Generalizability of Dutch Prediction Models for Low Hemoglobin Deferral: A Study on External Validation and Updating in Swiss Whole Blood Donors

A. Mireille Baart^a Stefano Fontana^b Anita Tschaggelar^b Martijn W. Heymans^c
Wim L.A.M. de Kort^a

^aSanquin Research, Department of Donor Studies, Amsterdam, The Netherlands;

^bInterregional Blood Transfusion Service SRC Ltd, Berne, Switzerland;

^cEMGO Institute for Health and Care Research, Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

Keywords

Blood donors · Donor deferral · External validation · Hemoglobin · Prediction model

Summary

Background: Sex-specific prediction models for low hemoglobin (Hb) deferral have been developed in Dutch whole blood donors. In this study, we validated and updated the models in a cohort of Swiss whole blood donors. **Methods:** Prospectively collected data from 53,772 Swiss whole blood donors were used. The predictive performance of the Dutch models was assessed in terms of calibration (agreement between predicted probabilities and observed frequencies) and discrimination (ability to discriminate between deferred and approved donors). The models were updated by revising the strength of the individual predictors in the models. **Results:** A total of 1,065 men (3.3%) and 2,063 women (9.7%) were deferred from donation because of a low Hb level. Validation in Swiss donors demonstrated underestimation of predicted risks and significantly lower discriminative ability. The predictive effects of most predictors were weaker in Swiss donors. Updating the models increased the calibration for both men and women, and slightly increased the discriminative ability in men. **Conclusion:** Validation of the Dutch prediction models in Swiss whole blood donors showed lower, though adequate performance. In general, the Dutch prediction models can reliably predict the risk of Hb deferral, although for application in other countries small adaptations are necessary.

© 2016 S. Karger GmbH, Freiburg

Introduction

Prior to a blood donation, donors are screened for hemoglobin (Hb) levels. Donors with low Hb levels are deferred from donation to protect them from developing iron deficiency and to guarantee that blood units for transfusion meet the required standards for Hb content [1].

Recently, sex-specific prediction models for Hb deferral risk have been developed in a large cohort of Dutch whole blood donors [2]. These prediction models include donor characteristics, visit characteristics, and characteristics from the donation history. With these prediction models, the risk of Hb deferral on the next visit to the blood collection center can be calculated. The model predictions may be helpful in the management of the blood donation program: donors with a low risk of Hb deferral can be invited with preference, whereas for donors with a high risk it is better to postpone the invitation for a next donation. For the application of the prediction models, a threshold value is required above which donors will be assigned in a high-risk group. At lower threshold values more deferrals can be prevented but at cost of more donors with appropriate Hb levels that are unnecessarily not invited for donation. Thus when such a decision threshold is determined the number of deferrals that can be prevented must be weighed against the number of unnecessarily postponements, and it depends on the available blood stock level and the number of replacement donors which threshold value should be used. Eventually, the prediction models could help to decrease the number of donor deferrals while the blood stock level remains adequate.

Before the prediction models can be applied in practice in the invitation process of donors, validation of the models is necessary [3, 4]. Initially, we performed a cross-validation within Dutch regions, which showed good model performance [2]. Such an inter-

nal validation gives a first indication of generalizability. The next step is to perform an external validation, in which the performance of the models is assessed in a group of independent donors [5]. To do so, we externally validated the prediction models in a cohort of Irish donors. Results of this study showed limited performance of the Dutch models [6]. An important difference between the Dutch development cohort and the Irish validation cohort was the lower Hb cutoff level for donation in the Irish cohort. For men, Hb cutoff levels for donation are 8.4 mmol/l in the Netherlands and 8.07 mmol/l (130 g/l) in Ireland, and for women Hb cutoff levels are 7.8 mmol/l in the Netherlands and 7.45 mmol/l (120 g/l) in Ireland. These lower Hb cutoff levels might have been a reason for the lower validity of the models in Irish donors.

In order to assess the external validity of the Dutch prediction models when the same Hb cutoff levels for donation are used in the validation cohort, we validated the models in a cohort of Swiss whole blood donors in this study.

