJ, Hemming K, Girling AJ, Taylor C a., Barach P. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. *BMJ*. 2010;341(aug27 1):c4413. PMID:20802000.

37. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350(24):2441–2451. PMID:15190136.

38. Carlisle JB, Stevenson CA. Drugs for preventing postoperative nausea and vomiting. *Cochrane Database Syst Rev*. 2006;3(1):CD004125. PMID:16856030.

39. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:a1655. PMID:18824488.

40. Astbury B, Leeuw FL. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *Am J Eval*. 2010;31(3):363–381. doi:10.1177/1098214010371972.

41. White H. Theory-based impact evaluation: principles and practice. *J Dev Eff*. 2009;1(June):271–284. doi:10.1080/19439340903114628.

42. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 2012;345:e5661-e5661. PMID:22951546.

43. Schaafsma JD, van der Graaf Y, Rinkel GJE, Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol*. 2009;62(12):1248–1252. PMID:19364636.

44. Eccles M, McColl E, Steen N, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *BMJ*. 2002;325:941. PMID:12399345.

45. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care*. 2003;12(1):47–52. PMID:12571345.

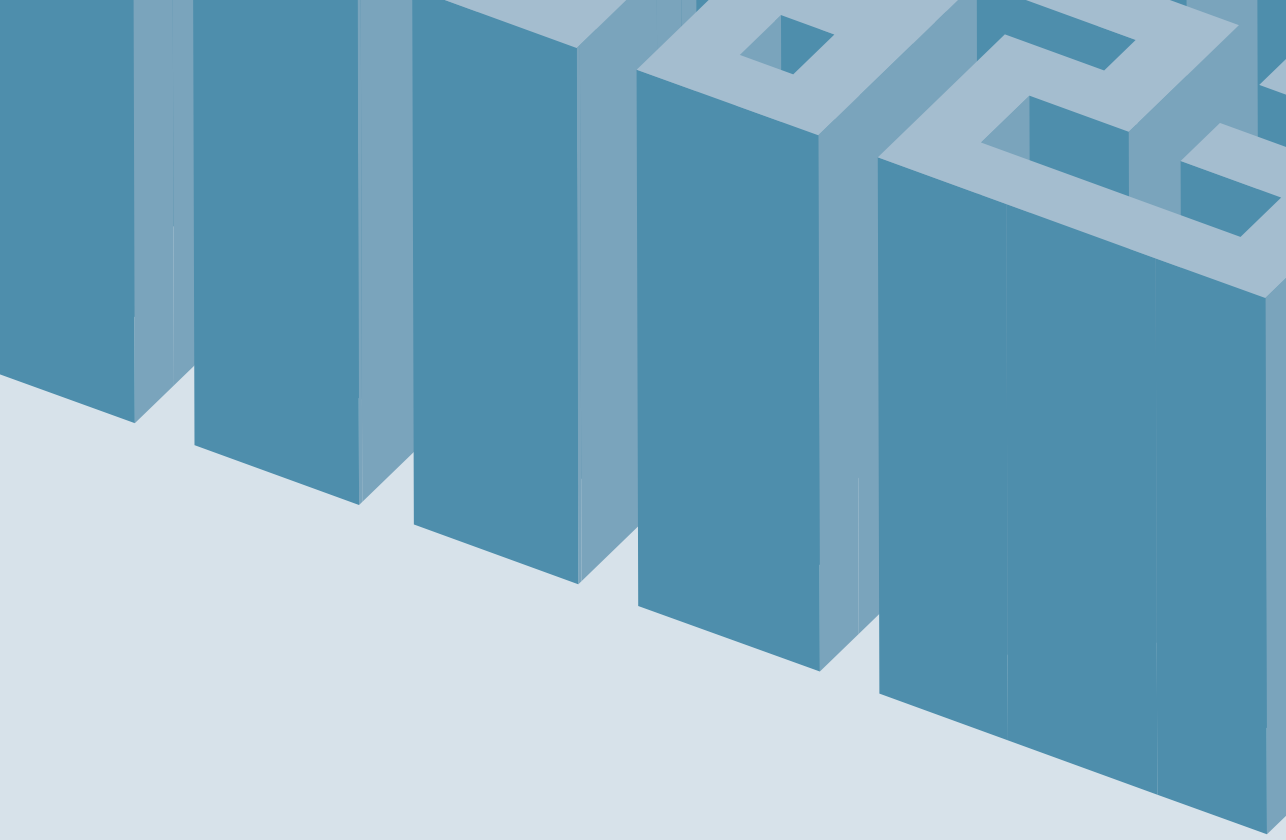
46. Wyatt J, Spiegelhalter D. Field trials of medical decision-aids: potential problems and solutions. *Proc Annu Symp Comput Appl Med Care*. 1991;5479:3–7. PMID:1807610.

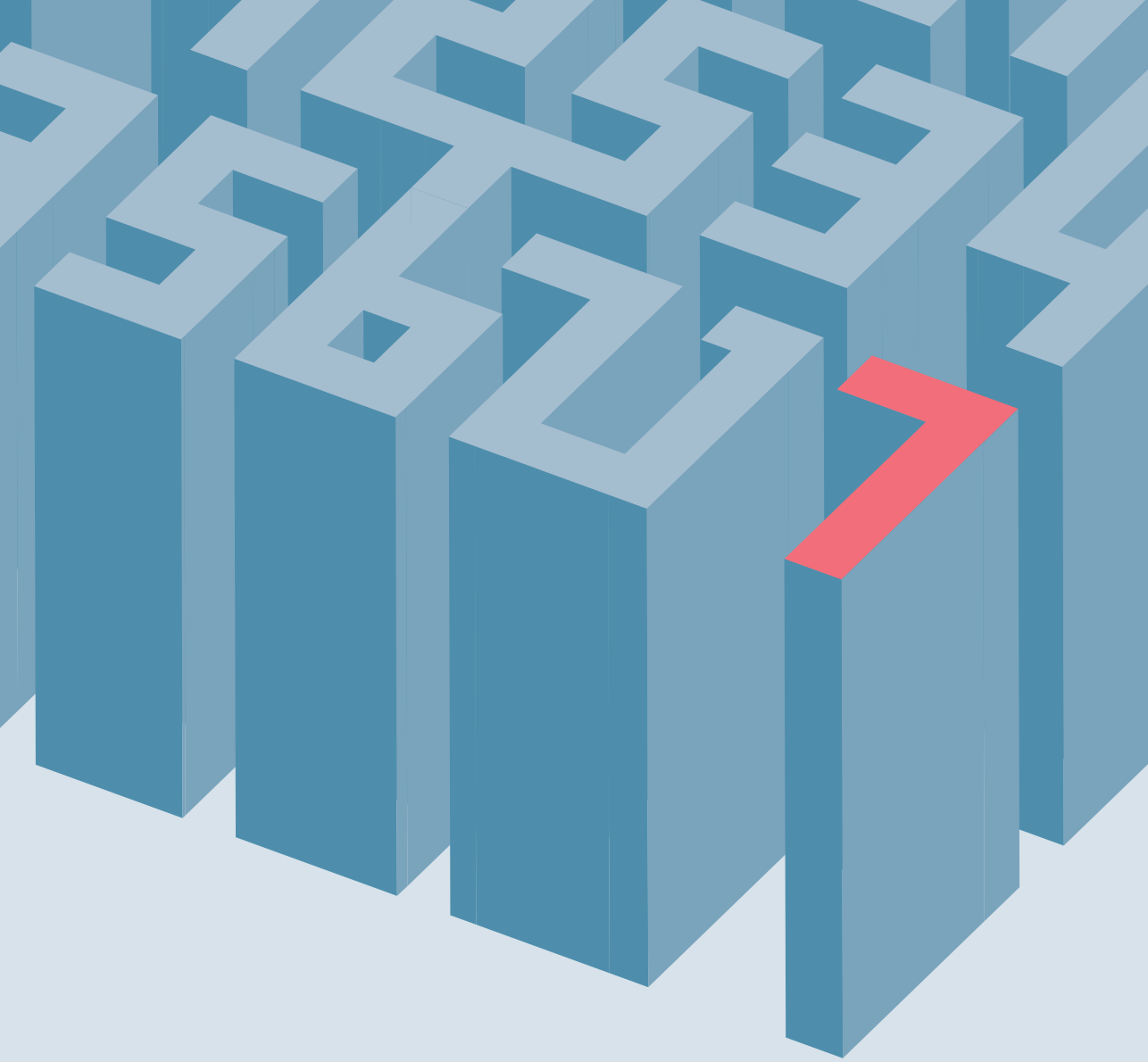
47. Marchal B, Westhorp G, Wong G, et al. Realist RCTs of complex interventions - an oxymoron. *Soc Sci Med*. 2013;94:124–128. PMID:23850482.

48. Karanicolas PJ, Montori VM, Devereaux PJ, Schünemann H, Guyatt GH. A new 'mechanistic-practical" framework for designing and interpreting randomized trials. *J Clin Epidemiol*. 2009;62(5):479–484. PMID:18468856.

49. Collins GS, Altman DG, Moons KG, Reitsma H. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. PMID:25569120.

50. Rudd N, Barr W. Satire X: Wrong Desire is the Source of Suffering. In: *Juvenal, The Satires*. Oxford: Oxford University Press; 1992.





GENERAL DISCUSSION II: DECISION SUPPORT IN THE CONTEXT OF A COMPLEX DECISION SITUATION

T.H. Kappen, M. Noordegraaf, W.A. van Klei, K.G.M. Moons, C.J. Kalkman

Prediction models in health care estimate patient-specific risks of current or future conditions in new patients. By presenting patient-specific risk predictions to physicians, prediction models provide personalized decision support that aims to improve clinical decision making.[1,2] However, the evidence indicating that such decision support tools improve the quality of clinical decision making and patient outcome, remains highly inconsistent.[3–7] Current literature on clinical decision support suggests that adding therapeutic recommendations to the risk predictions will improve the effectiveness of the decision support tool.[6–8] The patient-specific risk prediction is merely a single factor in a highly complex decision making process. Consequently, the addition of a recommendation to the risk prediction may provide the clinical context that is required to make a personalized decision for an individual patient.[6,9]

In a recent series of studies, we added therapeutic recommendations to risk predictions of postoperative nausea and vomiting, which improved both the decision making of physicians and patient outcome.[10–12] Despite the improved effectiveness of the decision support tool, physicians did not comply with the recommendations in 34% of their patients, and many patients still suffered from nausea and vomiting after their surgical procedure. One explanation would be that the physicians took individual patient factors into account such as patient comorbidities or even patient preferences, and deliberately decided not to comply with the recommendations in 34% of their patients. An alternative explanation would be that the addition of recommendations did not provide the physicians with sufficient insight to value the recommendations for individual patients. The physicians could not decide whether the recommendation was either right or wrong, because in their view that depends entirely on the clinical context. Simply presenting more information to physicians is not the same as placing the information from the decision support tool within the clinical context – or 'contextualizing' the decision support tool. In this article we will first argue that most current decision support tools are decontextualized. We then discuss how physicians rely on clinical context to handle the complexity of a clinical situation. Finally, we will argue that because of this complexity, decision support tools that contextualize predictions will be more effective in supporting physicians in decision making.

decision support

'Clinical decision support' generally refers to electronic systems that aim to help physicians make better clinical decisions, which in turn should improve the outcome of their patients.[13] The concept here is similar to that of physician decision making: decision support uses individual patient and disease characteristics to provide physicians with patient-specific risk predictions or risk-based recommendations on treatment or diagnostic testing.[6,14] The predictions and recommendations may help physicians to analyze and manage the complexity of decision situations and reduce the cognitive workload of their decision making.[6,15] Current literature suggests that, for optimal effectiveness, decision support needs to blend in with the process of routine care and the presented predictions and recommendations should be embedded in the clinical context.[5,6,16–18] Contextualized decision support then minimizes the attentional resources that are required from physicians to interpret and use the predictions and recommendations in their decisions.

However, these claims seem to be largely unfounded, as decision support has failed to demonstrate that it improves patient outcome on a large scale.[5,19,20] From the theory of socio-technical systems we know that the interaction between physicians and decision support tools will produce both 'designed' and 'unintended' changes in the existing decision making routines of physicians.[2,21–23] Such unintended consequences can paradoxically increase the cognitive workload of physicians or prevent the decision support from having a beneficial effect on patient outcome.[1,2] This does not mean that we should question the need for contextualized decision support. On the contrary, the unintended changes in decision routines and the inability to improve patient outcome indicate that many decision support tools are still decontextualized and do not necessarily manage complexity as well as they aim to.

decision complexity

A decision becomes complex when the outcome of the decision is difficult to predict, despite the fact that one is reasonably informed about the individual components of the decision and their interrelations.[24] From this rather abstract definition we can establish that the level of complexity of a decision is not simply determined by the number of components and interrelations that are involved in a decision. Rather, the complexity depends on the (un) predictability of the effects of a decision from the components and interrelations that are known at the time of decision making. Consequently, in addition to the information that is available to the physicians, we also need to consider the information that the physicians are missing.

the available information

A basic decision on a medical intervention requires at least three pieces of clinical information: 1) the future risk of harm from the natural course of the disease (i.e. the prognosis); 2) the expected benefit from the intervention (i.e. the effectiveness); and 3) the nature and likelihood of harm from the intervention (side effects or complications). When these three informational components are exactly known for an individual patient the decision is relatively simple. It is a choice between two prognoses: either the risk of harm from the disease – or the reduced risk of the disease after treatment, even after factoring the harms of the intervention. Since it is clear how the decision will affect outcome, the patient can weigh the two options according to his or her personal preferences and come to a decision.

Such a clear-cut decision situation is merely hypothetical. Even for the simplest clinical decision the exact risks and benefits are not known for an individual patient and can only be approximated by probability estimations. There are three major methods for acquiring medical knowledge that a physician can use to estimate the probabilities for an individual patient: clinical research, pathophysiologic understanding, and clinical experience.[25,26] It is important to note that these methods do not necessarily have to estimate the probability in a quantitative way. To estimate the probabilities of risks and benefits for an individual patient, the knowledge that is acquired through each method has to be made patient-specific. This requires specific characteristics of the individual patient and the current disease state – i.e. the clinical context – such as the patient's demographics, lifestyle factors, results of diagnostic testing and comorbidity. The complexity of a clinical decision then becomes apparent. Many

of these variables are interrelated, e.g. diagnostic testing depends on signs and symptoms, signs and symptoms may depend on comorbidity, etc. The greater the number of variables comprised in the clinical context, the greater number of interrelations will be involved. Further, each probability estimate for either a risk or an expected benefit is likely to be based on all three of above mentioned methods of knowledge acquisition, whereas all these methods use many of the same patient and disease variables. This greatly increases the number of possible interrelations in the decision. Even a basic medical decision is thus easily complicated by its clinical context.

the missing information: uncertainty and ambiguity

Notwithstanding the amount of information that is involved in the decision, it would still be possible to estimate the probabilities of risks and benefits for an individual patient when all the variables and interrelations are exactly known. It would be quite an undertaking for a physician – although not insuperable – whereas it would be very easy for a decision support tool. However, in most decision situations the variables and interrelations are not exactly known – there is missing information. Most diseases do not manifest themselves through textbook signs and symptoms, patients do not always report all aspects of their lifestyle, and most diagnostic tests do not have a perfect predictive value. Accordingly, in many clinical decisions complexity truly complicates decision making .

How much the missing information increases the complexity of a decision depends on what part of the information is not available. When one knows that a particular variable or one of the interrelations is involved in a decision, but we do not know its exact value, the information may either be *uncertain* or *ambiguous*. Uncertainty means that we are informed about the value but it lacks precision* – i.e. only the probability distribution of a variable or interrelation is known.[27–29] An imperfect diagnostic test is a good example of uncertainty, as we do know the probability of a diagnosis from the predictive value of the test, but we do not know whether in this instance the test is correct. Although uncertainty definitely adds to the complexity of a decision, with the right information it is still possible to calculate a probability estimate for the risks and benefits for an individual patient. The uncertainty in the variables and interrelations then results in uncertainty of the probability estimates for risks and benefits. A decision support tool is able to calculate the probability estimates with relative ease, through the use of sophisticated statistics and machine learning algorithms. Nonetheless, even under uncertainty, a rational patient would still be able to weigh the risks and benefits of treatment in a quantitative way and come to a decision according to his preferences.

