Holiday Hemoglobin: How the Vacations Affect Blood Donations Across Diverse Urban Sites

Filip Plesinger[1], Stanislava Koskova[2], Eniko Vargova[1], Martina Adamcova[2], Jan Pavlus[1], Gabriela Purdjakova[2], Radovan Smisek[1], Hana Lejdarova[2]

[1]Institute of Scientific Instruments of the CAS, Brno, Czech Republic

[2] Transfusion and Histic Treatment Department, University Hospital Brno, Czechia

Abstract

Sufficient blood supplies are essential for a wide variety of surgeries as well as for the treatment of a large number of diseases. In our scenario, three collaborating transfusion centers in different Czech cities acquire blood from voluntary donors. However, not every invited donor finally comes or can give usable blood due to unexpected events or non-compliance with pre-collection restrictions such as diet or limited sports.

In this study, we aimed to build predictive models to estimate the weekly amount of really acquired Transfusion Units (TU).

We combined evidence of invited donors and weekly acquired TUs in 2021-2024. We extended this dataset with evidence of national, regional, and company holidays. Finally, we trained predictive models using data from 2021, 2022, and the beginning of 2024. These predictive models were evaluated using data from the year 2023.

The overall prediction model for all three centers showed a mean absolute error of 26.9 TU and a Pearson correlation of 0.96 compared with the reality in 2023. Interestingly, city-specific models use different prediction features, pointing to diverse population behavior.

1. Introduction

Sufficient blood storage in a health system is required for planned surgeries (e.g., coronary artery bypass graft and others), trauma surgeries (accidents), or disease management (leukemia, anemia). Blood supplies in the Czech Republic are acquired from voluntary donors in transfusion centers. The amount of blood required in a short-term outlook can be forecasted [1], [2], but another part of the problem is achieving the proper donor count to donate blood. In transfuse centers from our study, donors usually schedule themselves using an online form. However, when the donor comes to donate, unexpected events or behavior in the preceding days may cause non-compliance with pre-collection restrictions, such as infections, allergic reactions, improper food, or just too much sports activity before donation.

Recent studies focused on predicting donors' behavior on a personal level [3]–[6]. However, another way to predict donor behavior could be on a population level, where events such as holidays could affect donor ability and willingness to give blood. This study will explore the effect of holidays and related events on regression models predicting the quantity of acquired transfusion units (TUs) in three specific transfuse centers.

2. Data

We analyzed retrospective data from three collaborating transfuse centers in different and anonymized Czech cities (City A, B, and C). Data were collected in 2021-2023 (whole) and in 2024 (first three weeks only). The data contain weekly-based evidence of invited donors (only City A and B) and acquired TUs.

In addition to supplied data, we extended the dataset to include events likely affecting donors and transfuse station activities. These events are:

- National holidays (NH), which are the same for the whole country. Only crucial jobs are operative.
- Regional & School holidays (RH) also apply for the whole country but partially differ in dates by region. Affects kinder gardens and elementary & high/secondary schools (therefore, affecting parents)
- Company holidays (CH) which are specific for transfuse centers.

For each week and center, we found the number of working days affected by each specific kind of holiday for the estimated, the previous, and the next week. The complete dataset contained information for 158 weeks.

City names were anonymized, but to discuss feature effects, we can state that the city population is > 200k (city A) and < 75k citizens (cities B and C).

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.075

3. Method

We split the data into training and testing sets by year: 2021, 2022, and 2024 formed a training set, while data from 2023 served as the test set. Therefore, the training set contained 106 samples (67%), and the testing set contained 52 (33%).

We iteratively trained linear regression models to find acquired TUs and applied manual backward feature selection. First, we found models for study centers in cities A, B, and C separately. Next, since city A processes raw blood from cities B and C, we created an overall model combining features from all sites.

The performance of all models was measured as the Mean Absolute Error (MAE). For comparison purposes, we normalized MAE to each center's average weekly production. Finally, we computed the Pearson correlation between predicted and real TU values in the test set.

The method was implemented in Python 3.10 [7], supported by Pandas [8] for managing data tables, Scikit-learn [9] for rapid experimenting with models, NumPy for computing correlations [10], and Matplotlib [11] packages to display results. Linear regression models were trained using the Statsmodels package [12].