Material and Methods

Donors

All Swiss whole blood donors who visited any blood collection center or mobile unit of the Regional Blood Transfusion Service of Berne in the years 2011 until 2013 were eligible for the study ($n = 102,436$; 56,745 men, 45,691 women). From these donors, data of all visits from January 2009 until December 2013 were extracted from the donor database. Data were anonymized for the analyses.

For each donor an 'intended visit' was defined. For donors whose last visit occurred in 2011 or 2012 and for donors with only one visit in 2013, their last visit was indicated as the intended visit. For donors with more than one visit in 2013, a randomly selected visit in 2013 was used as the intended visit. This random selection was done to allow for equal distributions across the seasons, that is to avoid a clustering of visits at the end of 2013. The distributions of intended visits from donors with a last visit in 2011 or 2012 and from donors with only one visit in 2013 were similar across seasons. The occurrence of Hb deferral at the intended visit was used as outcome in the analyses.

Inclusion and exclusion criteria were the same as those used for the development of the prediction model in Dutch donors. All donations since January 2009 should be whole blood donations, and not plasma or platelet donations. Donors should have donated whole blood at least twice prior to the intended visit since January 2009. Donors were excluded if the outcome, i.e. the Hb level at the intended visit, was unknown. A total of 29,706 donors were excluded because they had not given any whole blood donation prior to the intended visit since January 2009, 18,084 donors were excluded because they had donated only once before the intended visit, and an additional 874 donors were excluded because the Hb level at the intended visit was missing since it had not been registered. Finally, 53,772 Swiss whole blood donors (32,484 men; 21,288 women) were included in the study.

This research has been performed with the approval of the ethical advisory council of the Sanquin Blood Supply Foundation. Moreover, all donors have given their consent by stating that part or all of their donations can be used for research aiming at improving the blood supply chain. Our ethical advisory council includes members of both Sanquin and non-Sanquin affiliations. This committee includes members with the background training and experience required for such ethical committees.

Prediction Model

The existing prediction models were developed in 112,491 Dutch male whole blood donors and 108,455 Dutch female whole blood donors [2]. The

sex-specific models predict the risk of Hb deferral at the intended visit. Hb deferral was defined as having an Hb level below the sex-specific cutoff level for donation as described in the European Commission Directive [7], which is 8.4 mmol/l for men and 7.8 mmol/l for women.

Predictive factors of the models are age, seasonality, Hb level measured at the previous visit (previous Hb level), difference in Hb level between the previous visit and the second-last visit prior to the intended visit (ΔHb), time since the previous visit, total number of whole blood donations in the past 2 years, and deferral at the previous visit. Seasonality was defined as the four meteorological seasons: Winter (donations between December 1 and February 29), Spring (March 1 to May 31), Summer (June 1 to August 31), and Fall (September 1 to November 30). Previous Hb level and ΔHb were included in the model with a piecewise linear function based on two pieces. The breakpoint for previous Hb level was chosen at the sex-specific cutoff level for donation and for ΔHb at 0 mmol/l. Deferral at the previous visit was categorized as deferral because of a low Hb level, deferral because of reasons other than a low Hb level, and no deferral. The exact formulas to calculate the risk of Hb deferral are given in Appendix I (see Supplemental Material, available at <http://content.karger.com/ProdukteDB/produkte.asp?doi=446817>).

Data Collection

Data on Hb levels and the relevant predictors were obtained from the database of the Blood Transfusion Service of Berne. Hb levels were routinely measured during donor screening in fingerstick capillary samples using a photometer (HemoCue, Angelholm, Sweden). Swiss Hb levels were assessed in g/l rather than mmol/l. Cutoff levels for donation used in Switzerland are 135 g/l for men and 125 g/l for women, which are similar to the Dutch cutoff levels of 8.4 mmol/l for men and 7.8 mmol/l for women. For the analyses, Hb levels were converted from g/l into mmol/l using the conversion factor 0.06206.

Statistical Analysis

All statistical analyses were performed for men and women separately, and therefore the difference in gender distribution (male/female ratio of 50.9/49.1 in the Dutch development cohort and 60.4/39.6 in the Swiss validation cohort) could not influence the results.

