# The added value of ferritin levels and genetic markers for the prediction of hemoglobin deferral

**Marieke Vinkenoog[1,2], Jarkko Toivonen[3], Matthijs van Leeuwen[2], Mart P. Janssen[1], and Mikko Arvas[3]**

[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**

## ABSTRACT

**Background** - On-site hemoglobin deferral for blood donors is sometimes necessary for donor health, but demotivating for donors and inefficient for the blood bank. Deferral rates could be reduced by accurately predicting donors' hemoglobin status before they visit the blood bank. Although such predictive models have been published, there is ample room for improvement in predictive performance. We aim to assess the added value of ferritin levels or genetic markers as predictor variables in hemoglobin deferral prediction models.
**Methods** - Support vector machines with and without this information (the full and reduced model, respectively) are compared in Finland and the Netherlands. Genetic markers are available in the Finnish data; ferritin levels in the Dutch data.
**Results** - While there is a clear association with hemoglobin deferral for both ferritin levels and several genetic markers, predictive performance increases only marginally with their inclusion as predictors. The recall of deferrals increases from 68.6% to 69.9% with genetic markers and from 79.7% to 80.0% with ferritin levels included. Subgroup analyses show that the added value of these predictors is higher in specific subgroups: e.g., for donors with minor alleles on SNP 17:58358769, recall of deferral increases from 73.3% to 93.3%.
**Conclusions** - Including ferritin levels or genetic markers in hemoglobin deferral prediction models improves predictive performance. The increase in overall performance is small, but may be substantial for specific subgroups. We recommend including this information as predictor variables when available, but not to collect it for this purpose only.

## INTRODUCTION

Deferral of blood donors with low hemoglobin levels is necessary to prevent iron depletion. Currently, in Finland and the Netherlands, hemoglobin is measured before donation, and leads to on-site deferral if hemoglobin is below the donation threshold of 7.8 mmol/L (125 g/L) for women or 8.4 mmol/L (135 g/L) for men. On-site deferral is demotivating for donors and can be a reason to drop out of the donor pool permanently. [1] Hemoglobin deferral prediction models can help reduce the on-site deferral rate: for invitation-based donations, predictions can be included in the decision-making process of which donors to invite; for walk-in donations, the prediction could be communicated to the donor (e.g., shown on a donor dashboard or app that many blood banks offer), who can use this information to decide when to visit the blood bank.

Currently, hemoglobin deferral prediction models are not very accurate at predicting deferral on the specific day a donor may visit the blood bank. Although it is possible to correctly predict most deferrals as such (and therefore prevent them), this comes at the cost of incorrectly predicting some non-deferrals to be deferrals, which results in a large net loss of donations if these donors are then not invited to the blood bank based on this incorrect prediction. However, in a previous study we showed that

predicting hemoglobin deferral at different time points, and inviting a donor once the predicted outcome is 'non-deferral', results in non-deferred donors to be invited earlier and deferred donors to be invited later, thereby eliminating the loss of successful donations. [2] This tells us that hemoglobin deferral prediction models are useful, and it is worth the effort of trying to improve the predictions.

Multiple studies [3, 4, 5] have shown previous hemoglobin levels to be the most important predictor of future hemoglobin deferral. Researchers from blood services in different countries have investigated many different potential predictors of hemoglobin deferral, to assess whether the inclusion of these predictors improves prediction performance. Most of these predictors were found to not substantially improve the models: information on menstruation, diet, ethnicity, and smoking all only slightly improve model performance, even though they are known to be associated with iron stores. [4] One small-scale study on 261 donors did show that ferritin, soluble transferrin receptor, and hepcidin were associated with subsequent anemia. [5]

In this study we investigate the added value of including ferritin levels and genetic information in hemoglobin deferral prediction models. Ferritin is routinely measured at Sanquin, the Dutch national blood service, and therefore available for all donors. Genetic information for several iron-related SNPs is collected for many donors by the Finnish Red Cross blood service. Because the information in both countries is collected without targeting specific donors, our results provide a realistic indication of how much predictions would be improved if the prediction model was to be used in practice. Our results will therefore be useful for blood services that would like to collect additional donor information to improve hemoglobin deferral predictions.

## METHODS

### Data

Data on blood donation attempts by whole-blood donors from (almost) five recent years were extracted from the eProgesa database (MAK-SYSTEM, Paris, France) in Finland and the Netherlands. Only data from donors who explicitly provided informed consent for the use of their data for scientific research were used. This consent is given by more than 99% of all Dutch donors. All Finnish blood donors studied provided an informed consent for biobank research in accordance with the Finnish Biobank Act and the study was approved by the Blood Service Biobank (project 004-2019). In Finland, approximately 23% of active blood donors have given this consent since the founding of the Blood Service Biobank in 2017.

Finnish data reflects data entries from January 2016 through April 2020, Dutch data from January 2017 through December 2021. For each visit the following information was collected in both countries: donor sex, donor age, donation date, and hemoglobin level. Additionally, ferritin level is measured at every new donor intake and upon every fifth donation in repeat donors in the Netherlands.

In Finland, only donors participating in the Blood Service Biobank are included, as only for these donors, genetic information related to iron metabolism is available. [6] The four SNPs were identified as significantly associated with higher prevalence of iron deficiency anemia in an iron deficiency anemia meta-analysis on Finnish and UK data. Polygenic risk scores were derived for three related endpoints: iron deficiency anemia, ferritin, and hemoglobin. [7]

In total, complete information on the predictor variables (see Table 1) was available for 172 508 donation attempts by 42 255 donors in Finland, and 456 384 donation attempts by 157 423 donors in the Netherlands.

The variable of interest is 'HbOK', a dichotomous variable that indicates whether the result of the donation attempt was deferral (i.e., hemoglobin level below the eligibility threshold for donation) or non-deferral (i.e., hemoglobin level equal to or above the threshold).

