Received: 2 November 2022 | Revised: 4 February 2023 | Accepted: 1 March 2023

DOI: 10.1111/vox.13426

ORIGINAL ARTICLE

Vox Sanguinis ISBT International Society of Blood Transfusion

An international comparison of haemoglobin deferral prediction models for blood banking

Marieke Vinkenoog[1,2] | Jarkko Toivonen[3] | Tinus Brits[4] | Dorien de Clippel[5] | Veerle Compernolle[5,6] | Surendra Karki[7] | Marijke Welvaert[7] | Amber Meulenbeld[1] | Katja van den Hurk[1] | Joost van Rosmalen[8,9] | Emmanuel Lesaffre[10] | Mikko Arvas[3] | Mart Janssen[1]

[1]Donor Medicine Research, Sanquin Research, Amsterdam, The Netherlands

[2]Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

[3]Research and Development, Finnish Red Cross Blood Service, Helsinki, Finland

[4]Business Intelligence, South African National Blood Service, Johannesburg, South Africa

[5]Dienst voor het Bloed, Belgian Red Cross Ugent, Ghent, Belgium

[6]Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

[7]Research and Development, Australian Red Cross Lifeblood, Sydney, Australia

[8]Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands

[9]Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands

[10]L-Biostat, KU Leuven, Leuven, Belgium

Correspondence
Marieke Vinkenoog, Plesmanlaan 125Y, Amsterdam 1066CX, the Netherlands.
Email: m.vinkenoog@sanquin.nl

Funding information
Australian Government; Punainen Risti Veripalvelu; Stichting Sanquin Bloedvoorziening, Grant/Award Number: PPOC 18-14/L2337

Abstract

Background and Objectives: Blood banks use a haemoglobin (Hb) threshold before blood donation to minimize donors' risk of anaemia. Hb prediction models may guide decisions on which donors to invite, and should ideally also be generally applicable, thus in different countries and settings. In this paper, we compare the outcome of various prediction models in different settings and highlight differences and similarities.

Materials and Methods: Donation data of repeat donors from the past 5 years of Australia, Belgium, Finland, the Netherlands and South Africa were used to fit five identical prediction models: logistic regression, random forest, support vector machine, linear mixed model and dynamic linear mixed model. Only donors with five or more donation attempts were included to ensure having informative data from all donors. Analyses were performed for men and women separately and outcomes compared.

Results: Within countries and overall, different models perform similarly well. However, there are substantial differences in model performance between countries, and there is a positive association between the deferral rate in a country and the ability to predict donor deferral. Nonetheless, the importance of predictor variables across countries is similar and is highest for the previous Hb level.

Conclusion: The limited impact of model architecture and country indicates that all models show similar relationships between the predictor variables and donor deferral. Donor deferral is found to be better predictable in countries with high deferral rates. Therefore, such countries may benefit more from deferral prediction models than those with low deferral rates.

Marieke Vinkenoog and Jarkko Toivonen shared first authorship.

Mikko Arvas and Mart Janssen shared last authorship.

Keywords
donor health, haemoglobin deferral, haemoglobin measurement, prediction

Highlights

- Within countries, different haemoglobin deferral prediction models perform similarly well.
- The relative importance of predictor variables is very similar across countries.
- Performance of models in different settings is dependent on the deferral rate. As a result, prediction models may be of higher value in countries with higher deferral rates.

INTRODUCTION

To avoid blood donations by donors at risk of becoming anaemic, blood banks test the donors' haemoglobin (Hb) levels. In case of pre-donation testing, a low Hb level leads to on-site deferral, which is demotivating for donors and makes them less likely to return to the blood bank than non-deferred donors [1, 2]. Additionally, it is in the interest of blood banks to keep deferral rates low to save time and costs. The ability to accurately predict low Hb deferral and adjust donation intervals based on these predictions likely decreases deferral rates. In the last 15 years, various Hb deferral prediction models, such as multiple logistic regression models [3], Bayesian linear mixed models (LMM) [4, 5] and ensemble models [6], have been evaluated by blood banks. Most prediction models use donors' previous Hb measurements in combination with donor characteristics such as age and sex, but the prediction accuracy has been modest. Nonetheless, even models with modest accuracies could be beneficial in practice [5]. Accurate prediction of Hb levels and/or deferral remains a difficult task, as many factors affect Hb, and both intra- and inter-individual variation is large. Therefore, it stands to reason that machine learning models might improve the prediction accuracy over the traditional regression models, as they are capable of learning more complex associations between predictors and outcome variables. Support vector machines (SVMs) have been shown to predict Hb deferral in Dutch donors reasonably well [7], as do random forests (RFs) in Finnish donors [5].

Most prediction models are developed and validated on donation data of a single country [3, 6]. Between countries, sets of available predictor variables differ widely. Ferritin levels, genotyping data, smoking status and iron supplementation are examples of variables that are associated with Hb levels but are not systematically measured or recorded by most blood banks [8]. Therefore, prediction models using such variables cannot be applied to data from other blood banks. Additionally, differences in blood bank policies regarding donor deferral require models to be calibrated for each country separately.

The SanguinStats group is a collaboration of statisticians and epidemiologists from several countries carrying out research in the area of donor health. It currently consists of researchers from blood banks in Australia, Belgium, Denmark, Finland, the Netherlands, South Africa and the United Kingdom, as well as researchers with statistical expertise who are associated with research institutes other than blood banks. The aim of the SanguinStats group is to combine the available expertise and data sources to develop and evaluate the outcome of state-of-the-art models in various settings.