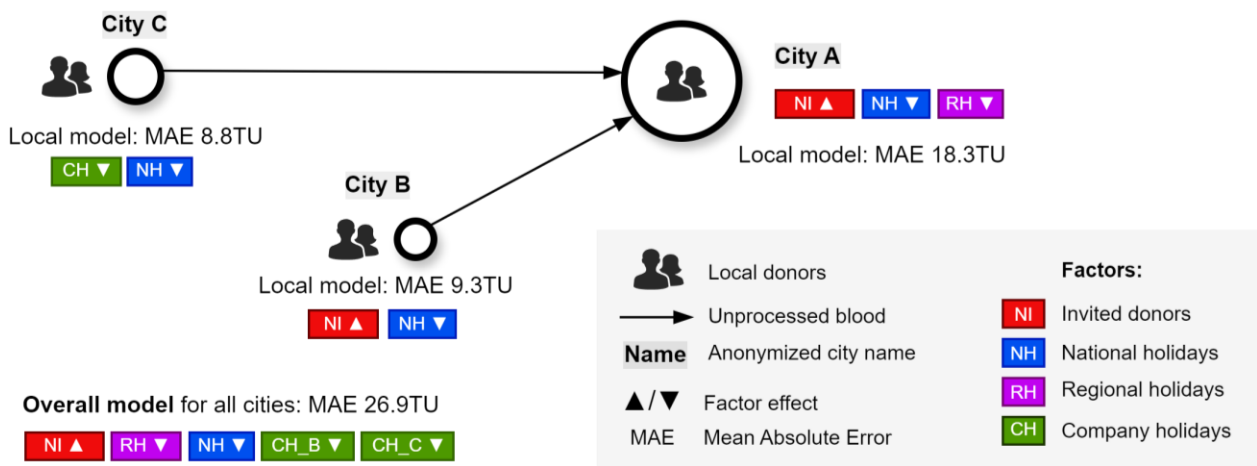

Figure 1. Resultant linear regression models for transfusion centers in three cities and the overall model.

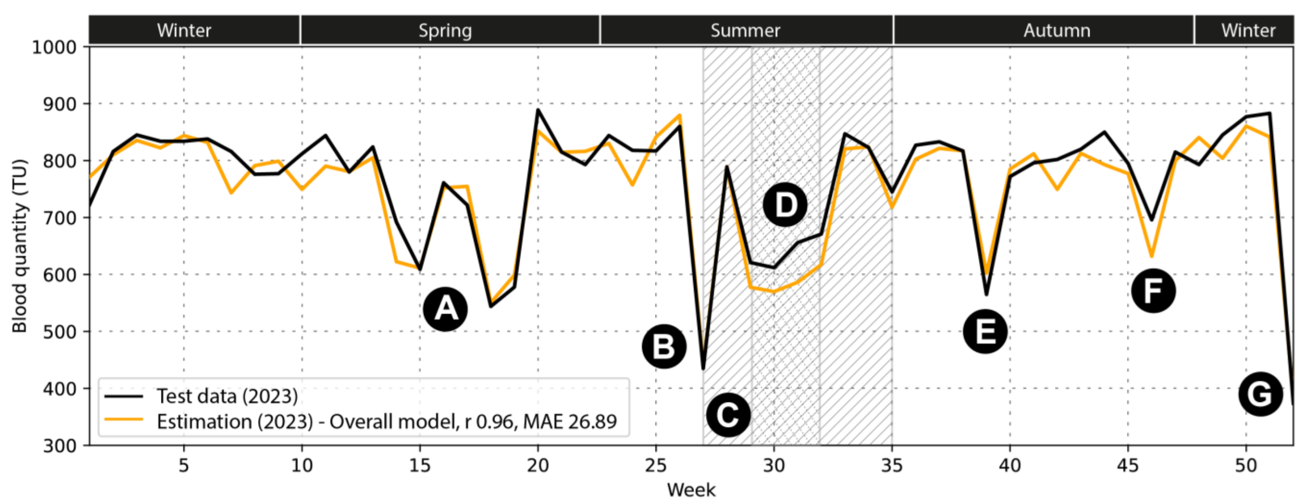

Figure 2. The overall prediction model: comparison of model estimate (orange) and reality (i.e. test set, year 2023, black line) for each week. The most important decreases are caused by Easter (A, left minimum) and two national holidays (A, right minimum, May 1st and 8th), a combination of two-day national holidays (B, July 5th and 6th) with summer holidays (C, diagonally hatched), company holidays (D, cross-hatched), Czech Statehood national holiday (28th September, E), Struggle for Freedom and Democracy Day (17th November, F), and Christmas (G).

