Data-driven donation strategies: understanding and predicting blood donor deferral

Vinkenoog, M.

CHAPTER

10

Conclusions, general discussion and anticipated future research

Conclusions, general discussion and anticipated future research

Throughout this thesis, several statistical and data science analysis techniques have been applied to blood donation data in order to explore how their application can improve different aspects of donor management strategies. In this chapter, we summarise the findings from these analyses and discuss challenges that occurred in multiple chapters and therefore deserve more in-depth explanation and attention.

10.1 Conclusions

In the Introduction, seven research questions were formulated. Here, each research question is answered based on the results as described in Chapters 3-9, and main conclusions for each question are given.

10.1.1 Hemoglobin and ferritin levels

Q1 **Does a ferritin-based donor deferral policy prevent donors from returning with iron deficiency?**

The vast majority of donors that are deferred for low ferritin levels returns with considerably increased ferritin levels. After a 6-month deferral, ferritin levels were $\geq$ 15 µg/L for 88% of returning female donors and for 99% of returning male donors, which is a positive result. After a 12-month deferral, this was the case for 74% and 95% of returning female and male donors, respectively. [1] Although comparisons to a control group would be needed to draw conclusions about causality, it is reasonable to assume that if these donors had returned to the blood bank sooner (i.e., had they not been deferred), their ferritin levels would have been lower. From observational data, we also showed that the implementation of the ferritin deferral policy was associated with a substantial decrease in deferral rates due to low hemoglobin. Before the implementation of the policy, the hemoglobin deferral rate was around 8% for women and 5% for men, and currently it is down to 3% and 1%, respectively. [96]

In the same study, the distribution of ferritin levels was compared between sexes and between first-time and repeat donors. We found that in first-time donors, 25% of women and 2% of men have ferritin levels below the deferral threshold of 30 µg/L. These percentages are considerably higher in repeat donors, where 53%

of women and 42% of men have ferritin levels below 30 µg/L. The distribution of ferritin levels among first-time donors is very different for men and women, with women having much lower ferritin levels, but this difference almost disappears when comparing repeat donors. This suggests that regular blood donation results in a decrease in iron stores, which impacts men more than women because their natural iron stores are generally higher, and because of their higher donation frequency. These numbers underline the necessity of the ferritin-based deferral policy, especially since the proportion of new female donors under the age of 25 has been increasing rapidly in the Netherlands [34], and this is the group most at risk of having a low ferritin level.

The above findings are all described in Chapter 3. The main conclusion to this research question is that the ferritin-based donor deferral policy is successful in preventing donors from returning to donate with ferritin levels below 15 µg/L.

Q2 **What are determinants of variations in ferritin levels?**

Distributions of ferritin levels were compared between sexes and between first-time and repeat donors in the study on the effect of the ferritin-based deferral policy. [96] Within these groups, considerable variation in ferritin levels between individuals was observed. By applying structural equation modelling, we found that 25% of ferritin variance in new donors and 40% in repeat donors could be explained by individual characteristics, donation history (for repeat donors only), and environmental factors. [36] We confirmed previous findings, both our own and from other donor populations, that ferritin levels are substantially higher in men than in women among first-time donors, and that repeated blood donation impacts men's ferritin levels more than women's, resulting in similar ferritin levels for both sexes among repeat donors.

The main determinants of variation in ferritin levels are individual characteristics and donation history, as expected. The association between ferritin levels and environmental factors was smaller but still substantial. A likely explanation for this association is that air pollution can cause low-grade inflammation, and ferritin levels are known to be correlated with inflammatory activity. [47, 9, 126] Interestingly, the association was twice as high for repeat donors as for first-time donors. This indicates that environmental factors are more associated with ferritin recovery after blood loss than with ferritin levels in a steady state.

The above findings are described in Chapter 4. The main conclusion is that combining multiple determinants in a single integrative model allows us to ex-

plain a considerable part of ferritin variation based on individual characteristics, donation history and environmental factors.