Donor deferral due to low hemoglobin is similar in Finland and the Netherlands. Hemoglobin is measured using a capillary skin-prick device before each donation, and eligibility thresholds for donation are 7.8 mmol/L for women and 8.4 mmol/L for men. However, in case the measurement is below the eligibility threshold in Finland, hemoglobin is measured again (using the same device) in a venous sample, and this measurement is used for the deferral decision. In the Netherlands two additional capillary hemoglobin measurements are taken when the first measurement outcome is below the eligibility threshold, and the donor is allowed to donate if any of the three measurement outcomes is above the eligibility threshold.

| Variable | Unit or values | Description | Country/-ies |
|---|---|---|---|
| Sex | male, female | Biological sex of the donor; separate models are trained for men and women | Both |
| Age | years | Donor age at time of visit | Both |
| Month | 1-12 | Month of the year of the visit | Both |
| NumDon | donations | Number of successful (collected volume $>$ 250 ml) whole-blood donations in the last 24 months | Both |
| DaysSinceFirstDon | days | Number of days since the donor's first visit to the blood bank | Both |
| HbPrev*i* | mmol/L | Hemoglobin level at *i*th previous visit, for *i* between 1 and 5 | Both |
| DaysSinceHb*i* | days | Time since related hemoglobin measurement at *i*th previous visit, for *i* between 1 and 5 | Both |
| FerritinPrev | µg/L | Most recent ferritin level measured in this donor | Netherlands |
| SNP 1:169549811 | 0, 1, 2 | Number of minor alleles in SNP rs6025 | Finland |
| SNP 6:32617727 | 0, 1, 2 | Number of minor alleles in SNP rs3129761 | Finland |
| SNP 15:45095352 | 0, 1, 2 | Number of minor alleles in SNP rs199138 | Finland |
| SNP 17:58358769 | 0, 1, 2 | Number of minor alleles in SNP rs199598395 | Finland |
| PRS_anemia | standard deviations | Standardised polygenic risk score for anemia | Finland |
| PRS_ferritin | standard deviations | Standardised polygenic risk score for ferritin | Finland |
| PRS_hemoglobin | standard deviations | Standardised polygenic risk score for hemoglobin | Finland |

**Table 1.** Predictor variables available in each country.

## Analyses

For both countries, two models were fitted for each sex: one with all predictor variables available (the full model), and one with only those predictor variables that are available in both countries (the reduced model). By comparing the full model with the reduced model in both countries, the added value of extra predictor variables (i.e., genetic information in Finland and ferritin information in the Netherlands) can be assessed. The prediction models used were based on models developed for an earlier study considering Dutch data only. [2] All models are based on support vector machines (SVMs), supervised machine learning models that learn a separation between outcome classes from a training set, after which the model can be used to predict donor deferral for observations in an unseen test set. Here the training set consists of blood bank visits in the first four years of data, whereas the test set consists of data collected in the final year.

Given a dataset and a set of predictor variables, a model consists of ten SVM sub-models. The sub-models are named SVM-sex-*n*, where sex indicates donor sex (m for male, f for female donors) and *n* indicates the number of previous blood bank visits that are used for prediction. That is, each sub-model includes HbPrev*i* and DaysSinceHbi for *i* ranging from 1 to *n* as predictor variables. If sex is omitted in the sub-model name, it refers to the combination of two sex-specific sub-models. The number of blood bank visits (*n*) considered in this study varies from one through five, and so five sub-models per sex are created. Donors can only be included in the SVM-sex-*n* sub-model if they have at least *n* previous visits, therefore the sizes of the datasets used for both training and testing decrease from SVM-1 to SVM-5. Hyperparameters were optimised separately for each sub-model, using stratified (on the outcome variable) five-fold cross-validation within the training set data only. Hyperparameters were optimised using grid search, using the balanced accuracy (defined as the weighted average of recall in both classes) as scoring method, which is suitable for datasets with imbalanced outcome sizes, as mistakes in the minority class are penalised more than those in the majority class.

During model training, the classification threshold is chosen again by optimizing the balanced accuracy. The predictive performance of the models is assessed using precision (also known as positive predictive value) and recall (also known as sensitivity) at this classification threshold. For non-deferral prediction, precision is defined as the proportion of true non-deferrals out of all predicted non-deferrals; recall is defined as the proportion of predicted non-deferrals out of all true non-deferrals. In this context, the complement of the precision is the hypothetical new deferral rate if the model would be used to choose which donors to invite, and the complement of the recall is the proportion of successful donations that would be missed by the model because the donors are incorrectly predicted to have a low hemoglobin level. Precision and recall can be calculated for both outcome classes ('deferral' and 'non-deferral').

The precision-recall curve is a graph in which the recall and the precision of a prediction model at varying classification thresholds is shown. The AUPR is the area under this curve, a number between 0 and 1, where 1 would indicate a perfect classifier. By subtracting the deferral rate from the AUPR, we get an adjusted AUPR, which reflects the improvement by the model over a strategy that would always predict non-deferral. Without this correction the improvement made by the model would be biased by the difference in deferral rate. The AUPR represents the ability of the model to distinguish between two classes at differing classification thresholds. It is possible for model A to have a higher AUPR than model B, even if precision and recall at the optimal classification threshold are the same in both models.

## Model explanations

Because SVMs do not provide model coefficients that can be directly interpreted, we use Shapley Additive exPlanations (SHAP) values to investigate the importance of different predictor variables. [8] SHAP is a model agnostic explainer that shows the contribution of each predictor variable to the predicted outcome. This contribution is calculated for each individual observation separately (in a subsample of the test set) and is therefore very informative.

## Subgroup analysis

To further investigate the value of including ferritin and genetic information in the models, we perform additional analyses in which donors are placed in groups defined by ferritin level or genotype. Deferral rate, model performance, and the difference between reduced and full model performance are calculated and compared to assess whether there are subgroups of donors for whom including the extra variables results in better predictions.

| Model | Women | | Men | |
|---|---|---|---|---|
| | Finland | Netherlands | Finland | Netherlands |
| SVM-1 | 83 628 (3216; 3.9%) | 236 994 (7724; 3.3%) | 88 880 (1480; 1.7%) | 219 390 (2411; 1.1%) |
| SVM-2 | 68 718 (2494; 3.6%) | 166 640 (5875; 3.5%) | 78 268 (1264; 1.6%) | 179 465 (2114; 1.2%) |
| SVM-3 | 55 011 (1859; 3.4%) | 123 171 (4370; 3.6%) | 68 225 (1054; 1.5%) | 150 396 (1889; 1.3%) |
| SVM-4 | 43 164 (1307; 3.0%) | 93 868 (3149; 3.4%) | 58 951 (896; 1.5%) | 127 807 (1667; 1.4%) |
| SVM-5 | 33 179 (868; 2.6%) | 72 165 (2112; 2.9%) | 50 540 (749; 1.5%) | 108 832 (1424; 1.3%) |

**Table 2.** Number of blood bank visits available per model for both countries; number of deferrals and deferral rates are given in brackets.

