Developing Risk Prediction Models for Postoperative Pancreatic Fistula: a Systematic Review of Methodology and Reporting Quality

Zhang Wen¹ · Ya Guo¹ · Banghao Xu¹ · Kaiyin Xiao¹ ·
Tao Peng¹ · Minhao Peng¹

Received: 19 July 2015 / Accepted: 28 December 2015 / Published online: 23 January 2016
© Association of Surgeons of India 2016

Abstract Postoperative pancreatic fistula is still a major complication after pancreatic surgery, despite improvements of surgical technique and perioperative management. We sought to systematically review and critically assess the conduct and reporting of methods used to develop risk prediction models for predicting postoperative pancreatic fistula. We conducted a systematic search of PubMed and EMBASE databases to identify articles published before January 1, 2015, which described the development of models to predict the risk of postoperative pancreatic fistula. We extracted information of developing a prediction model including study design, sample size and number of events, definition of postoperative pancreatic fistula, risk predictor selection, missing data, model-building strategies, and model performance. Seven studies of developing seven risk prediction models were included. In three studies (42 %), the number of events per variable was less than 10. The number of candidate risk predictors ranged from 9 to 32. Five studies (71 %) reported using univariate screening, which was not recommended in building a multivariate model, to reduce the number of risk predictors. Six risk prediction models (86 %) were developed by categorizing all continuous risk predictors. The treatment and handling of missing data were not mentioned in all studies. We found use of inappropriate methods that could endanger the development of model, including univariate pre-screening of variables, categorization of continuous risk predictors, and model

validation. The use of inappropriate methods affects the reliability and the accuracy of the probability estimates of predicting postoperative pancreatic fistula.

Keywords Prediction model · Postoperative pancreatic fistula · Systematic review · Methodology

Background

Both improvements in surgical and perioperative management have reduced surgical morbidity and mortality after resective pancreatic surgery in high-volume centers. However, postoperative pancreatic fistula (POPF) is still regarded as the most relevant complication after resective pancreatic surgery (distal pancreatectomy and pancreatoduodenectomy) because it potentially leads to deleterious secondary complications, increased health care costs, and prolonged hospital stay [1, 2].

More studies using the strict definition applied by the International Study Group of Pancreatic Surgery (ISGPS) [3] report that POPF rates vary between 3 % after pancreatic head resection [2, 4] and up to 30 % following distal pancreatectomy [5–7]. Many risk factors for POPF have been identified, such as gender [8], body mass index (BMI) [9, 10], status of the pancreatic parenchyma [11], diameter of the main pancreatic duct (MPD) [12], and kind of disease [13]. These risk factors lead to use some special techniques or perioperative precautions to minimize POPF rates [5, 14, 15]. However, it remains difficult to integrate various risk factors and make an accurate prediction of POPF. Recently, multiple studies have constructed perioperative scoring systems to predict POPF. Ideally, a predictive model should be based on easily available perioperative parameters allowing the surgeon to adopt strategies on individual basis.

Zhang Wen and Ya Guo contributed equally to this work.

✉ Minhao Peng
minhaopeng@163.com

¹ Department of Hepato-Biliary-Pancreatic Surgery and Liver Transplantation, the First Affiliated Hospital, Guangxi Medical University, Nanning 530021, China

The aim of this study was to review the perioperative predictive model for POPF based on the methodology and reporting quality and to inform and prompt further improvements in model building.

Methods

Search Strategy

We attempted to identify all observational studies that developed the predictive scoring system for postoperative pancreatic fistula, which were published in the Pubmed and EMBASE databases before January 1, 2015.

The search strategy for Pubmed was “postoperative pancreatic fistula” and (“risk prediction model” or “predictive model” or “predictive equation” or “prediction model” or “risk calculator” or “prediction rule” or “risk model” or “risk factor” or “scoring system” or “statistical model” or “Cox model” or “multivariable”) not (review [publication type] or bibliography [publication type] or editorial [publication type] or letter [publication type] or meta-analysis [publication type] or news [publication type]).