In this first joint paper, we present a comparison of various Hb deferral prediction models on data from five blood banks. The goal of this research is not to create the best performing predictor, but rather to use exactly the same models for all datasets and to compare the performance and importance of variables between countries. Therefore, only basic predictor variables that are available in all individual countries are included in the models. Comparing the importance of variables between countries will show whether models show the same relationships between the variables and Hb deferral.

This is the first study to compare multiple Hb deferral prediction models on datasets from multiple countries. The results can be used by other blood banks to anticipate benefits from collecting additional measurement data and the use of various predictors for the prediction of donor deferral.

MATERIALS AND METHODS

Data sources and variables

Within each country, data were extracted from the blood banks' database, selecting data from whole blood donors from the past 5 years. The exact years differ per country because of the availability of up-to-date datasets. For each country, the timeframe of data collection was carefully selected to minimize iron-related blood bank policy changes in the dataset. In Australia, Finland and the Netherlands, there is one national blood bank (Australian Red Cross Lifeblood, Finnish Red Cross Blood Service and Sanquin Blood Bank, respectively), and data from these blood banks were used. In Belgium, data from Red Cross Flanders were used, which covers the whole of Flanders. In South Africa, data from South Africa National Blood Service were used, which is the major blood bank in the country.

For this study, only donors with five or more donation attempts were included to balance the trade-off between prediction accuracy (which has been shown to decrease with shorter time series at least in LMM) and data availability, as data becomes scarcer with higher thresholds of minimum donation numbers [5].

The following donation-level variables are used in the prediction models:

- Donor age ('Age')
- Days to previous donation ('Days to previous whole blood donation')
- Time of day at the start of the donation ('Time')
- Hb level at first donation ('First Hb') (not used by dynamic linear mixed model [DLMM])
- Hb level at previous donation ('Previous Hb') (not used by linear mixed model [LMM])
- Low Hb at previous donation ('Previous visit low Hb')
- Warm season (April–September for Northern hemisphere and October–March for Southern hemisphere) ('Warm season')
- Number of consecutive deferrals since previous successful donation ('Consecutive deferrals')
- Number of successful donations in last 5 years ('Recent donations')
- Number of low Hb measurements in the last 2 years ('Recent low Hb')

Models were fitted separately for male and female donors. Unless otherwise specified, the analyses presented in this study were performed on a random subset of 10,000 donors per sex, to prevent differences in model performance between countries due to different dataset sizes. The outcome is a dichotomous variable: deferral or non-deferral.

Statistical methods

Five prediction models were compared in this study: a baseline model, RF, SVM, LMM and DLMM. Note that these models are fundamentally very different. Each of the models is briefly described below.

The baseline model is a simple logistic regression model that estimates the likelihood of deferral as a function of only the Hb level at the previous donation.

RF is a classification algorithm that consists of several decision trees, fitted on sub-samples of the data. It uses averaging to improve predictive accuracy and prevent overfitting. The prediction output of an RF is the class selected by the majority of the decision trees. The RF takes as input all predictor variables listed in the previous section.

SVM is a classification algorithm that aims to find the best hyperplane to separate both outcome classes in a multi-dimensional space. The SVM again takes all predictor variables listed in the previous section as input. Note that none of the three models mentioned above explicitly models the subsequent donations, but rather uses aggregated information on donation history (see list above). This is where these differ from LMM and DLMM, which include a donor-specific intercept as the only random effect.

LMM does not include previous Hb as a predictor, but instead uses the first Hb level. DLMM, however, does include the previous Hb as a predictor. Both LMM and DLMM are regression models that predict not Hb deferral but the actual Hb level. If this predicted Hb level is lower than the country-specific donation threshold, deferral is predicted. These LMMs were trained in a Bayesian setting with weakly informative conjugate priors. They are described in more detail in a previous article [5], and they are essentially simplified versions of the models proposed by Nasserinejad et al. [4], excluding the modelling of the temporary reduction in Hb after blood donation.

Model performance is assessed using the area under the precision–recall (AUPR) curve. As no perfect model exists, each model provides an estimate of the probability of deferring a donor. Depending on the probability that is applied as a classification threshold (so anyone with a higher probability of deferral is labelled 'deferral' and the others 'non-deferral'), a different number of correct and incorrect predictions will be found. The precision–recall curve is a graph in which the recall versus the precision of a prediction model at varying classification thresholds is shown, where precision is the proportion of correctly predicted deferrals of all predicted deferrals and recall is the proportion of all deferred donors that were correctly labelled as such. The higher the AUPR curve, the better the prediction model's performance. To fairly compare AUPR across countries, we adjusted the AUPR values by subtracting the countries' deferral rate. The adjusted value now indicates the improvement by the model over always predicting non-deferral.

SHapley Additive exPlanations (SHAP) values were used to quantify the contribution of each predictor variable to the prediction for each individual observation [9]. Because SHAP values are model-agnostic, they can be calculated and compared for each model. This results in variable importance measures even for models that do not have interpretable coefficients, such as RF and SVM.