Q3 **Can we find groups of donors whose hemoglobin levels change in a similar manner over the course of their donor career?**

By clustering donors' hemoglobin trajectories, we aimed to identify groups of donors with similar long-term responses to blood donation. This clustering could then be used to find associations with characteristics of the donors, and to find optimal donation intervals for the different groups. Both clustering methods resulted in distinct clusters of donors with clear differences in hemoglobin levels over time. However, the clusters differ mostly in the average hemoglobin value over time, as donors with similar hemoglobin levels at their first measurement are clustered together, and hemoglobin levels reduce slowly over time. With the clustering methods used, it was not possible to distinguish groups of donors with rapidly declining hemoglobin levels from those with relatively stable hemoglobin levels. [127] In later studies, the concept of clustering donors was replaced by making personalised predictions as described in research questions five through seven.

The results, along with an in-depth discussion on challenges in clustering these hemoglobin trajectories, are described in Chapter 5. The main conclusion is that the resulting clusters are based mostly on average hemoglobin value, therefore it seems more useful for the blood bank to focus on individual hemoglobin trajectories, rather than on characteristics that distinguish between clusters.

10.1.2 SARS-CoV-2 antibodies

Q4 **How are individual characteristics and symptoms associated with IgG antibody response in COVID-19 recovered donors?**

In this observational study into donors' IgG antibody response after a COVID-19 infection, we found higher age and BMI to be associated with higher antibody counts, indicating more severe illness. Antibody decay was found to be faster in male than in female donors, as well as for donors who had been hospitalised during their infection. We also identified associations between antibody counts and several self-reported symptoms that donors had experienced. The presence of nasal cold, headache and anosmia were associated with lower IgG levels, while dry cough, fatigue, fever, dyspnoea, diarrhoea, and muscle weakness were associated with higher IgG levels. [128] This was in line with findings from studies on

hospital cohorts, which found fatigue and dyspnoea to be prognostic for severe infection, while a stuffed nose (comparable to nasal cold in our data) was prognostic of mild infection. [91] At the time, our study was one of the largest studies concerning not only hospitalised patient cohorts, making it more representative of the total COVID-19 patient population.

These findings are described in Chapter 6. The main conclusion is that in addition to previously described associations with sex, age and BMI, SARS-CoV-2 antibody levels are also associated with several COVID-19 symptoms.

10.1.3 Prediction of hemoglobin deferral

Q5 **Can we accurately and reliably predict hemoglobin deferral based on historical data?**

We have presented a support vector machine (SVM) prediction model that predicts hemoglobin deferral based on several donor characteristics and up to five previous hemoglobin measurements. We found that although the model could correctly classify 80% of all deferrals, this comes at a cost of incorrectly classifying about 30% of donors with adequate hemoglobin levels as having to be deferred for low hemoglobin. This would imply a substantial net loss of donations for the blood bank. However, by using the model to predict deferral at different time points, we found that 64% of non-deferred donors would be invited earlier or on the same date, and 80% of deferred donors would be invited later. We assume that for some of these deferred donors, the extra recovery time would be enough to increase their hemoglobin level above the donation threshold. Using the prediction model to decide when to invite which donor, the deferral rate was estimated to decrease by 60% without decreasing the number of successful donations. [109]

SHAP values were used to see how predictor variables were related to the model prediction. These showed that previous hemoglobin levels are the most important predictors of future hemoglobin deferral, with low previous values being indicative of deferral. The use of SHAP values makes this model explainable, and we found that most predictor variables are related to the model predictions in ways that can be explained either by biological processes, or organisational policies.

These results are described in Chapter 6. The main conclusion is that using

prediction models to guide donor invitations may help to reduce donation intervals as well as deferral rates.

Q6 **How do country-specific blood bank policies and donor demographics affect hemoglobin deferral prediction models?**

By applying the same set of models to blood donation data of five different countries, we found that performance of the different models within countries is very similar. The most interesting result was that the relative importance of the predictor variables (again calculated using SHAP values) was very similar across countries. Previous hemoglobin remains undefeated as the best predictor variable for future hemoglobin deferral. Additionally, we found that model performance is highly dependent on the deferral rate, with higher deferral rates being associated with better model performance.

These findings are described in Chapter 8. The main conclusion is that model performance is more dependent on the deferral rate than on the model architecture, and that the relative importance of predictor variables is very similar across countries.

Q7 **Do ferritin measurements or genetic information add value to hemoglobin deferral prediction models?**

By comparing several simple models that only contain widely available predictor variables, we found that including ferritin as a predictor for hemoglobin deferral in Dutch donors increases model performance slightly, as does including genetic information as a predictor for Finnish donors. For certain subgroups of donors, including this extra information leads to a large increase in recall of deferrals. This is the case for donors with a rare minor allele on an iron-related single-nucleotide polymorphism (SNP) in Finland, and donors with ferritin levels just above the deferral threshold in the Netherlands.

