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

Vinkenoog, M.

CHAPTER

4

Individual and environmental determinants of serum ferritin levels: a structural equation model

Published in Transfusion Medicine 33(2): 113-122. doi:10.1111/tme.12902

Authors: M Vinkenoog, R de Groot, J Lakerveld, MP Janssen, K van den Hurk

Abstract

Background - Serum ferritin levels are increasingly being used to assess iron stores. Considerable variation in ferritin levels within and between individuals has been observed, but our current understanding of factors that explain this variation is far from complete. We aim to combine multiple potential determinants in an integrative model, and investigate their relative importance and potential interactions.

Methods - We use ferritin measurements collected by Sanquin Blood Bank on both prospective ($N = 59596$) and active blood donors ($N = 78318$) to fit a structural equation model with three latent constructs (individual characteristics, donation history, and environmental factors). Parameters were estimated separately by sex and donor status.

Results - The model explained 25% of ferritin variance in prospective donors, and 40% in active donors. Individual characteristics and donation history were the most important determinants of ferritin levels in active donors. The association between environmental factors and ferritin was smaller but still substantial; higher exposure to air pollution was associated with higher ferritin levels, and this association was considerably stronger for active blood donors than for prospective donors.

Conclusions - In active donors, individual characteristics explain 20% (17%) of ferritin variation, donation history explains 14% (25%) and environmental factors explain 5% (4%) for women (men). Our model presents known ferritin determinants in a broader perspective, allowing for comparison with other determinants as well as between new and active donors, or between men and women.

Introduction

Iron is essential for human life, but both iron deficiency and iron overload can cause various adverse health effects. Therefore, iron homeostasis is tightly regulated in humans. In case of insufficient availability of iron in the circulation, recycling of old red blood cells is increased and hepcidin is downregulated both to increase dietary iron absorption and release iron stored in ferritin. [8, 35] Hemoglobin levels have long been the standard method to assess iron status. However, hemoglobin levels can remain sufficient for some time, even when iron stores are dwindling; this is known as iron deficiency non-anemia. [8]

In contrast to hemoglobin, serum ferritin levels reflect the amount of stored iron. [8] Therefore, they are increasingly used to assess individuals' iron stores when these are at risk, for instance after traumatic blood loss, during pregnancy, or in blood donors. [21] Sanquin, the national blood service in the Netherlands, started measuring ferritin levels in each new donor, and subsequently after every fifth donation, in October 2017. Donating blood has a substantial impact on ferritin levels. Ferritin levels are lower among blood donors than in the general population: cross-sectional studies report lower ferritin levels in donors with a higher number of whole blood donations and a large randomised trial showed that ferritin levels indeed decline with more frequent blood donations. [36, 37] Among new donors, large variation in ferritin levels is observed. [36] It is well established that individual characteristics such as sex and age are relevant: women in general, but pre-menopausal women in particular, have considerably lower ferritin levels than men. [36, 2, 38] Higher body mass index (BMI) is associated with higher ferritin levels. [39] In recent decades, many other factors that affect iron status have been identified: diet, [40, 41] genetics, [42, 43] ethnicity, [44] and iron supplementation, which is mostly studied among blood donors. [20, 45]

Ferritin is also a known acute-phase protein that is elevated in inflammatory conditions, complicating its diagnostic value in individuals with conditions such as inflammatory bowel disease or chronic heart failure. [9] This could also explain the association between BMI and ferritin levels, as adipose tissue is known to promote systemic inflammation. [46] Additionally, exposure to environmental pollutants has been linked to disordered iron homeostasis, [47, 48] and ambient particle matter (PM) concentration is correlated with ferritin levels. [48] The biological mechanism behind this is still unclear, but it is postulated that iron attaches to the PM rather than to cell nuclei, effectively creating a functional deficiency. [47, 48] In turn, mechanisms start upregulating iron uptake and recycling in an attempt to meet the iron require-

ment of the cells, thereby altering iron homeostasis. Another suggested mechanism is that when pollutants enter the lungs, iron is transported away from the surface of the lung tissue and stored in ferritin complexes, in order to avoid chemical reactions between iron and the pollutant. [47] Other potential environmental determinants are neighbourhood characteristics, including population density and socio-economic status, which are consistently shown to be related to body weight and blood parameters. [49]

Previous studies on ferritin levels have focused on studying the association with variables in a limited setting, for example, characteristics such as age and BMI, donation-related variables, or environmental pollutants. In this paper, we propose a novel framework that integrates multiple settings, using a structural equation model. By grouping relevant explanatory variables into constructs, we describe relationships with ferritin on a more general level. This enhances the insight into various mechanisms that influence ferritin levels, which is valuable to those who use these as a diagnostic tool. We explore associations between ferritin levels and individual characteristics, donation behaviour and environmental factors, in a large group of newly registered and active whole blood donors.