Docker container

To ensure that all collaborators perform exactly the same analyses, but without having to export data outside of their organization or between jurisdictions, we implemented all models for Hb-deferral prediction in a Docker container whose development was started earlier [5]. The Docker platform is easy to install on all major operating systems. After installation, the Docker container image can be downloaded and the user can run all models presented in this paper in a secure environment (without requiring an internet connection). For this study, we added an implementation of the SVM to the container, in addition to some specific improvements to facilitate the comparison of outputs. Both the ready-to-use container image [10] and its source code [11] are freely available through Dockerhub and Github, respectively. All analyses presented in this paper were obtained using version 0.32 of the container. Analyses of the results were performed using the R language and environment for statistical computing (version 4.2.0) [12], using packages dplyr (version 1.0.9) [13] and tidyr (version 1.2.0) [14] to handle data, and ggplot2 (version 3.3.6) [15] to create graphs.

RESULTS

Table 1 shows the distribution of the predictor variables in all countries.

TABLE 1 Distributions of predictor variables in all four datasets.

Visits by male donors					
Variable	**Australia**	**Belgium**	**Finland**	**Netherlands**	**South Africa**
Number of donors	10,000	8552	10,000	10,000	10,000
Age in years	41 (29–54)	39 (25–52)	53 (41–60)	52 (39–60)	44 (33–54)
Mean consecutive deferrals	0.003	0.025	0.018	0.029	0.213
Days to previous donation	98 (84–167)	99 (90–182)	106 (77–168)	92 (70–147)	73 (59–118)
Hb in g/L	149 (142–157)	153 (147–159)	154 (147–162)	148 (142–156)	153 (142–163)
Proportion of Hb deferrals	0.004	0.022	0.018	0.029	0.129
First Hb level in g/L	150 (143–158)	154 (148–160)	155 (147–162)	150 (143–158)	153 (140–163)
Time of day as hour between 0 and 24	13.1 (10.8–15.6)	18.9 (17.8–19.7)	14.8 (13.1–16.4)	16.3 (13.1–18.7)	12.8 (11.2–14.6)
Hb level at previous visit in g/L	148 (139–156)	151 (143–158)	153 (144–161)	148 (140–155)	151 (137–162)
Proportion of low Hb at previous visit	0.003	0.020	0.018	0.030	0.124
Mean recent low Hb	0.008	0.066	0.074	0.127	0.553
Recent donations	4 (2–6)	4 (2–6)	5 (2–9)	5 (2–9)	4 (2–7)
Warm season proportion	0.500	0.477	0.491	0.494	0.524
Visits by female donors					
Variable	**Australia**	**Belgium**	**Finland**	**Netherlands**	**South Africa**
Number of donors	10,000	9028	10,000	10,000	10,000
Age in years	37 (25–50)	34 (21–47)	50 (35–58)	47 (31–57)	41 (31–52)
Mean consecutive deferrals	0.016	0.131	0.040	0.057	0.146
Days to previous donation	104 (87–183)	111 (91–196)	140 (106–224)	154 (132–222)	84 (62–156)
Hb in g/L	133 (127–140)	135 (130–141)	140 (133–147)	135 (129–142)	136 (128–144)
Proportion of Hb deferrals	0.021	0.106	0.038	0.054	0.141
First Hb level in g/L	134 (128–141)	137 (132–143)	141 (134–147)	135 (129–142)	135 (128–144)
Time of day as hour between 0 and 24	13.1 (11.0–15.4)	18.7 (17.7–19.6)	15.3 (13.5–16.6)	15.5 (13.1–18.3)	13.0 (11.4–14.7)
Hb level at previous visit in g/L	131 (124–138)	134 (126–140)	138 (130–146)	134 (127–142)	134 (125–143)
Proportion of low Hb at previous visit	0.017	0.096	0.040	0.057	0.127
Mean recent low Hb	0.039	0.287	0.110	0.151	0.390
Recent donations	3 (1–5)	3 (1–5)	4 (2–6)	3 (1–6)	3 (1–6)
Warm season proportion	0.506	0.472	0.504	0.491	0.523

Note: Numerical variables are described by their median and (interquartile range) unless otherwise stated. Dichotomous variables are described by the proportion of visits where the value was true.
Abbreviation: Hb, haemoglobin.

TABLE 2 Haemoglobin (Hb) measurement and donor deferral policies per country.

Country	When and how is Hb measured?	When is the donor deferred?
Australia	Capillary skin-prick Hb measurement by haemoglobinometer before each donation. If the Hb is below the threshold, a venous sample is taken from the non-donation arm and Hb is measured using the haemoglobinometer at the donation site to confirm.	Hb levels below 120 g/L (women) or below 130 g/L (men) as well as donors with a 20 g/L drop in Hb level relative to their previous donation.
Belgium	Haematology analyser Hb measurement from venous sample after every successful donation. Capillary skin-prick Hb measurement before donation for new donors and for donors with a venous Hb below the eligibility threshold at the previous donation.	Hb level below 125 g/L (women) or below 135 g/L (men) at previous and current donation.
Finland	Capillary skin-prick Hb measurement point of care (POC) before each donation. If the Hb is below threshold, venous sample is taken and Hb measured by POC device at donation site [19].	Hb level below 125 g/L (women) or below 135 g/L (men) as well as donors with a 20 g/L drop in Hb level relative to their previous donation.
The Netherlands	Capillary skin-prick Hb measurement before each donation. If a Hb level is below the threshold, the measurement is repeated (up to three times in total). The highest value is used for the deferral decision. Since late 2017, donors are also deferred for low ferritin levels.	Hb level below 125 g/L (women) or below 135 g/L (men).
South Africa	Capillary skin-prick Hb measurement before each donation.	Hb level below 120 g/L (women) or below 130 g/L (men). Before 2020, cut-off levels of 125 and 135 g/L were used.