External Validation

Missing values occurred in the following variables: previous Hb level (0.8%), ΔHb (1.4%), and deferral at the previous visit (0.8%). In order to be able to use the observed information of other known variables, we single imputed missing values using stochastic regression imputation [8].

For the external validation, the predictive performance of the models was assessed in terms of calibration and discrimination [9]. Calibration is the agreement between predicted probabilities and observed frequencies. Calibration was studied with a logistic regression model with Hb deferral as dichotomous outcome and the linear predictor as the only covariate. The regression coefficient of the linear predictor (the calibration slope, visualized in a calibration plot) reflects whether the effects of the predictors in the Swiss data are on average similar as the effects in the Dutch models, and is ideally 1. We also assessed calibration-in-the-large by fitting a logistic regression model with the linear predictor as an offset variable (setting the regression coefficient to 1). The intercept indicates whether predictions are in general correct, and is ideally 0. Discrimination is the ability of the model to differentiate between deferred and approved donors. Discrimination was determined with the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). The AUC indicates the percentage of randomly selected pairs of donors, in which one donor is deferred and the other donor is approved for donation, for which is correctly assigned a higher risk of Hb deferral to the deferred donor. A value of 1 indicates perfect discrimination; a value of 0.50 indicates poor discrimination, equivalent to flipping a coin.

The AUC contains information about the sensitivity and specificity over the entire range of decision thresholds to divide donors into groups with high versus low risk of Hb deferral. For a practical interpretation, we assessed the sensitivity and specificity at a number of specific decision thresholds.

Table 1. Distribution of predictors at the intended visit

Predictor	Swiss validation data		Dutch development data	
	men (n = 32,484)	women (n = 21,288)	men (n = 112,491)	women (n = 108,455)
Age, years	46 ± 12	43 ± 13	49 ± 12	44 ± 13
Seasonality, n (%)				
Winter	6,148 (18.9)	3,972 (18.7)	22,789 (20.3)	20,487 (18.9)
Spring	8,384 (25.8)	5,686 (26.7)	29,040 (25.8)	28,089 (25.9)
Summer	8,769 (27.0)	5,953 (28.0)	31,315 (27.8)	31,645 (29.2)
Fall	9,183 (28.3)	5,677 (26.7)	29,347 (26.1)	28,234 (26.0)
Previous Hb level, mmol/l	9.5 ± 0.7	8.5 ± 0.6	9.4 ± 0.7	8.5 ± 0.6
Median ΔHb, mmol/l (25th–75th percentile)	0 (–0.4 to 0.4)	0 (–0.4 to 0.4)	0 (–0.4 to 0.4)	0 (–0.4 to 0.4)
Median time since previous visit, days (25th–75th percentile)	177 (119–315)	204 (134–364)	112 (77–210)	154 (120–259)
Deferral at previous visit, n (%)				
Due to low Hb (based on mmol/l)	884 (2.7)	1,762 (8.3)	3,607 (3.2)	6,658 (6.1)
Due to other reason than low Hb	1,189 (3.7)	884 (4.2)	3,675 (3.3)	4,311 (4.0)
Median number of whole blood donations in the past 2 years (25th–75th percentile)	3 (2–4)	2 (2–3)	4 (2–6)	3 (2–4)
Hb intended visit, mmol/l	9.5 ± 0.7	8.5 ± 0.6	9.3 ± 0.7	8.4 ± 0.6
Median Hb deferral, based on mmol/l† (25th–75th percentile)	1,065 (3.3)	2,063 (9.7)	4,568 (4.1)	8,297 (7.7)

†Hb < 8.4 mmol/l (= 135 g/l) for men and < 7.8 mmol/l (=125 g/l) for women.

Model Updating

Usually, model performance is poorer in external validation compared to the performance in the development data. If this is the case, the models should be updated and adjusted to the conditions in the validation cohort to improve performance [5, 10, 11].

