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

Vinkenoog, M.

CHAPTER

1

Introduction

Introduction

A safe and steady blood supply is essential for a healthy society, and requires a blood bank to collect and manage blood products. Blood that is collected from donors can be used for transfusion or for manufacturing various medicines – in either case, blood saves lives. Patient safety must be ensured; therefore donated blood is extensively tested for infectious diseases, and donors are not allowed to donate if the pre-donation check indicates a potential presence of blood-transmissible disease. In the Netherlands, adverse events for patients after a blood transfusion are extremely rare. [1] In addition to patient safety, donor safety is a priority for blood banks. Both from an ethical and practical perspective, it is important that people only donate blood when it does not harm them. The pre-donation screening visit therefore consists of a questionnaire and blood tests that assess both patient and donor safety.

The biggest health risk in terms of prevalence for blood donors is iron deficiency. [2, 3] With every whole blood donation, donors lose about 250 mg of iron, which is 8 to 13% of their iron stores for men and non-menstruating women, and up to 81% for menstruating women. [4] In general, iron deficiency symptoms may begin mild and vague, including fatigue, increased irritability, and difficulties with concentration. As the deficiency progresses, it can evolve into iron deficiency anemia, meaning that there is a shortage of healthy red blood cells. Symptoms will intensify as the need for iron increases. [5] A recent systematic review investigated associations between iron deficiency in whole blood donors and several health consequences related to iron deficiency. [4] Although most included studies reported a high prevalence of iron deficiency among blood donors, no clear overall association for most iron deficiency-related symptoms was found, and only restless leg syndrome and pica (the act of eating non-food items) were associated with iron deficiency in blood donors. [4]

To prevent anemia in blood donors, most blood banks implement pre-donation checking of hemoglobin or ferritin levels, proteins that transport oxygen and carbon dioxide and store iron, respectively. Donors that do not meet the eligibility criteria for safe blood donation are deferred, i.e., sent home without donating blood. Although it protects donor health, deferral is demotivating for donors, and is often a reason for donors to stop donating altogether. However, a stable donor pool is needed to maintain a steady blood supply, and because retaining existing donors is less costly than recruiting new donors, it is in blood banks' interests to keep donors healthy and motivated.

The Dutch blood bank Sanquin has data on millions of donations, including data

relevant to donor health, and measurements of hemoglobin and ferritin levels in particular. Using statistical and machine learning methods, the information contained in this data can be used to develop algorithms that may help to improve blood bank policies. Combining expert biomedical knowledge gained over decades of blood banking with new insights obtained with data science will allow blood banks to move towards data-driven donation strategies. Similar to precision medicine, blood donation could become more data-driven as well, for instance with tailored donation intervals or eligibility thresholds.

1.1 Research questions and contributions

Sanquin's mission is rooted in its commitment to being a knowledge-driven organisation that supplies life-saving products while upholding careful, responsible, and efficient processing of the voluntary contributions made by donors. Over time, Sanquin has collected large amounts of data on donors and their donations. These data can be leveraged by proper analyses, which would allow Sanquin to advance towards more data-driven donation strategies. Accurately predicting various donor outcomes holds the potential to optimise the blood bank process, for example, by anticipating donor deferral and subsequently adapting donor invitation strategies to minimise such deferrals.

In this thesis, we explore the application of data science in enhancing donor management. By employing several statistical and data science analysis techniques on blood donation data, we address research questions that bear significance for donor health monitoring and protection. The studies focus on investigating a series of research questions that have been categorised into three primary areas. What follows is a list of research questions that are studied in this thesis and their contributions to current knowledge.

Research questions on hemoglobin and ferritin levels and recovery after donation:

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

In 2017, Sanquin implemented a new policy wherein donors' ferritin levels are measured routinely and donors with low ferritin levels are deferred for six or twelve months. We analysed changes in ferritin levels of deferred donors.

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

Many factors that affect iron stores, including ferritin levels, are known and have been extensively researched. Most studies investigate associations from a single perspective, e.g., focussing only on donation-related variables, or only on environmental variables. In Chapter 4, we present a statistical model that integrates multiple sets of variables to give a more comprehensive overview of determinants of ferritin. Differences between donors and non-donors are investigated.

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

Some donors exhibit very stable hemoglobin levels during their donor career, while others show a declining trend. In Chapter 5, we regard these hemoglobin trajectories as time series and use clustering methods to find groups of donors with similar trends. Clustering is complicated when time series are very irregular and sparse, as in this case, with only a few data points per donor per year and no information about hemoglobin levels in between data points. Two methods to tackle the irregularities in these time series are compared.