Hb measurement and deferral policies

All participating countries use Hb measurements to defer donors, but there are differences in how Hb is measured and when donors are deferred. Table 2 shows a summary of Hb-deferral-related policies per country.

Comparison of model performance

Figure 1 shows the AUPR values (adjusted for deferral rate) and their confidence intervals for all models for all countries. All models outperform the baseline model in all countries. Performance of different models does not differ greatly within one country, except for Australian female donors, for which RF and SVM clearly outperform the LMM and DLMM. The same pattern is visible in South African male donors, although less obvious, and slightly in Belgium. In general, variation in within-country model performance is much smaller than variation between countries. Belgium and South Africa obtain significantly higher AUPR values than the other three countries in all models, except for the high-performing RF and SVM on Australian female donors.

Tables 3 and 4 show the predicted versus observed outcomes of the model with the lowest AUPR (baseline model, female donors, Finland; unadjusted AUPR = 0.07) and the model with the highest AUPR (RF, male donors, South Africa; unadjusted AUPR = 0.69) to illustrate the AUPRs with actual case counts to make the results more tangible.

Figure 2 shows the deferral rate and AUPR for all countries and models. Even though the AUPR values are adjusted for the deferral rate, there is still a positive correlation between deferral rate and (adjusted) AUPR. All models show the same pattern for this association. Again, we see that for Australian female donors the RF and SVM obtain a much higher AUPR than expected based on the deferral rate.

To further investigate whether the low deferral rates indeed affect the ability of the models to predict deferral, we intentionally modified the deferral rate of the Belgian datasets by removing a varying proportion of the deferred donors from the dataset and refitting the models on these adapted datasets. The results are shown in Figure 3. This figure clearly shows the positive association between deferral rate and AUPR. There is no monotonically increasing association even though the datasets with lower deferral rates are subsets of the datasets with larger deferral rates. The fact that classification tasks are more difficult when there is a large imbalance between outcome classes is a well-known phenomenon in statistics [16].

Importance of individual variables

Figure 4 shows the variable importances derived from SHAP values calculated on a random subset of 1000 donors from the validation data. Variable importances are presented as mean absolute attribution (MAA) values. Variables are sorted by MAA over all countries and models (represented by the horizontal bars). For each individual country, the MAA values are provided and connected by a line.

RF and SVM

Comparing variable importances between countries within the same model allows identification of differences in predictive power of individual model parameters. In the RF and SVM models, previous Hb is the most important predictor for all countries and sexes and has almost twice the MAA of the second-most important predictor. The MAA for most variables is similar across countries. There are some

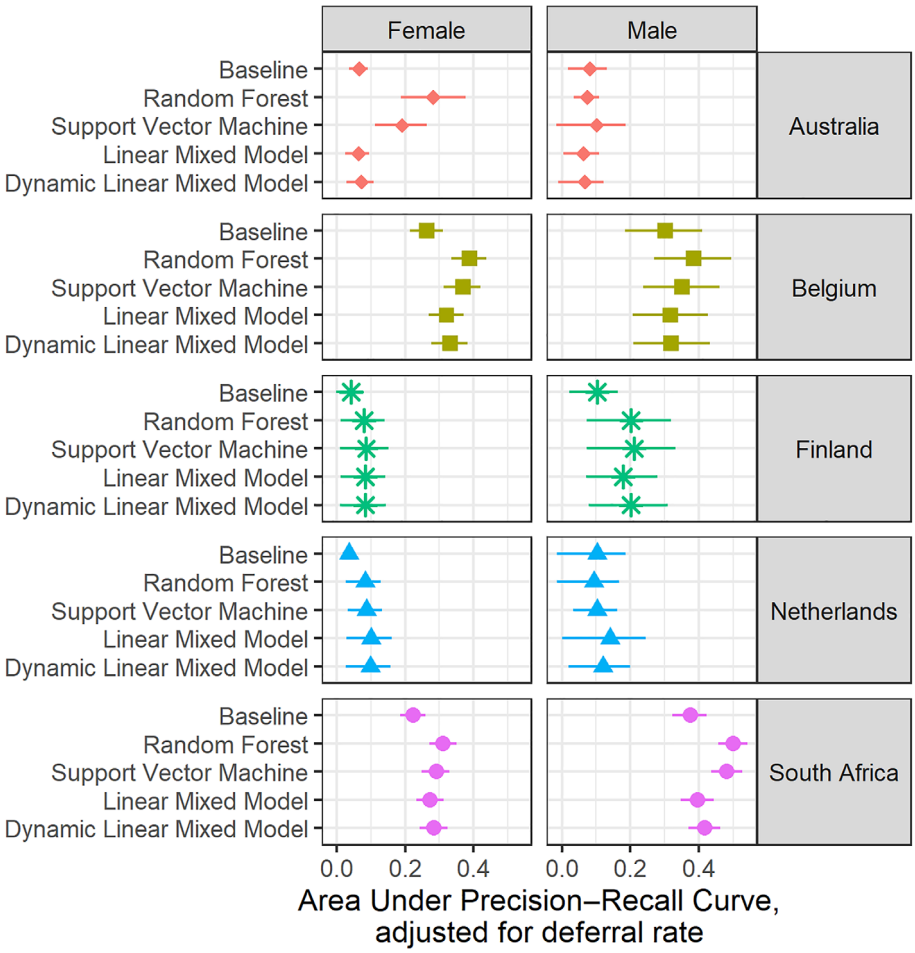

FIGURE 1 Area under the precision–recall (AUPR) curve for all countries and all models. Note that each AUPR curve is adjusted by subtraction of the country's deferral rate.

