Modelling the effect of conditionally repeating hemoglobin measurements prior to blood donation

Authors: Mart Pothast[1*], Katja van den Hurk[2,3], Mart P. Janssen[1]

(1) Transfusion Technology Assessment, Department of Research, Sanquin Blood Supply Foundation, Amsterdam, The Netherlands; (2) Donor Health, Department of Research, Sanquin Blood Supply Foundation, Amsterdam, The Netherlands; (3) Department of Public and Occupational Health and the Amsterdam Public Health research institute, Amsterdam UMC, Amsterdam, The Netherlands

*m.pothast@sanquin.nl

Abstract

Background: At many blood services it is policy to repeat low hemoglobin levels, and in the Netherlands, only the highest value is recorded. This policy introduces a bias and distorts the distribution of recorded hemoglobin values.

Study Design and Methods: We model the effect of conditionally repeating these measurements. We use data from the HAPPEN study to decompose the hemoglobin distribution into variation present in the population and the variation inherent to the measurement. Then, we use our model to predict the recorded hemoglobin distribution in new donors in The Netherlands.

Results and Discussion: The standard deviation of the population hemoglobin distribution is 0.54 mmol/L and the measurement uncertainty of the fingerprick measurement 0.38 mmol/L. Our model accurately describes the shape of the distribution of recorded hemoglobin levels in new donors, but additional effects around the threshold for donation are apparent. The current policy of repeating low measurements reduces the number of false deferrals but increases the

likelihood of allowing truly ineligible donors to donate. To make better use of recorded data, ensure donor health, and support rational deferral decisions, all measurements should be recorded.

Introduction

Hemoglobin (Hb) levels are routinely measured to screen blood donors prior to donation using a point-of-care test. Hb measurements below a predefined threshold lead to temporary deferral from donating. At many blood services, including Sanquin, the Dutch national blood service, it is policy to repeat measurements that are below the threshold in an attempt to reduce on-site deferrals that may be due to measurement error.[1] Unfortunately, at Sanquin only the *highest value* of (at most three) repeated measurements is registered, which results in an upward bias in the recorded Hb levels, as was already reported by Chung et al.[2] This increases the likelihood that a donor with a truly low Hb level is allowed to donate. Also, with these biased Hb measurements it is impossible to correctly monitor donor's Hb levels which are important for maintaining donor health[3] and for the improvement of on-site deferral strategies.[4] Understanding the effect of the repeated measurements is essential when using these data to train predictive models that can aid in reducing on-site deferrals and for deriving optimal donation intervals.[5,6]

In this paper, we model the effect of conditionally repeated measurements on the distribution of recorded Hb levels. This policy does not only induce a bias, but also distorts the whole Hb-level distribution. The approach we follow to understand this effect has two steps. First, we estimate the donor population Hb distribution (the inter-person variation) and the measurement variation from data from the HAPPEN study.[7] Second, we use the policy of conditionally repeating measurements given the population distribution and the measurement uncertainty to model the observed distribution of Hb levels. We show that our model correctly predicts the effect of conditionally repeating Hb measurements on the observed Hb distribution, and study its impact on the number of (non-)deferrals.

Methods

Data

Sanquin, responsible for the collection and supply of blood products in the Netherlands, records all data from donors and donations in the information system e-Progesa. We extract 424,575 Hb measurements from unique first-time donors that made a visit to the blood bank between 2017 and 2024. Additionally, 402,616 whole-blood donations made in 2024 from 221,351 unique donors were used to estimate the current deferral rate.

The HAPPEN dataset includes at least one and up to three Hb measurements acquired with a HemoCue201+ analyzer from capillary fingerprick samples.[7]

Statistical methods

We propose a simple model for the repeated Hb measurements which assumes that the donor population Hb distribution follows a normal distribution and that the measurement procedure introduces a normally distributed error. Given that the measurement is only repeated if the prior measurement is below the threshold, the probability that a measurement is repeated can be computed from a multivariate normal distribution. The resulting probability density function (*PDF*) of the conditionally repeated measurements can thus be calculated. The observed Hb distribution from routinely collected data at Sanquin—where low Hb measurements are repeated and the highest value is recorded—is then a combination of the PDFs of the first, second and third measurement.

We use the data from the HAPPEN study to decompose the Hb distribution into the distribution present in the population and the additional variation that stems from the measurement. In other words, we estimate the standard deviation (SD) of the measurement uncertainty and the mean and SD of the population distribution. The estimation procedure is:

1. Calculate the mean and SD of the first measurement. The mean is the same as the population mean (ignoring any possible bias of the measurement device).
2. Calculate the SD of the difference between the first and second measurement. This determines the uncertainty in the measurement with a correction factor that comes from the PDF of the conditionally repeated measurement.
3. The population SD can now be found by subtracting the measurement SD from the total SD of the first measurements, in quadrature and taking the root.