### Software

All analyses were performed in Python 3.10 using packages numpy and pandas for data processing, scikit-learn for model training and predictions, shap for calculating SHAP values, and matplotlib for creating graphs. All code is available on GitHub and is indexed on Zenodo at https://doi.org/10.5281/zenodo.7780718.

## RESULTS

Table 2 shows the number of donation attempts used for each model in both countries. Deferral counts and rates are given in brackets. Sample sizes are much larger in the Netherlands than in Finland. This is because the total number of blood donations is much higher in the Netherlands than in Finland due to a larger population (17.4 million versus 5.5 million in 2020); but also, because genetic information is available in Finland in only a subgroup of donors, whereas ferritin measurements are available for all Dutch donors.

Deferral rates are very similar in both countries, around 3% for women and 1% for men. The biggest difference in deferral rates is found in men with at least one previous hemoglobin measurement, where the deferral rate is 0.6 percentage points higher in Finland. In most cases deferral rates go down whenever more previous visits are included; this is most likely the result of self-selection, where donors with lower hemoglobin levels are less likely to return for subsequent donations than donors with higher hemoglobin levels. Surprisingly, for Dutch men this pattern seems to some extent to be reversed as their deferral rate goes up with an increasing number of donations.

Tables S1 and S2 in the Appendix show the marginal distribution of the predictor variables, combined for all sub-models. Donors in Finland are older than donors in the Netherlands (median age 46 vs 30 years in women, 52 vs 34 years in men) and the number of donations in the past two years ('NumDon') is also higher, with a difference in median donations of 2 for both sexes. This difference can be explained by the sample composition: the Finnish dataset consists of participants of the Blood Service Biobank, who have given consent for medical research and are typically regular, committed blood donors. Genetic information is only available for these donors.

Hemoglobin levels are slightly higher in Finland for both sexes for all variables HbPrev*i*, by 0.1-0.3 mmol/L. The time between subsequent donation attempts (variables DaysSinceHbi) is slightly shorter for Finnish women than for Dutch women, but almost identical for men. This difference can be partly explained by a difference in minimum donation interval between blood donations: for women, 91 days in Finland vs 122 days in the Netherlands; for men, 61 days in Finland vs 57 days in the Netherlands.

### Predictive performance

Predictive performance can be assessed for individual sub-models, or for all sub-models combined, by using the most complex sub-model possible to predict each outcome. When more previous blood bank visits are taken into consideration, more predictor variables are used, and we expect the performance of the sub-model to increase. Figure 1 shows that this is the case for both the full and reduced model

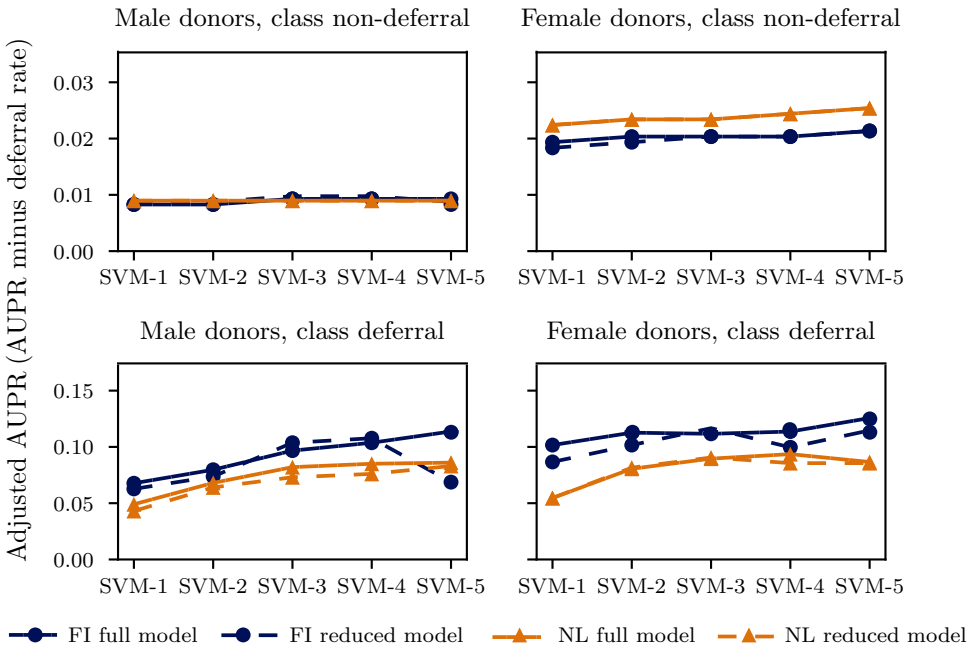


**Figure 1.** Adjusted AUPR by sub-model for both countries and both sets of predictor variables.

| | Baseline | | Reduced model | | Full model | |
|---|---|---|---|---|---|---|
| | FI | NL | FI | NL | FI | NL |
| Male donors, class non-deferral | 0.990 | 0.989 | 0.008 | 0.009 | 0.009 | 0.009 |
| Female donors, class non-deferral | 0.975 | 0.967 | 0.019 | 0.024 | 0.020 | 0.024 |
| Male donors, class deferral | 0.010 | 0.011 | 0.066 | 0.072 | 0.104 | 0.078 |
| Female donors, class deferral | 0.025 | 0.033 | 0.106 | 0.086 | 0.115 | 0.086 |

**Table 3.** AUPR values for all models. AUPR values for the reduced and full models have been adjusted by subtracting the baseline AUPR.

in both countries. The adjusted AUPR increases from SVM-1 through SVM-5 almost everywhere. An exception is the AUPR for class deferral in SVM-m-5, where the reduced model for Finnish donors shows an unexpected drop in the adjusted AUPR. For male donors, class non-deferral, the adjusted AUPR does not seem to change from SVM-m-1 through SVM-m-5.