The search strategy for EMBASE was risk prediction model or risk prediction model or predictive model or predictive model or predictive equation or predictive equation or prediction model or prediction model or risk calculator or risk calculator or prediction rule or prediction rule or risk model or risk factor or scoring system or statistical model or Cox model or multivariable and postoperative pancreatic fistula not letter not review not editorial not conference not book.

Additionally, the references of the primary and review articles were examined to identify publications not retrieved by electronic searches. Finally, we attempted to identify any imminent or unpublished material relevant to this topic using the clinical trial search and clinical trials. Titles and abstracts of all citations were screened independently by two reviewers. Discrepancies between reviewers’ opinions were resolved by a third reviewer. Articles were restricted to the English-language literatures.

Inclusion Criteria

A study was included in the systematic review if it provided the perioperative predictive scoring system or model for postoperative pancreatic fistula. Postoperative pancreatic fistula was diagnosed according to the International Study Group on Pancreatic Fistula (ISGPF) definition.

Exclusion Criteria

Articles were excluded if (1) they included only validation of a preexisting risk prediction model (that is, the article did not

develop a model), (2) participants were children, and (3) the authors developed a genetic risk prediction model.

Data Extraction, Analysis, and Reporting

Data were extracted by two reviewers and checked by a third. Data items extracted for this article include study design, sample size and number of events, outcome definition, risk predictor selection and coding, missing data, model-building strategies, and aspects of performance. The data extraction form for this article was based largely on two previous reviews of prognostic models of cancer and can be obtained on request from the first author. Any discrepancy was resolved by discussion with a fourth reviewer and reanalysis of publication.

We reported our systematic review in according with the PRISMA guideline [16], with the exception of items relating to meta-analysis, as our study includes no formal meta-analysis.

Results

The search string retrieved 199 articles in PubMed and 42 articles in EMBASE, and, after removing duplicates, our database search yields 216 articles (see Fig. 1). Seven articles met our inclusion criteria. Totally, seven studies were eligible for reviewing, published between September 2009 and October 2011 (Table 1) [8, 10–13, 17, 18]. Postoperative pancreatic fistula predicted by the models was defined by the International Study Group for Pancreatic Fistula. One studies described the development of a predictive model for POPF, based on perioperative risk factors [10]. Three studies derived a predictive model for POPF based on preoperative risk factors [8, 13, 17]. Intraoperative risk factors were used by one study to derive a predictive model of POPF [12]. Two studies built a predictive model for POPF in accordance with postoperative risk factors [11, 18].

In terms of geography, three articles were from Asia [8, 17, 18] and four articles were from Europe [10–13].

Number of Patients and Postoperative Pancreatic Fistula

The number of participants included in developing risk prediction model was clearly reported in all articles. The median number of participants included in model development was 244 (interquartile range (IQR) 50–387). The median number of events about postoperative pancreatic fistula was 69 (IQR 25–197). Five studies reported the number of events on postoperative pancreatic fistula grades based on ISGPF [5, 8, 11, 12, 18], the medium number of events on grade A 47 (IQR 6 to 76), grade B 17 (IQR 10 to 94), and grade C 8 (IQR 5 to 25). Two studies did not report the number of events on postoperative pancreatic fistula grades [10, 17].

Fig. 1 Selection of articles for review

Number of Risk Predictors

A median of 11 risk predictors (IQR 10 to 16, range 9 to 32) were considered candidate risk predictors. The rational or references for including risk predictors were in seven studies [8, 10–13, 17, 18]. The final reported prediction models included a median of four risk predictors (IQR 3 to 5, range 3 to 10). In total, 18 different risk predictors were included in the final risk prediction models (see Fig. 2). The most commonly identified risk predictors included in the final risk predictor model were fatty pancreas ($n=3$), pancreatitis ($n=2$), MPD

index ($n=2$), pancreatic fibrosis ($n=2$), pancreatic duct size ($n=2$), and diagnosis of pancreatic cancer ($n=2$). Other 12 risk predictors appeared only once in the final risk prediction model.