TABLE 3 Observed versus predicted outcomes of the baseline model applied to female Finnish donors.

Observed outcome		
Predicted outcome	**Accepted**	**Deferred**
Accepted	1146	10
Deferred	807	37

Note: This is the model with the lowest area under the precision–recall (0.07). The precision of class deferral is 0.04 and the recall is 0.79.

TABLE 4 Observed versus predicted outcomes of the random forest model applied to male South African donors.

Observed outcome		
Predicted outcome	**Accepted**	**Deferred**
Accepted	1433	108
Deferred	195	264

Note: This is the model with the highest area under the precision–recall (0.69). The precision of class deferral is 0.58 and the recall is 0.71.

exceptions, however: for South Africa, the number of recent low Hb measurements is much more important than in other countries, as well as the deferral status of the previous blood bank visit. For Belgium, whether the donation visit took place during the warm season is more important than in the other countries.

Linear and dynamic linear mixed models

For the LMMs, the MAA of variables show the highest similarity between countries. A donor's first Hb measurement is the most important predictor, and all other predictor variables have a relatively low MAA in comparison. Conversely, for DLMMs, there is much more variation in MAA values between countries and between sexes. For female donors, the most important predictor is age, and previous Hb is only the third-most important predictor, which deviates considerably from what was found for all other models. In both LMM and DLLM, the difference in MAA for age between sexes is much larger than in RF and SVM models.

Unlike the RF and SVM models, the LMM and DLMM estimate regression coefficients that may be compared across countries. For

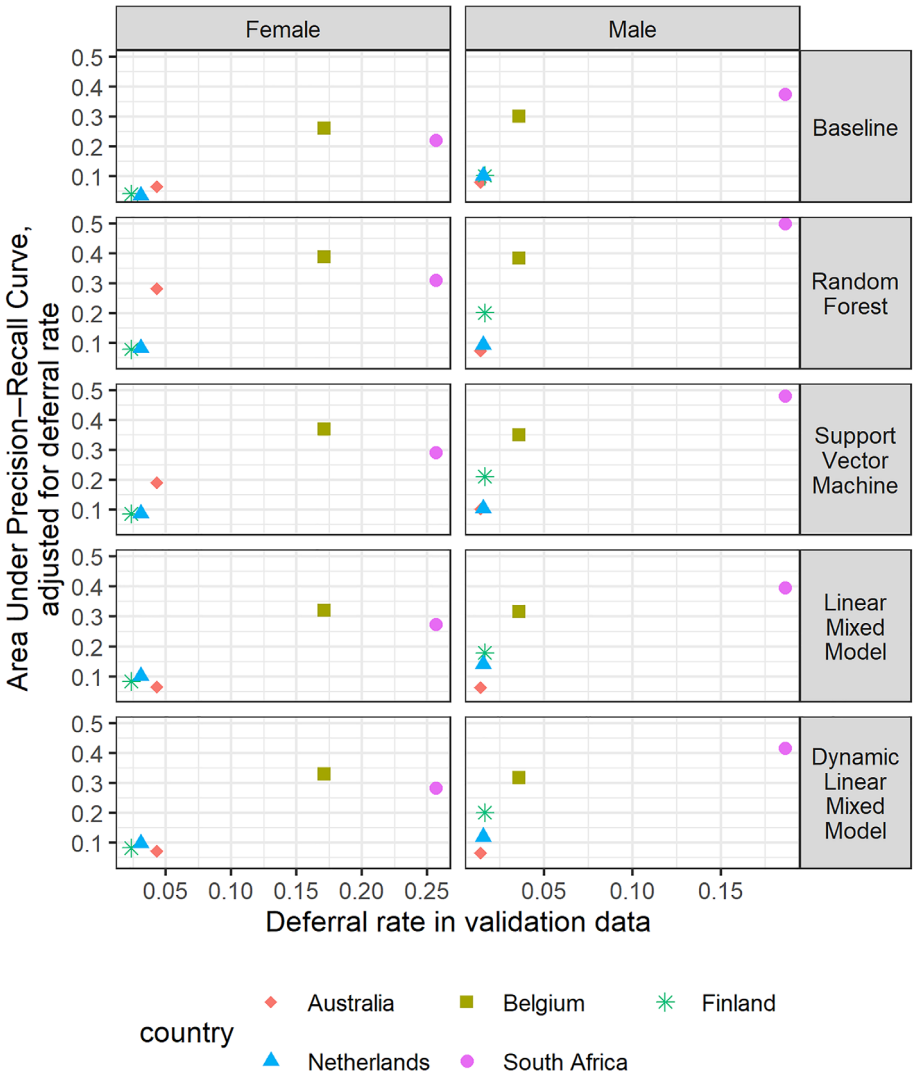

FIGURE 2 Adjusted area under the precision–recall value versus deferral rate in various settings for various models.

consistency with other model results, we compared the MAA output rather than regression coefficients. A comparison of regression coefficients can be found in Supplementary Material. For all variables except for 'Low Hb at previous visit' (which is the second to last most important predictor), coefficients are very similar between countries and 95% highest posterior density intervals mostly overlap.