Overall model performance and the difference in model performance between the full and reduced models are assessed by precision-recall curves and adjusted AUPR values as described in the Methods section. Figure 2 shows the precision-recall curves for various models (SVM-1 through SVM-5, using the model with the most predictor variables possible for each donation attempt) by sex and true outcome class. Table 3 shows the corresponding adjusted AUPR values for each model. In general, models are better at identifying non-deferrals (the most common outcome) than deferrals, even with scoring methods that weigh mistakes in both outcome classes proportionally. However, all curves are well above the baseline, indicating a structural improvement as compared to random guessing.

When comparing the reduced models to each other, one can observe that the performance is very similar in both countries. For women the AUPR is higher in Finland than in the Netherlands for the class deferral, but lower for the class non-deferral. This indicates that deferrals are more likely to be predicted correctly, but at the cost of more inaccuracies when predicting non-deferrals.

Moving from the reduced to the full model has virtually no effect on the AUPR for the class non-deferral: the AUPR of the full model is almost identical to the AUPR of the reduced model for both countries and sexes. For the class deferral, however, there is a difference: in Finland, AUPR increases by

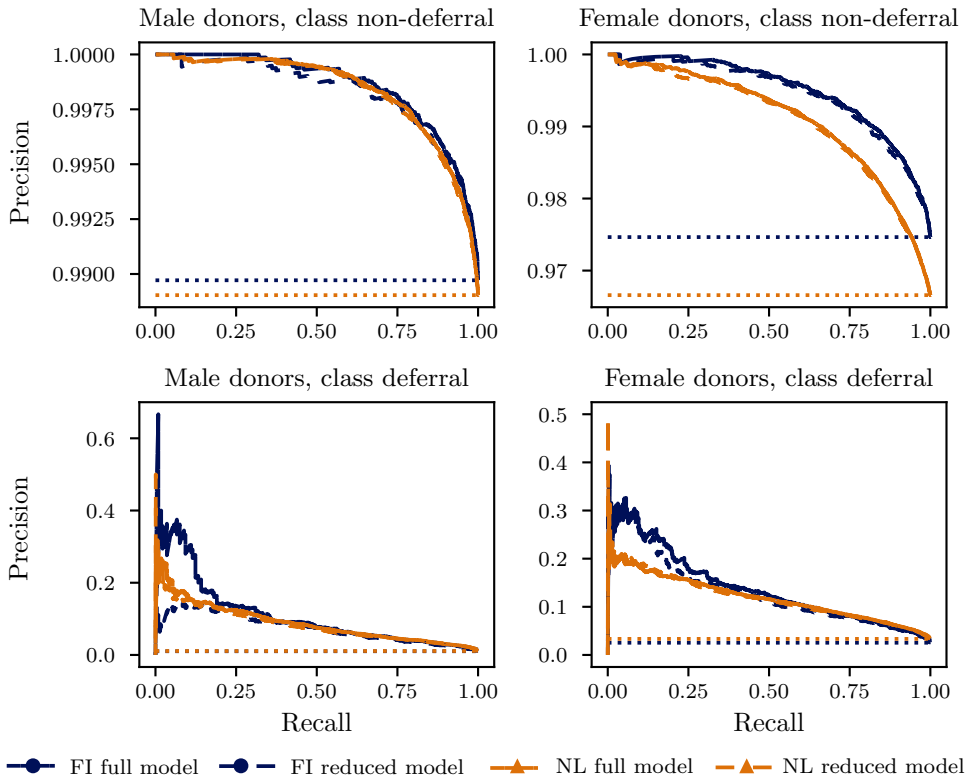


**Figure 2.** Precision-recall curves for the prediction models. For both countries, the curve is shown for the reduced and full prediction models. The baseline (proportion of observations belonging to this outcome class, i.e., for class deferral, the deferral rate) is shown as a dotted horizontal line.

| Finnish donors - reduced model | | |
|---|---|---|
| | Predicted deferral | Predicted non-deferral |
| True deferral | 363 | 166 |
| True non-deferral | 4573 | 18 713 |

| Finnish donors - full model | | |
|---|---|---|
| | Predicted deferral | Predicted non-deferral |
| True deferral | 370 (+7) | 159 (-7) |
| True non-deferral | 4662 (-59) | 18 624 (+59) |

| Dutch donors - reduced model | | |
|---|---|---|
| | Predicted deferral | Predicted non-deferral |
| True deferral | 3762 | 957 |
| True non-deferral | 56 676 | 145 549 |

| Dutch donors - full model | | |
|---|---|---|
| | Predicted deferral | Predicted non-deferral |
| True deferral | 3775 (+13) | 944 (-13) |
| True non-deferral | 55 203 (-1473) | 147 022 (+1473) |

**Table 4.** Confusion matrices of predictions by the reduced and full models. Numbers are summed over both sexes and over all sub-models SVM¬-1 through SVM-5. Observations that can be predicted with multiple sub-models are included the most complex sub-model.

58% (from 0.066 to 0.104) for men and by 8.5% (from 0.106 to 0.115) for women. In the Netherlands, AUPR remains the same for women (0.086 for both) but increases by 8.3% (from 0.072 to 0.078) for men.

Table 4 provides the confusion matrices of model predictions by the reduced and full models for both countries. In the Finnish data, going from the reduced to the full model causes 7 (1.9%) more deferrals to be predicted correctly, while 59 (0.3%) more non-deferrals are predicted correctly. These improvements were all for female donors; at the chosen threshold values, no net changes in the confusion matrix were seen for male donors. In the Dutch data, 13 (0.3%) more deferrals, as well as 1473 (1.0%) more non-deferrals are predicted correctly by the full model as compared to the reduced model.

Note that the large increase in AUPR for Finnish male donors, class deferral, is not reflected in the confusion matrices. The PR-curve in Figure 2 shows that the AUPR increase is due to higher precision in the full model between a recall of 0 and 0.2. However, the optimal classification threshold that is used by the models corresponds to a recall of 0.7, at which point precision in the full model is exactly equal to precision in the reduced model.

### Variable importance

For all sub-models, SHAP values show the importance of the different predictor variables on the predicted outcome. Figures 3 and 4 shows SHAP plots of sub-model SVM-5 of the full model, separately for both sexes and countries. These plots show that in both countries and for both sexes, the most important predictor variable is HbPrev1, the most recent hemoglobin measurement. The direction of the association between the impact on the model output and the feature value for all HbPrev*i* variables is sensible: a lower hemoglobin measurement is predictive of deferral. Age is a more important predictor variable for women than for men in both countries, which is known from previous studies: young women have the highest probability of being deferred due to low hemoglobin, due to monthly iron loss with menstruation.