Sample Size

Three risk prediction models (43 %) were developed in which the number of events per variable was <10 [10, 12, 13]. Overall, the medium number of events per variable was 10 (IQR 7 to 39, range 5 to 175).

Table 1 Models for predicting risks of postoperative pancreatic fistula

Study	Year	Country	Definition of postoperative pancreatic fistula as reported	Risk predictors in the model
Belyaev O et al [12]	2011	Germany	POPF was defined by ISGPF	Pancreatic duct size, MPD index, periductal fibrosis, interlobular fibrosis, intralobular fibrosis, interlobular fat, intralobular fat, signs of chronic pancreatitis, signs of acute pancreatitis, and tissue edema
Yamamoto Y. et al [8]	2011	Japan	POPF was defined by ISGPF	MPD index, relation to portal vein on CT, gender, intraabdominal fat thickness, diagnosis of pancreatic cancer or disease other than pancreatic cancer
Braga M et al [11]	2011	Italy	POPF was defined by ISGPF	Pancreatic texture, pancreatic duct diameter, operative blood loss, and ASA score
Lee SE et al [17]	2010	Korea	POPF was defined by ISGPF	RSID, intralobular fat, interlobular fat, and total fat
Gaujoux S et al [10]	2010	France	POPF was defined by ISGPF	BMI, fatty pancreas, and pancreatic fibrosis
Wellner UF et al [13]	2010	Germany	POPF was defined by ISGPF	Age, preoperative diagnosis other than pancreatic carcinoma or chronic pancreatitis, history of smoking, history of weight loss, and history of acute pancreatitis
Kawai M et al [18]	2009	Japan	POPF was defined by ISGPF	Leukocyte counts on POD4, increasing amylase level of drainage fluid, and amylase level of drainage fluid on POD1

EBL estimated blood loss, *MPD index* ratio between the size of the main pancreatic duct and the short axis of pancreatic body at the resection margin, *RSID* relative signal intensity decreases, *POD4* postoperative day 4, *POD1* postoperative day1, *BMI* body mass index

Fig. 2 Frequency of identified risk predictors in the final prediction models

* Other risk predictors appearing once in the final model: (1) tissue edema, (2) relation to portal vein on CT, (3) gender, (4) intra-abdominal fat thickness, (5) diagnosis of disease other than pancreatic cancer, (6) ASA score, (7) RSID, (8) BMI, (9) age, (10) Leukocyte counts on POD4, (11) amylase level of drainage fluid on POD1, (12) estimated blood loss
Abbreviation: MPD=main pancreatic duct;

Treatment of Continuous Risk Predictors

One prediction model (14 %) was developed retaining a continuous risk predictor as continuous [17], and six risk prediction models (86 %) dichotomized or categorized all continuous risk predictors [8, 10–13, 18].

Missing Data

All seven studies made no mention of missing data (Table 2); thus, it can only be assumed that a complete case analysis was conducted or that all data for all risk predictions (including candidate risk predictors) were available.

Model Building

Five studies (71 %) reported using univariate screening to reduce the number of risk predictors [8, 10, 11, 13, 18], while it was unclear how the risk predictors were reduced prior to development of the risk predictor model in two studies (29 %) [12, 17]. Two studies (29 %) included all risk predictors in the multivariable analysis [11, 17].

Two studies (29 %) reported using automatic variable selection procedures to derive the final multivariable model [8, 13] (Table 2). One study reported using stepwise backward

elimination [13]. One study reported using forward stepwise selection model [8].

All seven studies clearly demonstrated the type of model they used to derive the prediction model. The final models were based on logistic regression in four articles [8, 10, 13, 18], linear regression in one article [17], and univariate analysis in one article [11].