The results are described in Chapter 9. The main conclusion is that although the overall value of ferritin and genetic information for hemoglobin deferral prediction is low, for specific subgroups it appears very useful in increasing the accuracy of deferral predictions.

10.2 General discussion

Throughout the research presented in this thesis, many challenges were encountered that were not study-specific and deserve a more in-depth discussion. Many studies were performed on the same dataset (albeit on updated versions), extracted from Sanquin's blood bank database system eProgesa. This dataset contains information on all donations that take place at Sanquin, and the purpose of recording this information is to ensure the safety and traceability of all blood products. During the pre-donation screening, donors are asked for their consent to the use of their data for scientific research, which over 99% of donors grant. Although many researchers at Sanquin use these data, it is not collected for research and therefore not optimised for that purpose.

Notable limitations of this dataset are the variability of recorded hemoglobin levels and the presence of selection bias. Two limitations for hemoglobin and ferritin research in blood donors in general are the uncertainty about how these proteins are related to health, and the fact that reproducing the research is difficult due to the very specific study population. These four topics are discussed in the following sections.

10.2.1 Hemoglobin measurement variability

Measurement variability is the phenomenon that whenever the same measurement is repeated the result will never be exactly the same. For measurements such as hemoglobin levels, measurement variability occurs through three causes. First, there is biological variation, as hemoglobin levels vary naturally throughout the day and year, and due to changes in diet, lifestyle, or even illness. Second, variation can occur as a result of differences in pre-analysis conditions, for instance by differences in temperature or transport, or how the donor physician handles the finger that the blood is collected from. Third and last, there is variation in the test itself, depending on the reliability of the measurement method, the assay and the machine used.

Because Sanquin tests hemoglobin at the new donor intake, and again approximately three weeks later before the first donation, we have data that allows estimation of the variability of the hemoglobin measurement. Differences due to biological variation in hemoglobin levels measured three weeks apart should be very minimal: diet and lifestyle are unlikely to change drastically in such a short time, and blood donation, pregnancy or major blood loss (apart from menstrual blood loss in premenopausal women) are also unlikely to have occurred. [11] From these two measurements, it can be derived that the standard deviation of an individual hemoglobin measurement is 0.43 mmol/L for men and 0.38 mmol/L for women. [119] This means that donors

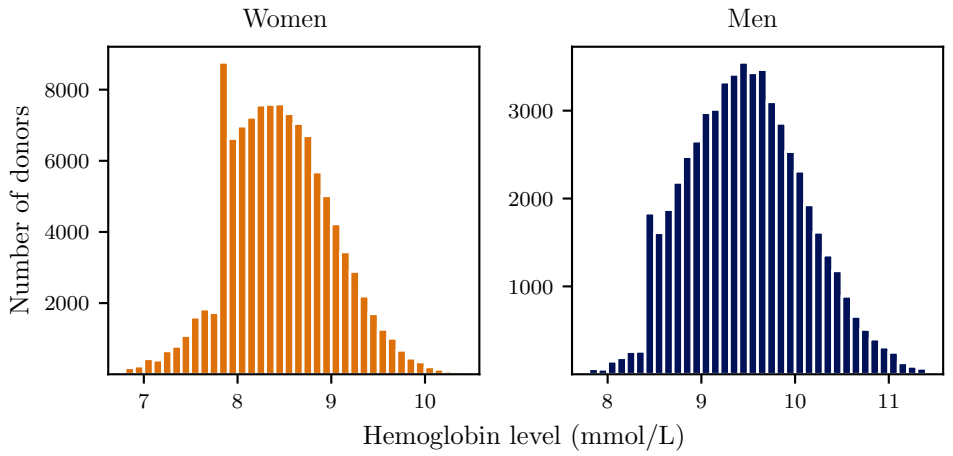

Figure 10.1: The distribution of recorded hemoglobin levels in all 114 459 female (left) and 58 511 male (right) prospective donors at donor intake between 2018 and 2020.

with hemoglobin levels around the deferral threshold will have a substantial chance of having a *measured* hemoglobin level below the threshold and therefore being deferred. The repeated measurement policy mitigates this effect, but at the same time introduces an upward bias in the data, as only the highest of three measurements is recorded. This bias is especially noticeable for donors around the donation threshold, and is illustrated by a simple histogram of all reported hemoglobin levels, as shown in Figure 10.1.

