ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 1296−1301

Journal of Clinical Epidemiology

Calibration of clinical prediction rules does not just assess bias

Werner Vach*

Clinical Epidemiology, Institute of Medical Biometry and Medical Informatics, Freiburg University Medical Center, Freiburg, Germany

Accepted 2 June 2013; Published online 8 September 2013

Abstract

Objectives: Calibration is often thought to assess the bias of a clinical prediction rule. In particular, if the rule is based on a linear logistic model, it is often assumed that an overestimation of all coefficients results in a calibration slope less than 1 and an underestimation in a slope larger than 1.

Study Design and Setting: We investigate the relation of the bias and the residual variation of clinical prediction rules with the typical behavior of calibration plots and calibration slopes, using some artificial examples.

Results: Calibration is not only sensitive to the bias of the clinical prediction rule but also to the residual variation. In some circumstances, the effects may cancel out, resulting in a misleading perfect calibration.

Conclusion: Poor calibration is a clear indication of limited usefulness of a clinical prediction rule. However, a perfect calibration should be interpreted with care as this may happen even for a biased prediction rule. © 2013 Elsevier Inc. All rights reserved.

Keywords: Bias; Calibration; External validation; Prognosis; Prognostic model; Residual variation

1. Introduction

Clinical prediction rules, also called prognostic models, are often the result of many efforts to use an available data set to select appropriate prognostic factors and develop a well-fitting model describing the relation between the factors and the event of interest. Hence, they often tend to be too extreme; that is, they overestimate the risk of high-risk patients and underestimate the risk of low-risk patients [1−3]. Moreover, the prognostic value of factors and their interrelation may vary even between similar patient populations, such that a rule may work well in one population but not necessarily in another [4−6]. For both reasons, clinical prediction rules should be validated in an external data set, which is not related to the data used to develop the prognostic model [7−10].

For the validation in an external data set, two basic principles are usually advocated and used in practice [8,11−13]: calibration and discrimination. Calibration aims to check whether the event probabilities according to the prediction rule coincide with the event rates that we can observe in the external validation data set [14,15]. Discrimination refers to the ability of the prognostic model to separate subjects with an event from subjects without an event. The latter is often approached by receiver operating characteristic (ROC) curves and related statistics like the c-index or the area under the ROC curve [1,16].

In this article, we focus on the first step, calibration. This is typically approached by a calibration plot, that is, some type of nonparametric regression relating the binary outcome Y to the probability values $\widehat{\pi}$ according to the prediction rule in the external validation data set. A popular choice is the division of the subjects into some risk groups according to the probability values from the prediction rule and plotting the relative frequency of $Y = 1$ against the mean of $\widehat{\pi}$ in each risk group. Sometimes, a smoothing method is used in addition. If a prediction rule is perfect, then the resulting points should be on the diagonal and the smoothed line should coincide with the diagonal. Any deviation from the diagonal indicates some imperfectness. Often, it is observed that the frequency of $Y = 1$ is smaller than that suggested by $\widehat{\pi}$ for high-risk patients and larger than that suggested by $\widehat{\pi}$ for low-risk patients, suggesting that the rule is indeed too extreme. This behavior can also be caught by computing a calibration slope $\widehat{\beta}_{\text{calib}}$ by a (logistic) regression of Y against $\widehat{\pi}$, with $\widehat{\beta}_{\text{calib}} < 1$ reflecting the situation of a rule with too extreme values. Janssen et al. [17] describe the typical interpretation of the calibration slope in the following way: ‘‘A calibration slope smaller than 1 indicates optimism; the regression coefficients of the original model were too large, which results in too

Conflict of interest: This material has neither been published nor is under consideration for publication elsewhere. There is no external funding involved in this study, and there are no possible conflicts of interest.

* Corresponding author. Tel.: +49 761 203 6722; fax: +49 761 203 6711.
E-mail address: wv@imbi.uni-freiburg.de

0895-4356/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jclinepi.2013.06.003

What is new?

- Calibration is often applied in the external validation of a clinical prediction rule.
- Poor calibration is often thought to reflect some bias of the prediction rule.
- Poor calibration can be due to variation or bias.
- Perfect calibration can occur even for a biased prediction rule.

extreme predictions in the new patients … A calibration slope that is larger than 1 indicates that the regression coefficients of the original model were too close to zero.''