Table 2 Issues in model of development

Variable	Data
Sample size, median (IQR)	
Development cohort	100 (50 to 279)
Validation cohort	108 (50 to 231)
Treatment of continuous risk predictors, <i>n</i> (%)	
All kept continuous	1 (14 %)
All categorized/dichotomized	6 (86 %)
Treatment of missing data	
Not mentioned	7 (100 %)
Model-building strategy, <i>n</i> (%)	
Stepwise, forward selection, and backward elimination	2 (29 %)
All significant in multivariable analysis	2 (29 %)
Other	3 (42 %)
Overfitting mentioned or discussed, <i>n</i> (%)	1 (14 %)

IQR interquartile range

Validation

Two studies (29 %) split cohort into development and validation cohorts [8, 13]. All studies conducted and published an internal validation of their risk prediction models within the same article and used two or more data sets in an attempt to demonstrate the internal validity of the risk prediction model [8, 10–13, 17, 18].

Model Performance

The type of performance measure was assessed in the risk prediction models (Table 3). Six studies (86 %) reported C-statistics, calculating C-statistics on internal validation data sets [8, 10–12, 17, 18]. Spearman's rank correlation in one study was used to validate the risk predictor model [13]. Only two studies (29 %) assessed how well the predicted risks compared to the observed risks (calibration); investigators in these two studies chose to calculate the Hosmer-Lemeshow test [10, 11].

Model Presentation

Five studies (62.5 %) derived simplified scoring systems from the risk predictor models [8, 10–13]. The combination of serum albumin level and leukocyte counts on postoperative day

4 was used as the risk predictor model in one study [18]. One study used preoperative measurements of pancreatic fat by magnetic resonance as the risk predictor model [17].

Discussion

Main Findings

In the present article, we have highlighted the methods currently used to develop risk prediction models for POPF and poor reporting of those methods. The quality of risk prediction models depends on study design and statistical methods. Also, developed models need to provide accurate and validated estimates of probabilities for POPF. Pancreatic surgeons should understand the principles of statistical methods, choose the appropriated risk prediction model, and subsequently improve individual outcomes and the cost-effective care.

When developing a risk prediction model, one of the problems researcher meet is overfitting. Overfitting generally occurs when a model is excessively complex, such as having too many candidate risk predictors relative to the number of events. A rule of thumb is that models should be developed with 10–20 events per variable (EPV) [19]. Of the studies included in this review, 43 % had fewer than 10 EPV. A risk prediction model which has been overfit will generally have poor predictive performance, as it would exaggerate minor fluctuation in the data [20]. Other investigators have reported similar findings (EPV < 10) when appraising the development of multivariable risk prediction models [21, 22].

Another key component affecting the performance of the final risk prediction model is how continuous variables are treated, whether they are kept as continuous measurements or whether they have been categorized into two or more categories [23]. Common approaches include dichotomizing at the median value or choosing an optimal cutoff point based on minimizing a *P* value. Regardless of the approach used, the practice of artificially treating a continuous risk predictor as categorical should be avoided [23, 24]; yet, this is often done in the development of risk prediction models [22, 24–27]. In this review, 86 % of studies were identified that all or some continuous risk predictors were dichotomized or categorized. Dichotomizing continuous variables causes an inevitable loss of information and power to detect real relationship, equivalent to losing a third of the data. If the predictor is exponentially distributed, the loss associated with dichotomization at the median is even larger [28]. Continuous risk predictors (like age) should be retained in the risk prediction model as continuous variables, and the risk predictor has a nonlinear relationship with the outcome; then, the use of splines or fractional polynomial functions is recommended [23].