Methods

For this cross-sectional study, data collected by Sanquin and the Geoscience and health cohort consortium (GECCO) were analysed. Sanquin is by law the only blood service in the Netherlands, collecting over 400 000 whole-blood donations each year, with collection sites geographically well-distributed throughout the country. Several eligibility criteria exist to ensure the safety of the donors and recipients and the quality of the blood product. Donors must be aged between 18 and 79 years old, and a pre-donation screening visit takes place before the first 500 mL whole blood donation, which includes blood sampling for blood type and infectious disease testing, as well as initial hemoglobin and ferritin measurements. We will refer to these prospective donors, who have not donated yet, as *new donors*.

Before every donation, a donor screening is performed, including a donor health questionnaire and measurements of blood pressure, pulse rate and hemoglobin levels to assess whether the donor is eligible to donate. Hemoglobin levels need to be at least 7.8 mmol/L for women and 8.4 mmol/L for men. This is measured by point-of-care testing with a photometer (HemoCue, Angelholm, Sweden). Ferritin levels, are measured in serum samples, using the Architect i2000 (Abbott Diagnostics, Chicago,

IL), after the pre-donation screening visit and after every fifth whole blood donation. As such, ferritin measurements are only available in case of successful whole blood donations, and for new donors whose venous samples are taken as part of the pre-donation screening visit.

Data

This study included all new and active whole blood donors who gave consent to the use of their data for scientific research (this consent is given by > 99% of all donors) and for whom ferritin measurements were available between 1 October 2017 and 31 December 2019. If multiple ferritin measurements were available for a donor, only the first measurement was used. Information on donors and donation histories was extracted from the blood bank information system (ePROGESA, MAK-SYSTEM International Group, Paris, France). Variables used were sex, age, height, weight, time since previous successful donation, the number of successful donations in the previous 2 years, donor status (new or active donor), and ferritin levels. BMI was calculated from self-reported donor height and weight. Sanquin does not register donor ethnicity, but Duffy negative phenotype was included to function as a proxy for sub-Saharan African descent.

Environmental exposure variables of various characteristics were obtained from the Geoscience and health cohort consortium (GECCO). [50] The exposure data were operationalised based on publicly available data. Data from 30 weather stations in the Netherlands—obtained from the Royal Netherlands Meteorological Institute (KNMI)—were used to estimate temperature at a spatial resolution of 1 km. Three options for the measurement level were considered (minimum, average, and maximum daily temperature), as well as three time spans (day, week or month before donation), resulting in nine options in total. The combination that showed the highest correlation with ferritin was included in the final model.

Daily concentrations for particulate matter (PM) 2.5, PM10, NO2, ozone and soot levels were obtained via the Dutch National Institute for Public Health and the Environment (RIVM), for the years 2017–2019. These variables were imputed on a spatial resolution of 1 by 1 km. Neighbourhood socio-economic status (SES) scores and population density from 2017–2019 were acquired from Statistics Netherlands (CBS), both available on 6-digit postal code level. SES scores are based on percentiles of income, education level and vocational history of households, with a score of 0 being exactly the national average, and positive scores being above average. All spatio-temporal variables were matched with donor and donation data based on donation date and donor postal code. Lastly, the date and time of each donation were included

as potential factors to account for seasonal and diurnal variation, as they are known to affect hemoglobin levels and may also affect ferritin levels.

To check for a possible confounding effect of smoking on environmental variables, we analysed the correlation between the percentage of smokers per municipality (data from Statistics Netherlands) and all environmental variables described in the above paragraph.

There were no missing data for environmental datasets from the RIVM and CBS. Donors with no ferritin measurement were excluded from the analysis. There were no missing data for the other donor or donation level variables.

Statistical analysis

Structural equation modelling (SEM) was used to investigate which variables relate to serum ferritin and to what extent. Briefly, observed variables and latent constructs are distinguished in SEM. Latent constructs cannot be measured or observed directly, but are inferred from the observed variables. One or more hypothesised sets of relationships and correlations between variables and constructs are specified a priori and shown in a path diagram. For each relationship, a parameter is estimated that indicates its strength. Estimates are obtained by numeric optimisation of a fit criterion, using maximum likelihood estimation. A more detailed overview of this method is provided in the Appendix.