Clearly, the increase in observations between 7.7 and 7.8 mmol/L for women (and 8.3 and 8.4 mmol/L for men) is not due to those values naturally occurring more often, but rather it is an artefact of selective repeated measuring, and recording only the highest hemoglobin value. A more extensive review of this problem and several possible solutions was recently published and is well worth a read. [119]

10

The consequence of recording hemoglobin levels in this way is that a bias is introduced in the data, especially for hemoglobin levels around the deferral threshold. This makes the class imbalance (see Section 2.2.2) more extreme, making accurate classification harder. Additionally, observations right around the donation threshold potentially contain the most crucial information for deferral prediction, and precisely these observations are most impacted by the bias. It is therefore reasonable to expect that our models would perform better on data without such a bias.

10.2.2 Selection bias

Research on blood donors is generally influenced by the *healthy donor effect*, a selection bias caused by health criteria imposed on prospective blood donors. [129] The main consequence of the healthy donor effect is that it is difficult to draw conclusions on health effects of blood donation. Because donors are selected based on health criteria, in general they are ill less frequently than non-donors. This may lead to the incorrect conclusion that donating blood is beneficial for your health, while in reality, this association is found as a result of selection bias. [130] The effect persists during the entire donor career, as healthier donors are more likely to keep returning for subsequent donations.

Depending on the associations researched, it may therefore not be possible to generalise findings in donor cohorts to the general population. Most topics studies in this thesis would only be generalised to other donor populations, and therefore the healthy donor effect is not such a big complication. However, in some cases we may be tempted to extrapolate our findings to the general population, such as in the study on ferritin determinants, or the SARS-CoV-2 antibody paper. In the study on ferritin determinants, the association we found between environmental factors and ferritin level may have been underestimated, as this association is likely mediated by inflammation, and people with inflammation are probably underrepresented in the donor population. In the study on SARS-CoV-2 antibodies, only donors who were healthy enough to regularly donate plasma could be included, which means the results are mainly applicable towards people with a mild disease outcome. The results may not hold for people suffering from long COVID or other chronic health issues.

Due to this selection bias, it is also possible to draw incorrect conclusions or to miss correct ones when generalizing outside of donor cohorts. In some cases, the distribution of a predictor variable may be much narrower among donors than in the general population. This may prevent a true association from being found by a model, while in other cases, associations may be found that exist only in blood donors, as they are mediated by organisational policies of blood banks.

There is more selection bias in the donation dataset than the healthy donor effect. Ideally, we would like each donor to have the same probability of being invited to the blood bank and having their hemoglobin levels measured. Instead, donor invitations are based on many different criteria, the two most important being the current need for their blood group, and their response history. Donors that fail to respond to invitations (either by visiting the blood bank, or by rejecting the invitation) will be less prioritised and eventually may not at all be invited anymore. This is one example

of how the available data is based on existing donor management strategies, and we may not be able to learn an optimal strategy from such data.

In practice, this means that the performance of our hemoglobin deferral prediction models is inherently limited. Precision of deferral prediction is quite low, meaning that many donors with adequate hemoglobin levels are incorrectly predicted to have hemoglobin levels below the deferral threshold, which will lead to fewer donations and a potentially insufficient blood supply if these donors are not invited to the blood bank as a result of this incorrect prediction. However, predictions are only made for donors that were invited and visited the blood bank, and these donors are a (non-random) subset of all registered donors. If all donors were invited to the blood bank at the same rate, there would be a wider pool of donors for the model to choose from to mitigate the missed successful donations in absolute numbers. The fact that the process by which the donation dataset is formed is far from random also means that it is harder to predict what would happen if something were to be changed in the invitation process, such as the inclusion of a hemoglobin deferral prediction model. It is therefore difficult to predict what impact the application of such a model would have on the deferral rate exactly. In the current situation, a selection of loyal donors is prioritised for donor invitations, which leads to lower iron stores and higher probability of deferral due to low hemoglobin for these donors. Since our models are developed on data mostly from prioritised donors, it is possible that predictive performance on non-prioritised donors is lower.