Absolute value of MAA per model

It should be noted that the MAA values for different models are on different scales. In the baseline and SVM, SHAP values are on the log-odds scale, while for the RF and (dynamic) LMM, these are expressed on the probability scale. Since only the relative size of MAA values within models are compared, the difference in scales has no effect on the interpretation of the results.

The effect of sample size

We fitted the same models as above on the full datasets from Finland, the Netherlands and Australia to see whether this improves performance. This experiment showed that using the full dataset increases performance only by a very small amount and within the size of the confidence interval for the subsample of 10,000 donors.

DISCUSSION

In this paper, various prediction models for Hb deferral were applied to blood bank visit data from five countries to investigate the performance of prediction models in different settings. In all countries, the baseline was outperformed by all other models, although the overall performance was quite low for all models in all countries. Model performance, however, varies considerably between countries, and a high deferral rate is associated with better model performance. The relative

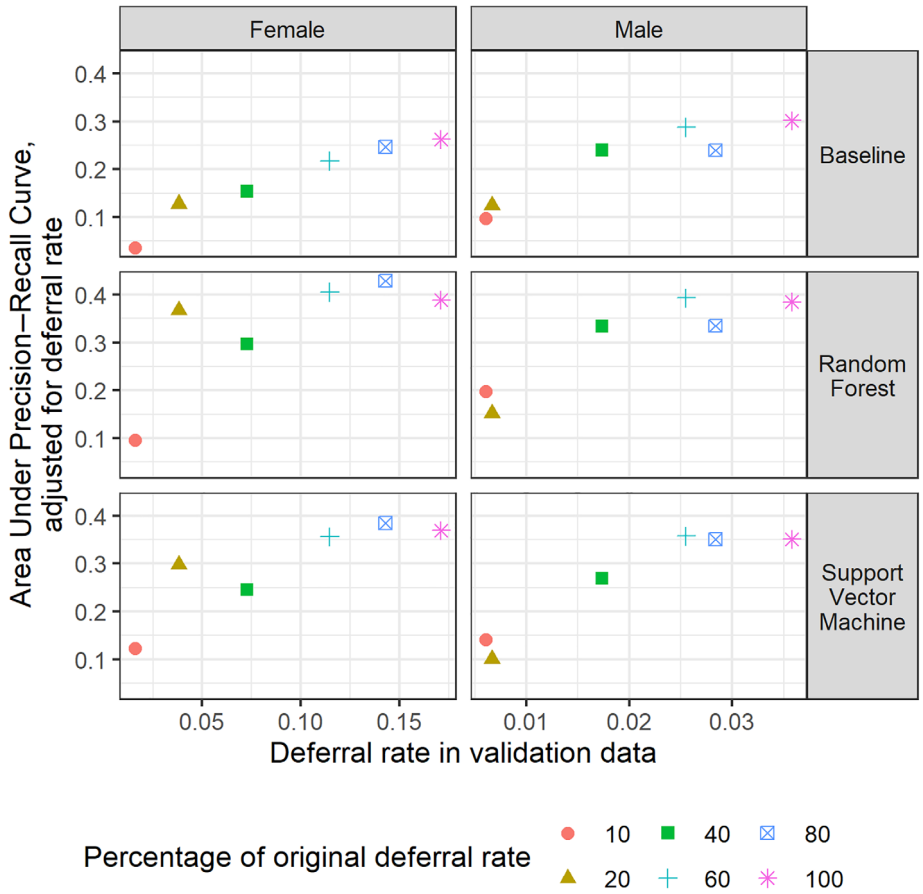

FIGURE 3 Adjusted area under the precision–recall as a function of the deferral rate for various deferral levels in the Belgian dataset. The reduction in deferral rate was obtained by sequentially removing an increasing number of deferred donations from the data.

importance of individual predictors is very similar in different countries. In particular, the Hb level at previous donation is an important predictor for donor deferral in almost all models. This indicates that models learn the same associations in different settings, which supports the idea that these associations are the result of similar biological processes underlying donor deferral.

The similarity of the relative importance of predictors also indicates that the differences in performance are not caused by different associations between predictors and Hb deferral. Rather, deferrals are more difficult to predict in countries with low deferral rates as there are fewer deferrals. The experiment with the Belgian data, which shows that the predictability collapses with a decrease in deferral rate, supports this finding. However, there appears to be an exception with the Australian data on female donors, where a relatively high AUPR is obtained for two models despite the very low deferral rate. Another possible explanation for the difference in performance could be that data collected in some countries is more informative than in others, for instance due to differences in the accuracy of Hb measurements and/or differences in deferral policies. However, we were unable to confirm this as a plausible hypothesis: Hb deferral is based on the same capillary measurement in South Africa and the Netherlands, and yet model performance on South African data is much higher than on Dutch data.

This study is the first to compare prediction models for Hb deferral across different settings. By focusing on the comparison of models between countries rather than optimizing model performance based on variables available within a single country, the effect of the setting on model performance becomes visible. We show that low deferral rates substantially limit model performance, although they do not hinder the model in learning the same associations as with higher deferral rates. Comparing results for male donors from Australia and South Africa illustrates this perfectly: the deferral rate in South Africa is more than 10-fold than in Australia (18.6% vs. 1.4%), resulting in a much higher AUPR (0.50 vs. 0.08 for RF), yet the variable importance is very similar.