We compared four ways to divide the 15 variables included in the analysis into latent constructs, as shown in Table 4.1. Date and time of the donation were added to the model separate of the constructs, and as such are not included in Table 4.1. Model A contains four latent constructs, and in models B, C and D different sets of constructs are combined. Confirmatory factor analysis (CFA) was used to test the validity of the specified measurement models, that is, the hypothesised relationships between the latent constructs and their observed variables. The overall fit of the models was assessed by the Tucker-Lewis Index (TLI) and the root mean square error of approximation (RMSEA). A rule of thumb is to exclude variables for which the absolute value of the standardised factor loading is below 0.4, but at sample sizes larger than 300, if the overall model fit is good, exclusion is not necessary and should be judged separately for each variable based on sensible background knowledge. [51]

Pairwise residual correlations between observed variables were calculated to identify whether any covariances needed to be added to the model. Of the four specified models, we continued our analysis with the best fit according to CFA, based on the TLI and RMSEA.

Variable	Model A	Model B	Model C	Model D
Age Height Weight BMI Duffy phenotype	Individual characteristics	Individual characteristics	Individual characteristics	Individual characteristics
Time since prev. donation Number of prev. donations	Donation history	Donation history		
Population density Temperature Socio-economic status	Environment	Environment	Environment	Environment
Ozone PM2.5 PM10 Soot NO2	Pollution		Pollution	

Table 4.1: Grouping of variables into constructs for each model. Note that variables *time since previous donation* and *number of previous donations* are only available for active donors.

To the model with the best fit, we added the structural component, which contains the relationships between the latent variables and ferritin, the outcome variable. A multiple group SEM was carried out with parameters estimated separately for male and female donors, and for new and active donors. Because the assumption of normality of the explanatory variables does not hold in our data, a different estimator than the default maximum likelihood estimator was used: the 'mean and covariance adjusted weighted least squares estimator', which is robust against violations of the normality assumptions in a multivariate setting. [52]

The same model was fitted in all four groups, although the variables belonging to the donation history construct (see Table 4.1) are not available for new donors, as they do not (yet) have a donation history. The overall fit of the SEM model was assessed using the TLI and RMSEA, as well as the R2 measure.

All analyses were conducted using R programming language and environment for statistical computing version 4.0.3, with package zoo for pre-processing environmental data, and lavaan for CFA and SEM analyses. Path diagrams were created with yEd Live Graph Editor.

Results

Sample composition

Table 4.2 shows descriptive statistics of the study population by sex and donor status. The size of each of the groups was comparable, except for the group of new male donors, which was only half the size of the other groups. Between new and active donors, age differed considerably, new donors being younger than active donors by 17 years on average ($p < 0.001$). In both new and active donors, men were older (by 6 years on average, $p < 0.001$) and heavier (by 13 kg on average, $p < 0.001$) than women. P-values were obtained using two-sample t-tests. The time since last donation is higher in women than in men, and the number of prior donations is higher in men than in women. These differences are due to differences in the minimum required donation interval: for women, there must be 122 days between two donations with a maximum of 3 donations per year, while for men, the minimum is 57 days between two donations with a maximum of 5 donations per year. Differences in ferritin levels between the groups are as expected from previous studies: men have higher ferritin levels than women, and repeat donors have lower ferritin levels than new donors.

For pollution and environmental variables, there was little difference between the

groups, any differences between new and active donors were most likely due to the different age and geographical distribution of the groups. None of these differences were statistically significant.

We found a weak correlation between the percentage of smokers and SES score (Pearson's $r = -0.4$) and a moderate correlation between the percentage of smokers and population density (Pearson's $r = 0.5$). No correlation was found for any of the other environmental variables.