Because we assume that the population Hb distribution and the measurement error follow a normal distribution, it is straightforward to calculate the expected deferral rate for different scenarios. Please see the Supplemental Material for mathematical details and verification of the procedure with simulated data. Confidence intervals on the results are calculated from 10,000 bootstrapped samples.

Results

HAPPEN data

A total of 2,607 donors had at least one Hb fingerprick measurement of which 219 (8.4%) had a second and 93 (3.6%) a third measurement. Not all repeated entries have a first measurement below the deferral threshold and, vice-versa, not all first (or second) measurements that were below the deferral threshold were repeated. For males and females, respectively, 86% and 91% of 74 and 141 first measurements below the deferral threshold were repeated and 70% and 67% of 30 and 70 second measurements. A Kolmogorov-Smirnov test did not reveal differences between the distributions of measurements that were repeated and those that should have been repeated but were not ($P > 0.33$). We used only those second measurements that had a first measurement below the deferral threshold and only those third measurements of which both previous measurements were below the threshold.

The mean Hb and number of repeat measurements are given in Table 1.The measurement uncertainty of the fingerprick Hb measurement is 0.37 mmol/L (95% CI 0.30, 0.45) for females and 0.38 mmol/L (95% CI 0.32, 0.44) for males. The population standard deviation is 0.52 mmol/L, (95% CI 0.45, 0.58) for females and 0.56 mmol/L (95% CI 0.50, 0.61) for males. The computed Hb level PDFs are shown in Figure 1, where they are compared to the actual observations in the HAPPEN data.

Donation data

Figure *2* shows that the distribution of Hb levels observed in new blood donors is distorted as a result of the conditionally repeated Hb measurements. The shape is well described by the PDF derived from the HAPPEN data. The probability that a measurement was repeated once or twice was multiplied by the percentage of actual repeated measurements from the HAPPEN data as described above (e.g. 86% for the second measurement for females). The parameter of the mean is adjusted by -0.04 for females and +0.04 for males to better fit the data.

Deferral rates

With the population and measurement parameters we can now calculate the deferral rate for different scenarios by using properties of the normal distribution. If the ability to donate would be judged on the first Hb measurement only, we estimate that the on-site deferral rate would increase from the current 4.4% to 13% for females and from 1.6% to 6.3% for males. On basis of the population distribution, however, we would conclude that 8.5% and 3.2% of female and male donors, respectively, would have “true” Hb levels below the threshold. By repeating the low measurements and using the highest value the misclassification of false deferrals is reduced from 7.2% to 0.8% (females) and 4.1% to 0.4% (males), but this comes at the cost of increasing the misclassification of true deferrals from 2.5% to 3.9% (females) and 1% to 1.5% (males).

Discussion

In this study we have shown that we can describe the effect of conditionally repeated fingerprick measurements on the observed Hb distribution. We decomposed the observed distribution into variation present in the population and additional variation from the measurement. We determined the standard deviation of the population (0.54 mmol/L) and the measurement uncertainty of the fingerprick measurement (0.38 mmol/L). Without the repeated fingerprick measurement the deferral rate would be substantially higher than it currently is.

The policy of repeating Hb measurements below the deferral threshold is understandable from the point of view of compensating for the high measurement uncertainty as it would otherwise result in substantially more deferrals. However, only recording the highest value introduces a bias which hinders the accurate interpretation of donor Hb levels, especially for low Hb levels. Moreover, the probability that donors with actual low Hb levels may donate is substantially increased. This situation is not unique to the Netherlands. A study by the BEST collaboration reported that in eight out of 34 countries a similar strategy is applied.[1]

It can be argued that a (first) fingerprick result may be affected by external factors like cold hands or when additional fluid is being purged from the finger which dilutes the blood drop, resulting in a reduced Hb measurement. The fact is, however, that Hb measurements from capillary fingerprick samples are noisy.[8–10] Even if there are external factors that would necessitate repeating a measurement, such actions should be recorded to enable consistent, data-driven decision-making. We show that the distributions of conditionally repeated measurements are predictable by population characteristics and measurement variation alone (Figure 1). The mean Hb levels of the second and third measurement as recorded in the HAPPEN study (Table 1) are fully in line with what is estimated by our model. This supports the hypothesis that Hb measurements just below the cutoff are not unusually low. Therefore, it is unlikely that

external factors influence the fingerprick measurement in a specific direction, which would justify repeating the measurement.

The distribution of measured Hb levels in first-time donors can be largely understood from the conditionally repeated measurements (Figure 2). However, it is also clear that there are additional effects that we cannot account for. The number of times that the exact deferral threshold Hb level has been measured can in no way be explained by the measurement procedure.