Missing data, which is a serious problem in studies deriving a risk prediction model, is a potential source of bias when

Table 3 Evaluating performance of risk predictor models

Parameter	Number of studies (%)
Validation	
Apparent	4 (57 %)
Internal	
Bootstrapping	0
Jack-knifing	0
Random split sample	3 (43 %)
Cross validation	0
Temporal	0
External	0
Performance metrics	
Discrimination	
C-statistic	6 (86 %)
Spearman's rank correlation	1 (14 %)
Calibration	2 (29 %)
Hosmer-Lemeshow statistic	2 (29 %)
Calibration plot	0
Classification	
Reclassification (NRI)	0
Other (for example, sensitivity and specificity)	6 (86 %)

NRI net reclassification index

analyzing clinical data sets. There are many possible reasons for missing data (e.g., patient refusal to continue in the study, treatment failures or successes, adverse events, and patients moving). Regardless of study design, collecting all data on all risk predictors for all individuals is a difficult task [29]. There can be no universally methodological approach for handling missing data. A common approach is to exclude individuals with missing values and conduct a complete case analysis. However, a complete case analysis, in addition to sacrificing and discarding useful information, is not recommended as it has been shown that it can yield biased results [30–32]. There is no missing data reported in all studies involved in our review. It is advised that the completeness of data should be reported so the reader can judge the quality of the data.

In the development of risk prediction model, many efforts have been made to use multivariable analysis in accessing the importance of risk factors on outcome. To identify the most significant risk factors associated with an outcome, bivariate analysis is frequently performed. In bivariate selection, if the statistical p value of a risk factor in bivariate analysis is greater than an arbitrary value (often $p=0.05$), then this factor will not be allowed to compete for inclusion in multivariable analysis. However, it is not simple to select predictors only based on statistical significance during model development, as it is crucial to retain those risk predictors known to be important from literatures, but which may not reach statistical significance in a data set. This has been shown to be inappropriate, as it can wrongly reject potentially important variables when the relationship between an outcome and a risk factor is confounded by any confounder and when this confounder is not properly controlled, thus leading to an unreliable model [30, 33]. Five studies in this review reduced the initial number of candidate risk predictor prior to the final model; however, two studies failed to provide sufficient detail on how the procedure was carried out.

Automated variable selection methods (forward selection, backward elimination, or stepwise) are frequently used to derive the final risk prediction models (29 % in our review). However, automated selection methods are data-driven approaches based on statistical significance without literature to clinical relevance, and it has been shown that these methods are unstable and not reproducible, have biased estimates of regression coefficient, and yield poor predictions [29, 34, 35].

After a risk prediction model has been derived, it is essential to access the performance of the model. To address this, several approaches have been suggested to estimate a risk prediction model's optimism: (1) internal validation using bootstrapping, cross validation, and split-sampling techniques [36] and (2) temporal validation to evaluate the performance of a model on subsequent patients from the same center(s). Temporal validation is no different in principle from splitting a single data set by time; however, temporal validation is a prospective evaluation of a model. Temporal validation can

be considered external in time and thus intermediate between internal validation and external validation [37], and (3) external validation aims to address the accuracy of a risk prediction model in patients from different centers or locations [38]. Investigators in all of the studies in our review reported an internal validation on cohorts.

Few prediction models are routinely used in clinical practice, probably because most have not been externally validated [39]. A risk score for prediction model should be clinically credible, accurate (well calibrated with good discriminative ability), have generality (be externally validated), and ideally be shown to be clinically effective, that is, provide useful additional information to clinicians that improves therapeutic decision making and thus patient outcome [39]. It is crucial to quantify the performance of a prognostic model on a new series of patients, ideally in a different location, before applying the model in daily practice to guide patient care. Although still rare, temporal and external validation studies do seem to be becoming more common. We have observed poor reporting in all aspects of developing the risk prediction models based on the data and detail provided in all steps in building the model.

Definition of POPF is also important for developing a risk predictor model of POPF. The classification system of ISGPF was used to definite POPF in all seven studies. However, because of its retrospective character, the ISGPF may have limitations in clinical decision making. To improve clinical decision making about management of patients, there is a need that the ISGPF classification system is merged with newer clinical data.