Information is ambiguous when we understand the concept of the variable or the interrelation† (we may even know the set of possible values), but we do not know the probability or the distribution of those values – at least not in a quantitative way.[27–29] Ambiguity exists

* Precision in this instance does not refer to the statistical terminology of 'precision' and 'accuracy', as at the time of decision making one cannot distinguish between a random or systematic error for a single value.

† Note that when a variable or interrelation is unknown, its concept is not known and there is no information, it does not directly contribute to the complexity, as such a variable contains information that is missing but not information that is available. Such an unknown factor is more likely to be a cause of ambiguity or uncertainty.

when a variable has multiple interpretations that coexist and may even oppose each other, or when its information is indistinct or ill-defined.[27,28] Ambiguous variables and interrelations are ample in clinical decision situations, e.g., atypical signs and symptoms, or interobserver differences for a diagnostic test. As a result of the ambiguity in variables and interrelations, the expected risks and benefits can no longer be quantified. Ambiguity thus greatly decreases the predictability of the effects of a decision on patient outcome, which makes it much more difficult for patients to weigh risks and benefits.

Moreover, the preferences of patients – or physicians – are often ambiguous as well. For example, postoperative nausea and vomiting is very unpleasant for patients undergoing surgery and up to three different antiemetic drugs may need to be administered for effective prophylaxis.[30] Although the risk of serious adverse drug events is probably very low (below one in one thousand), there is still a non-zero risk of a fatal cardiac arrhythmia or a serious infection of the surgical site. How does one value the risk of a fatal cardiac arrhythmia due to prophylactic treatment – even when its risk is one in one hundred thousand – against being nauseous and vomiting for 24 hours? Especially, when considering that the antiemetic drugs only reduce the risks of nausea and vomiting and cannot guarantee their prevention. Such ambiguity is more of an 'ethical' nature, in contrast to the more 'technical' ambiguity of the medical variables and interrelations. The ethical and technical ambiguity are also interrelated, as the difficulty for a patient to decide on such an ethical dilemma will depend on his or her ability to understand the risks and benefits that are presented by the physician. The complexity of a clinical decision is thus determined by the number of variables and interrelations that comprise the clinical context, the amount of uncertainty and ambiguity that is associated with them, and the ethical ambiguity of personal preferences. In the next sections, we will separately discuss the way that physicians and decision support tools handle complexity.

physicians and decision complexity

Regardless of their complexity, clinical decisions will have to be made by patients and physicians. A good decision can only be made when all important elements – variables, preferences, interrelations – of the decision are considered, even when these elements are ambiguous. One strategy for physicians to handle the ambiguous elements in their decisions would be to somehow interpret and weigh the value of the elements. The estimated values for the ambiguous elements can then be used in the calculation of the probability estimates of risks and benefits. Nonetheless, the time and effort that would be required to weigh the value of all ambiguous elements and to analyze their interrelations with other decision elements would impose a far too great workload on a physician.[31,32]

Instead of relying on the slow and more analytic type of reasoning – also known as Type 2 reasoning – physicians rely on a faster and more intuitive type of reasoning – known as Type 1 reasoning – to deal with everyday decision complexity.[33–36] Type 1 reasoning uses experience-based strategies for problem-solving – or heuristics – that allow people to judge frequencies or possibilities in a fast and frugal way.[37] A physician will try to recognize the context of a complex decision, rather than try to understand the context and the complexity.[22,38,39] A recognized context then becomes a frame of reference for all relevant information that

the physician has learned to associate with the context, such as typical and atypical signs and symptoms, diagnostic testing, treatment options, etc. As the context is recognized in an intuitive, imprecise yet reliable way, it allows physicians to accept the complexity of a decision as it is – contextualization of the decision situation thus simplifies the decision itself. It then becomes much easier to interpret specific elements of the decision, to reason on risks and benefits and to communicate clinical information to other healthcare workers and patients.[40] The 'fuzzy' recognition of complex patterns within the clinical context enables physicians to handle ambiguous information, which they may use in further analytic and reflective reasoning (Type 2 reasoning).[40,41] Physicians also acquire additional contextual information to understand how clinical information was produced, e.g. knowing who performed a diagnostic test may be informative of its reliability.[42] This production-contextual information may allow physicians to interpret otherwise ambiguous variables and weigh the value of those variables in their decisions. Contextualization of a decision situation thus provides a highly efficient heuristic for physicians to handle the complexity of clinical decisions while maintaining an acceptable workload.[40,43]

decision tools and decision complexity

Decision support tools rely on the use of algorithms to provide risk-based predictions and recommendations. The algorithms may use any of the sources of medical knowledge to estimate the risks and benefits: clinical research, pathophysiologic understanding, and clinical experience. Today's computers are so powerful that decision support tools that must quickly handle lots of data are no longer limited by processing power. It thus seems intuitive that decision support tools are well equipped to model risks and benefits for complex decision situations, as long as there is sufficient data to do the necessary calculations. However, as decision support tools use algorithmic approaches to model the risks and benefits, they require quantifiable variables and interrelations as inputs. They thus do not easily handle ambiguous information. The greater the amount of ambiguity in decision situations, the greater the need for heuristics, which allow physicians to consider all available information, including the ambiguous elements.

decision tools and physician interaction

The predictions and recommendations from decision support tools may still be very useful to physicians, even when there is ambiguity involved. For decision support tools to be useful in ambiguous decision situations, they have to effectively communicate their risk-based predictions in such way that physicians will be able to weigh the predictions in their decision making. When a decision support tool only presents quantitative risk predictions, it communicates decontextualized and analytic information to physicians, who may find it quite difficult to integrate in their more heuristic decision making process.[12]

The communication between a decision support tool and physicians is analogous to a native speaker who tries to communicate with a non-native speaker – as a decision tool can be considered 'native' in using probabilities and probability distributions. Even when the message is correctly phrased by the native speaker, it may require a lot of time and attention

from the non-native speaker to understand the message. The message may even be incorrectly understood by the non-native speaker. Without a proper context the predictions are thus likely to increase the workload of physicians,[1,2] or may have an unintended negative effect on the quality of decisions – and may even result in a worse patient outcome.[21–23] This would make physicians fall back on their clinical experience and thus rely more on heuristics. The difficulty to use the decontextualized predictions will prove greater for physicians when the decision complexity contains increased levels of ambiguity, as that would mean a greater gap between the predictions and the clinical context. Paradoxically, the more support a decision tool tries to provide in an ambiguous decision situation, the greater will be the risk of unintended changes in the physicians' routines, such as loss of situational awareness or degradation of clinical experience.[44] The negative consequences of a wrong decision on patient outcome are greater when a patient's risk of harm is higher. In other words, when the stakes of a decision are higher, the impact of a wrong decision will be greater, which requires a greater need to rely on physician heuristics.

contextual decision support

From the previous sections we may conclude that the need for physicians to rely on heuristics – rather than to trust computerized decision support – will be greater when: there is a lot of ambiguity, the stakes of the decision are high, and the decision support poses too much extra workload. The goals of contextualizing a decision support tool thus are: to make the tool acknowledge and 'factor in' ambiguity; to provide risk-based support to physicians; and to keep (or make) the workload acceptable. In the next sections, we discuss three ways to contextualize the predictions of decision support tools: the algorithms of the model may be changed to include more complexity; decision support tools may be more flexible to changes in clinical context; the communicative capabilities of the tool may be improved.

the complexity model

One way to contextualize a decision support tool is to improve the algorithms that the tool uses to calculate the predictions for the decision situation. The goal of the improvement would be to model all complexity. However, the challenge for the future will be to understand how we should deal with ambiguity in the algorithms of decision support tools. It is beyond the scope of this paper to discuss all available models and algorithms, or to speculate whether future artificial intelligence techniques will be able to handle ambiguity in a satisfactory way. Nevertheless, when ambiguity can be truly modeled by algorithms, the information is no longer ambiguous. A more promising approach is thus disambiguation of the decision, which will probably require specific research on the topic of the decision.

Alternatively, improving the data quality on which the algorithms are based would be another strategy, as it will improve the predictions of risks and benefits, despite the residual ambiguity. When we consider the example of postoperative nausea and vomiting, which was mentioned in the introduction, we would choose disambiguation of the decision. The risk predictions of which patients will become nauseous after surgery as well as the effectiveness of prophylactic treatment have been extensively quantified, albeit with a fair amount

of uncertainty.[45–48] From the results of our recent impact studies, we started questioning whether there is an unquantified interrelation between the predicted risk and the effectiveness of treatment.[9] Further study may help to solve the ambiguous relation between predictions and treatment, or how to predict which patients will respond to treatment. Nevertheless, when a larger part of complexity is modeled by the algorithm, there is also an increased risk of unintended consequences, such as unwanted changes in the decision making routines of physicians.

flexible heuristic space

A flexible heuristic space refers to the ratio between the use of heuristics and the use of predictive information in the decision making process. This concept is similar to that of the level of automation that is used in ergonomics.[44,49] The level of automation thus also depends on the stakes, the ambiguity, and the workload of the decision, as discussed in previous sections.

When the level of ambiguity and the stakes are both low or when they are both high, the choice will be the easiest. For an unambiguous decision for which the stakes are low, human involvement may be restricted to the minimum and the process may be fully automated if at all possible. High levels of ambiguity with very high stakes will require a fully autonomous physician, with the decision support tool functioning merely as an assisting instrument for information gathering and clinical reasoning. And perhaps – if properly designed and implemented – it can also reduce the physician's workload.

An unambiguous decision with high stakes poses a bigger challenge. As there is little ambiguity, it would make sense to let the decision support tool make the decisions to a great extent, with the physicians monitoring those decisions as the margin for error is small. Nonetheless, this situation is also prone to unintended effects of the decision support tool, as such a high level of automation may cause the physicians to lose situational awareness or degrade their clinical experience.[44] There are numerous examples from aviation where automation-related incidents could be traced back to loss of situational awareness by the crew.[50–52] It is thus not at all said that physicians will be able to keep the margin for error small when they are tasked with monitoring an almost fully automated process. As physicians will still be responsible for the final decision, it may prove more effective to lessen the level of automation. In fact, because the stakes are so high, the physicians may not even indulge such a high level of automation, especially when the decision support tool is perceived as counterproductive.[19,53,54]

A similar problem may occur when the decision is highly ambiguous, but when the stakes are low. The amount of ambiguity implies a greater need for heuristics. However, physicians are limited by their workload capacity.[36] Should this decision be the only problem that the physician faces at that moment, it would be highly efficient to simply let the physician make the decision, perhaps assisted by a computer. However, in a highly demanding environment, such as during surgery or treating a patient at the Intensive Care Unit, the physicians will not prioritize their attention to this relatively unimportant decision. In such an instance, it may be more efficient to let the decision support tool make a standardized recommendation, which directs the decision of the physician despite the presence of ambiguity.

These are four distinct situations. In clinical practice this will be more of a continuum. It is important to note that the level of automation does not automatically follow from the level of ambiguity and the stakes of the decision. The level of automation has to be chosen when designing the decision support tool. The ambiguity and the stakes of the decision may indicate what the risks of automating a decision will be, not whether the risks are acceptable. Furthermore, the workload of the physicians in the decision situation at hand – from both the decision itself and other tasks and decisions – will determine how much time and attention can be additionally spent on using the decision support tool. In our example of postoperative nausea and vomiting, physicians had to make ambiguous decisions on preventive treatment in a high-workload environment for a problem that was not considered to be of utmost importance.[12] A higher level of automation – providing treatment recommendations that only required confirmation – was more effective than a lower level of automation, i.e. providing only risk predictions.[10,11] However, the effects on decision making and patient outcome were still modest when the decision support tool provided treatment recommendations. Going to the next higher level of automation – perhaps even fully automated decision support – may further increase the effectiveness of the tool, but may also increase the risk of unintended negative effects. The decision to further automate a decision support tool for postoperative nausea and vomiting may thus depend on the willingness of patients to accept the increased risks for a possible reduction in discomfort from nausea and vomiting after surgery.

Even within similar decision situations the level of ambiguity, the stakes and the workload may fluctuate. Ideally, the level of automation should be flexible, responding to these fluctuations. A future perspective might be that decision support tools do not only provide risk-dependent predictions or recommendations, but its level of automation also depends on the predicted risks and the workload of physicians. In our example of postoperative nausea and vomiting, the decision support tool might be able to estimate the workload of the physician at any time point during the anesthetic case. A flexible decision support tool would then provide its predictions and recommendations when there is a low workload during the case. Alternatively, a flexible tool would adjust its level of automation when the workload becomes too high.

communicative capabilities

Many factors are reported in current literature that improve the capabilities of a decision support tool to communicate context to physicians together with its calculated predictions. The hypothesis is that this will free up the mind of physicians to pay more attention to other complex activities that are not covered by the decision support tool.[55,56]

For example, the addition of actionable recommendations to risk predictions already aims to provide context on what a risk prediction may imply for treatment.[6,8,10,11] Alternatively, the reasoning behind the predictions may be provided, including details about the variables on which the model is based, and what predictors dominate in a particular patient.[6,16,19] This would facilitate physicians to better interpret the predictions and weigh them against their own clinical experience. Although there are several proven facilitators that improve the decision making of physicians, those interventions hardly incorporate ambiguity as a

factor. As we argued in the section on Decision support, most decision support tools are still decontextualized. Ambiguity in clinical contexts causes information to be imprecise, thus true contextual decision support should not aim to provide precise information.[23,57] Decision support tools would be truly supportive when they communicate to physicians in the language of the physicians, instead of their own technical language of probabilities. Decision support should merely provide a description, to convey meaning instead of facts.

When it is possible for a decision support tool to communicate ambiguous information to the physicians, the physicians would be able to enter a response to the decision support tool. As the decision support tool knows the quantitative meaning behind the otherwise ambiguous description that it provided to the physician, the tool may use the response for further calculations in self-learning algorithms. For example, the decision support tool may communicate that a patient has a high risk of postoperative nausea and vomiting and has no direct contraindications for preventive treatment. The physician may respond that he does not want to administer one of the possible drugs because the patient has several comorbidities. Although the decision support tool does not have to directly understand what the physician exactly means, but the tool may learn in time to recognize patients which are considered to be too fragile by physicians to accept the risks of preventive treatment. When the decision support tool collected sufficient data on such patients it may either confirm or refute the risk and communicate that to the physicians. In this way, the algorithms of decision support may be able to include clinical intuition of physicians into their calculations.

patient preferences

Similar communicative capacities of decision support tools may facilitate communication with patients, either directly with patients or through the physicians. Patients should be able and allowed to decide whether and how their diseases will be treated. It is a physician's job to provide the patients with the right decision situation and decision options that are appropriate for the clinical context. The more complex the clinical context and the decision situation, the more difficult it will be to provide the right information and context to patients. Without context the predictions and recommendations of a decision support tool may not mean much to patients. Providing them with contextual information that is communicated in natural languages may improve the patients' understanding of the decision situation. Such communication may be bidirectional. For example, rather than asking a patient to weigh a risk of one in one hundred thousand of a serious adverse event, to a risk of one in three of postoperative nausea and vomiting, it may be more useful to explore the preferences of patients and provide the context that they require. This may help patients to straighten out their thoughts on the decision, but also help physician to understand their patients. A decision support tool may use a patient's input to select appropriate information about the clinical context and the decision situation, but the tool may also weigh patient preferences in its recommendations. Moreover, flexibility of a decision support tool may also be determined by patient preferences. After all, it is very common in (non-automated) clinical practice that patients express their preferences and priorities to their physicians. In our example, a flexible decision support tool for postoperative nausea and vomiting would draw more attention

from a physician when the patient indicated that preventing nausea and vomiting was highly important to him. In other words, a flexible decision support tool would allow physicians to prioritize their time and attention according to patient preferences.

conclusion

Clinical decision support aims to reduce the complexity of clinical decisions, yet if poorly implemented it may actually complicate care and confuse physicians. In this paper we argued that information from decision support tools is often removed from its clinical context. In contrast, physicians largely depend on clinical context to handle the complexity of their day-to-day decisions. Simply offering analytic information to physicians – without its context – may be similar to trying to get two people who speak different languages have an elaborate conversation.