Our findings are also in line with previously published work on Hb deferral prediction, which consistently shows that previous Hb measurements are by far the most important predictor [3, 5, 8]. Another interesting finding is that LMM, which is the only model to use a donor's first Hb instead of the previous Hb, performs just as well as the other models. This may indicate that most donors' Hb levels are quite stable over time, and that predictions of personalized donation intervals can already be made after a first Hb measurement at donor intake. To account for sudden drops in Hb level, inclusion of the previous Hb seems to be more relevant. The importance of first Hb levels is also shown by others [17], which indicates that iron dynamics

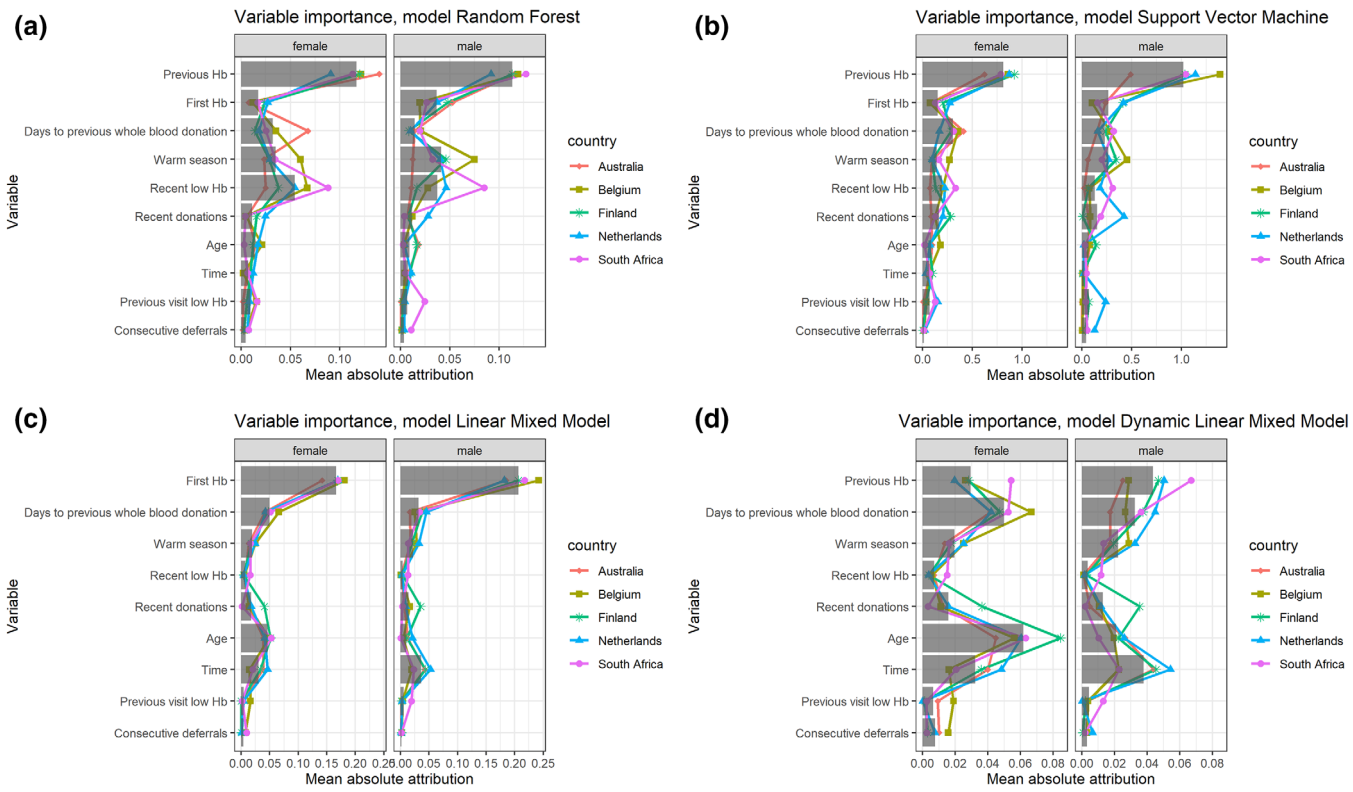

FIGURE 4 (a–d) Variable importance (mean value and per individual country) determined by the mean absolute attribution according to SHapley Additive exPlanations values for various models and sexes. The bars indicate the mean over all countries. Variables are ordered by the mean mean absolute attribution over both sexes and all models. Hb, haemoglobin.

(Hb and ferritin levels) in blood donors can be predicted over a longer period from the Hb and ferritin levels at donor intake.

Although this study offers new insights into the predictability of donor deferral in different settings, the actual predictive value of the models is low, which may be explained by the substantial variability in Hb measurement outcomes [18]. Note also that all analyses were done on donors with at least five donation attempts, which limits the generalizability of the models to the full donor population. Many blood banks collect more variables than were used in the predictions in this study and including those may improve model performance. Improved performance is paramount, as a model will create added value for the blood bank only when the benefits of the correctly predicted deferrals will outweigh the loss due to incorrectly predicted deferrals. The prediction of a potential reduction of donation intervals by some donors by the model may again add to the value of applying such prediction models.