An ideal prediction model of pancreatic fistula should include objective characteristics, which are identified preoperatively, intraoperatively, or postoperatively and be easily investigated. These characteristics include BMI, acoustic radiation force impulse, signal intensity of pancreas on magnetic resonance, the histological assessment of pancreatic steatosis and fibrosis, pancreatic duct width, inflammatory cytokines, and chemokines. The model should also be used simply and quickly to assure the surgeon intraoperative decision of performing an anastomosis in individual critical cases and validated by prospective multicenter clinical trials. An individual assessment of risk for postoperative pancreatic fistula enhances preoperative counseling and patient selection for pancreatic surgery. Furthermore, the model may permit a change to established clinical practice and strategies to decrease pancreatic fistula among those patients at high risk. Thus, further larger standardized studies are required.

Despite the systematic search and inclusion of the most recent publications, this systematic review was limited to English-language articles and did not consider grey literatures. Therefore, we may have missed some studies.

In conclusions, we found that published risk prediction models for PF were often characterized by both use of

inappropriate methods for development of multivariable models and poor reporting. Additionally, these models were limited by the lack of studies based on prospective data set of sufficient sample size to avoid overfitting. There is an urgent need for investigators to use appropriate method and to report effectively. An appropriate risk prediction model would guide the surgeons to predict the objective probability of POPF, treat POPF promptly, and also benefit patients.