'Believers' of artificial intelligence argue that, since computers are much more capable to handle large amounts of information than humans, future elaborate decision support tools can replace physicians.[58–60] Simply typing 'Will computers replace doctors' into Google yields millions of hits discussing the subject. However, information is not the same as knowledge.[61] Apart from the non-technical skills patients expect from their doctor, future *iDoctors* and *Doctor Algorithms* would need to be able to handle clinical context and decision complexity to a similar – or even better – degree as physicians. Obviously, computers are more capable to handle large number of variables and estimate levels of uncertainty, but physicians have the unique skill to use – albeit imperfectly – the entire clinical context in their decisions, taking into account all ambiguous information and patient preferences. Instead of arguing about who is ultimately better at clinical decision making, it may prove to be a more profitable strategy to use the best of both worlds and let doctors arrive at optimal patient-centered therapeutic decisions using tools on their computers for probability estimation and other applications requiring hardcore number crunching. To improve the quality of decision support tools, designers should focus not only on predictive accuracy, but also on optimizing the way in which such probabilities and therapeutic recommendations are communicated to doctors.

references

1. Varonen H, Kortteisto T, Kaila M. What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. *Fam Pract*. 2008;25(3):162–167. PMID:18504253.

2. Ahmad F, Norman C, O'Campo P. What is needed to implement a computer-assisted health risk assessment tool? An exploratory concept mapping study. *BMC Med Inform Decis Mak*. 2012;12(1):149. PMID:23253913.

3. Black AD, Car J, Pagliari C, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med*. 2011;8(1):e1000387. PMID:21267058.

4. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med*. 2012;157(1):29–43. PMID:22751758.

5. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. 2011;18(3):327–334. PMID:21422100.

6. Lobach D, Sanders GD, Bright TJ, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. *Evid Rep Technol Assess (Full Rep)*. 2012;(203):1–784. PMID:23126650.

7. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. PMID:23393430.

8. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209. PMID:16461965.

9. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the Impact of the Use of Prediction Models in Clinical Practice: Challenges and Recommendations. *<this thesis chapter 6>*.

10. Kappen TH, Moons KGM, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. *Anesthesiology*. 2014;120(2):343–354. PMID:24105403.

11. Kappen TH, Vergouwe Y, Wolfswinkel L Van, Kalkman CJ, Moons KGM, Klei WA Van. Impact of adding therapeutic recommendations to risk assessments from a prediction model for postoperative nausea and vomiting. *Br J Anaesth*. 2015;114(2):252–260. PMID:25274048.

12. Kappen TH, van Loon K, Kappen MAM, et al. How Clinical Prediction Models Impact Decision Making by Physicians: an Example Study. *<this thesis, submitted 2015>*.

13. Osheroff JA, Teich J, Levick D, et al. *Improving Outcomes with Clinical Decision Support:An Implementer's Guide*. 2nd ed. Chicago, IL: Healthcare Information and Management Systems Society (HIMSS); 2012.

14. Wyatt J, Spiegelhalter D. Field trials of medical decision-aids: potential problems and solutions. *Proc Annu Symp Comput Appl Med Care*. 1991;5479:3–7. PMID:1807610.

15. Perreault L, Metzger J. A pragmatic framework for understanding clinical decision support. *J Healthc Inf Manag*. 1999;13(2):5–21.

16. Downing GJ, Boyle SN, Brinner KM, Osheroff J a. Information management to enable personalized medicine: stakeholder roles in building clinical decision support. *BMC Med Inform Decis Mak*. 2009;9:44. PMID:19814826.

17. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *J Biomed Inform*. 2012;45(6):1202–1216. PMID:22995208.

18. McCormack JL, Ash JS. Clinician perspectives on the quality of patient data used for clinical decision support: a qualitative study. *AMIA Annu Symp Proc*. 2012;2012(Table 1):1302–1309. PMID:23304409.

19. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S-A. Computerized clinical decision support for prescribing: provision does not guarantee uptake. *J Am Med Inform Assoc*. 2010;17(1):25–33. PMID:20064798.

20. McGinn CA, Grenier S, Duplantie J, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. *BMC Med*. 2011;9(1):46. PMID:21524315.

21. Pentland BT, Feldman MS. Organizational routines as a unit of analysis. *Ind Corp Chang*. 2005;14(5):793–815. doi:10.1093/icc/dth070.

22. Patel VL, Arocha JF, Zhang J. Medical Reasoning and Thinking. In: Holyoak KJ, Morrison RG, eds. *The Oxford Handbook of Thinking and Reasoning*. Oxford: Oxford University Press; 2012:736–754.

23. Maguire M. Socio-technical systems and interaction design - 21st century relevance. *Appl Ergon*. 2014;45(2):162–170. PMID:23856116.

24. Edmonds B. Syntactic measures of complexity [PhD Thesis].http://www.researchgate.net/publication/2330748_Syntactic_Measures_of_Complexity/file/d912f50c4f400b4977.pdf. Published 1999. Accessed January 3, 2015.

25. Tonelli MR. Integrating evidence into clinical practice: an alternative to evidence-based approaches. *J Eval Clin Pract*. 2006;12(3):248–256. PMID:16722902.

26. Tonelli MR. The challenge of evidence in clinical medicine. *J Eval Clin Pract*. 2010;16(2):384–389. PMID:20367870.

27. March JG. *Primer on Decision Making: How Decisions Happen*. Simon and Schuster; 1994.

28. Noordegraaf M, Abma T. Management by Measurement? Public Management Practices Amidst Ambiguity. *Public Adm*. 2003;81(4):853–871. doi:10.1111/j.0033-3298.2003.00374.x.

29. Barham BL, Chavas JP, Fitz D, Salas VR, Schechter L. The roles of risk and ambiguity in technology adoption. *J Econ Behav Organ*. 2014;97:204–218. doi:10.1016/j.jebo.2013.06.014.

30. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg*. 2014;118(1):85–113. PMID:24356162.

31. Gonzalez C. Task workload and cognitive abilities in dynamic decision making. *Hum Factors*. 2005;47(1):92–101. PMID:15960089.

32. Teo G, Szalma JL. The Effects of Task Type and Source Complexity on Vigilance Performance, Workload, and Stress. *Proc Hum Factors Ergon Soc Annu Meet*. 2011;55:1180–1184. doi:10.1177/1071181311551246.

33. Stanovich KE, West RF. Individual differences in reasoning: implications for the rationality debate? *Behav Brain Sci*. 2000;23(5):645–665; discussion 665–726. PMID:11301544.

34. Kahneman D, Frederick S. Representativeness revisited: Attribute substitution in intuitive judgment. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, UK: Cambridge Univ. Press; 2002:49–81.

35. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022–1028. PMID:19638766.

36. Evans JSBT. Questions and challenges for the new psychology of reasoning. *Think Reason*. 2012;18(1):5–31. doi:10.1080/13546783.2011.637674.

37. Holyoak KJ, Morrison RG. Thinking and Reasoning: A Reader's Guide. In: Holyoak KJ, Morrison RG, eds. *The Oxford Handbook of Thinking and Reasoning*. Oxford: Oxford University Press; 2012:1–8. PMID:20328338.

38. Pretz JE. Intuition versus analysis: Strategy and experience in complex everyday problem solving. *Mem Cognit*. 2008;36(3):554–566. doi:10.3758/MC.36.3.554.

39. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ*. 2010;44(1):94–100. PMID:20078760.

40. Edmonds B. Complexity and Context-Dependency. *Found Sci*. 2012;18(4):745–755. doi:10.1007/s10699-012-9303-x.

41. Thornton T. Clinical judgement, expertise and skilled coping. *J Eval Clin Pract*. 2010;16(2):284–291. PMID:20367849.

42. Galster G. Modeling and Using Context. In: Brézillon P, Blackburn P, Dapoigny R, eds. *Modeling and Using Context*. Vol 8175. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013:248–258. doi:10.1007/978-3-642-40972-1.

43. Edmonds B, Norling E. Integrating learning and inference in multi-agent systems using cognitive context. In: Antunes L, Takadama K, eds. *Multi-Agent-Based Simulation VII*. Springer Berlin Heidelberg; 2007:142–155.

44. Onnasch L, Wickens CD, Li H, Manzey D. Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Hum Factors J Hum Factors Ergon Soc*. 2013;56(3):476–488. doi:10.1177/0018720813501549.

45. 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(3):693–700. PMID:10485781.

46. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350(24):2441–2451. PMID:15190136.

47. Carlisle JB, Stevenson CA. Drugs for preventing postoperative nausea and vomiting. *Cochrane Database Syst Rev*. 2006;3(1):CD004125. PMID:16856030.

48. Kappen TH, Vergouwe Y, van Klei W a, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of clinical prediction models for application in local settings. *Med Decis Making*. 2012;32(3):E1-E10. PMID:22427369.

49. Endsley MR, Kaber DB. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*. 1999;42(3):462–492. PMID:10048306.

50. Sethumadhavan A. Effects of automation types on air traffic controller situation awareness and performance. *Proc Hum Factors Ergon Soc Annu Meet*. 2009;53(1):1–5. doi:10.1177/154193120905300102.

51. Dao a. Q V, Brandt SL, Battiste V, Vu KPL, Strybel T, Johnson WW. The impact of automation assisted aircraft separation on situation awareness. *Lect Notes Comput Sci*. 2009;5618 LNCS:738–747. doi:10.1007/978-3-642-02559-4_80.

52. Chow S, Yortsos S, Meshkati N. Asiana Airlines Flight 214 Investigating Cockpit Automation and Culture Issues in Aviation Safety. 2014;4(2):113–121. doi:10.1027/2192-0923/a000066.

53. Feldman MS, Pentland BT. Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Adm Sci Q*. 2003;48(1):94–118. PMID:10818804.

54. Kortteisto T, Komulainen J, Mäkelä M, Kunnamo I, Kaila M. Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care. *BMC Health Serv Res*. 2012;12:349. PMID:23039113.

55. Horsky J, Kaufman DR, Oppenheim MI, Patel VL. A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. *J Biomed Inform*. 2003;36(1–2):4–22. doi:10.1016/S1532-0464(03)00062-5.

56. Patel VL, Zhang J, Yoskowitz N a, Green R, Sayan OR. Translational cognition for decision support in critical care environments: a review. *J Biomed Inform*. 2008;41(3):413–431. PMID:18343731.

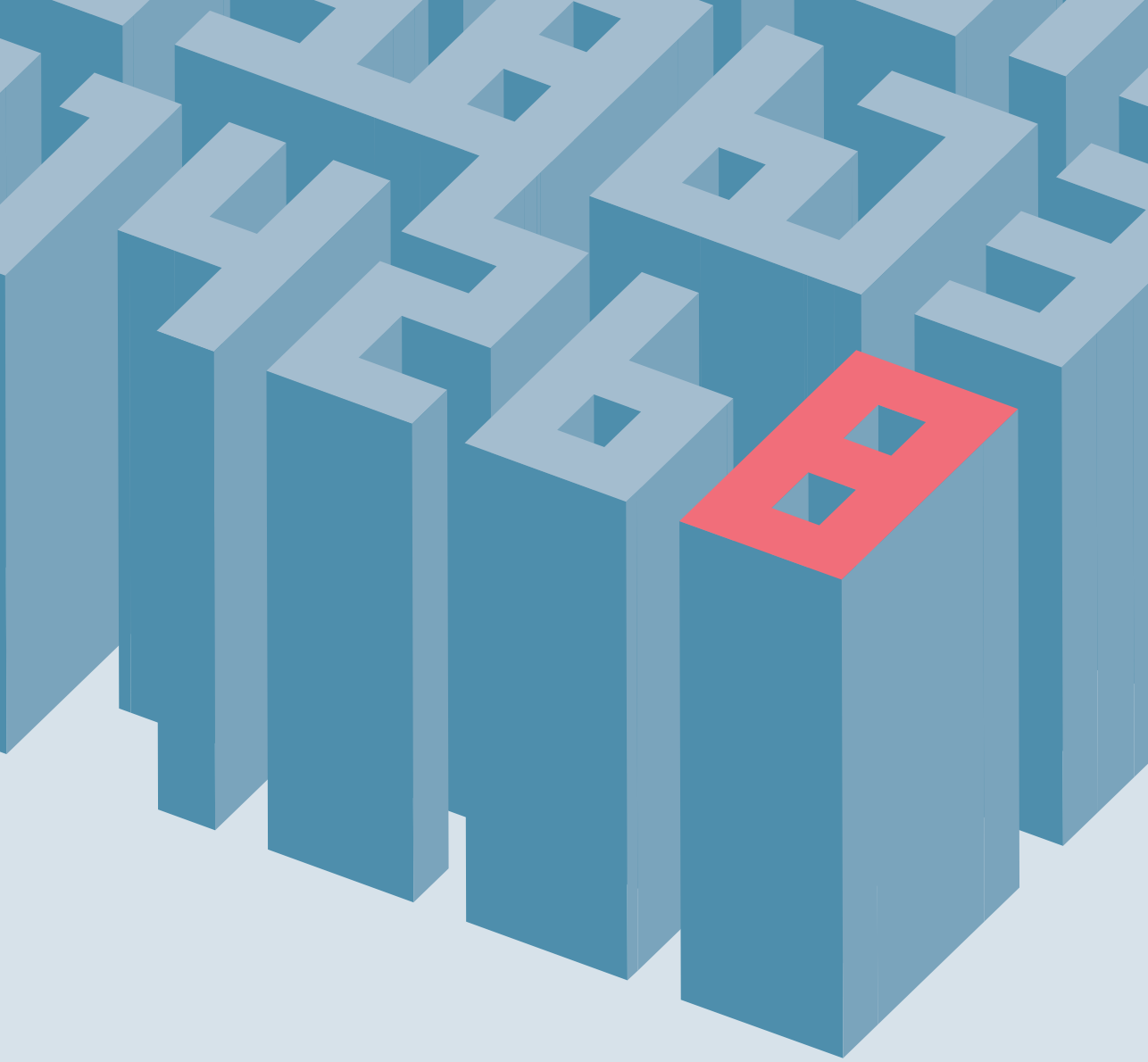
57. March JG. Ambiguity and accounting: The elusive link between information and decision making. *Accounting, Organ Soc*. 1987;12(2):153–168. PMID:5402797.

58. Khosla V. Technology will replace 80% of what doctors do. *FORTUNE*. http://fortune.com/2012/12/04/technology-will-replace-80-of-what-doctors-do/. Published December 4, 2012. Accessed January 27, 2015.

59. Wadhwa V. Why I'm excited about the promising future of medicine. *The Washington Post*. http://www.washingtonpost.com/blogs/innovations/wp/2014/04/07/why-im-excited-about-the-promising-future-of-medicine/. Published April 7, 2014. Accessed January 27, 2015.

60. Gobry P-E. How computers will replace your doctor. *The Week*. http://theweek.com/articles/441528/computers-replace-doctor. Published 2014. Accessed January 27, 2015.

61. Hilbert M. Big data for development: From information-to knowledge societies. *Available SSRN 2205145*. 2013:1–39. doi:10.2139/ssrn.2205145.



APPENDIX: ADDITIONAL METHODS AND RESULTS OF CHAPTER 4

appendix section 1

full description and analysis of the interviews

methods

study setting

Both the present study and the original cluster-randomized trial were performed at the anesthesiology department of the University Medical Center Utrecht, the Netherlands. The University Medical Center is a teaching hospital, and functions as a tertiary center, performing surgical procedures of high complexity. Management of the anesthesiology department was part of a larger division of perioperative care and emergency medicine, with both medical and non-medical executives. For every patient the 'anesthesia team' consisted of an attending anesthesiologist, a highly specialized anesthesia nurse (possibly with an additional trainee anesthesia nurse), and frequently a resident in anesthesiology. The interviews were taken after the cluster-randomized trial had ended, in the first half of January 2008, while the survey started in the second half of January 2008.