In this article, we try to check whether calibration plots and calibration slopes are indeed useful to detect the type of bias described previously; that is, whether they correctly reflect a bias in the prediction rule. For this purpose, we assume in the external data set that the true-event probabilities follow a linear logistic model with known regression coefficients and investigate how certain choices of the regression coefficients for a prognostic model translate into patterns in the calibration plot and into certain values of the calibration slope.

2. Methods

We consider the artificial situation of an external validation study for which we know both the distribution of the covariates and the true model relating the binary outcome Y to the covariates. The four covariates X_1, …, X_4 are assumed to be independent and each taking the values -1, 0, and 1 with a probability of 1/3. The true model is assumed to be of a linear logistic type, that is, with $\pi_0(x_1, \ldots, x_4) = P(Y = 1|X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)$ it can be written as

$$\text{logit}\pi_0(x_1,\ldots,x_4) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4.$$

The choice of the true values of the regression parameters is shown in the first line of Table 1, with effects of the covariates ranging from small effects of 0.2 to moderate effects in the magnitude of 0.8. Moreover, we consider seven different prediction rules $\widehat{\pi}_1,\ldots,\widehat{\pi}_7$, which have been developed in other studies. All these rules are again based on linear logistic models; that is, they can be expressed as

$$\text{logit}\widehat{\pi}_j(x_1,\ldots,x_4) = \widehat{\beta}_{0\text{j}} + \widehat{\beta}_{1j} x_1 + \widehat{\beta}_{2j} x_2 + \widehat{\beta}_{3j} x_3 + \widehat{\beta}_{4j} x_4.$$

The values of the regression coefficients chosen are shown in Table 1. For the first three choices, two coefficients are underestimated and two are overestimated, but with increasing magnitude of the difference to the true coefficients when moving from $\widehat{\pi}_1$ to $\widehat{\pi}_3$. In the forth choice, all coefficients are overestimated. The final three choices reflect the situation that all coefficients are underestimated. However, they differ in the variation of the extent of the underestimation. In $\widehat{\pi}_5$, two coefficients are estimated nearly correctly and two are estimated even with an incorrect sign, that is, with a bias of greater than 100%. In $\widehat{\pi}_6$, one is estimated correctly, two are underestimated by about 50%, and one is underestimated by 100%. In $\widehat{\pi}_7$, all coefficients are underestimated with a bias in the range between 25% and 75%.

For each of the seven prediction rules $\widehat{\pi}_j$, we consider the joint distribution of $\widehat{\pi}(X_1,\ldots,X_4)$ and $\pi_0(X_1,\ldots,X_4)$; that is, we consider for all $3^4 = 81$ possible values for $x_1,\ldots, x_4$ the pairs $[\widehat{\pi}(x_1,\ldots,x_4), \pi_0(x_1,\ldots,x_4)]$. We start with considering $\widehat{\pi}$ in dependence on π_0 in a corresponding scatter plot, and we fit a regression line to this scatter plot. As both for π_0 and all prediction rules, the average probability is close to 0.5, the bias of $\widehat{\pi}$ is directly described by the slope $\widehat{\beta}_{\text{bias}}$ of this regression line: a bias slope of 1 indicates no bias, a bias slope greater than 1 indicates a rule with too extreme probability values, and a bias slope less than 1 indicates a rule which is too pessimistic: the risk of high-risk patients is underestimated, and the risk of low-risk patients is overestimated. Then, we turn to the perspective of calibration; that is, we consider Y in dependence on $\widehat{\pi}$. For this, we draw a random sample of 1,600 observations following the true model, divide the values of $\widehat{\pi}$ into eight risk groups of equal size, and plot the observed frequency of $Y = 1$ in each risk group vs. the mean value of $\widehat{\pi}$ in each risk group. In addition, the lowess smoother [18] is used to obtain a smooth regression line. As the expectation of Y given $X_1 = x_1$, $X_2 = x_2$, $X_3 = x_3$, $X_4 = x_4$ is just given by $\pi_0(x_1, \ldots, x_4)$, we can study the typical behavior of calibration also by considering π_0 in dependence on $\widehat{\pi}$. Hence, we provide also a scatter plot of π_0 vs. $\widehat{\pi}$ together with the corresponding regression line. The slope $\widehat{\beta}_{\text{calib}}$ of this regression line describes the expectation for the slope of a regression line fitted to a calibration plot, and we refer to it as the calibration slope.