Results of the external validation prompted us to update the models. We adjusted the intercept and regression coefficients of the prediction models for the Swiss donors. Two methods were applied for updating: recalibration of the model and model revision [10]. Recalibration included adjustment of the intercept and adjustment of the individual regression coefficients with the same factor, i.e. the calibration slope. For the revised models, the individual regression coefficients were further adjusted. This was done by adding the seven predictors of the existing models once again in a step forward manner to a logistic regression model with the linear predictor of the recalibrated model as the only covariate, and to test with a likelihood ratio test ($p < 0.05$) if this addition had added value. If so, the regression coefficient for that predictor was adjusted further by adding the β of the added predictor to the β of that predictor in the recalibrated model.

Statistical analyses were performed with computer software (SPSS, Version 19; SPSS, Inc., Chicago, IL, USA; and R, Version 2.12.2, <http://cran.r-project.org/bin/windows/base/old/2.12.2/>).

Results

In the Swiss validation cohort, a total of 1,065 male donors (3.3%) and 2,063 female donors (9.7%) were deferred because of a low Hb level (table 1). Swiss men were less often deferred compared to Dutch men (4.1%), whereas Swiss women were more fre-

quently deferred compared to Dutch women (7.7%). In Swiss donors, mean Hb levels were higher than in Dutch donors at the intended visit (men and women) and at the previous visit (men). The difference was 0.1 mmol/l for men at the previous visit and for women at the intended visit, and 0.2 mmol/l for men at the intended visit. Swiss donors had donated less often in the past 2 years compared to Dutch donors (median value was 1 lower), and, accordingly, the median time interval since the previous visit was longer. For the other predictive factors, no substantial differences between Swiss and Dutch donors were observed. The distribution of predictors at the intended visit in deferred and approved donors is presented in Appendix II (see Supplemental Material, available at <http://content.karger.com/ProdukteDB/produkte.asp?doi=446817>).

Table 2 presents the results of the external validation. Predicted risks from the Dutch models were systematically too low for the Swiss donors as indicated with the average predicted risks compared to the observed proportions of donors with Hb deferral and with the calibration-in-the-large. The calibration slopes deviated from the ideal value of 1: 0.74 (95% CI 0.70–0.78) for men and 0.76 (95% CI 0.72–0.80) for women. Calibration plots for men and women are presented in figure 1a and b, respectively.

Discrimination of the models in the Swiss cohort was significantly lower than in the Dutch cohort: the AUC for men in the Swiss cohort was 0.84 (95% CI 0.83–0.85) versus 0.89 (95% CI 0.88–0.89) in the Dutch cohort; for women the AUC was 0.79 (95%

Table 2. Performance of the prediction models in the Dutch development data and the Swiss and Irish validation data

Model	Predicted Hb deferral, %	Observed Hb deferral, %	Calibration-in-the-large (95% CI)	Calibration slope (95% CI)	AUC (95% CI)
Men					
Dutch development data	4.1	4.1	0.00 (−0.06 to 0.06)	1.00 (0.97–1.03)	0.89 (0.88–0.89)
Swiss validation data	1.9	3.3	0.61 (0.55–0.68)	0.74 (0.70–0.78)	0.84 (0.83–0.85)
Irish validation data	1.6	2.4	0.47 (0.38–0.55)	0.65 (0.60–0.70)	0.82 (0.80–0.83)
Women					
Dutch development data	7.7	7.7	0.00 (−0.02–0.02)	1.00 (0.98–1.02)	0.84 (0.83–0.84)
Swiss validation data	5.9	9.7	0.65 (0.60–0.70)	0.76 (0.72–0.80)	0.79 (0.78–0.80)
Irish validation data	5.5	8.4	0.54 (0.48–0.60)	0.63 (0.59–0.67)	0.75 (0.74–0.76)

Fig. 1. Calibration plots for men (**A**) and women (**B**). 1. original model, 2. recalibrated model, 3. revised model. Triangles indicate the proportion of donors with low Hb level per pentile of the predicted probabilities. The solid line shows the relation between observed proportions and predicted probabilities. Ideally, this line equals the dotted line.

Fig. 2. ROC curves for men (**A**) and women (**B**). The ROC curves show the discriminative ability of the Dutch prediction models when validated in the Swiss donors cohort.