The authors of this article comprise of physicians (THK, KVL, LVW, WAVK, CJK), epidemiologists (YV, KGMM) and a layperson in medicine (MAMK). Most of the physicians followed extensive training in epidemiology (THK, KVL, WAVK, CJK) or are experienced in doing prediction research (THK, LVW, WAVK, CJK). Some of the authors, including the first author, have a predilection for information technology (THK, LVW, CJK). Two of the authors are related as son and father (THK and MAMK).

participants

All attending anesthesiologists and senior residents of the anesthesiology department of the University Medical Center Utrecht were randomized to receive individual PONV risks of patients as calculated by a prediction model, and the risks were presented in the anesthesia information management system during the case. Because of progressively more distant supervision from attending anesthesiologists during the 5-year residency, senior residents (4th or 5th year of training) were also considered for randomization in the trial, as they typically make decisions to administer antiemetic drugs independently. Anesthesiologists and senior residents (henceforth referred to as 'physicians') were excluded from the impact trial when they exclusively worked in one of the anesthetic subspecialties (e.g. pediatric or cardiac anesthesia, or pain medicine), or when they had treated less than 50 patients within the impact trial. At the start of the

study, or when a new physician enrolled during the study period, physicians were asked at what predicted risk they would administer PONV prophylaxis to a patient.

sampling

Physicians were selected for the interviews by stratified random sampling from the pool of enrolled physicians. Stratification allowed selection of physicians in different career stages (resident or junior attending versus senior attending) and different randomization status within the cluster-randomized trial (intervention versus care-as-usual). Two physicians per stratum was expected to provide sufficient variation in the conceptions and barriers of physicians to administer prophylactic antiemetics to their patients during general anesthesia. In total eight physicians were interviewed.

interview procedures

A semi-structured interview schedule was devised using both open and closed questions. Questions were formulated according to four prespecified topics, which were hypothesized to reflect the process of decision making: 1) perceived patient burden of PONV by the physician; 2) professional risk stratification: how does the physician identify patients at risk for PONV; 3) the decision-making process regarding preventive antiemetic strategies; 4) attitude towards the use of prediction models and decision support for prophylactic treatment of PONV.

The questions were phrased and ordered according to their expected dependency on other questions. For example, the answer to a question on the benefits and harms of a prophylactic antiemetic drug might be influenced by a preceding question on the perceived burden of PONV. The specific phrasing and order were only meant as a guide; interviewer and interviewees were encouraged to explore the different themes as came natural to them during the conversation.

All interviews were held in a quiet consulting room at the operating room facility. The interviewer (MAMK) did not have a medical background, but was experienced in conducting structured and semi-structured interviews. The interviewer was briefed on the concepts of PONV and on the backgrounds of prediction models and decision support by the first author (THK). The interviewer did not know the interviewees before the interview. The interviews were performed face-to-face and were audio recorded with permission of the interviewees. A pilot interview was performed with one, non-randomly selected interviewee. The pilot interview was not included in the analysis. The duration of the interviews ranged from approximately 20 to 60 minutes. Data saturation was achieved after eight interviews and additional interviews were not expected to provide new insights.

interview analysis

After verbatim transcription of the interviews, KVL first sorted each line of the interview transcript according to the questions of the interview model. The individual responses to each question were then coded and rearranged into themes. Axial coding was used to group the themes into broader thematic categories, using Microsoft Excel. To prevent selective perception

and biased interpretation, THK then repeated the coding process. The themes and categories were subsequently integrated into central themes. Differences in results were first discussed between KVL and THK. Remaining discrepancies were reviewed by the remaining authors and discussed until all authors agreed on the final coding.

results

The eight physicians who were interviewed consisted of four females and four males, equally distributed over the intervention group and the care-as-usual group. Three of the four females were senior attendings, whereas one of the four males was a senior attending. In total 233 minutes of speech were recorded, resulting in 3,569 transcription lines with a maximum line width of 100 characters.

The coded transcriptions were rearranged into 17 themes, distributed over four thematic categories (see also Table 4.1). Although the interview model was not used during the coding process of the interviews, the categories resulting from the axial coding process were very similar to the four overall topics of the interview model. Therefore, we named the categories after the four topics of the interview model. The themes and categories are described in this appendix, whereas the integrated central themes are discussed in Chapter 4.

category 1: perceived burden of PONV

1A. knowledge of the frequency of PONV

The interviewees from the trial's intervention group were able to give a numeric estimate of which percentage of their patients suffered from PONV, although some needed some encouragement from the interviewer. Most of the interviewees from the care-as-usual group were unable to provide a numeric estimate, not even after some encouragement by the interviewer.

> *I do not know the percentages...A number of patients, but not that many...Actually, I do not see it that often. No, I do not see many patients who get nauseous...The percentage is not that high, so might I say 1 in 10? Well, no, I don't know.* (male senior attending, care-as-usual group)

1B. perception of the frequency of PONV among their patients

After the request for a numeric estimate of the frequency of PONV, the interviewees were asked how they perceived the frequency of PONV among their patients. A similar distinction between the study groups was observed. Interviewees from the intervention group all felt that the percentage of PONV among their patients was considerable.

It occurs in more than one out of three patients. And yes, I do believe that is a lot. (female resident, intervention group)

In contrast, the interviewees from the care-as-usual group – except for one – felt that their patients did not frequently suffer from PONV. Although they were uncertain about the exact frequency, the interviewees considered the percentage of their patients with PONV to be low because of their personal case mix, i.e. the type of surgery for which they most commonly provided general anesthesia.

You see, that is my experience of late, as I now mostly provide anesthesia for neurosurgical procedures. They do become nauseous, but because of the type of procedure, in fact, I don't see it that often. (male senior attending, care-as-usual group)

Although interviewees from the care-as-usual group perceived PONV to be fairly uncommon among their patients, they indicated in following questions that they were aware that PONV is a frequent event after general anesthesia. For example, the same interviewee from the care-as-usual group responded:

The risk of PONV, because of the things we do, is quite large. (male senior attending, care-as-usual group)

1C. feedback on PONV occurrence increases awareness

Interviewees from the intervention group received individualized feedback by email after the first twelve months of study, which had been planned prior to the start of the study. This feedback included the incidence of PONV among the patients they had treated, the overall (hospital wide) incidence of PONV and the quantity of antiemetics administered. All interviewees from the intervention group reported that a higher percentage of their patients suffered from PONV than they had expected. Three interviewees perceived the high occurrence of PONV among their patients as confronting.

Well, I received those numbers from the researchers, and they were shockingly higher than I expected. (female senior attending, intervention group)

Two interviewees expressed frustration as they felt that their efforts to reduce PONV were not reflected by the reported incidences from the individualized feedback.

...I am more aware of PONV, I administered more prophylaxis and it did not change a thing. So I thought to myself, this is not getting us anywhere. (female senior attending, intervention group)

All interviewees said that the individualized feedback made them more aware of the problem of PONV and made them increase their administration of antiemetic prophylaxis.

> *I try to pay attention that I administer – to my opinion – adequate prophylaxis and adjust my anesthetic technique to lower the occurrence of PONV.* (female resident, intervention group)

One interviewee from the intervention group reported she was only able to provide a numeric estimate of the occurrence of PONV among her patients, because of this individualized feedback.

> *About 40 to 50% of my patients suffer from PONV. I know that from the feedback reports, otherwise I would not have known.* (female senior attending, intervention group)

In contrast, two interviewees from the care-as-usual group indicated that they receive little feedback in daily practice on which patients suffer from PONV.

> *Of course I see my patients at the postanesthesia care unit, but I rarely hear back from the nursing ward when my patients get nauseous.* (female senior attending, care-as-usual group)

1D. PONV is a side effect of general anesthesia

All interviewees considered PONV a side effect of general anesthesia and not a complication, because it is an expected and common event after general anesthesia.

> *As an anesthesiologist you know that, because of the things you do, you structurally increase the patient's risk of PONV.* (male senior attending, care-as-usual group)

> *It is inherent to our procedures, I mean, when you are on the bus for a holiday, and you get nauseous because it is a bumpy ride, do you consider that a complication or is it just a part of the bus trip because you are sensitive to motion sickness?* (female senior attending, care-as-usual group)

Some interviewees also considered PONV a side effect because its influence on a patient's well-being is generally not severe enough for it to be a complication.

> *It is different when the vomiting is that severe that it damages the esophagus, or something like that. Then it could turn into a complication.* (female senior attending, care-as-usual group)

Although the interviewees generally considered PONV a side effect, several interviewees indicated that it should not be taken for granted by anesthesiologists. Moreover, PONV may become a complication if it is not treated appropriately.

> *PONV is a side effect of anesthesia and surgery, but if it is too easily overlooked and its prevention is insufficiently being addressed, then one might call it a medical error and consider it a complication.* (male junior attending, intervention group)

When asked whether the problem of PONV is a subject of regular discussion among the medical staff of the anesthesia department, the interviewees reported that discussion of PONV is fairly limited.

> *Because of the study it [author's note: awareness of PONV] has increased a little bit. However, it is still not a problem which is regularly brought up during case discussions... Perhaps it is just something that slips your mind quite easily.* (female resident, intervention group)

category 2: professional risk stratification: how does a physician identify patients at risk for PONV?

2A. variation in the use of risk factors

The interviewees reported that during the anesthetic case they make some form of an assessment of which patients are at risk for PONV and should receive PONV prophylaxis, regardless of their exposure to the results of the prediction model. The interviewees based their assessment on a variety of factors: demographics, lifestyle, related to the surgical procedure, or the current emotional state.

> *When you see a very stressed and anxious young woman who is scheduled for a laparoscopy, then you immediately think: this will probably be Bingo.* (female senior attending, intervention group)

However, the number of factors and which specific factors they used in their assessment varied per interviewee. One specific factor was commonly mentioned by the interviewees: a history of PONV, especially when patients had actively reported it at the preanesthesia clinic.

> *Yes, people with a history of PONV. People that claim: 'I am always terribly nauseous'. It almost seems like a self-fulfilling prophecy.* (female senior attending, intervention group)

2B. risk factors may be experience-based or evidence-based

All interviewees reported several factors which are predictive factors in one of the commonly known risk scores (e.g. gender, non-smokers). Most interviewees used a combination of experience-based and evidence-based factors. However, both extremes were present: one interviewee did not want to rely on his own experience, whereas another interviewee knew the risk factors from literature but did not use them in his assessment on whom to give PONV prophylaxis.

> *Yes, my assessment is based on literature...Relying on your own experience can be quite treacherous, because most patients get nauseous at the nursing ward, not at the postanesthesia care unit, and I do not regularly visit all my patients at the nursing ward.* (male resident, care-as-usual group)

> *So these are risk factors known from literature, but I do not take them into account to decide on who will become nauseous...For example smoking, if it has any consequences for my anesthetic case, I prefer to consider its pulmonary effects.* (male senior attending, care-as-usual group)

2C. weighing factors: no clear-cut risk appraisal

Whichever factors are used by the individual interviewees, they did not consciously weigh the factors to arrive at a risk percentage for PONV. They seemed to use the factors more in a qualitative way than in a quantitative way.

> *No, it is not a conscious process. It is more that you roughly know the risk factors, you have an overall view of the patient, and then you start thinking: Does this patient has a high risk or not? But in my mind, I am not actually calculating the risk.* (female resident, intervention group)

> *Just an intuitive feeling: this person will probably suffer from PONV.* (male senior attending, care-as-usual group)

category 3: making decisions on prophylactic antiemetic strategies

3A. perceived benefits and harms of prophylactic antiemetic strategies

The interviewees were asked about three prophylactic antiemetic drugs, which are part of the local protocol on PONV. All interviewees considered ondansetron to be a suitable drug for PONV prophylaxis in terms of effectiveness and with little side effects.

> *The only drug that I actually administer as prophylaxis and for which I don't feel any reluctance is ondansetron.* (male junior attending, care-as-usual group)

Interviewees from the intervention group were better able to quantify the perceived effectiveness of ondansetron, although their quantification was mostly intuitive.

> *Well, when your risk is 60% at first and afterwards a risk of 40%, then it is about 20% effective.* (female resident, intervention group)

The overall positive attitude towards ondansetron contrasted with the more divided and negative attitude towards the other two prophylactic drugs, droperidol and dexamethasone. Most of the interviewees considered at least one of these drugs to be quite effective as prophylaxis, although most of the interviewees were unable to quantify their effectiveness. In addition, the opinion on the severity of side effects, especially for dexamethasone, varied among the interviewees.

> *Dexamethasone? Definitely not as prophylaxis. That stuff has too many side effects. And I do not believe that it actually works that well.* (female senior attending, care-as-usual group)

> *Patients who always become nauseous after surgery, I always just give them dexamethasone, and it works, yes it works really well....No, I don't see many side effects of that.* (female senior attending, intervention group)

One interviewee actually reported the difference in opinion on dexamethasone.

> *There are people who say that a single dose of dexamethasone is not so bad, whereas other people claim that even a single dose may cause serious side effects.* (male junior attending, intervention group)

Several interviewees reported to rarely administer either dexamethasone, droperidol or both as prophylaxis. However, the interviewees were unable to explain why they were reluctant to administer those drugs. For example, one of the interviewees even made several conflicting statements on the use of droperidol as prophylaxis during the interview.

> *There is definitely added value for droperidol, I think that is a good drug....I rarely administer droperidol intraoperatively...It provides some drowsiness, but in the low dose that we give, it is hardly a problem.* (female senior attending, intervention group)

3B. decisions on PONV prophylaxis are made implicitly

Although the interviewees were not explicit in how they decide to administer or withhold specific prophylactic antiemetic drugs, they seemed to weigh benefits (numbers needed to treat) over side effects (numbers needed to harm) in a qualitative – non-numerical – way.

You see, you can prevent nausea for a number of people, but when opposed by one person suffering from a surgical site infection [author's note: a perceived risk of dexamethasone PONV prophylaxis], your net benefit is lost. (male junior attending, care-as-usual group)

3C. PONV prophylaxis: only when indicated

The interviewees agreed that because of the possible side effects, PONV prophylaxis should not be considered a standard procedure for all patients.

Let's put it this way: to systematically provide PONV prophylaxis to all patients is impossible, because of the side effects. Although, when indicated for a specific patient, I believe you can accept the side effects. (male senior attending, care-as-usual group)

category 4: attitude towards the use of prediction models and decision support

4A. intervention group: the predicted risk is higher than their own assessment

As interviewees from the intervention group received the individual predicted PONV risks of their patients during the cluster-randomized trial, these interviewees were asked about their experiences with the use of the predicted risks.

Three out of four interviewees reported that they were surprised by the magnitude of the predicted risks; for most patients their own risk assessment would have been lower than the risk as predicted by the prediction model. One interviewee reported that this discrepancy resulted from a more accurate risk assessment by the model.

And I believe the predicted risks were higher because the prediction model actually calculates the risk from all factors associated with PONV, whereas we anesthesiologists lean towards particular touchstone factors such as history of PONV and type of surgery. (male junior attending, intervention group)

4B. intervention group: no conscious roadmap for using the predicted risk

All interviewees from the intervention group reported to use the presented risks in their decision to administer PONV prophylaxis. The interviewees were unable to explicate how they use the predicted risk to arrive at a decision on PONV prophylaxis.

When you open the anesthesia information management system, you immediately see it. Well, it is not that I immediately am consciously aware of it, sometimes I am, but not always. It occasionally happens that at the end of the procedure you think: 'Oh yes, the risk was high, shall I... (female senior attending, intervention group)

One interviewee reported a decrease in the use of the predicted risks over time. The interviewee was not able to pinpoint the cause for this decrease.

> *Yes, well, when I entered the study and I received these predicted risks, I did use them for a while...I used the numbers to decide whether or not I would administer prophylaxis. Lately, I obviously eased on doing that, and I cannot clearly explain why that happened.* (male junior attending, intervention group)

4C. care-as-usual group: the predicted risk is mainly a reminder

Interviewees from the care-as-usual group were all positive on automatically receiving a patient's predicted PONV risk during the anesthetic case. They all considered presentation of the predicted risk a good way to prompt them to consider PONV prophylaxis for each patient.

> *You know, it is foremost a reminder...Something in your screen that says: 'Remember!'... You are supposed to think of everything, sometimes you just fall short. So when it is on by default...* (female senior attending, care-as-usual group)

One of the interviewees said specifically that the actual predicted risk would not be that helpful.

> *You do not benefit that much from a predicted risk. It is more a way to direct your attention to PONV, because to actually solve the problem itself you need more details than just a predicted risk.* (male senior attending, care-as-usual group)

4D. authority in making decisions is more important than decision support

The predicted risk was implemented without an actionable recommendation added to the predicted risk (assistive implementation). When the interviewees were asked whether they would like a recommendation on PONV prophylaxis to be added to the risk (directive implementation), the interviewees would all like to have such a recommendation, but not at the cost of their authority to make case-by-case decisions.

> *There is nothing wrong with an advice, as long as it does not become compulsory.* (male junior attending, care-as-usual group)

> *Of course it is nice to get some advice. On the other hand, all doctors, including me, feel that they can decide for themselves... I am in charge and I am also responsible.* (female senior attending, intervention group)

4E. prediction models do not consider other comorbidity

The interviewees prefer to retain their authority in medical decision making, because the recommendation only targets the risk of PONV and does not take into account the full clinical condition of the patient.