In the literature, calibration plots are typically presented on the probability scale, whereas calibration slopes are considered on the logit scale. As Appendix at www.jclinepi.com, we provide also calibration plots on the logit scale

Table 1. Regression coefficients in the true model and for seven clinical prediction rules

	Regression coefficients				
	β_0	β_1	β_2	β_3	β_4
True model π_0	0.0	0.21	0.37	0.64	0.77
Prediction rule $\widehat{\pi}_1$	0.0	0.25	0.30	0.51	0.92
Prediction rule $\widehat{\pi}_2$	0.0	0.32	0.19	0.32	1.16
Prediction rule $\widehat{\pi}_3$	0.0	0.40	0.04	−0.06	1.62
Prediction rule $\widehat{\pi}_4$	0.0	0.29	0.59	1.20	0.85
Prediction rule $\widehat{\pi}_5$	0.0	0.20	−0.09	−0.16	0.73
Prediction rule $\widehat{\pi}_6$	0.0	0.11	0.19	0.64	0.00
Prediction rule $\widehat{\pi}_7$	0.0	0.08	0.27	0.38	0.19

together with calibration slopes from a logistic model using the probability values expressed on the logit scale as the only covariate. This model is fitted to a data set with all possible combinations of the covariate values and the two outcomes $Y = 1$ and $Y = 0$, with the outcomes weighted according to π_0. The bias slope is determined accordingly. All computations were performed with Stata 12.1 (Stata Corp., College Station, TX).

3. Results

Results for the three rules $\widehat{\pi}_1$, $\widehat{\pi}_2$, $\widehat{\pi}_3$ are shown in Fig. 1. As for all three rules, the average difference between the estimated regression coefficients and the true ones is roughly 0. Consequently, we can observe that the rules are nearly unbiased, as also indicated by bias slopes close to 1. The calibration plots and the calibration slopes indicate that we never obtain a perfect calibration, and we observe that the calibration becomes increasingly poor when moving from $\widehat{\pi}_1$ to $\widehat{\pi}_3$. This can be explained by the increasing residual variation of $\widehat{\pi}_j$ that we can observe in the scatter plots of $\widehat{\pi}_j$ vs. π_0 and hence in a decreasing correlation between $\widehat{\pi}_j$ and $\widehat{\pi}_0$: from the theory of simple linear regression involving two variables Z_1 and Z_2, it is well known that the slope $\widehat{\beta}_{Z_1|Z_2}$ of regressing Z_1 vs. Z_2 and the slope $\widehat{\beta}_{Z_2|Z_1}$ of regressing Z_2 vs. Z_1 are related to the correlation $\widehat{\rho}_{Z_1,Z_2}$ between Z_1 and Z_2 by the simple relation

$$\widehat{\beta}_{Z_1|Z_2} \times \widehat{\beta}_{Z_2|Z_1} = \widehat{\rho}^2_{Z_1,Z_2}.$$

In our context, this implies

$$\widehat{\beta}_{\text{calib}} \times \widehat{\beta}_{\text{bias}} = \widehat{\beta}_{\pi_0|\widehat{\pi}} \times \widehat{\beta}_{\widehat{\pi}|\pi_0} = \rho^2_{\pi_0\widehat{\pi}}. \quad (1)$$

So, if the bias slope is close to 1, the calibration slope is close to the square of the correlation between π_0 and $\widehat{\pi}$, and hence the calibration slope is decreasing with decreasing correlation.

Fig. 2 presents the results for the prediction rule $\widehat{\pi}_4$ for which all regression coefficients are overestimated. Consequently, the bias slope is greater than 1. The calibration slope is less than 1, and this is exactly what we have to expect from the relation (7): if the bias slope is greater than 1, the calibration slope must be less than 1 as squared correlations never exceed 1.