Table 1. Linear regression models to estimate really acquired transfuse units using specific features (holidays and invited donors count). Models' performance is shown as Mean Absolute Error (MAE) in raw and normalized form. CH: company holidays; RH: regional holidays; Invited: invited donors; Norm MAE is computed using average week quantity of received TU. Significance: ns (p-value > 0.05), * (0.01, 0.05 >, ** (0.001, 0.01>, *** <0.001

Model	Features									Results		
		National holidays		City A		City B		City C	Cities A+B	Test set performance		
	Intercept (-)	Current week	Next week	RH	Invited	CH	Invited	CH	Inivited	Peas. Corr. (r)	MAE (TU)	Norm MAE (%)
City A	98.8 ***	-18.5 *		-3.7 **	0.7 ***					0.93	**18.3**	**4.6**
City B	5.4 ns		-6.0 *				0.9 ***			0.96	**9.3**	**4.9**
City C	185.9 ***	-42.5 ***						-37.2 ***		0.94	**8.8**	**5.1**
Overall	236.3 ***	-51.4 ***		-4.5 *		-5.2 ns			0.9 ***	0.96	**26.9**	**3.5**

4. Results and Discussion

Using training data, we built four linear regression models. Fig. 1 shows the test results of models specific to cities A, B, and C and the overall model performance. Developed models achieve mean average errors of 18.3, 9.3, 8.8, and 26.9 TUs for models related to city A, city B, city C, and the overall prediction model, respectively. Furthermore, Fig. 1 depicts features (as rectangles) and their trends. The overall model's performance is shown in greater detail in Fig. 2 and Fig. 3, showing model estimates for each week in the testing set and specific events during a year (Fig. 2). Table 1 provides a detailed description of all models and used factors with their statistical significance. The Mean Absolute Error (MAE) in raw and normalized form demonstrates the usability of models. The normalization was computed using the average amount of weekly acquired TUs in 2023: 398.52 (A), 189.23 (B), 174.17 (C), and 761.92 (overall).

4.1. The Effect of National Holidays

National holidays in a week when a blood collection (BC) is planned for have a downward effect in all models, but it is the smallest in the biggest city (A). Interestingly, City B donations are affected by holidays in the week following the BC. National holidays are usually single days, but with two exceptions after Christmas (25th and 26th December) and on the 5th and 6th of July. Tab. 1, column National Holidays, shows a decrease of TUs per each day of NH in the respective week; Fig. 2B shows a strong effect of these days when combined with school holidays.

4.2. Regional and School Holidays

These holidays mostly affect students, their parents, and teachers (common donors). School holidays are during July and August (Fig.2C), sometimes before Christmas, and between the 27th of December and the 1st of January (Fig.2E). Specific regions in Czechia have specific dates for spring (Fig.2A) holidays. Our results show that the effect of these holidays is significant in city A only (and in the overall model) but not in cities B and C. However, the effect is much smaller than the effect of national holidays.

4.4. Effect of Company Holidays

Company holidays (Fig. 2D) are specific only for transfusion centers in cities B and C, meaning that donors cannot donate blood during these days. In both cases, there is a decreasing trend in TU estimation, although, in city B, it is non-significant (Tab. 1), probably due to its redundancy with the number of invited donors.

4.5. Effect of Invited Donors

Data related to the count of invited donors were only available in centers in cities A and B. Model coefficients in Tab. 1 suggest that donors in a smaller city B follow these direct calls more efficiently or that donors in the bigger city A come unexpectedly more often.