The recommendation is probably only aimed at PONV, so I do believe that you should continue to consider which patient you have in front of you and what drugs you are going to administer. (female resident, intervention group)

You may consider giving the patient dexamethasone as prophylaxis, but it may be contraindicated for specific patients or specific procedures. (male senior attending, care-as-usual group)

4F. the predicted risk does not reflect the underlying mechanisms of PONV

Moreover, not only should other comorbidity be considered according to the interviewees. The mechanism which causes nausea or vomiting (e.g. abdominal surgery, opioids, vestibular disturbances) may determine which antiemetic drug is expected to be prophylactically effective. Consequently, a physician should have insight in which factors contributed to the recommendation for an individual patient.

You can only treat a patient when you know the actual cause of the problem. Without that, the predicted risk will not do you any good, you need to know which factors contributed to that particular risk. (male senior attending, care-as-usual group)

Even when understanding the expected – i.e. predicted – mechanism for PONV, the physician cannot fully rely on the recommendation for the decision to administer particular prophylactic antiemetic drugs.

You need to know which factors contributed to that particular risk and then a recommendation may follow. However, as the treating physician you may regard this recommendation solely as an advice. (male senior attending, care-as-usual group)

4G. willingness to use decision support may be dependent on familiarity

When answering the question whether the interviewees would follow such a recommendation, interviewees of both study groups would consider it an advice and not standard practice. However, interviewees from the care-as-usual group considered the recommendation to be foremost a reminder, and doubted whether such a recommendation would have an added value to other reminders, such as the predicted risk alone.

We already have a protocol which antiemetic drugs to administer, so in its essence a predicted probability should suffice. (male junior attending, care-as-usual group)

When you are sufficiently informed on its background, you can use the recommendation to your advantage. The question remains whether all the efforts which precede the implementation are actually worth their while. (male senior attending, care-as-usual group)

In contrast to the care-as-usual group, interviewees from the intervention group expected that they would follow the recommendation on a regular basis.

> *I believe I would be inclined to administer more PONV prophylaxis, because you know that the predicted risk is based on an extensive calculation using several factors.* (male junior attending, intervention group)

The difference between the care-as-usual and intervention group suggests that using decision support increases the willingness to rely on decision support. One interviewee actually reported that it might be a matter of 'just getting used to it'.

> *On one hand it is nice to treat a patient based on a true evidence-based recommendation, on the other hand all doctors want to decide for themselves...So I have mixed feelings about this. I believe it will be a matter of getting used to it. At a certain point it will become a custom, and especially when it is has been proven to work, you would be a fool not to follow the advice.* (female senior attending, intervention group)

appendix section 2

full description of the structured survey

methods

The survey was developed using the same four topics as the semi-structured interview schedule: 1) perceived burden of PONV by the physician; 2) professional risk stratification: how does the physician identify patients at risk for PONV; 3) the decision-making process regarding preventive antiemetic strategies; 4) attitude towards the use of prediction models and decision support for prophylactic treatment of PONV.

These four topics were considered the most important factors contributing to the willingness of physicians to use a formal prediction model regarding clinical decision making in the prevention of PONV. For each topic several questions were formulated in Dutch, as discussed in the following paragraphs. As individual questions were expected to be highly related to other questions within their topic, a summary z-score was calculated for each topic and standardized to the mean score and standard deviation of all physicians, i.e. both allocations groups combined (henceforth referred to as the total group of physicians).

perceived burden of PONV by the physician

Physicians were asked to grade – on a 100-point visual analogue scale (VAS) – ten postoperative complications including PONV by their perceived burden (see Table 8.1).[1] Zero was considered to be 'no problem at all' and hundred was considered 'equal to death'. For each physician, nine standardized ratios were calculated between the perceived burden of PONV and the perceived burden of the other postoperative complications. Standardized ratios (z-scores) were calculated by subtracting each individual postoperative complication score from the individual PONV score minus the mean of the same difference for the total group of physicians, divided by the standard deviation of the same difference within the group. The sum of all nine z-scores for each physician was again standardized to the mean and standard deviation of the total group of physicians, and this standardized score was used as the summary z-score for this topic.

professional risk stratification: how does the physician identify patients at risk for PONV?

For fourteen prespecified patient or procedure characteristics the physicians were asked, whether they used the characteristics as risk factors, protective factors or no factors at all in their PONV risk assessment for an individual patient. These fourteen variables included the seven predictors of the formal PONV prediction model. In addition, physicians could specify remaining variables as risk factors or protective factors, in order to complete their personal prediction model. For each physician, we calculated the number of predictors of the formal PONV prediction model which the physician used for their personal risk assessment and had scored in the right predictive direction, subtracted by the number of other predictive factors in their personal prediction model. The resulting individual score was then standardized (z-score) by subtracting the mean score of the total group of physicians and dividing the difference by the standard deviation within the total group.

making decisions on preventive antiemetic strategies

Physicians were asked for their opinion on the three prophylactic antiemetic drugs: ondansetron, droperidol, and dexamethasone. For each drug the physicians were asked: 1) whether they use the drug as prophylaxis; 2) whether they consider the drug to be effective; 3) whether they believe the drug has many contraindications; and 4) whether they believe the drug has many side effects. Their opinion was scored on a 100 point VAS, ranging from 'never' to 'always' for statement 1, and from 'I strongly disagree' to 'I strongly agree' for statements 2–4. Standardized scores were calculated as the difference from the mean of the total group of physicians divided by the standard deviation (z-scores). Scores were standardized in the same direction: a positive score indicated a positive attitude towards to the use of the drug. The sum of all twelve z-scores for each physician was again standardized to the mean and standard deviation of the total group of physicians, and this standardized score was used as the summary z-score for this topic.

attitude towards the use of prediction models and decision support for prophylactic treatment of PONV

Eleven pairs of statements on regarding the use of a prediction model and how it is implemented as decision support were presented to the physicians. For each pair it was prespecified which statement was considered to be the most favorable towards the use of a prediction model for PONV prophylaxis (for all statements, see Table 8.4). Physicians were asked to choose which of the two statements they considered most relevant when using the prediction model in their clinical decision making. The total number of favorable statements was then standardized by taking its difference from the mean of the total group of physicians divided by the standard deviation (z-score).

additional questions for the intervention group

Two specific questions were added for physicians of the intervention group: 1) whether the PONV risks as predicted by the model were similar to their personal assessment of the patient's PONV risk; 2) whether the individualized feedback on the incidence of PONV among their patients was according to their expectancy.

statistical analysis

Each summary z-score was visually assessed for a normal distribution using histograms and Q-Q plots. For each allocation group, the means and standard deviations of the summary z-scores per topic were calculated. Multiple testing was avoided by summing all four summary z-scores for each physician into a total summation z-score, which was tested using a two-sample Student's t-test. Results for the two additional questions for the intervention group physicians were presented as absolute frequencies with percentages. A two-sided alpha of <0.05 was considered statistically significant.

post-hoc analysis of prediction model versus recommendation

During the interview analysis the question arose whether physicians from the intervention group preferred an actionable recommendation instead of the results of the prediction model alone, whereas care-as-usual physicians preferred only the predicted risk over the recommendation. This question was addressed in the survey by pooling questions which addressed either the predicted risk, a recommendation or both. One question directly asked for a preference for a recommendation over a predicted risk alone. Additionally, one set of questions compared the recommendation and the predicted risk to ease of use of the decision support, and a second set compared the recommendation and the predicted risk to receiving individualized feedback on the incidence of PONV. When in a set of questions a physician scored positive for the recommendation and negative for the predicted risk, this was counted positive for a preference for the recommendation. A positive score for the predicted risk and a negative score for the recommendation were counted negative for a preference for the recommendation. Two positive scores for a set of questions or two negative scores were counted as indifferent. The direct question and the results of the two indirect sets of questions were summed and testing using a Wilcoxon rank test.

adjusting the original trial's analysis

The original trial studied the PONV incidence (primary outcome) and the number of prophylactic antiemetics administered per patient (secondary outcome) for risk-dependent differences between allocation groups using mixed effects regression analyses.[2] As the total summation z-score from the web-based survey was hypothesized to reflect the inclination for physicians to treat PONV prophylactically, the original analyses were additionally adjusted for the total summation z-score. Thus, the adjusted mixed effects regression models – logistic for the primary outcome, Poisson for the secondary outcome – included as fixed effects the allocation group, the predicted risk, their interaction term, time in months and the total summation z-score. The random effects were the same as the original analysis: an intercept, predicted

risk, their interaction term, and time in months. The total summation z-score was tested for nonlinearity using restricted cubic splines.[3] For the original analysis missing data were multiply imputed using 10 imputation datasets. Survey results were merged with each of the original 10 imputation datasets, to be able to compare the results with the original trial. Results of regression analyses were presented as odds ratios with 95% confidence intervals (CI) or rate ratios with 95% CI.

results

Of the 57 physicians analyzed in the original trial, 53 (93%) completed the survey (intervention group 29 (94%); care-as-usual 24 (92%). Two of the four physicians who did not complete the survey emigrated during the study period and we were unable to contact them within a reasonable time frame. The other two physicians did not complete the survey despite repeated requests from the first author.

perceived burden of PONV by the physician

Compared to other postoperative complications, physicians considered the burden of PONV to be moderately important. Physicians of the intervention group considered PONV more important than physicians of the care-as-usual group (Table 8.1). In the intervention group it was the 5th most important complication (comparable to postspinal headache), and 6th most important complication for the care-as-usual group (comparable to urinary retention). Despite these group differences, there was a large variation among individual physicians as to how they valued PONV as a postoperative complication, as seen from the wide interquartile

TABLE 8.1 SURVEY RESULTS ON THE PERCEIVED BURDEN OF PONV

	Care-as-Usual Group	Intervention Group
	(*n* = 24)*	(*n* = 29)*
Anastomotic leakage	74 (61–90)	75 (62–85)
Peripheral nerve injury	62 (48–80)	75 (57–83)
Wound infection	60 (48–77)	59 (48–75)
Postspinal headache	49 (32–70)	51 (45–69)
PONV	34 (24–59)	49 (40–75)
Urinary retention	36 (21–46)	37 (28–62)
Pruritus	26 (19–46)	31 (17–51)
Obstipation	24 (19–50)	25 (14–48)
Sore throat	24 (13–40)	24 (9–30)
Shivering	22 (16–39)	19 (11–43)

Data represent medians (interquartile ranges) for a perceived burden score on a 0–100 visual analogue scale for each postoperative complication

* The number of physicians who completed the survey

PONV = postoperative nausea and vomiting

ranges. The standardized summary z-scores were also different for both groups (see Chapter 4, Table 4.4).

predictive professional risk stratification: how does the physician identify patients at risk for PONV?

There was a large variation between physicians in which factors they considered to be important predictive factors for PONV (Table 8.2). Two factors from the formal prediction model were considered important risk factors by almost all physicians: history of PONV or motion sickness; and abdominal or middle ear surgery. On the other hand, almost none of the physicians considered outpatient surgery to be an important predictive factor for PONV, which was inconsistent with the prediction model, as outpatient surgery is a major protective factor in the model. Almost none of the factors from the prediction model were scored in the opposite direction (i.e. scored as protective while risk factor in the model or scored as risk factors while protective in the model). Of the other factors, not used in the prediction model, postoperative use of morphine (53%) and preoperative anxiety (53%) were often considered as important risk factors by physicians.

TABLE 8.2 SURVEY RESULTS ON THE PROFESSIONAL RISK STRATIFICATION: HOW PHYSICIANS IDENTIFY HIGH PONV-RISK PATIENTS

	Care-as-Usual Group ($n = 24$)*	Intervention Group ($n = 29$)*
Risk factors from the formal prediction model		
History of PONV / Motion sickness	24 (92)	28 (90)
Surgery with a high PONV risk	24 (92)	26 (84)
Female sex	17 (65)	25 (81)
Current smoking	9 (35)	16 (52)
Inhalational anesthesia	8 (31)	19 (61)
Age	7 (27)	11 (35)
Ambulatory surgery	0 (0)	2 (7)
Characteristics not included in the formal prediction model		
Postoperative use of morphine	14 (54)	14 (45)
Preoperative anxiety	13 (50)	15 (48)
Intolerance to fasting	7 (27)	7 (23)
Prolonged operation duration	3 (12)	5 (16)
Obesity	4 (15)	3 (10)
Reflux disease	4 (15)	1 (3)
Prolonged preoperative fasting	2 (8)	3 (10)

Data represent absolute number of physicians (percentages) that scored the patient or procedure characteristic as a risk factor for PONV.

* The number of physicians who completed the survey

PONV = postoperative nausea and vomiting

TABLE 8.3 SURVEY RESULTS ON MAKING DECISIONS ON PROPHYLACTIC ANTIEMETIC STRATEGIES

	Care-as-Usual Group	Intervention Group
	(*n* = 24)*	(*n* = 29)*
I often use ondansetron as prophylaxis	56 (49–85)	75 (67–86)
Ondansetron is effective as prophylaxis	65 (62–76)	68 (59–81)
Ondansetron has many side effects	15 (8–50)	25 (13–40)
Ondansetron has many contraindications	20 (10–41)	29 (16–36)
I often use droperidol as prophylaxis	50 (20–60)	50 (22–61)
Droperidol is effective as prophylaxis	62 (50–70)	60 (50–68)
Droperidol has many side effects	52 (47–68)	50 (28–60)
Droperidol has many contraindications	50 (41–59)	50 (30–57)
I often use dexamethasone as prophylaxis	16 (0–40)	52 (15–62)
Dexamethasone is effective as prophylaxis	56 (42–68)	64 (50–74)
Dexamethasone has many side effects	60 (50–70)	50 (37–60)
Dexamethasone has many contraindications	70 (50–76)	56 (44–67)

Data represent medians (interquartile ranges) for each statement, scored on a 0–100 visual analogue scale with 0 = 'I strongly disagree' and 100 = 'I strongly agree'

* The number of physicians who completed the survey

Physicians of the intervention group considered more factors from the formal prediction rule to be important predictive factors (median 3, interquartile range (IQR) 1–4) as compared to physicians of the care-as-usual group (median 1, IQR 0.75–2) Overall, they considered a similar number of other factors – not used in the prediction model – to be predictive of PONV (median of 2 for both groups, overall IQR 1–3). The differences between the allocation groups were reflected by differences in the standardized summary z-score (see Chapter 4, Table 4.4).

making decisions on preventive antiemetic strategies

Physicians of both allocation groups considered ondansetron, droperidol and dexamethasone equally effective as prophylactic antiemetic drugs (Table 8.3). Physicians of the intervention group reported to use more prophylactic ondansetron (median 75 vs. 56, on a 0–100 VAS) and dexamethasone (median 52 vs. 16) than physicians of the care-as-usual group, whereas for droperidol they reported it to be the same (median 50 vs. 50). Ondansetron was considered the drug with the least side effects and contraindications, especially by care-as-usual group physicians. In contrast, care-as-usual group physicians considered dexamethasone to be the antiemetic with the most side effects (median 60, IQR range 50–70) and contraindications (median 70, IQR 50–76), whereas the intervention group physicians considered droperidol and dexamethasone equally harmful (medians 50 to 56). Overall, intervention group physicians had a more positive standardized summary z-score than care-as-usual physicians (see Chapter 4, Table 4.4).

attitude towards the use of prediction models and decision support for prophylactic treatment of PONV

The risk-tailored use of PONV prophylaxis was highly supported by all participating physicians (intervention group 90%, care-as-usual group 96%). Moreover, all physicians agreed that it is better to give PONV prophylaxis based on a patient's actual risk, instead of to every patient, ignoring the actual risk (Table 8.4). A majority of physicians believed that an evidence-based prediction model would render a more accurate estimation of a patient's individual PONV risk than a clinician solely clinical judgment (intervention group 69%, care-as-usual group 67%). When unaware of the predicted PONV risk, physicians of the intervention group would be more willing to administer PONV prophylaxis to everybody than physicians of the care-as-usual group (intervention group 47%, care-as-usual group 28%). For this topic, the standardized summary z-scores were similar for both allocation groups.

As already indicated by physicians during the interviews, physicians of the intervention group had a greater preference for a recommendation added to the predicted risk than physicians of the care-as-usual group (intervention group 59%, care-as-usual group 42%).