Acknowledgments The study was supported by the National Natural Science Foundation of China (Grant No. 81560387).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Pratt WB, Maithel SK, Vanounou T, Huang ZS, Callery MP, Vollmer CM Jr (2007) Clinical and economic validation of the international study group of pancreatic fistula (ISGPF) classification scheme. *Ann Surg* 245(3):443–451
- Reid-Lombardo KM, Farnell MB, Crippa S, Barnett M, Maupin G, Bassi C et al (2007) Pancreatic anastomotic leakage after pancreaticoduodenectomy in 1,507 patients: a report from the pancreatic anastomotic leak study group. *J Gastrointest Surg* 11(11):1451–1458, discussion 59
- Bassi C, Dervenis C, Butturini G, Fingerhut A, Yeo C, Izbicki J et al (2005) Postoperative pancreatic fistula: an international study group (ISGPF) definition. *Surgery* 138(1):8–13
- Buchler MW, Friess H, Wagner M, Kulli C, Wagnener V, Z'Graggen K (2000) Pancreatic fistula after pancreatic head resection. *Br J Surg* 87(7):883–889
- Wellner U, Makowiec F, Fischer E, Hopt UT, Keck T (2009) Reduced postoperative pancreatic fistula rate after pancreatogastrostomy versus pancreaticojejunostomy. *J Gastrointest Surg* 13(4):745–751
- Ferrone CR, Warshaw AL, Rattner DW, Berger D, Zheng H, Rawal B et al (2008) Pancreatic fistula rates after 462 distal pancreatectomies: staplers do not decrease fistula rates. *J Gastrointest Surg* 12(10):1691–1697, Discussion 97–8
- Goh BK, Tan YM, Chung YF, Cheow PC, Ong HS, Chan WH et al (2008) Critical appraisal of 232 consecutive distal pancreatectomies with emphasis on risk factors, outcome, and management of the postoperative pancreatic fistula: a 21-year experience at a single institution. *Arch Surg* 143(10):956–965
- Yamamoto Y, Sakamoto Y, Nara S, Esaki M, Shimada K, Kosuge T (2011) A preoperative predictive scoring system for postoperative pancreatic fistula after pancreaticoduodenectomy. *World J Surg* 35(12):2747–2755
- Hubbard TJ, Lawson-McLean A, Fearon KC (2011) Nutritional predictors of postoperative outcome in pancreatic cancer. *Br J Surg* 98:268–274, *Br J Surg* 2011;98(7):1032; author reply 32–3
- Gaujoux S, Cortes A, Couvelard A, Noullet S, Clavel L, Rebours V et al (2010) Fatty pancreas and increased body mass index are risk factors of pancreatic fistula after pancreaticoduodenectomy. *Surgery* 148(1):15–23
- Braga M, Capretti G, Pecorelli N, Balzano G, Doglioni C, Ariotti R et al (2011) A prognostic score to predict major complications after pancreaticoduodenectomy. *Ann Surg* 254(5):702–707, Discussion 07–8
- Belyaev O, Munding J, Herzog T, Suelberg D, Tannapfel A, Schmidt WE et al (2011) Histomorphological features of the pancreatic remnant as independent risk factors for postoperative pancreatic fistula: a matched-pairs analysis. *Pancreatology* 11(5):516–524
- Wellner UF, Kayser G, Lapshyn H, Sick O, Makowiec F, Hoppner J et al (2010) A simple scoring system based on clinical factors related to pancreatic texture predicts postoperative pancreatic fistula preoperatively. *HPB Oxford* 12(10):696–702
- Konstadoulakis MM, Filippakis GM, Lagoudianakis E, Antonakis PT, Dervenis C, Bramis J (2005) Intra-arterial bolus octreotide administration during Whipple procedure in patients with fragile pancreas: a novel technique for safer pancreaticojejunostomy. *J Surg Oncol* 89(4):268–272
- Alghamdi AA, Jawas AM, Hart RS (2007) Use of octreotide for the prevention of pancreatic fistula after elective pancreatic surgery: a systematic review and meta-analysis. *Can J Surg* 50(6):459–466
- Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 339:b2535
- Lee SE, Jang JY, Lim CS, Kang MJ, Kim SH, Kim MA et al (2010) Measurement of pancreatic fat by magnetic resonance imaging: predicting the occurrence of pancreatic fistula after pancreatoduodenectomy. *Ann Surg* 251(5):932–936
- Kawai M, Tani M, Hirono S, Ina S, Miyazawa M, Yamaue H (2009) How do we predict the clinically relevant pancreatic fistula after pancreaticoduodenectomy?—an analysis in 244 consecutive patients. *World J Surg* 33(12):2670–2678
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 49(12):1373–1379
- Babyak MA (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med* 66(3):411–421
- Altman DG, Vergouwe Y, Royston P, Moons KG (2009) Prognosis and prognostic research: validating a prognostic model. *BMJ* 338:b605
- Mallett S, Royston P, Dutton S, Waters R, Altman DG (2010) Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med* 8:20
- Royston P, Altman DG, Sauerbrei W (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 25(1):127–141
- Counsell C, Dennis M (2001) Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis* 12(3):159–170
- Jacob M, Lewsey JD, Sharpin C, Gimson A, Rela M, van der Meulen JH (2005) Systematic review and validation of prognostic models in liver transplantation. *Liver Transpl* 11(7):814–825
- Hukkelhoven CW, Rampen AJ, Maas AI, Farace E, Habbema JD, Marmarou A et al (2006) Some prognostic models for traumatic brain injury were not valid. *J Clin Epidemiol* 59(2):132–143
- Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI et al (2008) A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol* 61(4):331–343
- Lagakos SW (1988) Effects of misspecification and mismeasuring explanatory variables on tests of their association with a response variable. *Stat Med* 7(1–2):257–274
- Collins GS, Mallett S, Omar O, Yu LM (2011) Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 9:103

30. Harrell FE Jr, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15 (4):361–387
31. Marshall A, Altman DG, Royston P, Holder RL (2010) Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol* 10:7
32. Vergouwe Y, Royston P, Moons KG, Altman DG (2010) Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* 63 (2):205–214
33. Sun GW, Shook TL, Kay GL (1996) Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 49(8):907–916
34. Austin PC, Tu JV (2004) Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J Clin Epidemiol* 57(11):1138–1146
35. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 19(8):1059–1079
36. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD (2001) Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 54(8):774–781
37. Altman DG, Royston P (2000) What do we mean by validating a prognostic model? *Stat Med* 19(4):453–473
38. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE et al (2003) External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 56(9): 826–832
39. Reilly BM, Evans AT (2006) Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 144(3):201–209