10.2.3 Hemoglobin, ferritin and health

We monitor hemoglobin and ferritin levels in blood donors as an indication of their iron status, but research on the relation between these proteins and health is not entirely conclusive. Threshold values exist to diagnose anemia based on hemoglobin levels, but there are no clear threshold levels for hemoglobin and ferritin to diagnose iron deficiency without anemia. [9, 131] At Sanquin, donors are deferred for six months if their ferritin level is between 15 and 30 µg/L. This deferral is meant to prevent donors from returning with ferritin levels below 15 µg/L. However, among prospective female donors (women who have never donated blood before) 5% already have ferritin levels below 15 µg/L. [96] People generally only apply to become a blood donor when they feel healthy enough to do so, so these women are unlikely to experience symptoms of iron deficiency. On the other hand, the fact that they feel healthy enough to apply to become a blood donor does not exclude the possibility that they are already iron deficient or even anemic, as many women in the general population have a level of

iron deficiency, for example due to regular heavy menstrual blood loss, pregnancy, or breastfeeding. [132] This makes decisions on reference ranges of ferritin levels particularly challenging, and although those decisions are not in the scope of this thesis, it does complicate the interpretation of our study results to a wider health-related context. This is another reason that the results of our research are rarely generalisable outside of donation-related contexts, although it would be difficult to conceive of a relevant context where blood is regularly drawn without medical indication outside of blood banks.

Iron supplementation is often mentioned when discussing research concerning iron levels in blood donors. Would the best way to decrease deferral rates not be to provide donors with iron supplements? Some blood banks encourage all donors to take iron supplements, others encourage, or provide supplements to those most at risk for low iron (mainly young women, or donors donating at high frequencies). [133, 33] Sanquin does not recommend donors to take iron supplements, although of course donors are completely free to do so.

Even though iron supplementation is not current practice at Sanquin, its potential as a policy to enhance recovery after donating with low ferritin levels is currently being investigated. A randomised controlled trial is being conducted where donors with ferritin levels below 30 µg/L are given varying dosages of iron supplements or placebo pills. [134] Donor perceptions and changes therein are also important and being studied: as more donors are choosing to follow a vegetarian or vegan diet, their views on the necessity of iron supplements may also change. Furthermore, the success of iron supplementation policies is largely dependent on donors' willingness to take supplements, and their compliance.

10.2.4 Reproducibility of study results

Ideally, published research is reproducible by other research groups to be validated or challenged. Sanquin is the only blood bank in the Netherlands, making reproduction of our research by others difficult. Also, the data are considered privacy sensitive and therefore not easily shareable. It therefore makes sense to look across borders and compare our research results to those of other blood banks. The topics covered in this thesis are also studied in blood banks of other countries, and this allowed us to perform the comparison study presented in Chapter 8.

Collaboration with researchers from blood banks in Australia, Belgium, Finland, and South Africa showed that even though it appears that we do the same thing (hemoglobin testing and deferring donors below a certain cut-off value), small dif-

ferences in policies exist that make it hard to compare outcomes found in different countries. [20] Some sources of variation along with several (non-exhaustive) implementation alternatives are shown in Table 10.1, and an even wider range of alternative policies is described in a study by the BEST Collaborative Study Group. [55] All these factors are important to consider when comparing study results obtained in different settings, and it becomes even more complicated when we consider policy changes over the years. For instance, even only looking at Sanquin, ferritin testing was implemented in 2017, and hemoglobin deferral rates are now drastically lower than before 2017. What are the implications of this change in policy when comparing study outcomes from the Netherlands with those of other countries? Similarities found between countries may suggest similar associations, but any conclusions should be accompanied with words of caution for potential biases as a result of differences as specified in Table 10.1.

10.3 Anticipated developments and future research

Views on hemoglobin deferral and donor iron management are gradually changing. Researchers, policy-makers and health organisations are increasingly convinced that the most frequently used method of hemoglobin testing is suboptimal. One small change could be simply to record all hemoglobin measurements; even without changing the deferral policy, recording these extra measurements would allow obtaining unbiased hemoglobin estimates and better data for research (and decision-making in general). In general, it would be beneficial to move towards more individualised donation intervals rather than inviting donors back after a set amount of time and checking their hemoglobin and ferritin levels. Ferritin-guided donation intervals have been shown to increase ferritin and hemoglobin levels and thus decrease deferral rates. [135] These results are obtained with the same ferritin thresholds for each donor, but in the future donation intervals could even be guided individually for each donor, based on their own donation history.