The additional genetic and ferritin variables for either country end up rather low in the variable importance ranking. The importance of all polygenic risk score and SNP variables in the Finnish models is very low. However, having the minor allele present in either SNP 6:32617727, SNP 15:45095354 or SNP 17:58358769 impacts the model output negatively. This effect is more pronounced in male than

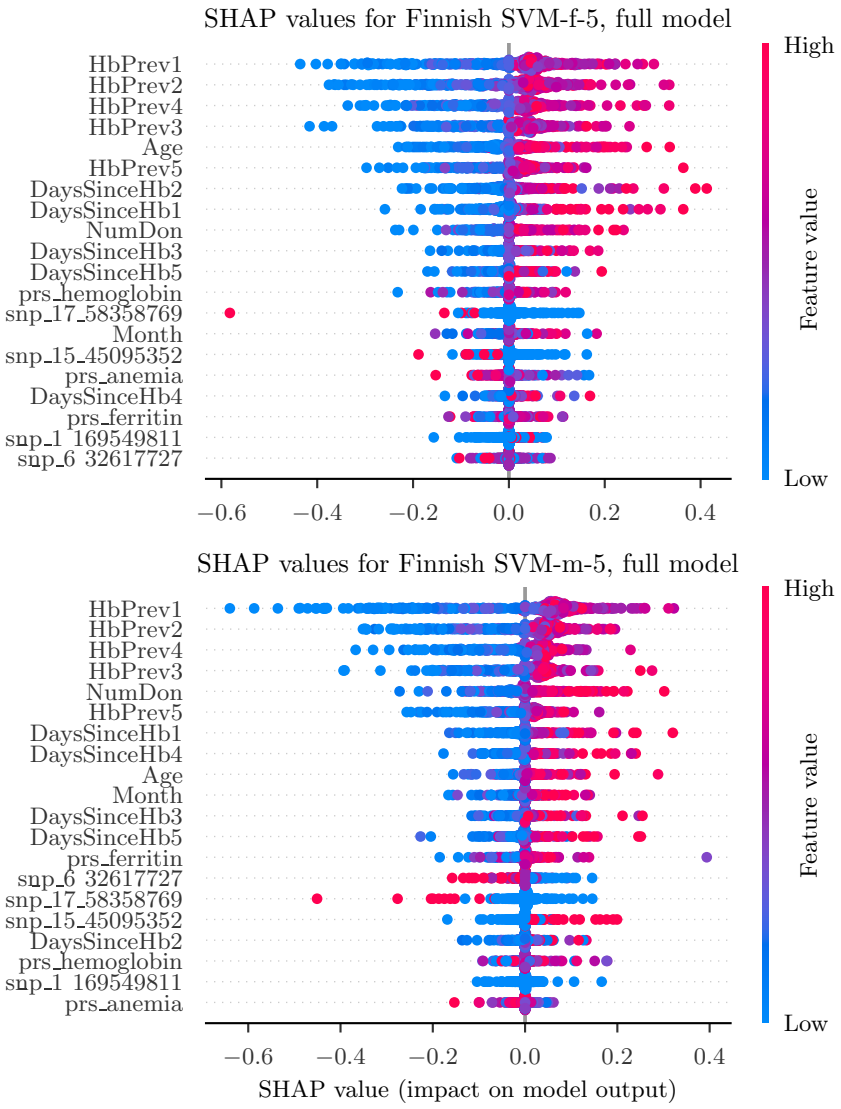


**Figure 3.** SHAP summary plots for the full Finnish model, for women (top) and men (bottom).

female donors.

### Subgroup analysis in Finnish data

To further investigate the effect of the SNPs on deferral prediction, model performance was calculated for groups of donors with the same value for one SNP at a time. Donors with value 1 and 2 are grouped together, as the proportion of donors with value 2 is extremely low, except for the SNP on chromosome 6.

Table 4 shows that for the SNPs on chromosomes 1, 6 and 17, deferral rates are higher amongst donors with one or two minor alleles than in donors with only major alleles. As these SNPs are selected because of their association with iron deficiency or anemia, this is to be expected. Additionally, precision and recall of class deferral are generally higher for donors with minor alleles than for those without, for both the reduced and full models. The SNP 17:58358769 shows this same trend, but the difference between donors with and without minor alleles is much larger. Precision in this subgroup is about twice as high as the overall precision in both the reduced and full model. The increase in recall between the full and reduced model (which changes from 0.733 to 0.933) is the highest of all subgroups.

An additional analysis on the distribution of hemoglobin measurement per donor showed that the higher deferral rate among donors with minor alleles on SNP 17:58358769 can be explained through a