More interesting case is given by an underestimation of the regression coefficients and correspondingly a bias slope less than 1. Then, according to Eq. (7), the calibration slope can be greater than and less than 1. And the three examples,

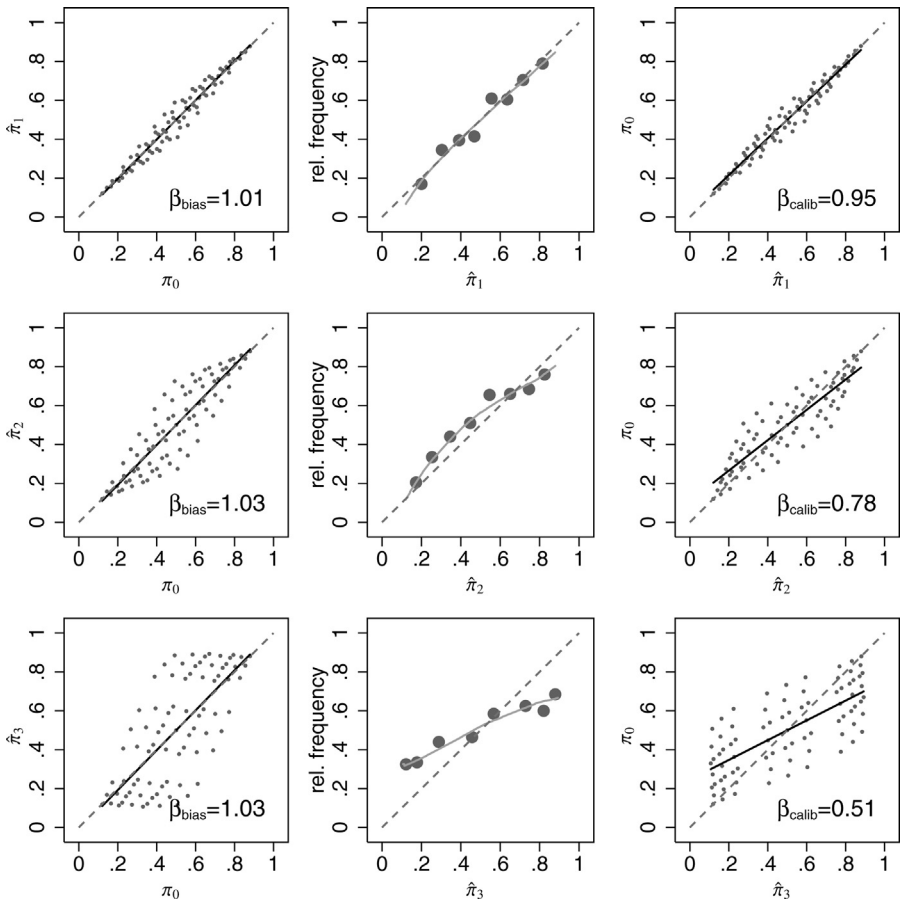

Fig. 1. A scatter plot of $\widehat{\pi}$ vs. π_0 with corresponding regression line and bias slope, an example of a calibration plot, and a scatter plot of π_0 vs. $\widehat{\pi}$ with corresponding regression line and calibration slope for the three prediction rules $\widehat{\pi}_1$ (upper row), $\widehat{\pi}_2$ (middle row), and $\widehat{\pi}_3$ (lower row). The diagonal is indicated by a dashed line.

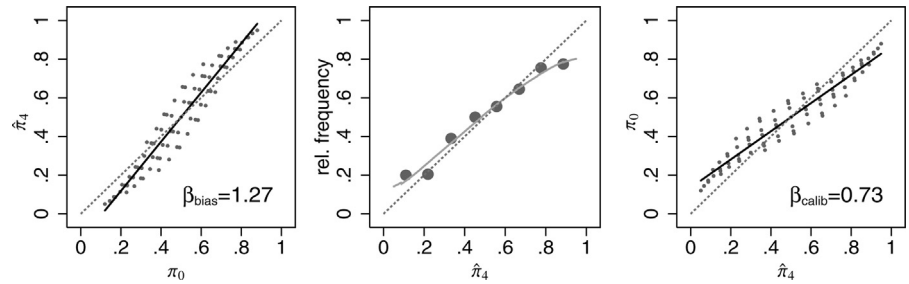

Fig. 2. A scatter plot of $\hat{\pi}$ vs. π_0 with corresponding regression line and bias slope, an example of a calibration plot, and a scatter plot of π_0 vs. $\hat{\pi}$ with corresponding regression line and calibration slope for the prediction rule $\hat{\pi}_4$. The diagonal is indicated by a dashed line.