The prediction models presented in this thesis are all strongly data-driven without any prior specification of how variables should theoretically be related to the outcome variable of interest. Currently, colleagues at Sanquin are obtaining great results with a prediction model based on ordinary differential equations with different states. [118] These states and equations are based on biological pathways for iron metabolism and erythropoiesis in the human body, and it turns out that this model captures changes in ferritin and hemoglobin very well. The current model requires only one

Source of variation	Implementation alternatives
Policy change	No policy change During (gradual) implementation of new policy After a change in policy
Timing of sampling for hemoglobin measurement	Before donation After donation
Method of sampling	Capillary, by finger-prick Venous sample, by venipuncture or via sampling bag
Repeated measurements	No repeated measurements Repeat same method if measurement is below threshold level Measure with different method if measurement is below threshold level
Hemoglobin deferral threshold	7.8 mmol/L for women, 8.4 mmol/L for men 7.4 mmol/L for women, 8.1 mmol/L for men
Additional iron and/or hemoglobin-related requirements	None Ferritin measurement Threshold for frop in hemoglobin relative to previous measurement
Maximum number of donations per year	Three for women, five for men One for women under 25, three for women over 25, five for men
Iron supplementation	Yes, provided or prescribed Yes, recommended No
Trigger to donate	Invitation-based Walk-in Mix of invitation-based and walk-in

Table 10.1: Sources of variation in donation policies and a non-exhaustive list of implementation differences between countries.

initial hemoglobin and ferritin measurement to predict subsequent levels and could potentially be improved by updating the model when new measurements are taken.

In our support vector machine models, we incorporated time by including up to five previous hemoglobin measurements as predictor variables, together with the time passed since these measurements. This way of including time-related variables is not optimal because the model does not allow linking the measurement results explicitly to the actual times when these measurements were obtained. It is therefore reasonable to expect that models that do incorporate such dependencies would perform better. Therefore, a potential other type of model worthwhile further exploration is the transformer network, which is a type of neural network architecture that has recently been used widely due to its high performance. [136] Although the most popular applications of transformer networks are in natural language processing (e.g., ChatGPT), they are very suitable for time series forecasting applications as well. [137] Transformer networks can model the relationship between measurements that are further apart, unlike recurrent neural networks. [136] However, a potential handicap may be the fact that the number of donation events from an individual donor are small relative to the length of the number of events for usual applications of transformer networks, which may limit the improvement in performance they may bring.

Hemoglobin deferral rates are currently very low in the Netherlands: about 3% and 1% of blood bank visits, by women and men respectively (reduced from 8% and 5% before implementation of the ferritin-based donor deferral policy). Any decrease in deferral rate is of course a good thing, and even a decrease of 0.5 percentage points would mean that 2000 on-site deferrals are prevented on a yearly basis. This potentially saves the blood bank the recruitment of several hundred new donors, as on-site deferral is known to be associated with donors dropping out of the donor pool. [29] However, although hemoglobin deferral rates currently are low, many donors are now deferred for low ferritin levels (approximately 10% of blood bank visits [135]), and therefore understanding how hemoglobin and ferritin are affected by blood donation remains extremely important.

The world of blood bank research has many opportunities for data science. More and more people see that data science can bring advances and improvements, but the actual implementation in day-to-day blood banking is far from easy. The primary task of blood banks will always be to ensure a safe and steady blood supply, and even a large increase in efficiency is not worth a small decrease in safety. Meanwhile, efforts are being made to make more room for data science research: Sanquin is preparing to set up the donor biobank *Sanquin Future Health*, which will process and store remainders

of blood donations to be used for research purposes. With repeated sampling and questionnaires from donors throughout the Netherlands, this biobank will be a treasure trove of data in a few years. As more data is collected, more complex models can be used to find new insights to enhance hemoglobin and ferritin predictions as well as inspire completely new research.

All blood banks struggle with the same balancing act: collecting enough donations to ensure a sufficient blood supply, while preventing iron deficiency and anemia in donors. As more insight is obtained in iron metabolism and how it is affected by blood donation, it will become clearer where there are opportunities to optimise donation strategies. Often, these are small steps that over the course of the coming years are likely to add up to substantial changes. In the future, this increased knowledge can lead to data-driven donation strategies, making optimal use of the information present in our data, resulting in a sufficient blood supply, maintained by healthy, motivated donors.