TABLE 8.4 SURVEY RESULTS ON THE ATTITUDE TOWARDS THE USE OF PREDICTION MODELS AND DECISION SUPPORT

	Care-as-Usual Group	Intervention Group
	(*n* = 24)*	(*n* = 29)*
Risk-based PONV prophylaxis[1st] vs. nobody[2nd]	23 (96)	26 (90)
Everybody PONV prophylaxis[1st] vs. nobody[2nd]	7 (29)	14 (48)
Risk-based PONV prophylaxis[1st] vs. everybody[2nd]	24 (100)	29 (100)
Prediction model[1st] predicts better than physicians[2nd]	16 (67)	20 (69)
Prefer presentation of the risk at the OR[1st] over the OPE[2nd]	16 (67)	12 (41)
Prefer a recommendation[1st] over a risk[2nd]	10 (42)	17 (59)
Prefer a protocol[1st] over a recommendation[2nd]	4 (17)	4 (14)
Presented risk[1st] more important than ease of use[2nd]	7 (29)	4 (14)
Recommendation[1st] more important than ease of use[2nd]	12 (50)	15 (52)
Presented risk[1st] more important than feedback on results[2nd]	11 (46)	12 (41)
Recommendation[1st] more important than feedback on results[2nd]	9 (38)	14 (48)

Data represent absolute number of physicians (percentages) that preferred the first statement ([1st]) over the second statement ([2nd]). The first statement was considered most favorable towards the use of a prediction model for PONV prophylaxis.

* The number of physicians who completed the survey

OR = operating room; OPE = outpatient evaluation; PONV = postoperative nausea and vomiting

When compared to the ease of use of decision support, intervention group physicians had a greater preference for the recommendation (preference for recommendation: intervention group 38%; care-as-usual group 21%. None of the physicians preferred the predicted risk alone). When compared to receiving individualized feedback on PONV the differences were smaller, intervention group physicians had a slightly greater preference for the recommendation (intervention group 10%; care-as-usual group 8%), whereas care-as-usual group physicians had a greater preference for the predicted risk alone (intervention group 3%; care-as-usual group 17%). For the combined result the intervention group also had a greater preference for the recommendation, which was statistically significant (median score intervention group 1; care-as-usual group 0; $p<0.05$).

Overall, the mean of the total summation z-score was 0.66 (SD 2.4) for the intervention group and was –0.80 (SD 2.4) for the care-as-usual group ($p<0.05$). As there was no difference between groups for 'Attitudes towards prediction models and decision support', the combined z-score for the other three topics without 'Attitudes towards prediction models and decision support' was also reported in Table 4.4 of Chapter 4. A density plot was graphed for all four summary z-scores and the total summation z-score to facilitate interpretation of the differences between the two study groups (see Chapter 4, Figure 4.1). Gaussian kernel estimates were used for the density plots, using the function *beanplot* from the *beanplot*

TABLE 8.5 MIXED EFFECTS REGRESSION ANALYSIS ADJUSTED FOR THE TOTAL SUMMATION Z-SCORE FROM THE SURVEY, COMPARED WITH THE RESULTS FROM THE ORIGINAL ANALYSIS

	Adjusted for total summation z-score*	Original analysis*,†
	PONV incidence (primary outcome)	
Intervention group	0.95 (0.84–1.06)	0.97 (0.87–1.10)
Predicted risk‡	2.30 (2.08–2.55)	2.29 (2.07–2.54)
Interaction: study group and predicted risk‡	0.92 (0.80–1.05)	0.92 (0.80–1.05)
Total summation z-score	1.00 (1.00–1.01)	N/A
	No. of prophylactic antiemetics (secondary outcome)	
Intervention group	1.86 (1.51–2.30)	2.03 (1.64–2.50)
Predicted risk§	2.21 (1.88–2.60)	2.16 (1.84–2.53)
Interaction: study group and predicted risk§	1.55 (1.23–1.96)	1.63 (1.30–2.03)
Total summation z-score	1.00 (1.00–1.01)	N/A

Numbers for the primary outcome represent odds ratios with 95% confidence intervals for the fixed effects.
Numbers for the secondary outcome represent rate ratios with 95% confidence intervals for the fixed effects.
All analyses included time as a covariable.
* After multiple imputation using 10 imputation datasets
† The results of the original analysis of the cluster-randomized impact trial
‡ Odds ratios represent predicted risks of 100% (a probability of 1)
§ Rate ratios represent predicted risks of 100% (a probability of 1)
PONV = Postoperative Nausea and Vomiting

package in R software, version 3.0.1 (the *beanplot* function uses the *density* function from the *stats* package for kernel estimates).

additional questions for the intervention group

Of the physicians of the intervention group, 7 (24%) reported that the predicted risk as presented to them was most of the time similar to their own assessment of the PONV risk for that particular patient. For 13 physicians (45%) the predicted risk was mostly higher than their own assessment, whereas the remaining 9 physicians (31%) reported that for most patients the predicted risk was either lower or higher than their own assessment, but rarely the same. Moreover, only two physicians of the intervention group (7%) had expected a higher incidence of PONV among their own patients than was reported to them by the individualized feedback. For 7 physicians (24%) their individual PONV incidences were according to their expectancy, whereas 20 physicians (69%) reported that their individual PONV incidences were higher – or even much higher – than they had expected.

adjusting the original trial's analysis

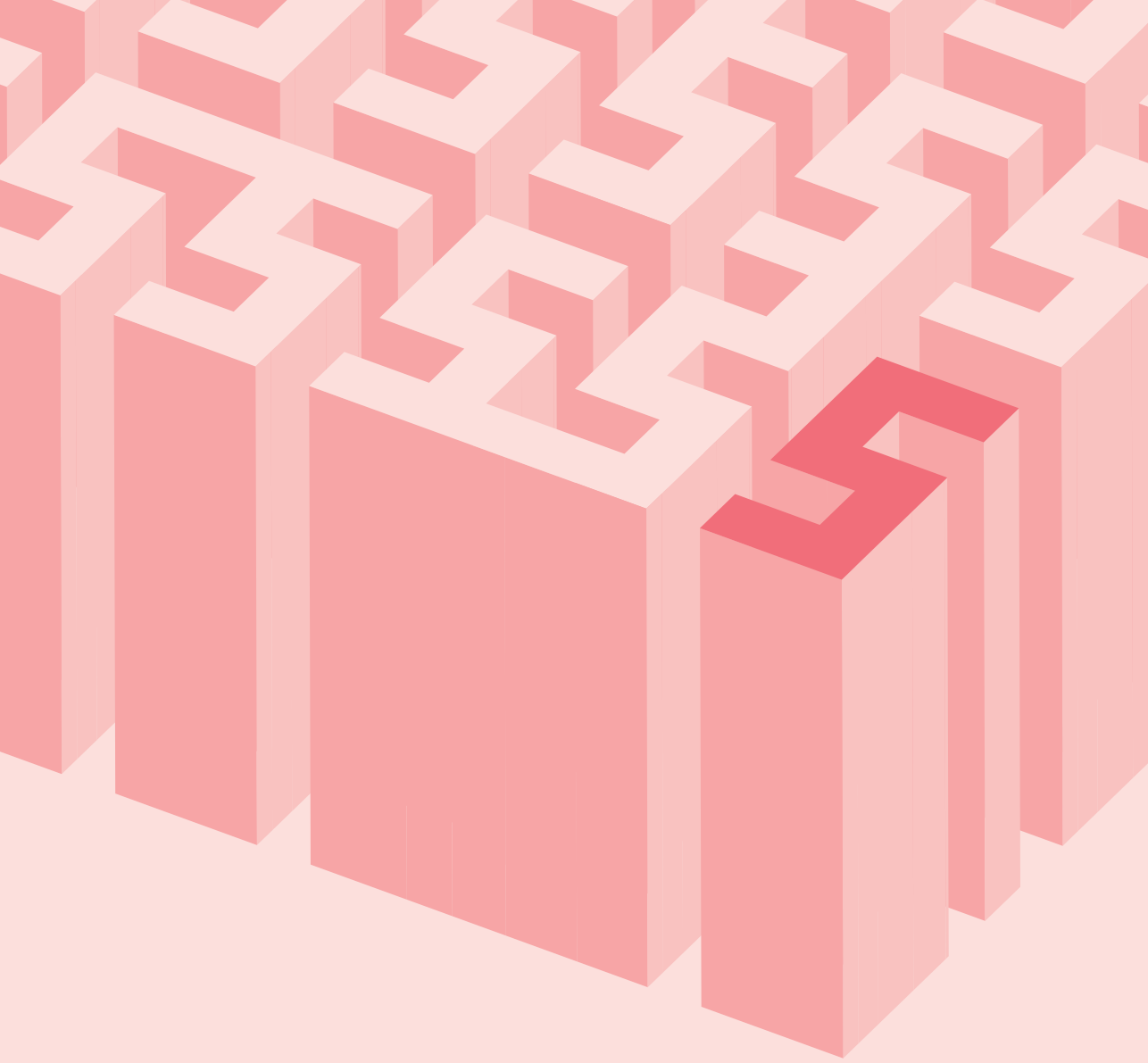
When the total summation z-score of the survey was used to adjust the original mixed effects regression analyses,[2] results were similar to the original results for both the primary and secondary outcome (Table 8.5). There was no association between the total summation z-score itself and the incidence of PONV (odds ratio 1.0, 95% CI 1.0–1.0) or the administration of prophylactic antiemetics (rate ratio 1.0, 95% CI 1.0–1.0).

references

1. Macario a, Weinger M, Truong P, Lee M. Which clinical anesthesia outcomes are both common and important to avoid? The perspective of a panel of expert anesthesiologists. *Anesth Analg*. 1999;88(5):1085–1091. PMID:10320175.

2. Kappen TH, Moons KGM, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. *Anesthesiology*. 2014;120(2):343–354. PMID:24105403.

3. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387. PMID:8668867.



SUMMARY

SAMENVATTING

summary

When presenting their doctor with signs and symptoms, patients first want to know the diagnosis. Once a diagnosis has been established, they would like their doctor to explain the available treatment options. They also want to know which course the disease will run – with and without treatment – to decide whether they will accept a proposed treatment or refrain from doing so. In other words: they are interested in their prognosis. Thus, finding the correct diagnosis and estimating the prognosis are important conditions for taking a medical decision.

A clinical prediction model can assist doctors in arriving at the most likely diagnosis or estimating the prognosis. By utilizing various patient and disease related properties, such models can yield objective estimations of the risk of a disease or the probability of a certain disease course for individual patients. This individual probability can then be used by doctors in their decision making process, thereby (hopefully) improving patient outcome.

A prediction model can only improve medical decision making when the probability estimates are both accurate and valid in 'new' patients. The model's predictive value can be verified in an 'external' validation study, i.e., in a different clinical setting. However, even if a model's predictive value is good, that does not guarantee that using such a model in clinical practice will enhance a doctor's decision making. In so-called 'impact studies' the prediction model is implemented in daily practice and the effect on clinical outcomes is compared to care-as-usual – i.e., care without using the model.

Chapter 2 describes the process of preparing a prediction model for postoperative nausea and vomiting (PONV) for implementation in the anesthesiology department of an academic hospital (UMC Utrecht). The prediction model was previously developed using patient data from another academic hospital (AMC, Amsterdam) and subsequently validated in Utrecht. Due to differences between the Amsterdam and Utrecht practices, and ongoing changes in the field of anesthesiology, the original model was not deemed suitable for implementation in its original form. Guided by these clinical differences and practice changes we were able to update the model, which improved the predictive value for the clinical practice in Utrecht.

In **chapter 3** we studied the effects of the implementation of the PONV prediction model by means of a large-scale randomized trial. The research question was whether the prediction model improved doctors decisions as to which patients should be administered PONV prophylaxis during surgery. We also determined if the prophylaxis indeed reduced PONV. For the study, the prediction model was implemented in the form of decision support. This decision support was then randomly presented to half of the doctors of the anesthesiology department of the Utrecht hospital (the intervention group). The decision support tool presented the predicted PONV probability for an individual patient to his / her anesthesiologist in the electronic patient record during surgery. Since we only presented the calculated probability of PONV, this was an 'assistive' (not 'directive') form of decision support. The latter would have meant presenting the doctor also with a corresponding treatment advice. Compared to the control group of doctors to whom the decision support was not offered, the doctors in the intervention group administered more PONV prophylaxis to their patients, particularly to patients with a high predicted risk. However, this positive effect on doctors' behavior did not result in a corresponding beneficial effect on patient outcome, as there was no difference in the incidence of PONV among patients from doctors in both study groups. We concluded that the increase in PONV prophylaxis prompted by the risk-based decision support was not sufficient to yield a clear advantage for patients.

Through interviews and surveys, the discrepant results of chapter 3 were explored in **chapter 4**. A viable explanation for these discrepancies is that the increased prescription of PONV prophylaxis caused by using the prediction model is still insufficient to achieve a positive effect on patients. The absolute increase in PONV prophylaxis was small in the randomized trial, and doctors could have administered more PONV prophylaxis to high-risk patients. Using interviews and surveys, the doctors who participated in the randomized study were questioned about their decision making regarding PONV prophylaxis, how they interpreted model-based PONV predictions, and subsequently, how the presented probabilities influenced their decisions.

Four barriers for the use of the prediction model in decision making on PONV prophylaxis emerged. First, it appears that preventing PONV is not a major concern for anesthesiologists during the surgical procedure. Secondly, decisions on PONV prophylaxis are mainly taken intuitively, rather than being deduced from conscious and analytic considerations. Thirdly, simply presenting predicted PONV probabilities was not perceived as an easy-to-use tool in decision making regarding prophylaxis, especially when not accompanied by treatment recommendations. Finally, some respondents argued that a prediction model for PONV does not take into account the risks of PONV prophylaxis for individual patients or their comorbidity.

Together, these impeding factors could explain the limited increase in the doctors' use of PONV prophylaxis in the randomized trial. Using probability estimates during a decision can be problematic, because the probability estimates need to be consciously integrated with the clinical experience of the doctor to properly assess the pros and cons of PONV prophylaxis. Thus, in a high workload environment – such as an operating room – the conscious use of a probability may require extra attention from doctors, especially when the decision is usually made intuitively. When the decision is related to a low-priority subject for doctors – as in

the case of PONV – then it is understandable why doctors did not always spend the extra effort. Therefore, when a prediction model is set to be used as a decision support tool, it is important to ensure that the decision support is as closely tailored to the doctors' workflow as possible, so as to minimize the need for extra attention. Adding an actionable recommendation on treatment or further diagnostic testing to the predicted probability – i.e., directive decision support – could well be a first step towards reducing the attention required for the use of decision support.

Chapter 5 documents how the prediction model for PONV was subsequently implemented in Utrecht in the same anesthesiology practice, but now in the form of 'directive' decision support. In this before-after study, the presentation of the predicted PONV probability – now with a corresponding advice for the administration of PONV prophylaxis – led to a larger increase in the actual administration of PONV prophylaxis by doctors than in the randomized trial of chapter 3. In contrast to the randomized trial, the increased administration of PONV prophylaxis in the before-after study lowered the incidence of PONV in patients with a high predicted risk of PONV. Despite this decrease, a substantial number of patients was still suffering from PONV. For a truly 'PONV-free' hospital, even more liberal administration of PONV prophylaxis would be required.

Chapter 6 summarizes the research team's experiences gained from the design, analysis and interpretation of the implementation studies regarding the PONV prediction model. These experiences may help guide further development of studies into the implementation of prediction models (*Impact Studies*). A prerequisite for such implementation research is to realize that a prediction model is a complex intervention. Consequently, it is necessary, at all stages of the research, to take into account the clinical problem for which the model is going to be used, as well as the clinical practice in which the model is to be implemented. Often, the prediction model requires an update before it can be implemented, so that the predictive value is guaranteed in the clinical practice. A format for effective implementation of the prediction model as a decision support tool must be designed. For example, decision support could be implemented as an automated tool in the electronic patient record. Background information on how the model calculates the predicted probability might be provided to the user. Finally, one must decide if 'actionable' treatment advice will be added to the predicted probability.

One should also bear in mind that the implementation of a prediction model is a 'complex intervention' that can have both intended and unintended effects. Doctors obtain information from the prediction model, which they then need to interpret and use in their decisions. Exactly how they will do this is not always clear in advance. Because of the possible side effects, the ultimate consequence for the patient is not simply the sum of the predictive value of the model supplemented with the change in the doctor's behavior. It is important to consider in advance, whether patient outcome should be studied in the implementation study, or whether studying the effect of the prediction model on doctors' behavior will suffice. Additionally, one can take unintended and unexpected effects into account in the study design. Randomization at doctor level can prevent unintended learning effects and measures can be taken to avoid

so-called contamination between study groups. Both learning effects and contamination could dilute the apparent effects of the prediction model on the doctors' behavior. Gathering additional information – through both qualitative and quantitative methods – may improve understanding of possible unintended effects of the prediction model. When these aspects are considered in advance, one can better estimate the 'feasibility' of a prediction model implementation study.