$\hat{\pi}_5$, $\hat{\pi}_6$, and $\hat{\pi}_7$, illustrate that indeed all these results are possible (Fig. 3). For $\hat{\pi}_5$, the underestimation of the coefficients is highly varying, hence the residual variance is large and the correlation between $\hat{\pi}$ and π_0 is low. Consequently, the calibration slope is less than 1. For $\hat{\pi}_6$, the underestimation is less varying, implying a smaller residual variation and hence a higher correlation, and hence a larger calibration slope. And in this specific example, the calibration slope is close to 1, indicating perfect calibration. The final example, $\hat{\pi}_7$, indicates that the calibration slope can be indeed greater than 1, if the residual variance is limited. This happens here because the extent of underestimation of the regression coefficients is rather uniform.

The corresponding Figs. 5–7 in the Appendix at www.jclinepi.com indicate that we come to very similar conclusions if all analyses are performed on the logit scale.

4. Discussion

We have demonstrated that calibration plots and calibration slopes mix a true bias of a prognostic model with the residual variance of the prognostic model. This is in line

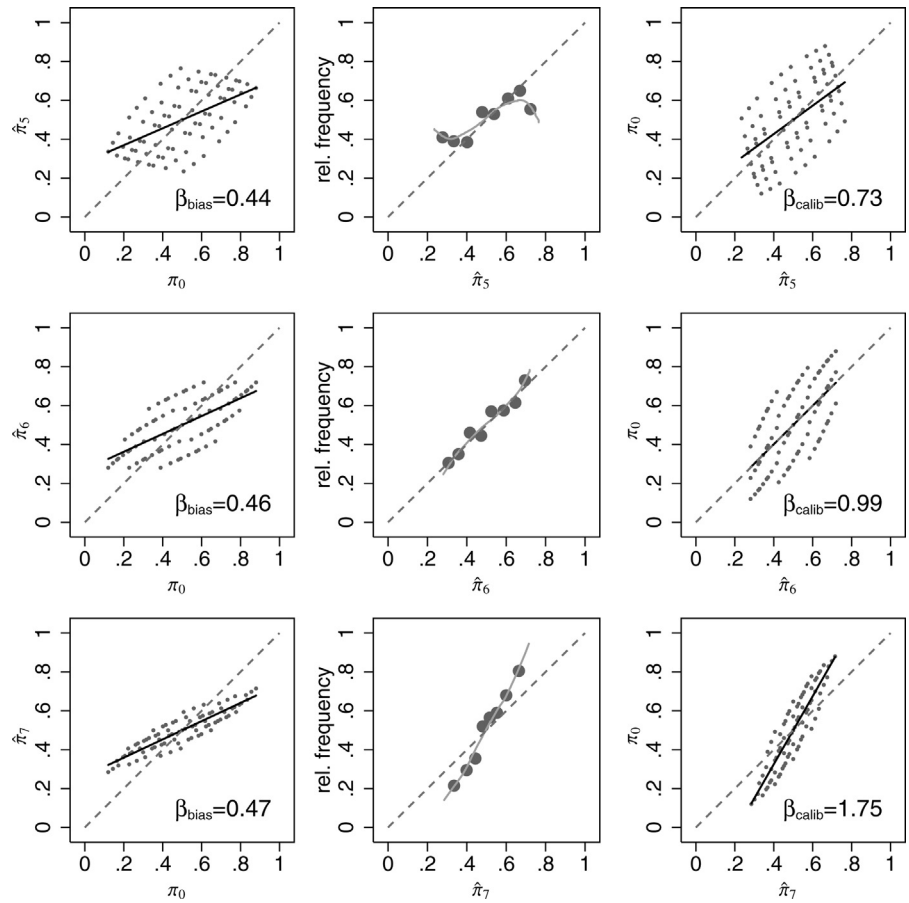

Fig. 3. A scatter plot of $\hat{\pi}$ vs. π_0 with corresponding regression line and bias slope, an example of a calibration plot, and a scatter plot of π_0 vs. $\hat{\pi}$ with corresponding regression line and calibration slope for the three prediction rules $\hat{\pi}_5$ (upper row), $\hat{\pi}_6$ (middle row), and $\hat{\pi}_7$ (lower row). The diagonal is indicated by a dashed line.

with the results of Vergouwe et al. [19] who observed that a calibration slope less than 1 if half of the regression coefficients are underestimated and half of them are overestimated to the same degree. In particular, calibration plots and slopes may indicate a perfect calibration even if a clinical prediction rule systematically underestimates the risk of high-risk patients and overestimates the risk of low-risk patients. Hence, calibration plots and slopes should be interpreted with care, in particular, if they are intended to prove the validity of a clinical prediction rule.