Table 3. Comparison of the accuracy of different threshold values for the predicted probability for low Hb deferral in Dutch and Swiss women and men

Probability	Donors, %	Low Hb levels*, % (95% CI)	Negative predictive value, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
<i>Men</i>					
Dutch development data					
≥5%	22.9	14.5 (14.1–14.9)	99.1 (99.0–99.2)	82.2 (81.1–83.3)	79.6 (79.4–79.8)
≥10%	13.3	20.1 (19.5–20.7)	98.4 (98.3–98.5)	66.0 (64.6–67.4)	88.9 (88.7–89.1)
≥15%	8.4	25.2 (24.3–26.1)	97.9 (97.8–98.0)	52.3 (50.1–53.7)	93.4 (93.3–93.5)
≥20%	5.3	29.6 (28.4–30.8)	97.4 (97.3–97.5)	38.4 (37.0–39.8)	96.1 (96.0–96.2)
Swiss validation data					
≥5%	10.9	15.5 (14.3–16.7)	98.2 (98.0–98.4)	51.7 (48.7–54.7)	90.4 (90.1–90.7)
≥10%	4.9	21.3 (19.3–23.3)	97.7 (97.5–97.9)	31.9 (29.1–34.7)	96.0 (95.8–96.2)
≥15%	2.4	26.4 (23.3–29.5)	97.3 (97.1–97.5)	19.2 (16.8–21.6)	98.2 (98.1–98.3)
≥20%	1.2	32.0 (27.4–36.6)	97.1 (96.9–97.3)	11.7 (9.8–13.6)	99.2 (99.1–99.3)
<i>Women</i>					
Dutch development data					
≥5%	40.7	16.4 (16.1–16.7)	98.4 (98.3–98.5)	87.3 (86.6–88.0)	63.1 (62.8–63.4)
≥10%	24.7	22.2 (21.7–22.7)	97.1 (97.0–97.2)	71.5 (70.5–72.5)	79.2 (78.9–79.5)
≥15%	15.7	27.3 (26.6–28.0)	96.0 (95.9–96.1)	56.0 (54.9–57.1)	87.7 (87.5–87.9)
≥20%	10.0	32.2 (31.3–33.1)	95.1 (95.0–95.2)	42.2 (41.1–43.3)	92.7 (92.5–92.9)
Swiss validation data					
≥5%	34.3	20.9 (20.0–21.8)	96.2 (95.9–96.5)	73.9 (72.0–75.8)	69.9 (69.3–70.1)
≥10%	18.4	26.8 (25.4–28.2)	94.2 (93.9–94.5)	50.9 (48.7–53.1)	85.1 (84.6–85.6)
≥15%	10.6	31.4 (29.5–33.3)	92.9 (92.5–93.3)	34.4 (32.3–36.5)	91.9 (91.5–92.3)
≥20%	6.5	34.8 (32.3–37.3)	92.1 (91.7–92.5)	23.3 (21.5–25.1)	95.3 (95.0–95.6)

*Positive predictive value.

CI 0.78–0.80) versus 0.84 (95% CI 0.83–0.84) in the Swiss and Dutch cohort, respectively. An AUC of 0.84 for men means that for 84% of randomly selected pairs of male donors, in which one donor is deferred and the other donor is approved for donation, is correctly assigned a higher risk of Hb deferral to the deferred donor. The ROC curves for men and women are presented in figure 2a and b respectively.

Tables 3 presents measures of accuracy of the models at different decision thresholds to divide donors into groups with high versus

low risk of Hb deferral. At all threshold levels, the sensitivity and the negative predictive value (NPV) were lower, whereas the specificity and the positive predictive value (PPV) were higher in the Swiss cohort compared to the Dutch cohort. This applies to both men and women. When the prediction models were developed, we concluded that a decision threshold of 10% risk would be suitable for application of the models [2]. Using a threshold level of 10% risk in the Dutch donor cohort, 66.0% of all deferrals in men can be prevented. At the same time, for 10.6% (i.e., $100 - 20.1 = 79.9\%$) of