	New donors		Active donors	
	Women	Men	Women	Men
N	40172	19424	39085	39233
Age (years)	26 (21–37)	28 (23–37)	47 (31–58)	53 (39–62)
Height (cm)	170 (166–175)	183 (178–188)	170 (166–175)	183 (178–188)
Weight (kg)	68 (62–77)	82 (74–90)	70 (64–80)	85 (78–93)
BMI (kg/m2)	24 (21–26)	24 (22–27)	24 (22–27)	25 (23–27)
Time since prev. donation (days)	NA	NA	154 (132–217)	139 (71–147)
Number of prev. donations	NA	NA	3 (2–4)	5 (4–7)
Population density (per km2)	1173 (425–2617)	1246 (477–2936)	827 (322–1855)	814 (320–1824)
Duffy phenotype (proportion)	0.25	0.17	0.28	0.16
Temperature (°C)	11.4 (6.4–16.6)	11.7 (6.6–16.7)	10.4 (6.0–16.0)	10.4 (5.9–16.0)
Socio-economic status	0.04 (-0.21 to 0.22)	0.02 (-0.24 to 0.22)	0.10 (-0.10 to 0.25)	0.12 (-0.07 to 0.26)
Ozone (µg/m3)	46.9 (45.6–48.8)	46.8 (45.5–48.7)	47.2 (45.9–49.2)	47.2 (45.9–49.1)
PM2.5 (µg/m3)	10.7 (9.7-11.6)	10.7 (9.8–11.6)	10.5 (9.6–11.5)	10.6 (9.7–11.6)
PM10 (µg/m3)	18.2 (16.8–19.3)	18.2 (16.9–19.3)	18.0 (16.6–19.0)	18.0 (16.7–19.1)
Soot (µg/m3)	0.66 (0.54–0.78)	0.66 (0.55–0.78)	0.63 (0.52–0.75)	0.65 (0.54–0.76)
NO2 (µg/m3)	17.6 (14.9–21.6)	17.8 (15.1–21.8)	16.8 (14.2–19.7)	16.9 (14.3–19.6)
Ferritin (µg/L)	47 (28–75)	118 (79–170)	30 (17–47)	34 (20–56)

Table 4.2: Distribution of explanatory variables by donor status and sex.

Model selection

CFA did not provide support for the environment construct as defined by the three variables temperature, population density and socio-economic status. These variables did not share a high proportion of their variance and consequently there was no convergent validity, effectively ruling out models A and C. In models B and D, variables Duffy phenotype, temperature, SES and height were omitted due to very low factor loadings (< 0.05). The factor loading for variable age was also low (0.35) but this variable was not excluded, as it is expected that this factor loading would be small, considering the other variables in the construct (weight and BMI) are much more closely related. All other factor loadings were above the suggested threshold of 0.6. All latent constructs (individual characteristics, donation history and environment) showed convergent and discriminant validity in models B and D. Variables time and day of year, which were added to the model outside the constructs, were also dropped due to very low factor loadings (< 0.05).

The presence of a donation history construct was the only difference between models B and D, and since new donors do not yet have a donation history, the models only differed for active donors. Model B had a TLI of 0.961 and RMSEA of 0.063, while model D had a TLI of 0.932 and RMSEA of 0.083. Based on these fit measures, model B fit the data best, and was therefore used in the remainder of the analyses.

Based on inspection of the pairwise residual correlations between all observed variables, two covariance terms were added to the model: one for PM2.5 and PM10 (residual correlation 0.092 to 0.102, depending on sex/donor status), and one for age and population density (residual correlation -0.151 to -0.149, depending on sex/donor status). We also added one covariance term for weight and BMI, as BMI was calculated using weight and was therefore inherently dependent.

Parameter estimates

Figure 4.1 shows the structure of the final model and the parameter estimates for new donors. Parameter estimates were similar for both sexes, but factor loadings for variables belonging to the individual characteristics construct were higher for women than for men, indicating more shared variance. Factor loadings in the environment construct did not differ between sexes, showing that the covariance structure of those variables was not dependent on sex. The parameter estimates for the regression coefficients show the relative importance of each latent construct for the outcome variable. Table 4.3 shows the percentage of variance in ferritin levels that is explained by each

Construct	New donors		Active donors	
	Women	Men	Women	Men
Individual characteristics	23%	23%	20%	17%
Donation history	NA	NA	14%	25%
Environment	2%	2%	5%	4%
Total variance explained	25%	25%	39%	46%

Table 4.3: Relative contribution to explanation of variance of ferritin levels per model.

construct for each model, adding up to the total percentage of variance explained.

Figure 4.2 shows the final model for active donors. As in new donors, factor loadings in the individual characteristics construct were higher for women than for men, and they were also higher for new donors than for active donors. The relative importance of individual characteristics and donation history was opposite for both sexes: for men, donation history was correlated with ferritin levels more strongly than individual characteristics (0.66 vs. 0.45), while this was reversed for women (0.43 vs. 0.61). The regression coefficient of the environment construct is 0.15 for women and 0.10 for men. The environment construct explains twice as much variation in ferritin levels in active donors as in new donors.

As for overall model fit, with a TLI of 0.981 and 0.979 and RMSEA of 0.052 and 0.042, for new and active donors respectively, both models fit very well when compared to commonly used thresholds (TLI > 0.95, RMSEA < 0.06). [53] R2 was calculated separately by sex: for new donors, R2 was 0.251 for men and 0.252 for women, and for active donors, 0.458 for men and 0.393 for women.