This somewhat surprising result stems from the perspective taken in a calibration, which is different from the usual perspective taken if we talk about bias. When talking about bias, we typically ask whether estimates tend to be on average below or above the true values given the true values. In our context, this means that we ask whether the probabilities $\widehat{\pi}$ according to the prediction rule tend to be above or below π_0 for subjects with identical or similar values of π_0. Calibration takes the opposite perspective and asks whether the true values tend to be on average above or below the estimates given the estimates; that is, whether the true event rates π_0 tend to be above or below the values $\widehat{\pi}$ according to the prediction rule for subjects with identical or similar values of $\widehat{\pi}$. As already pointed out by Vach [20], these two perspectives can lead to very different conclusions, which is again demonstrated in this article in the context of calibration.

The two perspectives also correspond to two conceptually different questions. In the bias perspective, we ask whether patients with a certain covariate pattern $x_1, \ldots, x_p$ and hence a certain true event probability $\pi_0(x_1, \ldots, x_p)$ will obtain by the prediction rule an unbiased estimate of this probability. In the calibration perspective, we ask whether patients with a certain estimated probability $\widehat{\pi}$ can expect on average to experience an event rate equal to this value. In my opinion, there can be little doubt that the first perspective is more relevant in a clinical context. The aim of a prediction rule should be to inform a patient (and the treating clinician) about the prognosis of the patient in dependence on his or her individual characteristics. The calibration perspective is not relevant for the single patient as it is not related to his or her individual risk. The calibration perspective may be relevant for clinicians in a broader perspective: they should be aware of that in patients with an event probability of, for example, 0.9 according to the prediction rule, they may experience the event only for 80% of the patients, although the rule is unbiased. There may be other, non clinical settings, where the calibration perspective may correctly reflect our expectations. Already, Cox [15] referred to weather forecasts as an example: here it may be reasonable to claim that the rate of rain falls on days with a predicted rain fall probability of 90% should be indeed about 90%.

In spite of the deficiency demonstrated, calibration plots and slopes are still useful tools for an initial investigation of the validity of a prognostic model in an external data set: if calibration is poor, then there can be no doubt that the prediction rule is not useful in its present form. Whether this is because of a bias or a large residual variation remains, however, unclear. For the first interpretation of a poor calibration, this is no problem as in both cases the rule is not useful. However, a poor calibration does not imply that simple corrections of the type $\widehat{\pi}_{new} = \widehat{\alpha}_{calib} + \widehat{\beta}_{calib}\widehat{\pi}$ yield a better prediction rule in any case as they may introduce bias instead of removing bias. This is illustrated for the rule $\widehat{\pi}_3$ in Fig. 4. Instead of using the calibration bias and slope, any correction of this type should be based on inverting the line according to the bias intercept and the bias slope. Unfortunately, to the best of my knowledge, estimation of the bias slope has not been addressed in the statistical literature until now.

On the other side, estimation of the bias slope does not seem to be out of reach. For example, we can express the bias slope as

$$\widehat{\beta}_{\hat{\pi}|\pi_0} = \frac{\widehat{s}^2_{\hat{\pi}}}{\widehat{s}^2_{\pi_0}} \widehat{\beta}_{\pi_0|\hat{\pi}},$$

that is, as a corrected version of the calibration slope. The correction factor is the ratio between the empirical variance of the probability values according to the prediction rule in the external validation data set and the corresponding empirical variance of the true probabilities. The first can be simply estimated from the data, whereas the latter is not observable directly. However, the latter can be estimated using a bootstrap bias correction of an estimate based on the empirical predicted probabilities of a model fitted in the external data set [21]. It will be a task for future research to find efficient and feasible procedures to estimate the bias slope.