13.3%) of the male donors, the invitation is unnecessarily postponed. For women, 71.5% of the deferrals can be prevented, and 19.2% of women are unnecessarily not invited. When a decision threshold of 10% is used in the Swiss donor cohort, the number of deferrals that can be prevented is lower; however, the percentage of donors that is unnecessarily not invited is also lower. For men, 31.9% of all deferrals can be prevented and only 3.9% is unnecessarily not invited. For Swiss women, the respective percentages are 50.9% and 13.5%. When a decision threshold of 5% risk is used in the Swiss donor cohort, 51.7% of deferrals in men can be prevented against 9.2% unnecessary postponements. For women, 73.9% of deferrals can be prevented against 27.1% unnecessary postponements. The sensitivity at a 5% threshold level in the Swiss cohort is comparable with the sensitivity at a 10% threshold in the Dutch cohort; however, the percentages of unnecessary postponements are larger.

For comparison of the results from the current study with the Irish validation study, the main performance measures of the latter study are presented in table 2 as well. Compared to the validity of the Dutch models in Irish donors, calibration and discrimination were better in the Swiss donor cohort: the calibration slopes in the Swiss donor cohort were closer to the ideal value of 1, and the AUCs were higher than in the Irish donor cohort.

The lower performance of the Dutch models in Swiss donors compared to Dutch donors prompted us to update the models. We updated the models using recalibration and model revision methods. For the recalibrated models, all regression coefficients were multiplied by the slope of the calibration model (0.74 for men and 0.76 for women). The intercept was adjusted by multiplying the original value by the calibration slope and adding the accompanying intercept of the calibration model (−0.18 for men and 0.08 for women). To derive the revised models, regression coefficients of predictors that had added value in the recalibrated model were further adjusted. For men, only the regression coefficient for the predictor time since the previous visit was further adjusted. For women, regression coefficients of all predictors were further adjusted. The exact formulas of the recalibrated and revised models to calculate the risk of Hb deferral are given in Appendix III (see Supplemental Material, available at <http://content.karger.com/ProdukteDB/produkte.asp?doi=446817>). The adjusted regression coefficients in the revised models were generally lower than in the original models. This was especially true for previous Hb level. After updating, the models were (by definition) well calibrated (fig. 1a, b), and the model for men had slightly better discriminative ability than the original model (0.85 vs. 0.84).

Discussion

We assessed the validity of sex-specific Dutch prediction models for low Hb deferral in a cohort of Swiss whole blood donors. Validation of the models in Swiss donors showed lower, though adequate performance. Updating the models for the Swiss donor population improved calibration in both men and women although discrimination improved only slightly in men.

The discriminative ability of the Dutch prediction models was lower in the Swiss donor cohort. An explanation for a lower discriminative ability could be a difference in the distribution of predictive factors. If the distribution in the Swiss validation cohort would be more homogenous, the discriminative ability of the model would be automatically lower [12].

However, no substantial differences in the distribution of predictors between the Swiss and the Dutch donor cohorts could be observed. The lower discriminative ability in Swiss donors is therefore more likely the result of different predictor effects. Indeed, updating the prediction models by means of model revision resulted in new regression coefficients and thus different predictor effects. The predictive effects of most predictors, and particularly of previous Hb level, were weaker in Swiss donors. The lower calibration was also related to the weaker effect of previous Hb level.

Some other differences between the two cohorts may have an influence on the model validity. Mean Hb levels were higher among Swiss donors compared to Dutch donors. Although the difference in mean Hb levels was relatively small, this could affect the model validity. The higher Hb levels in Swiss donors may be due to a smaller donation volume and a lower donation frequency among Swiss donors. In Switzerland, a standard whole blood donation involves a donation of 450 ml whole blood, whereas in the Netherlands 500 ml whole blood is donated. In both countries another 30–50 ml whole blood is collected for diagnostic purposes. A difference of approximately 50 ml whole blood corresponds to a difference in iron loss of 20–25 mg. As the daily amount of iron loss is 1–2 mg, this difference is substantial. Beside this difference in donation volume, the maximum number of whole blood donations allowed per year for men is 4 in Switzerland and 5 in the Netherlands. For women however, the maximum number of whole blood donation allowed per year is 3 in both countries. Though, we observed a lower donation frequency, and accordingly longer time intervals between visits for both Swiss men and women compared to Dutch men and women. There may also be differences between the two cohorts for variables that are not included in the model but are related to Hb levels and thus Hb deferral. Such differences could also affect the model validity. This may for example be the case for dietary factors [13]. Another factor may be that Swiss donors live at higher altitude in the Alps than Dutch donors who live at sea level. Studies have shown that people living at high altitude have higher Hb levels than people living at sea level [14]. However, >90% of Swiss donors live between 500 m and 1000 m above sea level and virtually all donors live below 1500 m. At this altitude an hypoxia-related Hb increase is unlikely. Furthermore, although Hb levels were measured in fingerstick capillary samples with the same measurement device in both countries, there may be subtle differences in measurement methodology that could affect the measurement, e.g., ambient temperature, the exact place on the finger where the blood drop is collected, posture, and length of time the posture has been held.