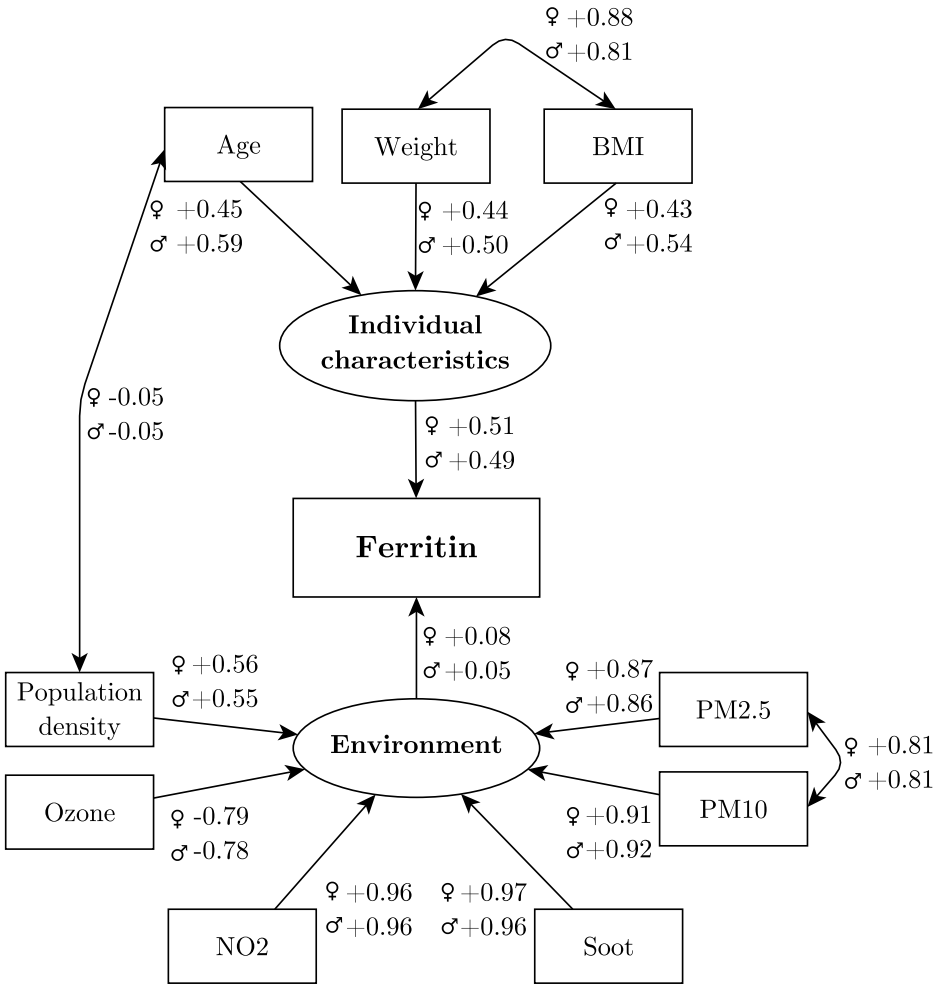

Figure 4.1: Final structural equation model for ferritin determinants in new donors, with parameters estimated separately for men and women. All parameter estimates are standardised so that the variance of each observed variable and latent construct equals 1.