As a result of the substantial uncertainty in fingerprick Hb measurements, truly ineligible donors may be accepted to donate and, vice versa, perfectly eligible donors may be deferred. But repeating the measurement and accepting the highest value has a striking effect: a female donor with a true Hb of 7.7 mmol/L has a probability of 39% to be allowed to donate (measured Hb ≥ 7.8) from only the first measurement, but from two measurements the probability to donate increases to 69% and even to 83% after three measurements.

Recording all measurements instead of only the highest measurement results would enable: (1) better assessment of the impact of repeated measurements and potential other external factors that might affect the Hb measurement, (2) better estimates of the population Hb distribution in blood donors, (3) creating unbiased prediction models that can aid in optimizing donation intervals, and (4) designing and evaluating alternative deferral strategies that may increase blood donor availability and better safeguard donor health.

To conclude, we demonstrated that conditionally repeated Hb measurements distort the observed distribution of Hb levels in blood donors and that it is possible to model the effects of this policy. A future-proof blood service should be able to rely on meaningful and interpretable data collected to safeguard donor health and support rational deferral decisions which implies an unbiased recording of all measurements performed.

References

1. Zalpuri, S. *et al*. Variations in hemoglobin measurement and eligibility criteria across blood donation services are associated with differing low-hemoglobin deferral rates: a BEST Collaborative study. *Transfusion* **60**, 544–552 (2020).

2. Chung, R. K., Wood, A. M. & Sweeting, M. J. Biases incurred from nonrandom repeat testing of haemoglobin levels in blood donors: Selective testing and its implications. *Biom. J.* **61**, 454–466 (2019).

3. Prinsze, F. J., De Groot, R., Timmer, T. C., Zalpuri, S. & Van Den Hurk, K. Donation-induced iron depletion is significantly associated with low hemoglobin at subsequent donations. *Transfusion* **61**, 3344–3352 (2021).

4. Janssen, M. P. Why the majority of on-site repeat donor deferrals are completely unwarranted…. *Transfusion* **62**, 2068–2075 (2022).

5. Toivonen, J. *et al*. Prediction and impact of personalized donation intervals. *Vox Sang.* **117**, 504–512 (2022).

6. Vinkenoog, M. *et al*. An international comparison of haemoglobin deferral prediction models for blood banking. *Vox Sang.* **118**, 430–439 (2023).

7. Baart, A. M., de Kort, W. L. A. M., van den Hurk, K. & Pasker-de Jong, P. C. M. Hemoglobin assessment: precision and practicability evaluated in the Netherlands—the HAPPEN study. *Transfusion* **56**, 1984–1993 (2016).

8. Bond, M. M. & Richards-Kortum, R. R. Drop-to-Drop Variation in the Cellular Components of Fingerprick Blood: Implications for Point-of-Care Diagnostic Development. *Am. J. Clin. Pathol.* **144**, 885–894 (2015).

9. Killilea, D. W., Kuypers, F. A., Larkin, S. K. & Schultz, K. Blood draw site and analytic device influence hemoglobin measurements. *PLOS ONE* **17**, e0278350 (2022).

10. Bell, S. *et al.* Comparison of four methods to measure haemoglobin concentrations in whole blood donors (COMPARE): A diagnostic accuracy study. *Transfus. Med.* **31**, 94–103 (2021).

Table 1: Mean Hb levels and number of measurements of the first, second and third repeated measurements from the HAPPEN data. The computed mean for repeated measurements is calculated given the mean of the first measurement and the population and measurement standard deviation given in the text.

	Females			Males		
	Mean [95% CI] (mmol/L)	Computed mean for repeated	Number of measurements	Mean [95% CI] (mmol/L)	Computed mean for repeated	Number of measurements
1st	8.47 [8.43, 8.50]	-	1230	9.39 [9.35, 9.42]	-	1377
2nd	7.7 [7.6, 7.8]	7.8 [7.7, 8.0]	128	8.5 [8.3, 8.6]	8.5 [8.3, 8.6]	64
3rd	7.4 [7.2, 7.5]	7.6 [7.5, 7.7]	47	8.2 [8.0, 8.4]	8.2 [8.1, 8.4]	20