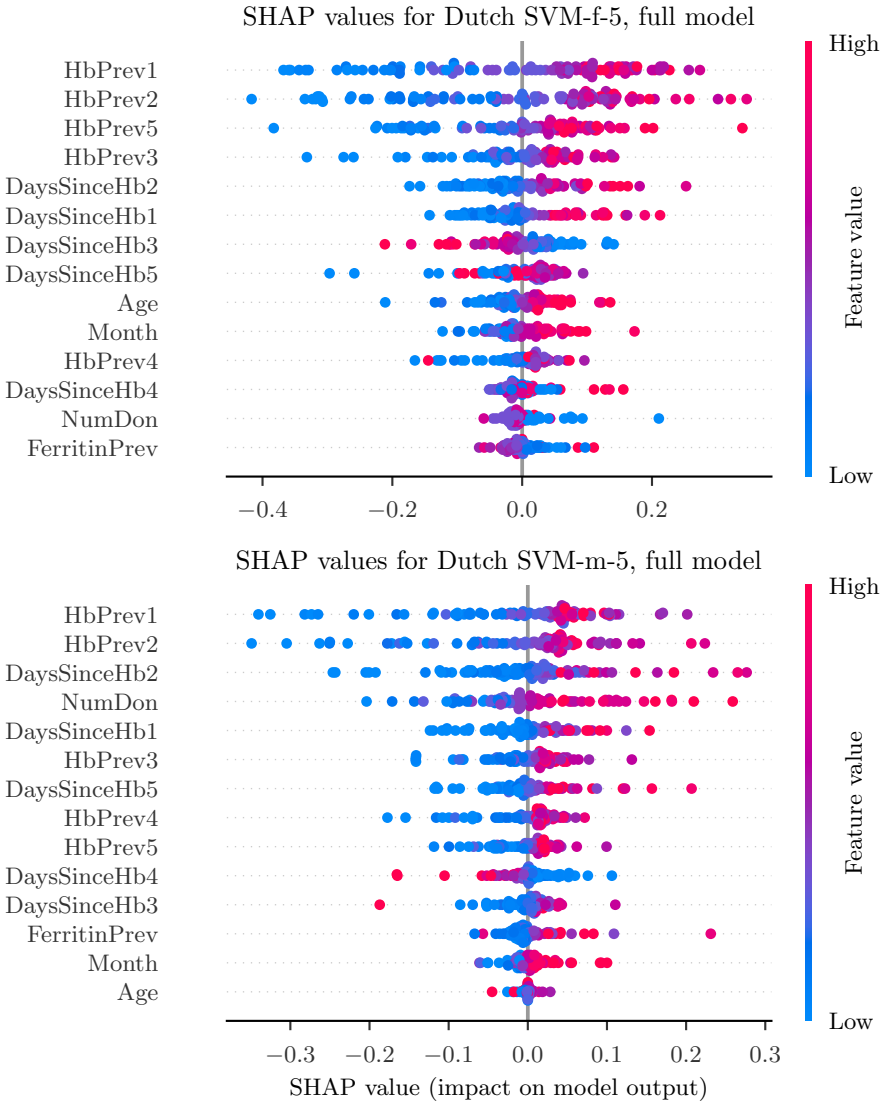


**Figure 4.** SHAP summary plots for the full Dutch model, for women (top) and men (bottom).

combination of a slightly lower average hemoglobin level and a slightly higher variance. This causes these donors to have a slightly higher deferral probability (median 32.6% for donors without minor alleles, median 36.6% for those with minor alleles). This difference was not observed for the other SNPs.

| SNP | Minor alleles | N | Deferral rate | Precision (class deferral) | | Recall (class deferral) | |
|---|---|---|---|---|---|---|---|
| | | | | Reduced model | Full model | Reduced model | Full model |
| SNP 1:169549811 | 0 | 22810 | 0.022 | 0.073 | 0.073 | 0.686 | 0.702 |
| | 1 or 2 | 1005 | 0.026 | 0.087 | 0.095 | 0.692 | 0.692 |
| SNP 6:32617727 | 0 | 7268 | 0.021 | 0.063 | 0.067 | 0.573 | 0.587 |
| | 1 | 11908 | 0.022 | 0.072 | 0.074 | 0.704 | 0.742 |
| | 2 | 4639 | 0.026 | 0.092 | 0.081 | 0.790 | 0.756 |
| SNP 15:45095352 | 0 | 20831 | 0.022 | 0.073 | 0.073 | 0.676 | 0.691 |
| | 1 or 2 | 2984 | 0.022 | 0.080 | 0.080 | 0.758 | 0.773 |
| SNP 17:58358769 | 0 | 23427 | 0.021 | 0.071 | 0.071 | 0.683 | 0.687 |
| | 1 or 2 | 388 | 0.077 | 0.156 | 0.129 | 0.733 | 0.933 |
| Total | - | 23815 | 0.022 | 0.074 | 0.074 | 0.686 | 0.701 |

**Table 5.** Sample sizes, deferral rates, and precision and recall of outcome class deferral for subsets of donors based on values for four SNPs.

### Subgroup analysis in Dutch data

Similar to the subgroup analysis in Finnish data, model performance was calculated for groups of donors with similar ferritin levels: $<$ 15 µg/L, 15-30 µg/L, 30-50 µg/L, 50-100 µg/L, and $>$ 100 µg/L. The first two groups are those that would be deferred for 12 or 6 months, respectively, in accordance with Sanquin's ferritin deferral policy.

Table 5 shows that precision and recall are highest for donors with ferritin levels between 30 and 50 µg/L. This is also the group of donors with the highest deferral rate: 3.2%, versus an overall deferral rate of 2.3%. The fact that this group has the highest deferral rate, and not donors with lower ferritin levels, can be explained by the fact that donors with ferritin levels below 30 µg/L were deferred for six months (twelve months for ferritin levels below 15 µg/L). This delay for the next donation provides the donors with sufficient time to replenish their iron stores and therefore reduces the deferral probability. Hence, donors with ferritin levels just above the ferritin-deferral threshold will have the highest hemoglobin-deferral rate, as they have neither the advantage of the donation break, nor that of a very high ferritin level, which also protects against low hemoglobin levels.

| Ferritin level | N | Deferral rate | Precision (class deferral) | | Recall (class deferral) | |
|---|---|---|---|---|---|---|
| | | | Reduced model | Full model | Reduced model | Full model |
| $<$ 15 µg/L | 7172 | 0.022 | 0.054 | 0.054 | 0.700 | 0.681 |
| 15 - 30 µg/L | 19903 | 0.022 | 0.058 | 0.056 | 0.744 | 0.783 |
| 30 – 50 µg/L | 62140 | 0.032 | 0.082 | 0.079 | 0.815 | 0.833 |
| 50 – 100 µg/L | 65141 | 0.024 | 0.064 | 0.063 | 0.798 | 0.799 |
| $>$ 100 µg/L | 52588 | 0.010 | 0.033 | 0.040 | 0.801 | 0.730 |
| Total | 206944 | 0.023 | 0.062 | 0.064 | 0.797 | 0.800 |

**Table 6.** Sample sizes, deferral rates, and precision and recall of outcome class deferral for various subsets of donors based on their ferritin level.

## DISCUSSION

Predicting deferral for low hemoglobin levels is a topic of interest to many blood banks, as accurate predictions could aid in decreasing deferral rates. This study investigates the added value of including information on the donor's ferritin level or iron-related genetic information to improve hemoglobin deferral prediction. This is done by comparing prediction models with and without information on genetic markers and ferritin levels for the Finnish and Dutch blood bank respectively. The reduced models (i.e., without the additional information) use the exact same predictor variables in both countries. The increase in AUPR is larger for adding genetic markers than it is for adding ferritin levels. Especially for the Finnish male donors, including genetic markers in the prediction model improves the ability of the model to distinguish between the two outcome classes, although at the optimal classification threshold precision and recall do not increase from the reduced model. The SHAP values of the predictions by the full models in both countries show that both genetic markers and ferritin levels have a much smaller impact on the prediction than the variables included in the reduced models, as confirmed by the modest increase in AUPR between the reduced and full models.