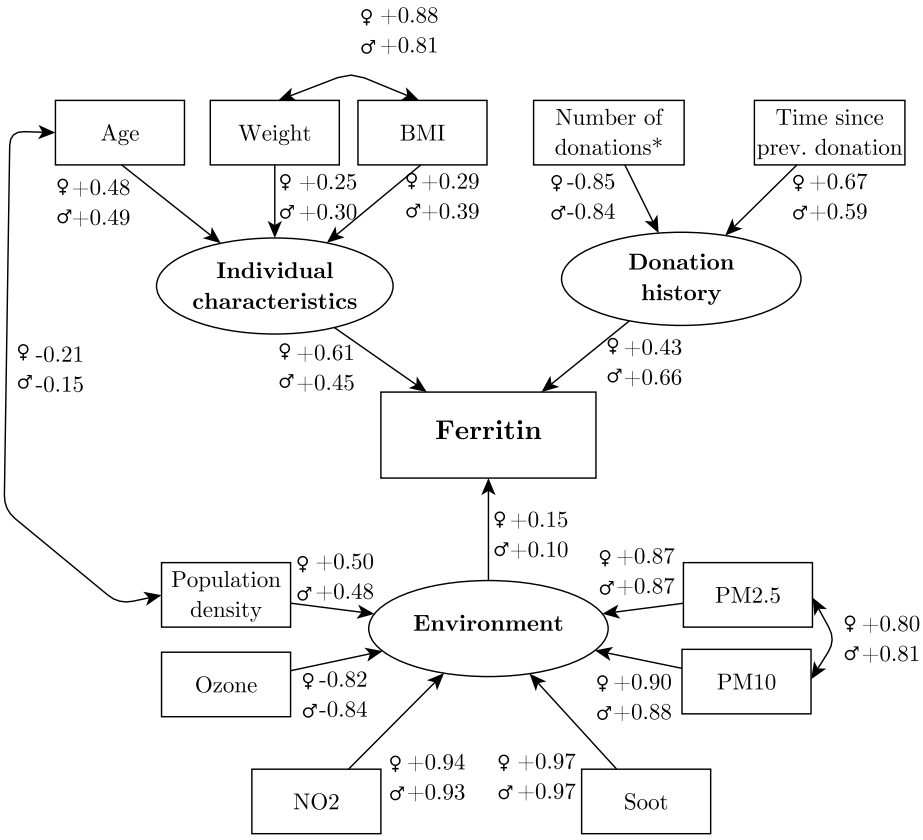

Figure 4.2: Final structural equation model for ferritin determinants in active donors, with parameters estimated separately for men and women. All parameter estimates were standardised so that the variance of each observed variable and latent construct equals 1.

Discussion

This study investigated the impact of individual and environmental determinants on ferritin levels in Dutch individuals, using SEM. The model was able to explain 25% of ferritin level variance in new donors for both sexes, and 46% and 39% in active donors for male and female donors, respectively.

We found the construct composed of individual characteristics (age, weight, and BMI) to be the most important determinant of ferritin in female active donors, followed by donation history (time since previous donation, number of donations in the past 2 years). For male active donors, this was the opposite: donation history was a more important determinant than individual characteristics. In both sexes, environmental factors are associated with ferritin levels, albeit to a lesser degree than individual characteristics and donation history.

The relationship between ferritin levels and anthropometric characteristics is well-documented, and the positive correlations we found for ferritin with age, weight and BMI are consistent with those found in other studies. [36, 45, 54] Men have much higher ferritin levels than women in general and show a larger decrease in ferritin levels after repeated donations. As a result, ferritin levels in active donors are similarly low for women and men. [36] The donation history construct explained more variance in ferritin levels in men than in women. Although often not explicitly mentioned, this discrepancy is also found in previous studies, with stronger relationships between variables regarding donation history and ferritin for men than for women. [45] A reasonable explanation for this is that men commonly display more variation in donation history variables due to the possibility of more frequent donations: in many blood services, men are allowed to donate more often than women and are usually less frequently deferred for low hemoglobin levels. [55]

From previous epidemiological studies, we know that environmental factors may play a role in iron metabolism, and that certain pollutants can disrupt iron homeostasis. [56] Our study shows that although environmental factors are less strongly associated with ferritin levels than individual characteristics and donation history, their effects are far from negligible. Because of the wide reach of environmental exposures over geographic areas, even a relatively small influence on individuals can result in a large effect on the population level. As this study includes only data from the Netherlands, which is a relatively small country, associations between environmental variables and ferritin levels were not very strong, as was expected. Repeating this study on a larger, or even global, scale may result in finding a more substantial effect.

Higher values for all but one environmental factor (ozone) were positively correlated with higher ferritin levels. These findings support the hypothesis that air pollution causes higher ferritin levels. The underlying mechanism may be that when certain pollutants enter the lungs, iron is transported away from the lung tissue surface and stored in ferritin complexes to avoid chemical reactions between iron and the pollutant. [47, 57] This would imply that using serum ferritin as a proxy for total body iron is less reliable when there is significant air pollution.

The environment construct was more strongly associated with ferritin level in active donors than in new donors. In new donors, environmental factors explain 2% of variance in ferritin levels, while in active donors this increases to 4% to 5% depending on sex. This indicates that environmental factors are more important for ferritin recovery after blood loss than for naive ferritin level. A plausible explanation for this difference is that since both exposure to air pollution and donating blood causes significant disruptions to iron homeostasis, these disruptions may interact and together have a larger effect than simply additive.

SEM is a technique well-suited to test hypotheses on how different factors interact and correlate with a specific outcome like ferritin levels, especially when there are many factors to consider. Compared to multiple (linear) regression, more complex models can be tested, and for each variable measurement error is taken into account. [58] Moreover, the percentage of variance explained by groups of related variables can be calculated and compared. The stratified approach in this study also adds to the model validity: parameter estimates can be compared across groups, allowing discovery of implausible results. Our analyses show that the convergent validity of the individual characteristics construct is lower for active donors than for new donors. This may indicate that new donors are a more homogenous group than active donors, which is likely due to the more narrow age range of new donors. Other strengths of this study are its large sample size and collection of data throughout the country.