We considered in this article particular choices of the given clinical prediction rule, which may raise the question about the relevance of the choices. The literature on external validation is often focusing on the situation that a clinical prediction rule is too extreme, corresponding to our choice $\widehat{\pi}_4$. This may reflect that in former times, extensive model building was the usual case in constructing clinical prediction rules resulting typically in overfitting. However, the widespread recommendations on careful development and internal validation of clinical prediction rules should have reduced the risk of overfitting. Hence, today it is more

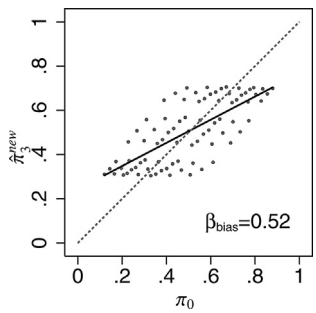

Fig. 4. A scatter plot of $\widehat{\pi}_3^{new}$ vs. π_0 with corresponding regression line and bias slope. $\widehat{\pi}_3^{new}$ is based on the calibration intercept and calibration slope according to Fig. 1, that is, $\widehat{\pi}_3^{new} = 0.25 + 0.51\widehat{\pi}_3$. The diagonal is indicated by a dashed line.

likely that any deviation between a given prediction rule and the true model in the external data set reflects the random imprecision of the effect estimates and hence can go in any direction. So, constellations like $\widehat{\pi}_6$ or $\widehat{\pi}_7$, for which calibration may give rise to misleading conclusions, are not unlikely today.

We have used the term ''bias'' in this article in a somewhat unusual manner. In the statistical literature, the term bias refers to the sampling properties of an estimator, reflecting a tendency to overestimate or underestimate the true quantity. We use the term bias to characterize properties of a single estimate, namely a clinical prediction rule. We have chosen this for two reasons. First, there is a common (mis) conception of calibration as an indicator that regression coefficients are too large (or too small), as explicitly expressed in Ref. [17]. So, this conception is related to a bias of the estimator of these regression coefficients. Second, we use the term bias as an opposite to calibration. Calibration is based on conditioning on estimated probabilities, whereas bias in our sense is based on conditioning on true probabilities, and hence closer to the traditional view on bias, asking whether an estimator tends to under- or overestimate, given the true value of the parameter to be estimated.

It may be regarded as a limitation of our investigation that we only considered the case of continuous covariates taking the values −1, 0, and 1. This choice was because of pedagogical reasons, namely to be able to visualize the joint distribution of $\widehat{\pi}$ and π_0 by a simple scatter plot. However, the relation (7) does not rely on any assumption about the covariate distribution, and hence our results can be assumed to hold for all covariate distributions.

5. Conclusions

Calibration of a clinical prediction rule mixes the bias of the rule with its residual variance. Sometimes, these effects may be canceled out. In particular, a calibration can look perfect even if all regression coefficients are underestimated. Hence, calibration should not be used as a proof for the validity of a clinical prediction rule. It is still useful as an initial step to identify rules that are completely useless.

Appendix

Supplementary material

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2013.06.003.

References

[1] Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361−87.
[2] Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. J R Stat Soc Ser C Applied Stat 1999; 48:313−29.
[3] Steyerberg E, Eijkemans M, Harrell F, Habbema J. Prognostic modeling with logistic regression analysis: In search of a sensible strategy in small data sets. Med Decis Making 2001;21:45−56.
[4] Bleeker S, Moll H, Steyerberg E, Donders A, Derksen-Lubsen G, Grobbee D, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003;56:826−32.
[5] Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98:691−8.
[6] Schuetz P, Koller M, Christ-Crain M, Steyerberg E, Stolz D, Mueller C, et al. Predicting mortality with pneumonia severity scores: importance of model recalibration to local settings. Epidemiol Infect 2008;136: 1628−37.
[7] Altman D, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19:453−73.
[8] Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338: b605.
[9] Justice A, Covinsky K, Berlin J. Assessing the generalizability of prognostic information. Ann Intern Med 1999;130:515−24.
[10] Moons KGM. Criteria for scientific evaluation of novel markers: a perspective. Clin Chem 2010;56:537−41.
[11] Bartfay E, Bartfay WJ. Accuracy assessment of prediction in patient outcomes. J Eval Clin Pract 2008;14:1−10.
[12] Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:e1001221.
[13] Vergouwe Y, Steyerberg E, Eijkemans M, Habbema J. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005;58: 475−83.
[14] Miller M, Langefeld C, Tierney W, Hui S, McDonald C. Validation of probabilistic predictions. Med Decis Making 1993;13:49−58.
[15] Cox D. Two further applications of a model for binary regression. Biometrika 1958;45:562−5.
[16] Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29−36.
[17] 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;61:76−86.
[18] Cleveland W. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74:829−36.
[19] Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol 2010;172:971−80.
[20] Vach W. On the relation between the shrinkage effect and a shrinkage method. Comput Stat 1997;12:279−92.
[21] Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;23:723−48.