We observed a lower deferral rate for Swiss men compared to Dutch men. This is in agreement with the observed higher mean Hb levels in Swiss men. However, mean Hb levels were also higher

in Swiss women, but Swiss women were deferred more frequently compared to Dutch women. An explanation for this apparent contradiction is a somewhat skewed distribution of Hb levels among Swiss women. This implies that despite the higher mean Hb levels, relatively more Swiss women have Hb levels below the cutoff level of 7.8 mmol/l compared to Dutch women.

Compared to the validity of the Dutch models in Irish donors, the models performed better in Swiss donors. An important difference between the Dutch development cohort and the Irish validation cohort was the lower Hb cutoff level for donation in the Irish cohort. In the Swiss donor cohort, the Hb cutoff level was the same as in the Dutch cohort, which may explain the better performance.

Despite a better performance compared to the Irish donor cohort, the performance in Swiss donors was worse compared to the performance in Dutch donors. This prompted us to update the Dutch models for the Swiss donors. We updated the models rather than fitting new prediction models to be able to use the information captured in the large development study [5]. Updating resulted in notably weaker effects for previous Hb level compared to the original Dutch models but only slightly weaker or similar effects for most other predictors. Updating the models improved calibration in both men and women. Discrimination improved only slightly in men: the AUC increased by 0.01 to 0.85 and for women the AUC remained at 0.79. For comparison purposes, we also refitted new models containing the same set of predictive factors in Swiss donors. AUCs for the refitted models were the same as for the updated models.

In validation studies, calibration and discrimination are the measures that are usually examined. The AUC contains information about the sensitivity and specificity over the entire range of decision thresholds to divide donors into groups with high versus low risk of Hb deferral. Application of the risk models for Hb deferral requires a choice of decision threshold of predicted risk above which donors are classified in the high risk group [15]. Thus, for a practical interpretation of the model validity other measures of accuracy, such as sensitivity and specificity at a given decision threshold, are also important. Using a low threshold value, i.e. a slightly increased risk is already considered as high (and donors are preventively not yet invited), more deferrals will be prevented (higher sensitivity), but at the cost that a lot of donors who could have donated are not invited for a donation (lower specificity). In the Swiss donor cohort, at different decision thresholds, the sensitivity was lower, whereas the specificity was higher compared to the Dutch cohort. This means that in general, among Swiss donors less deferrals can be prevented; however, there are also less donors unnecessarily postponed for donation. Based on the percentages of deferrals that could be prevented and the percentages of unnecessary postponements in the Dutch development cohort, we concluded that a decision threshold of 10% would be suitable for practical application of the models [2]. To prevent the same percentage of deferrals in Swiss donors, a decision threshold of 5% is neces-

sary. However, this is at cost of more donors who are unnecessarily postponed. It depends on the available blood stock level and the number of replacement donors which decision threshold should be used in practice in either the Netherlands or in Switzerland.

A strength of this study is the large number of donors in the validation cohort. A limitation is that the models were developed and validated only in blood donors who had given at least two whole blood donations prior to the intended visit. This criterion was used in order to study the effect of blood donation. However, first-time donors and donors who donated only once were not included, and the performance of the models in this group of donors remains therefore unknown.