Two main limitations of this study should be noted: its generalisability and its restricted scope. One might be tempted to generalise the results of new donors to the general Dutch population, as these donors have never donated blood before. However, even new donors form a very specific, generally healthier subgroup of the general population, which means that selection bias has likely been introduced. We can speculate that less healthy individuals would show a higher rate of inflammation, which may cause higher serum ferritin levels. On the other hand, iron deficient or anaemic individuals are likely underrepresented in our sample. As this selection bias most likely reduced variance in ferritin levels, this may have attenuated our results.

Regarding the scope, data on some other potentially important determinants of ferritin levels were not available in this study, the two most important being genetics and diet. [40, 41] Several genetic polymorphisms that have an effect on iron pathways have been identified, and these are likely to play a role in the recovery speed of ferritin levels after blood donation. [43, 59, 60, 61] Dietary behaviour, and in particular heme iron intake, is also a determinant of iron status in donors. [40, 45] Information on iron supplementation was also not available for this study. Sanquin does not prescribe oral supplementation of iron to donors, and only a small minority (8.7%) uses iron supplements. [40] Information on donors' smoking status is also expected to add value to the model. Had these determinants been available for our analysis, the proportion of variance explained in donor ferritin levels would likely have increased.

This study presents a model to explain variance in ferritin levels in individuals with or without donation history, based on three types of determinants. The model explained a relatively large part of the variance, especially in active donors. Individual characteristics and donation history form the most important determinants of ferritin levels. Although environmental factors accounted for less variance than the individual and donation history constructs, their contribution is meaningful and statistically significant. When clinicians or researchers use serum ferritin as a proxy for total body iron, they should be aware of this potentially confounding effect.

For blood services that are considering implementing ferritin testing for their donors, these results are of particular value. The results can be of use while the blood service is deciding on a sensible threshold for donation: rather than implementing a one-size-fits-all threshold, environmental conditions in the country can be taken into account. If there is a high level of air pollution, ferritin levels are likely to be overestimated, and thus a higher threshold for donation may be desired. It could even be taken further to make ferritin thresholds more tailored to a specific donor, by taking into account a donor's individual characteristics.

Appendix

Structural equation modelling (SEM) comprises a set of statistical methods that enables researchers to assess the support for hypothesised relationships between variables of interest. Its purpose is to account for variation and covariation of the variables in the model. Many different techniques are included in SEM, this appendix explains the approach taken in this particular study. In SEM, observed variables and latent constructs are distinguished. Observed variables are variables in the traditional sense, which are observations in the data set that have been collected by the researcher. Latent constructs are theoretical concepts that cannot be measured, but must be inferred from the observed variables; a well-known example is the latent construct intelligence that cannot be measured directly, but can be inferred from observed variables such as scores for an IQ test. Intuitively, observed variables that belong to a latent construct represent the same underlying concept, and latent constructs form in a way a dimensionality reduction of the observed variables. Mathematically, latent constructs represent shared variance of the observed variables related to the construct they belong to.

SEM is composed of two main model components: the measurement model, which shows how observed variables are divided among latent constructs, and the structural model, which shows the relationships between latent constructs and outcome variable(s). First, the measurement model is specified, and test its validity using confirmatory factor analysis (CFA). Often, several measurement models are tested and compared to see which division into latent constructs best fits the data. When the measurement model is considered to have a good fit, the structural part of the model is added, and the model fit is assessed for the full SEM model.

Measurement model

The validity of the latent constructs must be measured in two ways: each construct must have convergent and discriminant validity. Convergent validity occurs when the observed variables belonging to the latent construct share a high proportion of their variance. This is assessed by the factor loadings of the observed variables onto the latent construct: the higher the (absolute value of the) factor loading, the stronger the indication that this variable belongs to this construct. Very generally speaking, factor loadings greater than 0.4 are acceptable for including a variable within a construct, but this threshold depends greatly on the hypothesised interpretation of the latent variable. Variables with low factor loadings are excluded from the construct.

The discriminant validity of a latent construct is a measure for how well the construct can be distinguished from the other constructs in the model. It is measured by the covariances between latent constructs. A high covariance between two constructs can indicate that these constructs are (partly) overlapping, and thus have no discriminant validity.