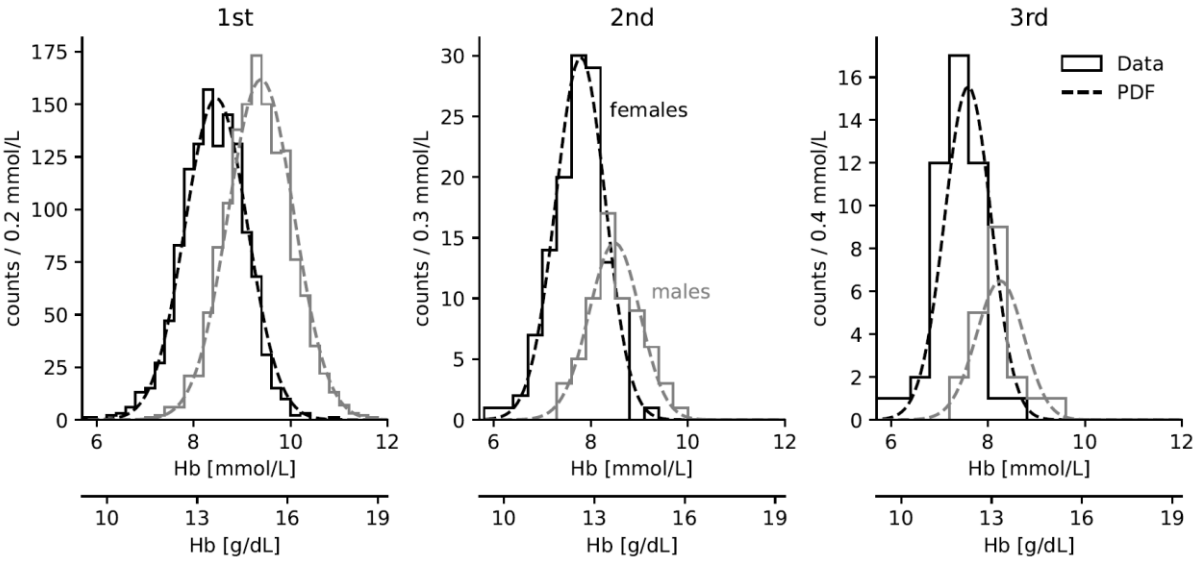

Figure 1: Distributions of the repeated Hb measurements from the HAPPEN data compared to the estimated probability density functions.

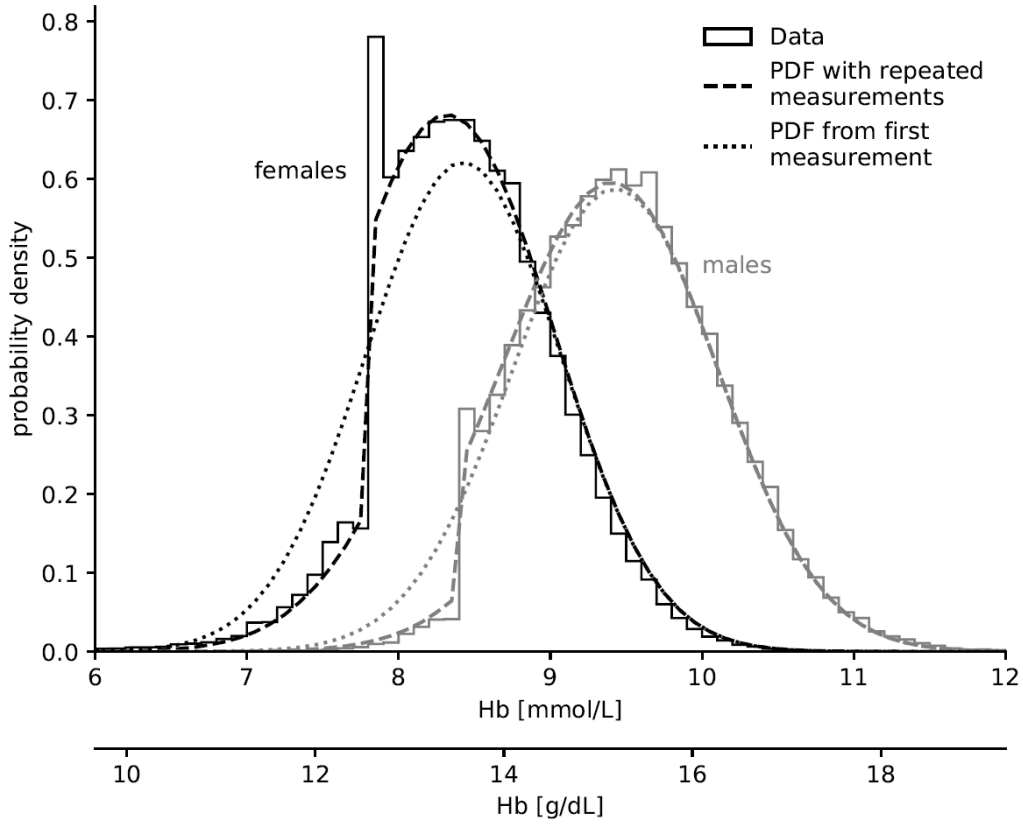

Figure 2: Observed and estimated probability distribution of new blood donor Hb levels.