The numbers of deferrals that can be prevented and the unnecessary postponements that are presented in this paper are calculated with historical data. To assess the impact of using the models in practice on the management of the blood donation program, an impact study should be carried out in the near future [5, 16–18]. In an impact study, the effect of using the models and applying interventions for donors with a high risk of Hb deferral on the number of Hb deferrals, retained donors, blood stock levels, and costs can be compared with a setting in which the prediction models are not used. Positive results of an impact study will bring a strong support for implementation of the prediction models in practice.

Conclusion

The performance of Dutch prediction models in Swiss whole blood donors was lower than in Dutch donors, although the performance was still adequate. Updating the prediction models for the Swiss donor population improved the performance. In general, the Dutch prediction models may be a useful tool that could be helpful in the management of the blood donation program. However, donor characteristics or predictive effect of the predictors may differ between countries. Therefore, small adaptations of the Dutch models are necessary for application in other countries. The models could be adjusted by revision of the Dutch models, or, when the donor population is larger than the Dutch donor population used for the development of the models, models could be refitted. Also, the decision threshold of predicted risk above which donors are classified in the high risk group could be different in different countries. To investigate the practical impact on blood donor management, an impact study should be performed, preferably in different countries. If results are positive, the prediction models may be implemented in practice in tailoring donation intervals in the invitation process of blood donors.

Disclosure Statement

The authors declare that they have no conflicts of interest.

References

- 1 European Directorate for the Quality of Medicines and Healthcare: Guide to the Preparation, Use and Quality Assurance of Blood Components, European Committee (partial agreement) on Blood Transfusion (CD-P-TS), Recommendation No. R(95) 15, 17th e. Strasbourg, Council of Europe, 2013, pp 265–298.
- 2 Baart AM, de Kort WL, Atsma F, Moons KG, Vergouwe Y: Development and validation of a prediction model for low hemoglobin deferral in a large cohort of whole blood donors. *Transfusion* 2012;52: 2559–2569.
- 3 Altman DG, Vergouwe Y, Royston P, Moons KG: Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
- 4 Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE: Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98:683–690.
- 5 Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M: Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691–698.
- 6 Baart AM, Atsma F, McSweeney EN, Moons KG, Vergouwe Y, de Kort WL: External validation and updating of a Dutch prediction model for low hemoglobin deferral in Irish whole blood donors. *Transfusion* 2014;54:762–769.
- 7 The Commission of the European Communities: European Commission Directive 2004/33/EC. 2004. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:091:0025:0039:EN:PDF> (last accessed September 29, 2016).
- 8 Marshall A, Altman DG, Royston P, Holder RL: Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol* 2010;10:7.
- 9 Justice AC, Covinsky KE, Berlin JA: Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515–524.
- 10 Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD: Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;23:2567–2586.
- 11 Janssen KJ, Moons KG, 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.
- 12 Vergouwe Y, Moons KG, Steyerberg EW: External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol* 2010;172:971–980.
- 13 Zimmermann MB, Hurrell RF: Nutritional iron deficiency. *Lancet* 2007;370:511–520.
- 14 Beall CM, Brittenham GM, Strohl KP, Blangero J, Williams-Blangero S, Goldstein MC, Decker MJ, Vargas E, Villena M, Soria R, Alarcon AM, Gonzales C: Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. *Am J Phys Anthropol* 1998;106:385–400.
- 15 Steyerberg EW: *Clinical Prediction Models*. New York, Springer, 2009.
- 16 Reilly BM, Evans AT: Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144:201–209.
- 17 Toll DB, Janssen KJ, Vergouwe Y, Moons KG: Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085–1094.
- 18 Moons KG, Altman DG, Vergouwe Y, Royston P: Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.
- 19 Radtke H, Tegtmeyer J, Rocker L, Salama A, Kiesewetter H: Daily doses of 20 mg of elemental iron compensate for iron loss in regular blood donors: a randomized, double-blind, placebo-controlled study. *Transfusion* 2004;44:1427–1432.
- 20 Magnussen K, Bork N, Asmussen L: The effect of a standardized protocol for iron supplementation to blood donors low in hemoglobin concentration. *Transfusion* 2008;48:749–754.
- 21 Hoekstra T, Veldhuizen I, van Noord PA, de Kort WL: Seasonal influences on hemoglobin levels and deferral rates in whole-blood and plasma donors. *Transfusion* 2007;47:895–900.