If convergent and discriminant validity are satisfactory, model fit indices can be calculated for the measurement model. Commonly used indices are the chi-square test, comparative fit index (CFI), Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA). The CFI and TLI are both relative measures of fit, and compare the fit of the tested model against a null model, which in CFA means that the means and variances of each variable are freely estimated, but no correlations are included. CFI and TLI are on a scale from 0 to 1, with higher values indicating a better fit of the hypothesised model relative to the null model. The TLI is always more conservative (lower value) than the CFI, because the TLI includes a harsher penalty for the number of parameters estimated. Because the two fit indices are highly correlated, only one should be reported. We chose the TLI because of its more elegant penalty for complexity. Values higher than 0.95 indicate good fit.

The RMSEA is an absolute measure of fit that is not sensitive to large sample sizes, unlike the chi-square test. It uses the covariance matrix of the entire data set and of the fitted hypothesised model, and calculates the differences between these two. This results in a measure between 0 and 1, with lower values indicating smaller differences and better model fit. Cut-offs of 0.08, 0.05, and 0.01 indicate mediocre, good, and excellent fits, respectively.

If multiple measurement models are compared, as in this study, the best fitting model is selected, based on the fit indices described above. If these indicate sufficient model fit, the analysis can be continued with inspection of residual correlation between observed variables. If the pairwise residual correlation between two variables is high (absolute value of 0.1 or higher is a common cut-off), this indicates that these two variables share more variance than is currently captured in the model. If this occurs, the researcher needs to decide whether a covariance term for these two variables should be included in the model. However, this should only be done if there is sufficient theoretical support for an interpretable correlation between these variables. Otherwise there is a risk of overfitting the model to the data; after all, in confirmatory factor analysis we build upon a set of relationships that are hypothesised by the researcher. It is not a data-driven method of finding the best set of relationships. If such an approach is desired, exploratory factor analysis (EFA) can be applied instead of CFA.

Structural model

The structural component is added to the model once the latent constructs are defined, variables with low factor loadings are removed, and necessary covariance terms are added. The structural component consists of the relationships between latent constructs, or between latent constructs and outcome variable(s). With this, we now have three types of parameters for which an estimate must be calculated:

- Factor loadings (observed variable → latent construct);
- Covariances (observed variable ↔ observed variable);
- Regression coefficients (latent construct → latent construct or outcome variable).

Each parameter adds one degree of freedom to the model, and the number of parameters determines the identifiability of the model. Parameter estimates can only be obtained when the number of free parameters (the number of *unknowns*) is equal to or smaller than the number of independent elements in the covariance matrix of the data (the number of *knowns*), which is equal to $k(k+1)/2$, where k is the number of observed variables in the model. If there are more unknowns than knowns, the model is under-identified and no solution can be found. If the numbers are the same, the model is just identified, and a unique solution can be obtained. If there are fewer unknowns than knowns, we have an over-identified model, which means that there is no unique solution but multiple, and we can select the best solution based on fit measures. An over-identified model is desired.

In most software packages parameter estimates are obtained by a maximum likelihood estimator by default, but alternative estimators can be chosen as well. In this study most observed variables did not follow a normal distribution, which violates maximum likelihood estimator assumptions. Therefore, the diagonally weighted least squares (DWLS) method was used instead, which is more robust and provides more accurate parameter estimates in case the normality assumption is violated.

If the model is over-identified, fit measures can be reported along with the parameter estimates. Again, TLI and RMSEA are used to assess model fit, with the same thresholds as seen in the CFA (TLI > 0.9, RMSEA < 0.08). If the model fit is acceptable the parameter estimates can be interpreted. The interpretation of the parameter estimates depends on the specification of the model. By default, one factor loading in each latent construct is set to 1, to fix the scale of the latent construct. However, in order to compare factor loadings across constructs it is useful to consider standardised parameter estimates. The variance of the latent construct is then set to 1

and factor loadings are interpreted in terms of a change in variance. In this study, we look only at the standardised parameter estimates, as we are interested in the relative importance of each observed variable and latent construct.

Factor loadings indicate how much variance of an observed variable is shared with the variance of its latent construct. Higher absolute values indicate more shared variance, and the sign of the factor loading specifies the direction of the association. Covariance terms provide the same information for two observed variables, which can belong to the same construct or to different constructs. If they belong to the same construct, a high covariance term indicates that these two variables share more variance with each other than can be explained by the latent construct. Regression coefficients indicate how much variance of the outcome variable is explained by the variance of the latent construct. To find the relative effect of a single observed variable on the outcome variable, its factor loading must be multiplied by the regression coefficient that connects the latent construct to the outcome.