In **chapter 7** we argue that, as of yet, clinical prediction models and decision support do not sufficiently take into account the context of the problem they are targeting. In their daily decision making, doctors depend on the context in which a patient and a problem are presented. Approaching the problem within the context is an essential method for humans to deal with the complexity of a problem. On the other hand, prediction models and decision support use a mathematical approach and relatively simple logic. Decision support should help doctors in making appropriate medical decisions. It is therefore of great importance that they are able to interpret decision support information within the context of the problem. If such interpretation is in any way impeded, decision support can actually be confusing, complicating the decision process of doctors. It is likely that the joint forces of doctors and technology will produce better decisions than either of them exclusively: after all, they do have different ways of dealing with the complexity of a decision and are thus complementary. Therefore, the future challenges of decision support do not reside in the optimization of the predictive value of the underlying models and algorithms, but equally in the effective communication of information and its context to doctors.

samenvatting

In de spreekkamer willen patiënten graag van hun arts weten welke diagnose het beste bij hun klachten past en welke behandelingsmogelijkheden daarbij horen. Daarnaast willen patiënten bij de keuze voor behandeling of juist het afzien van behandeling, graag vooraf weten hoe volgens de arts het beloop van de ziekte zal zijn. Met andere woorden: ze willen weten wat de prognose is. Het vinden van de juiste diagnose en het inschatten van de prognose zijn dus belangrijke voorwaarden voor het nemen van een medische beslissing.

Een klinisch voorspelmodel (predictiemodel) kan artsen helpen bij het stellen van de juiste diagnose of het inschatten van een prognose. Doordat het model gebruik maakt van diverse patiënt- en ziektegerelateerde kenmerken, geeft het een objectieve schatting van de kans op ziekte of de kans op een bepaald ziektebeloop voor een individuele patiënt. Deze individuele kans kan vervolgens door artsen gebruikt worden om een betere beslissing te nemen en daarmee de uitkomst voor de patiënt te verbeteren.

Een voorspelmodel kan alleen maar leiden tot betere medische beslissingen, wanneer de kansschattingen ook accuraat en valide zijn voor "nieuwe" patiënten. Deze voorspellende waarde van het model kan geverifieerd worden in een extern validatie-onderzoek. Maar zelfs wanneer de voorspellende waarde van een model is aangetoond, is het nog niet zeker dat dit ook daadwerkelijk de beslissingen van artsen verbetert. Dit kan onderzocht worden in zogenaamde 'impact studies', waarbij het voorspelmodel geïmplementeerd wordt in de dagelijkse praktijk en de verandering in zorg door het gebruik van het model wordt vergeleken met de reguliere zorg door artsen die het model niet gebruiken.

Hoofdstuk 2 beschrijft hoe een voorspelmodel voor misselijkheid en braken na een operatie (POMB, postoperatieve misselijkheid en braken) werd voorbereid voor implementatie op een afdeling anesthesiologie van een academisch ziekenhuis (UMC Utrecht). Dit voorspelmodel werd eerder ontwikkeld met behulp van patiëntgegevens uit een ander academisch ziekenhuis (AMC, Amsterdam) en gevalideerd in Utrecht. Vanwege enkele verschillen tussen de Amsterdamse en Utrechtse praktijk en enkele veranderingen binnen de anesthesiologie was het oorspronkelijke model niet toereikend om direct geïmplementeerd te worden. Op geleide van deze klinische verschillen en veranderingen werd het model aangepast, waardoor de voorspellende waarde van het model voor de Utrechtse praktijk verbeterde.

In **hoofdstuk 3** wordt het effect van de implementatie van het voorspelmodel voor POMB bestudeerd door middel van een grootschalig gerandomiseerd onderzoek. Het onderzoek brengt in kaart of het voorspelmodel artsen beter in staat stelt om te beslissen of hun patiënten medicatie als misselijkheidsprofylaxe moeten krijgen tijdens de ingreep. Vervolgens wordt gekeken of deze profylaxe dan ook tot minder misselijkheid en braken leidt. Voor het onderzoek werd het voorspelmodel in de vorm van zogenaamde beslissingsondersteuning geïmplementeerd – ook wel bekend onder de Engelse term 'Decision Support'. Deze beslissingsondersteuning werd vervolgens at random aangeboden aan de helft van de artsen van de afdeling anesthesiologie van het Utrechtse ziekenhuis (de interventiegroep). De beslissingsondersteuning presenteert de voorspelde kans op POMB van een individuele patiënt gedurende de ingreep aan diens behandelend arts in het elektronische verslagleggingssysteem. Het aanbieden van alleen deze kans is 'assisterend' en niet 'directief', wat betekent dat er geen bijbehorend behandeladvies wordt gepresenteerd aan de artsen. In vergelijking met de andere helft van de artsen aan wie de beslissingsondersteuning niet werd aangeboden (de controlegroep), gaven de artsen van de interventiegroep meer misselijkheidsprofylaxe aan hun patiënten, en dan met name aan patiënten met een hoog voorspeld risico. Dit positieve effect van het voorspelmodel op het risicogestuurd voorschrijfgedrag van de artsen kwam echter niet overeen met het effect op de uitkomst van patiënten. De toename in misselijkheidsprofylaxe leverde de patiënten namelijk geen duidelijk voordeel op, aangezien er geen verschil was tussen de onderzoeksgroepen in het vóórkomen van misselijkheid en braken.

De discrepantie in de resultaten van hoofdstuk 3 wordt in **hoofdstuk 4** nader onderzocht door middel van interviews en enquêtes. Een mogelijke verklaring voor deze discrepantie is dat de toename in het gebruik van misselijkheidsprofylaxe die veroorzaakt wordt door het voorspelmodel nog steeds onvoldoende is om een positief effect op de patiënt te kunnen bewerkstelligen. De toename in misselijkheidsprofylaxe is in het gerandomiseerd onderzoek beperkt, en artsen zouden nog meer misselijkheidsprofylaxe aan hoog-risico patiënten kunnen geven. Middels de interviews en enquêtes werd de artsen die deel namen aan het gerandomiseerd onderzoek gevraagd naar: hun besluitvorming omtrent misselijkheidsprofylaxe, hoe voorspelde kansen voor POMB worden geïnterpreteerd, en vervolgens hoe deze kansen hun besluitvorming beïnvloeden.

Uit de interviews en enquêtes komen vier belemmerende factoren naar voren voor het gebruik van het voorspelmodel. Allereerst blijkt dat POMB geen belangrijk aandachtspunt is voor de artsen tijdens de chirurgische procedure. Ten tweede worden beslissingen over misselijkheidsprofylaxe vooral intuïtief genomen, in plaats van dat het bewuste en analytische afwegingen zijn. Ten derde zijn voorspelde kansen niet makkelijk te gebruiken in de besluitvorming omtrent misselijkheidsprofylaxe wanneer er geen advies over behandeling aan de kansen gekoppeld is. En als laatste neemt een voorspelmodel voor POMB niet de risico's van misselijkheidsprofylaxe voor individuele patiënten en hun comorbiditeit in overweging.

Tezamen vormen deze belemmerende factoren een mogelijke verklaring voor de beperkte toename in het gebruik van misselijkheidsprofylaxe door artsen in het gerandomiseerd onderzoek. Het gebruik van een voorspelde kans in een beslissing kan lastig zijn, omdat de

kans bewust geïntegreerd moet worden met de klinische ervaring van de arts om een goede afweging van voor- en nadelen van misselijkheidsprofylaxe te kunnen maken. In een omgeving met een hoge werkdruk– zoals op een operatiekamer – kan het bewust gebruiken van een kans extra aandacht vergen van artsen, zeker wanneer de beslissing normaliter intuïtief verloopt. Betreft de beslissing dan ook nog een onderwerp dat niet een hoge prioriteit heeft voor artsen – zoals in het geval van POMB – dan is het wel te begrijpen waarom artsen de extra aandacht niet altijd opbrengen. Wanneer een voorspelmodel als beslissingsondersteuning gebruikt gaat worden, is het dus zaak om ervoor te zorgen dat de beslissingsondersteuning zo goed mogelijk aansluit bij het werkproces van de arts en daardoor zo min mogelijk extra aandacht vergt. Een advies over behandeling of vervolgonderzoek toevoegen aan de voorspelde kans – d.w.z. directieve beslissingsondersteuning – kan een eerste stap zijn voor het reduceren van de aandacht die nodig is om de beslissingsondersteuning te gebruiken.

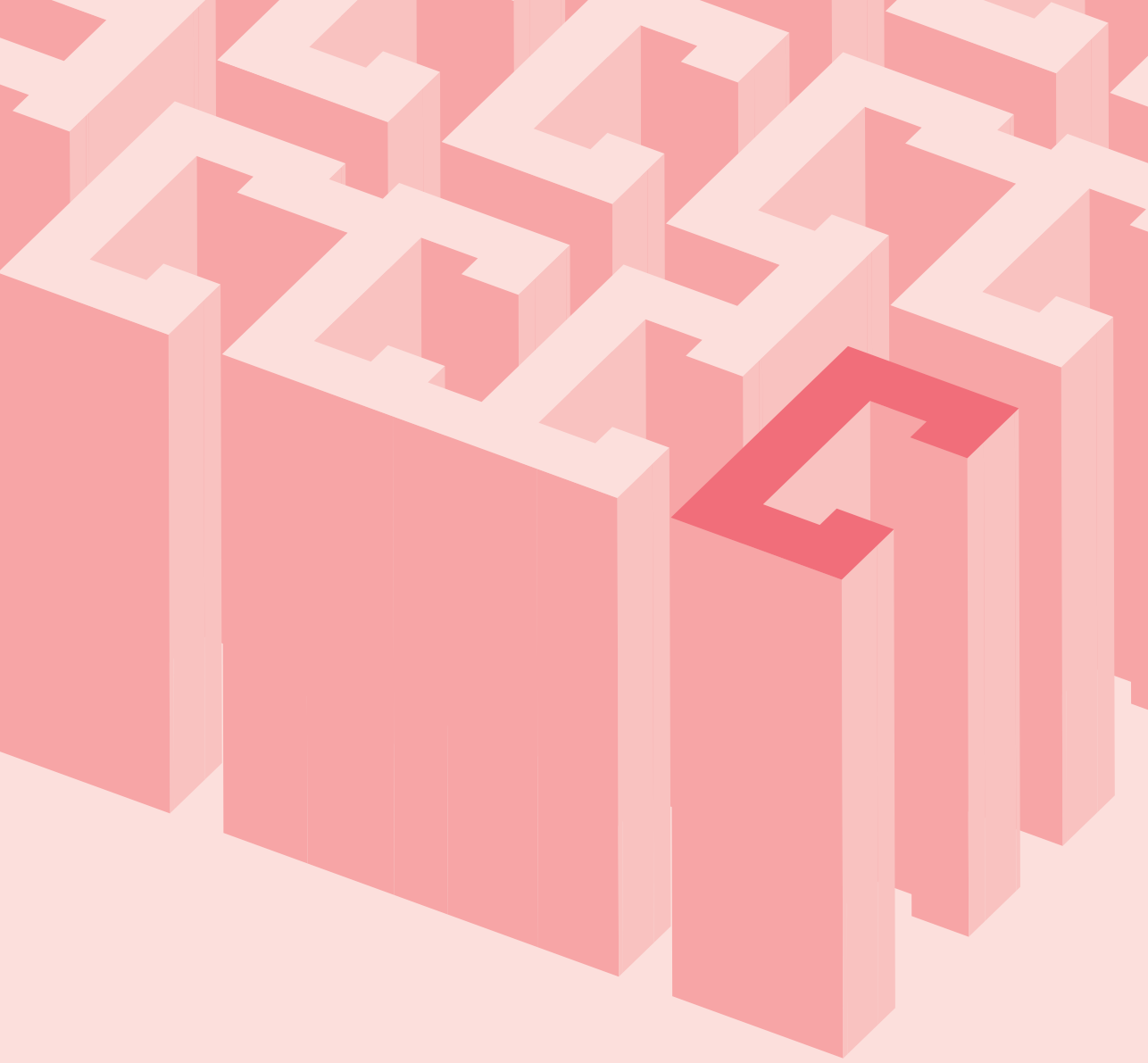
Hoofdstuk 5 beschrijft hoe het voorspelmodel voor POMB opnieuw werd geïmplementeerd in dezelfde Utrechtse anesthesiologische praktijk, maar dan in de vorm van 'directieve' beslissingsondersteuning. In dit voor-na onderzoek leidde het presenteren van de voorspelde kans op POMB met een daarbij horend advies voor het toedienen van misselijkheidsprofylaxe tot een grotere toename in de daadwerkelijke toediening van misselijkheidsprofylaxe door artsen dan in het gerandomiseerde onderzoek van hoofdstuk 3. In tegenstelling tot het eerdere gerandomiseerde onderzoek, leidde het ruimer toedienen van misselijkheidsprofylaxe in dit onderzoek wel tot een afname in het vóórkomen van POMB onder patiënten met een hoog risico op POMB. Ondanks deze afname blijft nog steeds een groot deel van de patiënten last houden van POMB. Voor een echt 'POMB-vrij' ziekenhuis zal het nog ruimhartiger toedienen van misselijkheidsprofylaxe nodig zijn.

Hoofdstuk 6 geeft een samenvatting van de ervaringen die het onderzoeksteam heeft het opgedaan met het opzetten, analyseren en interpreteren van het implementatieonderzoek naar het voorspelmodel voor POMB. Deze ervaringen kunnen een leidraad vormen voor het opzetten van andere onderzoeken naar de implementatie van voorspelmodellen (*Impact Studies*). Een randvoorwaarde voor het opzetten van een dergelijk implementatieonderzoek is te beseffen dat het voorspelmodel een complexe interventie is. Het is daarom noodzakelijk om in alle fases van het onderzoek rekening te houden met het klinische probleem waarvoor het model gebruikt gaat worden en de klinische praktijk waarin het model geïmplementeerd wordt. Het voorspelmodel behoeft vaak een update voordat het geïmplementeerd kan worden, zodat de voorspellende waarde in de klinische praktijk gewaarborgd blijft. Er moet een vorm ontworpen worden waarin het voorspelmodel daadwerkelijk als beslissingsondersteuning geïmplementeerd kan worden. Hierbij kan gedacht worden aan het toevoegen van een behandeladvies aan de kans, maar ook aan het geautomatiseerd aanbieden van de informatie in het elektronisch patiëntendossier en aan het verstrekken van achtergrondinformatie over hoe het model tot de voorspelde kans komt.

Vervolgens moet er rekening mee gehouden worden dat het implementeren van een voorspelmodel een complexe interventie is die ook onvoorziene neveneffecten kan hebben.

Artsen krijgen informatie van het voorspelmodel die zij vervolgens moeten interpreteren en gebruiken in hun beslissingen. Hoe zij dat zullen doen, is zeker niet altijd van te voren duidelijk. Vanwege de mogelijke neveneffecten is het uiteindelijke gevolg voor de patiënt niet simpelweg de optelsom van de voorspellende waarde van het model en de verandering in het gedrag van de arts. Het is daarom van belang om van te voren te overwegen of in het implementatieonderzoek het gevolg voor de patiënt moet worden onderzocht, of dat kan worden volstaan met het in kaart brengen van het effect van het voorspelmodel op de beslissingen van de artsen. Ook in de onderzoeksopzet kan rekening gehouden worden met onbedoelde en onverwachte effecten. Randomisatie op arts-niveau kan een onbedoeld leereffect voorkómen en er kunnen maatregelen getroffen worden om zogenoemde contaminatie tussen de onderzoeksgroepen te vermijden. Beiden zouden immers vertroebeling van de onderzoeksresultaten kunnen veroorzaken. Het verzamelen van extra informatie – door middel van zowel kwalitatieve als kwantitatieve methoden – kan helpen om meer inzicht te krijgen in de onbedoelde neveneffecten. Wanneer deze aspecten van te voren worden overwogen, kan een inschatting gemaakt worden van de 'levensvatbaarheid' van een onderzoek naar de implementatie van een voorspelmodel.