Overall, including either genetic or ferritin information has little effect on the predictions made by the models. Both increase the proportion of deferrals that are predicted correctly: 1.9% and 0.3% more deferrals are correctly identified in the Finnish and Dutch setting respectively when the full model is used rather than the reduced model. However, we found that in both countries, there is a subgroup of donors for which the full model performs substantially better than the reduced model. These are Finnish donors with minor alleles on SNP 17:58358769, and Dutch donors with ferritin levels between 30-50 µg/L. In both cases, these are subgroups of donors with a higher-than-average deferral rate. Performance for these subgroups is already higher than average in the reduced model, but when using the full model this difference increases even further.

Other studies have shown that previous hemoglobin measurements are the most influential predictors for hemoglobin deferral. Including lifestyle behavior, smoking, ethnicity, or menstruation in prediction models also improves performance, but only marginally. [4] A Finnish study showed that genetic information does not improve the predictive performance of hemoglobin levels (as opposed to hemoglobin deferral). [9] This study confirms that the performance of prediction models increases slightly when either ferritin or genetic information is added. Still, considering the large number of donation visits blood banks receive yearly, even a small increase could potentially prevent hundreds of deferrals. It should be noted that the Finnish population is more genetically homogenous than other countries, and that they are also genetically distinct from other countries due to several historic population bottlenecks and geographical isolation. [10] According to the Genome Aggregation Database (gnomAD) [11], the SNP 17:58358769 minor allele frequency in the Finnish population is 0.0147, and only 0.0007 in the European (non-Finnish) population. It is not found in any other populations and was discovered by an iron deficiency GWAS in the FinnGen project. [7] This means that findings on Finnish genetic data may not be representative for other countries, but analyses in other populations may discover similar population-specific variations that may make the use of genetic data more beneficial.

The main limitation of this study is that the effect of including ferritin and genetic information is studied in two different countries, rather than in a single population. By comparing against the reduced model and reporting the relative increase in performance, we attempt to mitigate this limitation. The very similar adjusted AUPRs of the reduced models and the similarity in SHAP values of the models indicate that the countries are rather comparable. A second limitation is that all Dutch donors could be included in this study, but only Finnish donors from the Blood Service Biobank, as genetic information is not available for other donors.

In general, we again confirm that accurately distinguishing deferrals from non-deferrals by predictive modelling is a complex task that comes at the cost of losing a substantial number of successful donations by incorrectly predicting them to be deferrals. A major reason for the low performance of our prediction models is the measurement variability, partly caused by the (pre-) analytical variability of the capillary hemoglobin measurements. [12] As long as we try to predict an outcome that is highly variable, the performance of any prediction model will remain unsatisfactory, regardless the number of predictor variables included.

However, in the absence of a better measurement or decision strategy, it is worthwhile investigating which information would lead to better hemoglobin deferral predictions as it still leads to a better understanding of the underlying process(es). Based on our results, we would recommend including

ferritin and genetic information in prediction models in case these are readily available. Compared to the reduced model, including genetic information would have resulted in seven fewer deferrals and 59 more donations in one year, at a cost of genotyping approximately 24 000 donors. Including ferritin levels results in 13 fewer deferrals and 1473 more donations in one year, and although measuring ferritin levels is less expensive than genotyping, this measurement must be repeated regularly whereas genotyping only has to be performed once for each donor. We would therefore not recommend collecting this information explicitly for the use in hemoglobin deferral prediction, as the marginal increase in performance is not likely to be worthwhile the investment of both time and money.

## APPENDIX

| | Women | | | |
|---|---|---|---|---|
| | Finland | | Netherlands | |
| Number of donations | 83 628 | | 236 994 | |
| Age | 46 | (29 - 57) | 30 | (23 - 47) |
| Month | 6 | (3 - 10) | 7 | (4 - 10) |
| NumDon | 3 | (2 - 5) | 1 | (0 - 3) |
| SNP_1_169549811 | 0: 79 991<br>1: 3567<br>2: 70 | | NA | |
| SNP_6_32617727 | 0: 26 241<br>1: 41 282<br>2: 16 105 | | NA | |
| SNP_15_45095352 | 0: 73 159<br>1: 10 101<br>2: 368 | | NA | |
| SNP_17_58358769 | 0: 82 336<br>1: 1287<br>2: 5 | | NA | |
| PRS_anemia ($*10^6$) | -0.002 | (-0.847 - 0.828) | NA | |
| PRS_ferritin ($*10^6$) | 0.032 | (-1.191 - 1.280) | NA | |
| PRS_hemoglobin ($*10^6$) | 0.039 | (-3.010 - 3.105) | NA | |
| FerritinPrev | NA | | 47 | (33 - 47) |
| HbPrev1 | 8.7 | (8.3 - 9.1) | 8.5 | (8.1 - 8.9) |
| DaysSinceHb1 | 131 | (104 - 203) | 135 | (104 - 194) |
| HbPrev2 | 8.7 | (8.3 - 9.1) | 8.5 | (8.1 - 8.9) |
| DaysSinceHb2 | 280 | (221 - 391) | 301 | (254 - 405) |
| HbPrev3 | 8.7 | (8.3 - 9.1) | 8.5 | (8.1 - 8.8) |
| DaysSinceHb3 | 419 | (338 - 558) | 475 | (396 - 627) |
| HbPrev4 | 8.7 | (8.3 - 9.1) | 8.4 | (8.1 - 8.8) |
| DaysSinceHb4 | 546 | (453 - 701) | 653 | (546 - 822) |
| HbPrev5 | 8.7 | (8.3 - 9.1) | 8.4 | (8.1 - 8.8) |
| DaysSinceHb5 | 666 | (561 - 825) | 831 | (703 - 1004) |
| Deferral rate | 0.0385 | | 0.0326 | |

**Table S1.** Marginal distribution of predictor variables in both countries for female donors. Variables are described by their median and 1st and 3rd quartiles, except for SNP variables, for which the allele count distributions are shown. Each donation attempt is included only once in this description and is given for the prediction using the highest number of previous visits only (e.g., a visit by a female donor with three previous visits could be included in SVM-f-1 through SVM-f-3 but is only included in SVM-f-3).