Currently, the development of prediction models requires extensive expertise and data to enable prediction of donor deferral. Ideally, the work and insights developed by this collaboration would result in strategies that could also be of use to countries with limited resources.

In conclusion, this study shows that model architecture in most cases has a limited impact on the performance of prediction models for donor deferral, but in some cases, exemplified by Australia, certain model architectures can capture the data better than others. It would be recommended for any new country starting with Hb deferral prediction to try several architectures if possible. Adding better predictor variables to the different model could considerably improve predictive performance. Performance is strongly affected by the donor deferral rate. For most countries with low deferral rates, prediction models are unlikely to contribute to an effective reduction of donor deferral rates. Conversely, deferral prediction models may be applied in countries with high deferral rates to reduce on-site deferral of donors. Hb deferral remains a relevant topic, as it negatively affects both donors and blood services. By joining efforts, we can enhance our understanding of which generic factors affect donor deferral and to what extent. Also, only by studying the performance in different settings, organization-specific and operational characteristics may be identified that enhance or deteriorate prediction models' performance, which may indicate directions for further research and meaningful policy changes.

ACKNOWLEDGEMENTS

This work was funded by the Sanquin Blood Supply Foundation, PPOC grant 18-14/L2337 and Valtion tutkimusrahoitus (VTR) funding from the Finnish Government. Australian governments fund the Australian Red Cross Lifeblood to provide blood, blood products and services to the Australian community.

M.J., M.A., J.T., M.V., E.L., J.v.R., K.v.d.H., V.C., T.B., D.d.C. and S.K. designed the study; J.T. and M.V. developed the software; J.T.,

M.V., T.B., D.d.C., S.K. and M.W. analysed the data; M.V. aggregated the results; M.V. wrote the paper; M.V., J.T., T.B., D.d.C., V.C., S.K., M.W., A.M., K.v.d.H., J.v.R., E.L., M.A. and M.J. reviewed and edited the paper.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

ORCID

Marieke Vinkenoog https://orcid.org/0000-0001-5653-8078
Jarkko Toivonen https://orcid.org/0000-0002-6843-5831
Surendra Karki https://orcid.org/0000-0003-1561-4171
Joost van Rosmalen https://orcid.org/0000-0002-9187-244X
Mikko Arvas https://orcid.org/0000-0002-6902-8488
Mart Janssen https://orcid.org/0000-0002-1682-7817

REFERENCES

1. Spekman MLC, van Tilburg TG, Merz E-M. Do deferred donors continue their donations? A large-scale register study on whole blood donor return in The Netherlands. Transfusion. 2019;59:3657–65.
2. Custer B, Chinn A, Hirschler NV, Busch MP, Murphy EL. The consequences of temporary deferral on future whole blood donation. Transfusion. 2007;47:1514–23.
3. Baart AM, de Kort WLAM, Moons KGM, Vergouwe Y. Prediction of low haemoglobin levels in whole blood donors. Vox Sang. 2011;100:204–11.
4. Nasserinejad K, van Rosmalen J, de Kort W, Rizopoulos D, Lesaffre E. Prediction of hemoglobin in blood donors using a latent class mixed-effects transition model. Stat Med. 2016;35:581–94.
5. Toivonen J, Koski Y, Turkulainen E, Prinsze F, Parolo PDB, Heinonen M, et al. Prediction and impact of personalized donation intervals. Vox Sang. 2022;117:504–12.
6. Russell WA, Scheinker D, Custer B. Individualized risk trajectories for iron-related adverse outcomes in repeat blood donors. Transfusion. 2022;62:116–24.
7. Vinkenoog M, van Leeuwen M, Janssen MP. Explainable haemoglobin deferral predictions using machine learning models: interpretation and consequences for the blood supply. Vox Sang. 2022;117:1262–70.
8. Baart AM, Timmer T, de Kort WLAM, van den Hurk K. Lifestyle behaviours, ethnicity and menstruation have little added value in prediction models for low haemoglobin deferral in whole blood donors. Transfus Med. 2020;30:16–22.
9. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2:56–67.
10. Hb-predictor Docker Image via Docker Hub [cited 2022 Sep 27]. Available from https://hub.docker.com/r/toivoja/hb-predictor
11. GitHub source for Hb deferral prediction. 2022.
12. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
13. Wickham H, François R, Henry L, Müller K. RStudio. dplyr: A Grammar of Data Manipulation. 2022.
14. Wickham H, Girlich M. RStudio. tidyr: Tidy Messy Data. 2022.
15. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. 2022.
16. Anderson JA. Separate sample logistic discrimination. Biometrika. 1972;59:19–35.
17. Paalvast Y, Moazzen S, Sweegers M, Hogema B, Janssen M, van den Hurk K. A computational model for prediction of ferritin and haemoglobin levels in blood donors. Br J Haematol. 2022;199:143–52.
18. Janssen MP. Why the majority of on-site repeat donor deferrals are completely unwarranted…. Transfusion. 2022;62:2068–75.
19. Bäckman S, Valkeajärvi A, Korkalainen P, Arvas M, Castrén J. Venous sample is superior to repeated skin-prick testing in blood donor haemoglobin second-line screening. Vox Sang. 2020;115:617–23.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Vinkenoog M, Toivonen J, Brits T, de Clippel D, Compernolle V, Karki S, et al. An international comparison of haemoglobin deferral prediction models for blood banking. Vox Sang. 2023.