In **hoofdstuk 7** wordt beargumenteerd dat klinische voorspelmodellen en beslissingsondersteuning op dit moment onvoldoende rekening houden met de context van het probleem waarop ze gericht zijn. Artsen zijn in hun dagelijkse besluitvorming afhankelijk van de context waarin patiënt en probleem gepresenteerd worden. Het probleem vanuit de context benaderen is voor mensen een essentiële methode om met de complexiteit van het probleem om te gaan. Voorspelmodellen en beslissingsondersteuning gebruiken daarentegen een wiskundige benadering en relatief eenvoudige logica. Beslissingsondersteuning moet artsen helpen bij het nemen van juiste medische beslissingen. Het is dan ook van groot belang dat zij de informatie van de beslissingsondersteuning binnen de context van het probleem kunnen interpreteren. Wanneer dat niet mogelijk is, kan beslissingsondersteuning juist verwarrend werken en het beslisproces van artsen bemoeilijken. Het is aannemelijk dat door bundeling van krachten, artsen en technologie samen tot betere beslissingen kunnen komen dan één van beide op zichzelf: ze hebben immers een verschillende manier van omgaan met de complexiteit van een beslissing en zijn dus complementair. De uitdaging voor de toekomst van beslissingsondersteuning ligt daarom niet alleen in het optimaliseren van de voorspellende waarde van de onderliggende modellen en algoritmes, maar ook in het effectief communiceren van informatie en diens context naar artsen.



CREDITS

— CREW —

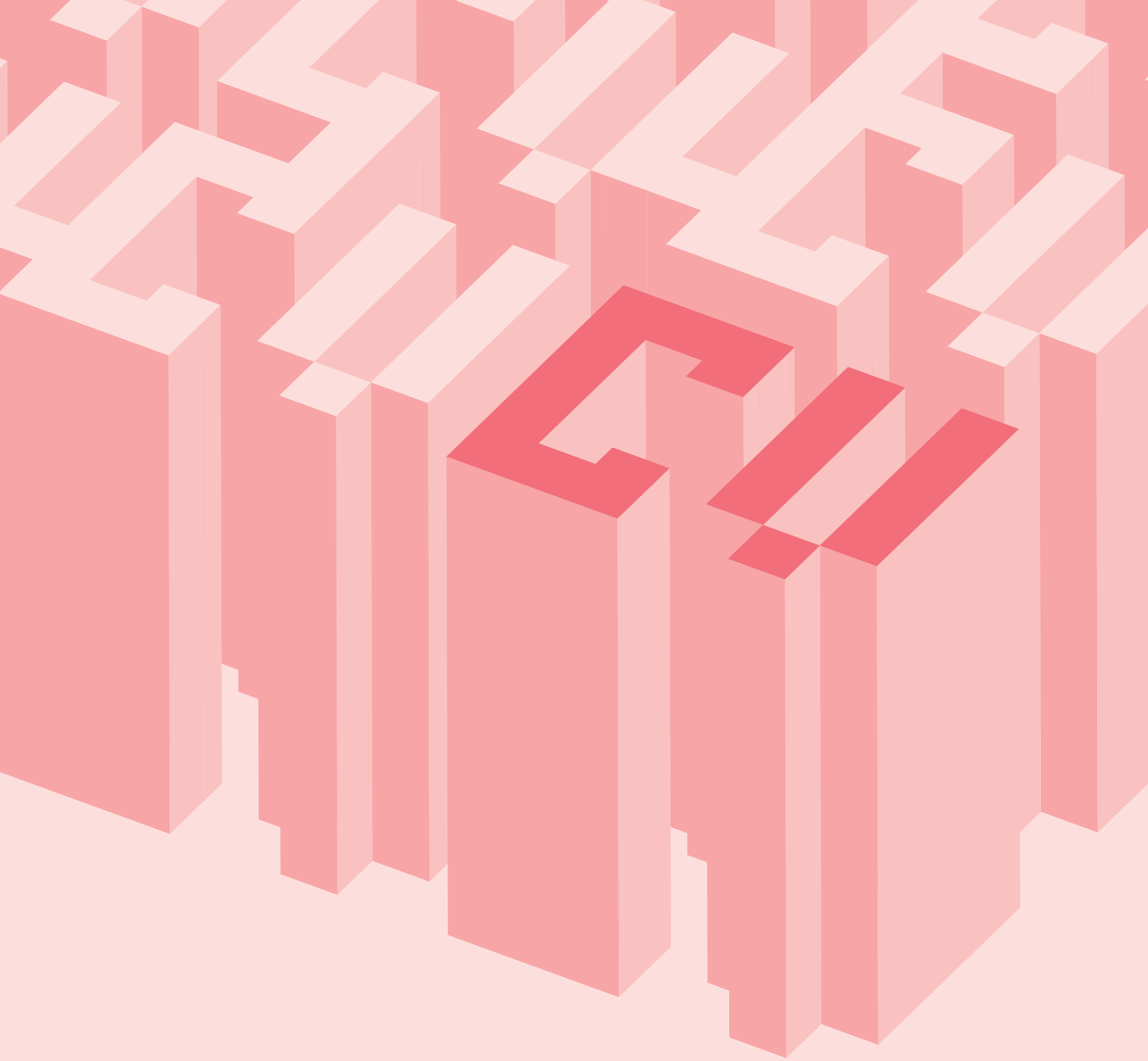
director & sound designer	COR KALKMAN
producer & director of photography	CARL MOONS
line producer	WILTON VAN KLEI
production designer	YVONNE VERGOUWE
gaffer	LEO VAN WOLFSWINKEL
public-executive producer	MIRKO NOORDEGRAAF
qualitative effects designer	KIM VAN LOON
decision makers interviewer	TINUS KAPPEN
trial management	ADA VAN KAMPEN, MARCEL BRUENS, ETIENNE SLUIS, JOANNA VELDKAMP, SELMA VAN GINKEL, MARGARET NICHOLLS, SANDRA NUMAN, HANS ALAERDS, JACQUELINE VAN DEN WIJNGAARD AND OTHERS
PACU data collection	ALL PACU NURSES OF THE UMC UTRECHT
decision makers	ALL ATTENDING ANESTHESIOLOGISTS AND RESIDENTS THAT PARTICIPATED IN THE TRIAL
assistant decision makers	ALL NURSE ANESTHETISTS OF THE UMC UTRECHT
POS data collection	ALL NURSES OF THE OUTPATIENT EVALUATION CLINIC OF THE UMC UTRECHT
first assistant director	ELINE DE GRAAFF
other assistant directors	ALL SECRETARIES TO PROF. KALKMAN AND TO PROF. MOONS
scientific advisory board	LINDA PEELEN MARIJE OSNABRUGGE JUDY VELDHUIJZEN THE JULIUS' PREDICTION CLUB
script consultant	SJOERD ROMBOUT
casting director	HANS KNAPE
assistant casting director	REINIER HOFF
costume designers for Mr. Kappen	MARJOS VERBEEK ANNE-METTE SAUËR SUZANNE FLIER VESNA SVIRČEVIĆ
the digital four-colored pen	ANSTAT

— CAST —

the girlfriend	OLIVIA LIEM
the parents	MARGREET KAPPEN
	TINUS KAPPEN
the brother	SANDER KAPPEN
the sister-in-law	NYNKE VAN BERKUM
the nephew	IMME KAPPEN
the paranymphs	VESNA SVIRČEVIĆ
	KIM VAN LOON
the parsecians	BART ROEFFEN
	MICHIEL KREMERS
	NIEK PENNINGS
	SANDER KAPPEN
	SJOERD ROMBOUT
the associate parsecians	MANON DERIKX
	ROY DERIKX
	(AND THEIR FIVE CHILDREN)
the procrastination team	ANNE-METTE, AUKJE, BAS, BERNARD, EMILY, JILLES, JUDITH, KIM, LEON, LIDWIEN, MARIANNE, MARLOES, MAURICE, MIENKE, PATTY, STEFAN, SÜLEYMAN, SUZANNE, THOMAS, AND VESNA
special thanks to	ALL RESIDENTS, NURSES AND STAFF OF THE ANESTHESIOLOGY DEPARTMENT OF THE UMC UTRECHT
	ALL RESIDENTS, NURSES AND STAFF OF THE PEDIATRIC INTENSIVE CARE OF THE WILHELMINA CHILDREN'S HOSPITAL

to all my friends and family whom I neglected over the past years

no data were harmed during the making of this thesis
no PhD students were (seriously) harmed during the making of this thesis



CURRICULUM VITAE

Teus Kappen was born on July 19, 1981 in the former Saint Cunera chapel in the small community of Kaathoven, Berlicum. He grew up in the village of Nuland and graduated from high school at the Rodenborch College, Rosmalen in 1999. He entered medical school at Utrecht University in the year 1999–2000.

During his medical study he went to Cape Town, South Africa for his clinical rotation in pediatrics at Tygerberg Hospital, University of Stellenbosch. His interests in clinical research started during an elective research project on medically unexplained symptoms at the Netherlands Institute for Health Services Research (NIVEL) under supervision of prof. dr. S. van Dulmen. In 2003 he did a short elective at the anesthesiology department of the University Medical Center Utrecht (UMC Utrecht). Intrigued by the field of anesthesiology, he did his final-year electives of medical school (2004–2005) at the pediatric intensive care unit of the Wilhelmina Children's Hospital (WKZ) and at the anesthesiology department of the UMC Utrecht. He combined his final anesthesiology elective with a research project on intraoperative hypertension and baseline blood pressure under supervision of prof. dr. C.J. Kalkman and dr. J.B. Bijker.

After his graduation from medical school in 2005, he worked as a senior house officer at the pediatric intensive care unit of the WKZ. In January 2006 he began his residency in anesthesiology at the anesthesiology department of the UMC Utrecht under supervision of prof. dr. J.T.A. Knape. Simultaneously, he started his Ph.D. training and research project as described in this thesis (promotores: prof. dr. C.J. Kalkman, prof. dr. K.G.M. Moons, prof. dr. W.A. van Klei; copromotor: dr. Y. Vergouwe). He obtained his Master of Science degree in Clinical Epidemiology at Utrecht University in 2009. Just before his registration as an anesthesiologist in January 2013, he started a fellowship in pediatric intensive care at the WKZ in October 2012 under supervision of dr. N.J.G. Jansen and drs. E. Koomen. After the completion of his fellowship in July 2015, he returned to the anesthesiology department of the UMC Utrecht to work as an attending anesthesiologist for both adults and children.

list of publications

this thesis

1. Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of Clinical Prediction Models for Application in Local Settings. *Med Decis Making*. 2012;32(3):E1-E10. PMID:22427369.

2. Kappen TH, Moons KGM, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of Risk Assessments on Prophylactic Antiemetic Prescription and the Incidence of Postoperative Nausea and Vomiting: a Cluster-Randomized trial. *Anesthesiology*. 2014;120(2):343-354. PMID:24105403.

3. Kappen TH, Vergouwe Y, Wolfswinkel L Van, Kalkman CJ, Moons KGM, Klei WA Van. Impact of Adding Therapeutic Recommendations to Risk Assessments from a Prediction Model for Postoperative Nausea and Vomiting. *Br J Anaesth*. 2015;114(2):252-260. PMID:25274048.

4. Kappen TH, van Loon K, Kappen MAM, et al. How Clinical Prediction Models Impact Decision Making by Physicians: an Example Study. *Under revision at Journal of Clinical Epidemiology*.

other

5. Bijker JB, van Klei WA, Kappen TH, van Wolfswinkel L, Moons KGM, Kalkman CJ. Incidence of Intraoperative Hypotension as a Function of the Chosen Definition. *Anesthesiology*. 2007;107(2):213-220. PMID:17667564.

6. Kappen T, van Dulmen S. General Practitioners' Responses to the Initial Presentation of Medically Unexplained Symptoms: a Quantitative Analysis. *Biopsychosoc Med*. 2008;2(1):22. PMID:19014616.

7. Ahmadian L, Cornet R, Kalkman C, de Keizer NF. Development of a National Core Dataset for Preoperative Assessment. *Methods Inf Med*. 2009;48(2):155-161. PMID:19283313.

8. Flier S, Post J, Concepcion AN, Kappen TH, Kalkman CJ, Buhre WF. Influence of Propofol-Opioid vs Isoflurane-Opioid Anaesthesia on Postoperative Troponin Release in Patients Undergoing Coronary Artery Bypass Grafting. *Br J Anaesth*. 2010;105(2):122-130. PMID:20573633.

9. Van Dijk JFM, van Wijck AJM, Kappen TH, Peelen LM, Kalkman CJ, Schuurmans MJ. Postoperative pain assessment based on numeric ratings is not the same for patients and professionals: A cross-sectional study. *Int J Nurs Stud*. 2012;49(1):65-71. PMID:21840522.

10. Van Klei WA, Hoff RG, van Aarnhem EEHL, et al. Effects of the Introduction of the WHO "Surgical Safety Checklist" on In-Hospital Mortality. *Ann Surg*. 2012;255(1):44-49. PMID:22123159.

11. Luykx JJ, Bakker SC, Lentjes E, et al. Season of Sampling and Season of Birth Influence Serotonin Metabolite Levels in Human Cerebrospinal Fluid. Tosini G, ed. *PLoS One*. 2012;7(2):e30497. PMID:22312427.

12. Luykx JJ, Vinkers CH, Bakker SC, et al. A common variant in ERBB4 regulates GABA concentrations in human cerebrospinal fluid. *Neuropsychopharmacology.* 2012;37(9):2088-2092. PMID:22549119.

13. Van Dijk JF, Kappen TH, van Wijck AJ, Kalkman CJ, Schuurmans MJ. The diagnostic value of the numeric pain rating scale in older postoperative patients. *J Clin Nurs.* 2012;21(21-22):3018-3024. PMID:23083383.

14. Luykx JJ, Bakker SC, Lentjes E, et al. Genome-wide association study of monoamine metabolite levels in human cerebrospinal fluid. *Mol Psychiatry.* 2013;19(2):228-234. PMID:23319000.

15. Bouwmeester W, Twisk JW, Kappen TH, van Klei WA, Moons KG, Vergouwe Y. Prediction models for clustered data: comparison of a random intercept and standard regression model. *BMC Med Res Methodol.* 2013;13(1):19. PMID:23414436.

16. Bouwmeester W, Moons KGM, Kappen TH, et al. Internal Validation of Risk Models in Clustered Data: A Comparison of Bootstrap Schemes. *Am J Epidemiol.* 2013;177(11):1209-1217. PMID:23660796.

17. De Graaff JC, Bijker JB, Kappen TH, van Wolfswinkel L, Zuithoff NPA, Kalkman CJ. Incidence of Intraoperative Hypoxemia in Children in Relation to Age. *Anesth Analg.* 2013;117(1):169-175. PMID:23687233.

18. Van der Kooi AW, Kappen TH, Raijmakers RJ, Zaal IJ, Slooter AJC. Temperature variability during delirium in ICU patients: an observational study. *PLoS One.* 2013;8(10):e78923. PMID:24194955.

19. Gerbershagen HJ, Pogatzki-Zahn E, Aduckathil S, et al. Procedure-specific risk factor analysis for the development of severe postoperative pain. *Anesthesiology.* 2014;120(5):1237-1245. PMID:24356102.

20. Van Dijk JFM, Kappen TH, Schuurmans MJ, van Wijck AJM. The Relation Between Patients' NRS Pain Scores and Their Desire for Additional Opioids after Surgery. *Pain Pract.* 2014:n/a - n/a. PMID:24735082.

21. Van Dijk JFM, van Wijck AJM, Kappen TH, Peelen LM, Kalkman CJ, Schuurmans MJ. The Effect of a Preoperative Educational Film on Patients' Postoperative Pain in Relation to their Request for Opioids. *Pain Manag Nurs.* 2014. PMID:25246325.

22. Flier S, Concepcion AN, Versteeg D, et al. Monocyte hyporesponsiveness and Toll-like receptor expression profiles in coronary artery bypass grafting and its clinical implications for postoperative inflammatory response and pneumonia. *Eur J Anaesthesiol.* 2015;32(3):177-188. PMID:25405275.

23. Marcus H, Gerbershagen HJ, Peelen LM, et al. Quality of pain treatment after caesarean section: Results of a multicentre cohort study. *Eur J Pain.* 2014. PMID:25413847.

24. Luykx JJ, Bakker SC, Visser WF, et al. Genome-wide association study of NMDA receptor coagonists in human cerebrospinal fluid and plasma. *Mol Psychiatry.* 2015. PMID:25666758.

25. Kalkman CJ, Kappen TH. Patient-centered Endpoints for Perioperative Outcomes Research. *Anesthesiology.* 2015;122(3):481-483. PMID:25689756.