One research question does not concern hemoglobin or ferritin at all, but rather focuses on SARS-CoV-2 antibodies. Before (and after) the pandemic, the main reasons for donor deferral were low hemoglobin or low ferritin levels. During the pandemic, however, the most common reason donors could not donate was the presence of a COVID-19 infection. The following research question is studied:

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

During the COVID-19 pandemic, Sanquin monitored antibodies in regular donations, but also specifically repeatedly measured antibodies in donors who had undergone a COVID-19 infection. Chapter 6 presents a linear mixed-effects model relating antibody decay to characteristics such as sex, BMI, and age, as well as the presence of various COVID-19 symptoms. At the time of publication, this was the largest study describing these associations, and one of the few studying antibodies in a non-hospitalised cohort.

Research questions on hemoglobin deferral prediction:

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

Chapter 7 presents the main hemoglobin deferral prediction model that was developed, using support vector machines. We assessed the consequences for the blood supply if these models were used to guide donor invitations, as well as prediction performance. Additionally, it focuses on explaining why the model makes certain predictions, opening the 'black box' and analysing if associations learned by the model are consistent with the physiology behind hemoglobin metabolism. This research question is also relevant in Chapters 8 and 9.

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

Although blood banks from many countries are working on hemoglobin deferral prediction, exchange and comparison of results is rarely done. Chapter 8 is the first publication from the international research group SanguinStats, a collaborative effort across five countries: Australia, Belgium, Finland, the Netherlands, and South Africa. In this chapter, researchers from all countries fit the same five prediction models to their blood bank data. Both prediction performance and variable importance are analysed for differences and similarities.

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

Many blood banks collect additional information that may improve predictions: specifically, the Finnish blood bank has collected information on iron-related genetic markers and in the Netherlands, ferritin measurement data are available. In Chapter 9, the added value of these predictor variables is investigated and compared.

1.2 Outline of this thesis

Following this introduction, this thesis continues with Chapter 2 – *Preliminaries*. In this chapter, all necessary background information needed to understand the work presented in this thesis is provided: both from a blood donation and a data science perspective.

Chapters 3 through 9 contain the research papers as published, starting with three papers on iron marker levels and their recovery after donations. Right in the middle, Chapter 6 interrupts the studies on hemoglobin and ferritin levels with a research paper on SARS-CoV-2 antibodies, just as the COVID-19 pandemic interrupted me in the middle of my PhD research. Chapters 7 through 9 focus on hemoglobin deferral prediction models.

The thesis is wrapped up with Chapter 10 – *Conclusions, general discussion and anticipated future research*, which summarises the results from Chapters 3 through 9, discusses overarching challenges, and proposes potential directions for future research and policies.

1.3 List of publications

The chapters in this thesis are based on the following publications. Publications are edited only for style cohesion; all content remains unchanged from the published versions.

Chapter	Publication
3	Vinkenoog M, van den Hurk K, van Kraaij M, van Leeuwen M, & Janssen MP (2020). *First results of a ferritin-based blood donor deferral policy in the Netherlands.* Transfusion 60(8), 1785-1792.
4	Vinkenoog M, de Groot R, Lakerveld J, Janssen MP, & van den Hurk K (2023). *Individual and environmental determinants of serum ferritin levels: A structural equation model.* Transfusion Medicine, 33(2), 113-122.
5	Vinkenoog M, Janssen MP, & van Leeuwen M (2019). *Challenges and limitations in clustering blood donor hemoglobin trajectories.* International workshop on advanced analysis and learning on temporal data, 72-84.
6	Vinkenoog M, Steenhuis M, Brinke AT, van Hasselt JG, Janssen MP, van Leeuwen M, Swaneveld FH, Vrielink H, van de Watering L, Quee F, van den Hurk K, Rispens T, Hogema B & van der Schoot CE (2022). *Associations between symptoms, donor characteristics and IgG antibody response in 2082 COVID-19 convalescent plasma donors.* Frontiers in immunology, 13.
7	Vinkenoog M, van Leeuwen M, & Janssen, MP (2022). *Explainable hemoglobin deferral predictions using machine learning models: Interpretation and consequences for the blood supply.* Vox Sanguinis, 117(11), 1262-1270.
8	Vinkenoog M, Toivonen J, Brits T, de Clippel D, Compernolle V, Karki S, Welvaert M, Meulenbeld A, van den Hurk K, van Rosmalen J, Lesaffre E, Arvas M & Janssen MP (2023). *An international comparison of hemoglobin deferral prediction models for blood banking.* Vox Sanguinis, 118(6), 430-439.
9	Vinkenoog M, Toivonen J, van Leeuwen M, Janssen MP & Arvas M (2023). *The added value of ferritin levels and genetic markers for the prediction of hemoglobin deferral.* Vox Sanguinis, 118(10), 825-834.