| | Men | | | |
|---|---|---|---|---|
| | Finland | | Netherlands | |
| Number of donations | 88 880 | | 219 390 | |
| Age | 52 | (38 - 60) | 34 | (26 - 48) |
| Month | 6 | (3 - 10) | 7 | (4 - 10) |
| NumDon | 5 | (3 - 7) | 3 | (1 - 5) |
| SNP_1_169549811 | 0: 85 358<br>1: 3487<br>2: 35 | | NA | |
| SNP_6_32617727 | 0: 26 779<br>1: 43 714<br>2: 18 387 | | NA | |
| SNP_15_45095352 | 0: 78 223<br>1: 10 168<br>2: 489 | | NA | |
| SNP_17_58358769 | 0: 87 358<br>1: 1522<br>2: 0 | | NA | |
| PRS_anemia ($*10^6$) | -0.040 | (-0.877 - 0.792) | NA | |
| PRS_ferritin ($*10^6$) | -0.023 | (-1.272 - 1.243) | NA | |
| PRS_hemoglobin ($*10^6$) | -0.019 | (-3.095 - 3.256) | NA | |
| FerritinPrev | NA | | 77 | (44 - 141) |
| HbPrev1 | 9.6 | (9.1 - 10.0) | 9.4 | (9.0 - 9.9) |
| DaysSinceHb1 | 98 | (71 - 147) | 81 | (63 - 133) |
| HbPrev2 | 9.6 | (9.1 - 10.0) | 9.4 | (9.0 - 9.8) |
| DaysSinceHb2 | 204 | (154 - 293) | 184 | (138 - 287) |
| HbPrev3 | 9.6 | (9.0 - 10.0) | 9.4 | (9.0 - 9.8) |
| DaysSinceHb3 | 306 | (235 - 419) | 300 | (224 - 434) |
| HbPrev4 | 9.5 | (9.0 - 10.0) | 9.4 | (8.9 - 9.8) |
| DaysSinceHb4 | 399 | (314 - 535) | 418 | (314 - 581) |
| HbPrev5 | 9.5 | (9.0 - 10.0) | 9.4 | (8.9 - 9.8) |
| DaysSinceHb5 | 489 | (389 - 639) | 535 | (409 - 714) |
| Deferral rate | 0.0167 | | 0.0110 | |

**Table S2.** Marginal distribution of predictor variables in both countries for male donors. Variables are described by their median and 1st and 3rd quartiles, except for SNP variables, for which the allele count distributions are shown. Each donation attempt is included only once in this description and is given for the prediction using the highest number of previous visits only (e.g., a visit by a male donor with three previous visits could be included in SVM-m-1 through SVM-m-3 but is only included in SVM-m-3).

## REFERENCES


[1] Marloes LC Spekman, Theo G van Tilburg, and Eva-Maria Merz. "Do deferred donors continue their donations? A large-scale register study on whole blood donor return in the Netherlands". In: *Transfusion* 59.12 (2019), pp. 3657–3665.

[2] Marieke Vinkenoog, Matthijs van Leeuwen, and Mart P. Janssen. "Explainable haemoglobin deferral predictions using machine learning models: Interpretation and consequences for the blood supply". In: *Vox Sanguinis* 117.11 (2022), pp. 1262–1270.

[3] Marieke Vinkenoog, Jarkko Toivonen, Tinus Brits, Dorien de Clippel, Veerle Compernolle, Surendra Karki, Amber Meulenbeld, Marijke Welvaert, Katja van den Hurk, Joost van Rosmalen, Emmanuel Lesaffre, Mikko Arvas, and Mart P Janssen. "An international comparison of hemoglobin deferral prediction models for blood banking". In: *Vox Sanguinis* (2023).

[4] A. Mireille Baart, Tiffany Timmer, Wim L. A. M. de Kort, and Katja van den Hurk. "Lifestyle behaviours, ethnicity and menstruation have little added value in prediction models for low haemoglobin deferral in whole blood donors". In: *Transfusion Medicine* 30.1 (2020), pp. 16–22.

[5] Sant-Rayn Pasricha, Zoe K. McQuilten, Anthony J. Keller, and Erica M. Wood. "Hemoglobin and iron indices in nonanemic premenopausal blood donors predict future deferral from whole blood donation". In: *Transfusion* 51.12 (2011), pp. 2709–2713.

[6] Mitja I. Kurki et al. *FinnGen: Unique genetic insights from combining isolated population and national health register data*. Mar. 2022.

[7] Jarkko Toivonen, Johanna Castrén, FinnGen, and Mikko Arvas. "The Value of Genetic Data from 665,460 Individuals in Predicting Anemia and Suitability to Donate Blood". In: *Genetic Epidemiology* 46.7 (), pp. 477–477.

[8] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

[9] Jarkko Toivonen, Yrjö Koski, Esa Turkulainen, Femmeke Prinsze, Pietro della Briotta Parolo, Markus Heinonen, and Mikko Arvas. "Prediction and impact of personalized donation intervals". In: *Vox Sanguinis* 117.4 (2022), pp. 504–512.

[10] Sanni Översti, Kerttu Majander, Elina Salmela, Kati Salo, Laura Arppe, Stanislav Belskiy, Heli Etu-Sihvola, Ville Laakso, Esa Mikkola, Saskia Pfrengle, Mikko Putkonen, Jussi-Pekka Taavitsainen, Katja Vuoristo, Anna Wessman, Antti Sajantila, Markku Oinonen, Wolfgang Haak, Verena J. Schuenemann, Johannes Krause, Jukka U. Palo, and Päivi Onkamo. "Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry". In: *Scientific Reports* 9 (Nov. 2019), p. 16883.

[11] Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Laura D. Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Daniel Rhodes, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X. Chong, Kaitlin E. Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S. Ware, Christopher Vittal, Irina M. Armean, Louis Bergelson, Kristian Cibulskis, Kristen M. Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E. Talkowski, Benjamin M. Neale, Mark J. Daly, and Daniel G. MacArthur. "The mutational constraint spectrum quantified from variation in 141,456 humans". In: *Nature* 581.7809 (May 2020), pp. 434–443.

[12] Mart P. Janssen. "Why the majority of on-site repeat donor deferrals are completely unwarranted. . ." In: *Transfusion* 62.10 (2022), pp. 2068–2075.