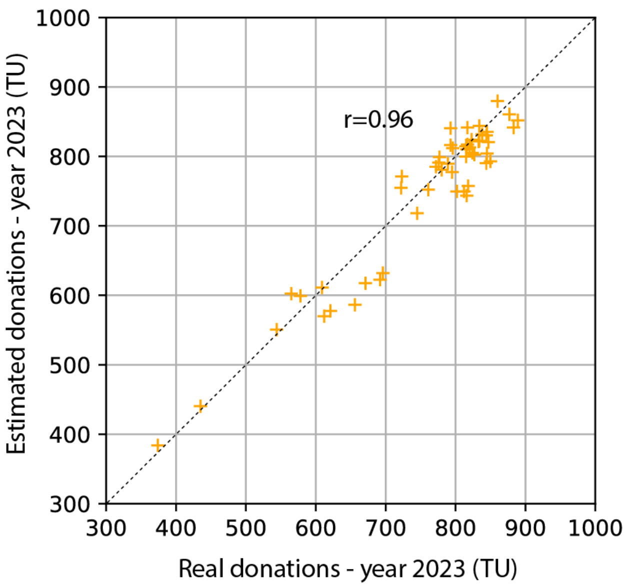

Figure 3. The overall prediction model: comparison of model estimates and reality for the test set (year 2023) with a Pearson correlation coefficient of 0.96. The ideal model is shown as the dotted black diagonal.

4.6. Effects not covered by this study

Although the performance of prediction models showed a very small error at approximately 5% of weekly blood production, we did not cover further events probably affecting donor behavior on the population level: popularization events in schools or social networks [13], the effect of pandemics (only the beginning of training data interacted with COVID-19), or the activity of other transfuse centers in the same cities.

5. Conclusion

We designed models to estimate the total number of donated transfusion units per week. The presented models work at a population level and do not use donor-specific information. They are based on linear regression and allow easy implementation in common planning tools, such as table processors. However, our results show strong but different effects of features in all analyzed centers, suggesting that models are specific for a given locality and should be retrained for any new center.

Acknowledgments

This research was supported by the project RVO:68081731 by the Czech Academy of Sciences.

References

[1] Y. Wei, T. Xia, J. Ma, X. Gao, and Z. Ge, "Machine Learning Applications for Prediction of Blood Transfusion and Survival in Acute Myeloid Leukemia," *Blood*, vol. 140, no. Supplement 1, pp. 2831–2832, Nov. 2022.

[2] C. Twumasi and J. Twumasi, "Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of Tema General Hospital of Ghana," *Int. J. Forecast.*, vol. 38, no. 3, pp. 1258–1277, Jul. 2022.

[3] I. Epifani, E. Lanzarone, and A. Guglielmi, "Predicting donations and profiling donors in a blood collection center: a Bayesian approach," *Flex. Serv. Manuf. J.*, pp. 1–24, Nov. 2023.

[4] C. Kauten, A. Gupta, X. Qin, and G. Richey, "Predicting Blood Donors Using Machine Learning Techniques," *Inf. Syst. Front.*, vol. 24, no. 5, pp. 1547–1562, Oct. 2022.

[5] N. El-rashidy, A. El-Ghamry, and N. E. ElSayed, "Machine Learning for Blood Donors Classification Model Using Ensemble Learning," *Lect. Notes Networks Syst.*, vol. 753 LNNS, pp. 173–181, 2023.

[6] C. Salazar-Concha *et al.*, "Predicting the Intention to Donate Blood among Blood Donors Using a Decision Tree Algorithm," *Symmetry 2021, Vol. 13, Page 1460*, vol. 13, no. 8, p. 1460, Aug. 2021.

[7] G. Van Rossum *et al.*, "Python 3 Reference Manual," *Nature*, vol. 585, no. 7825, pp. 357–362, 2009.

[8] W. McKinney, "Data Structures for Statistical Computing in Python," *Proc. 9th Python Sci. Conf.*, pp. 56–61, 2010.

[9] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, 2011.

[10] C. R. Harris *et al.*, "Array programming with NumPy," *Nat. 2020 5857825*, vol. 585, no. 7825, pp. 357–362, Sep. 2020.

[11] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.

[12] S. Seabold and J. Perktold, "Statsmodels: Econometric and Statistical Modeling with Python," in *Proceedings of the 9th Python in Science Conference*, 2010.

[13] S. Harrell, A. M. Simons, and P. Clasen, "Promoting blood donation through social media: Evidence from Brazil, India and the USA," *Soc. Sci. Med.*, vol. 315, p. 115485, Dec. 2022.

Address for correspondence:
Filip Plesinger
Kralovopolska 147, Brno, 612 00, Czech Republic
fplesinger@isibrno.cz