Predicting blood donor arrival

Vidar Bosnes, Magne Aldrin, and Hans Erik Heier

BACKGROUND: Keeping waiting time at blood donation short is important for making donation a good experience for the donors and hence to motivate for repeat donations. At the Blood Bank of Oslo, fixed appointments are used, and few donors arrive without appointments. On average, 59 percent of scheduled donors arrive, but day-to-day variations are large. Methods for predicting the number of donors that will arrive on a given day would be valuable in reducing waiting times.

STUDY DESIGN AND METHODS: Information about candidate explanatory variables was collected for all appointments made in a 971-day period (179,121 appointments). A logistic regression model for the prediction of blood donor arrival was fitted.

RESULTS: Among 18 explanatory variables, the most important were the time from appointment making to appointment date; the contact medium used; the donor age and total number of donations; and the number of no-shows, arrivals, and deferrals during the preceding 2 years. Compared to taking only the average arrival rate into account, prediction intervals were reduced by 43 percent.

CONCLUSION: Statistical modeling can provide useful estimates of blood donor arrival, allowing for better planning of donation sessions.

Scarcity of blood donors is a problem worldwide. Research in our department has shown that blood donors are themselves the most effective recruiters of new donors.[1] Making each donor's donation experience as enjoyable as possible is therefore probably a key factor for the recruitment and retention of donors.[2] Suboptimal scheduling inevitably leads to prolonged waiting time for donors and a stressful working situation for the blood bank staff. Accordingly, there is good reason to believe that suboptimal scheduling contributes significantly to the current problems of blood supply.

The scheduling strategies used at blood collection centers worldwide vary considerably. Many centers have open blood donation sessions and notify their donors only about opening hours. At the Blood Bank of Oslo, for many years, fixed appointments have been used. A major advantage of fixed appointments is that the blood bank can collect blood throughout the day, reducing the problem of hectic morning and afternoon hours, with low activity in between. The use of fixed appointments may also provide a more predictable blood supply ahead of public holidays. Simulation experiments have shown that scheduling systems with fixed appointments result in shorter waiting times before donation.[3] A donor with an appointment, however, will be less willing to accept long waiting times than a walk-in donor. Therefore, when using fixed

ABBREVIATIONS: ALR = additive logistic regression; OLR = ordinary logistic regression; RMSE(s) = root mean squared error(s).

From the Department of Immunology and Transfusion Medicine, Ullevål University Hospital, Oslo; and The Norwegian Computing Center, Oslo, Norway.

Address reprint requests to: Vidar Bosnes, MD, PhD, The Blood Bank of Oslo, Department of Immunology and Transfusion Medicine, Ullevål University Hospital, N-0407 Oslo, Norway; e-mail: vidar.bosnes@uus.no.

The work of M.A. was partially sponsored by the Norwegian Research Council, Project 154079/420.

Received for publication May 6, 2004; revision received June 28, 2004, and accepted July 15, 2004.

TRANSFUSION 2005;45:162-170.

appointments, it is essential that waiting times be kept to a minimum. At our center, an average of 59 percent of scheduled donors arrive. Overbooking is therefore necessary, and because day-to-day variations in donor arrival are large, we have not been able to eliminate the problem of long waiting times before donation. To optimize donor scheduling, methods for predicting the number of donors who will arrive on a given day are needed.

The aim of this study was to investigate if arrival of scheduled blood donors can be predicted with statistical models, and if so, which variables that are important for prediction. Previous studies on donor return behavior have shown that factors such as donor age,[4] previous donation pattern,[4-6] and previous short-time temporary deferrals[7,8] are useful in explaining the variation in donor return. These and other explanatory variables have been investigated and a logistic regression model for predicting the response to scheduled donation appointments is presented.

MATERIALS AND METHODS

Appointment data

We have collected information about appointments made during a 971-day period between April 2001 and November 2003. Only appointments on ordinary working days (Monday-Friday) that had not been canceled by the end of the previous working day were included. A total of 179,121 individual appointments were recorded. This period had 668 working days. The number of appointments per day ranged from 30 to 424, with a mean of 268. Appointments with repeat donors were usually arranged by sending an invitation by ordinary mail to donate at a given time, approximately 10 days in advance. These donors were requested to contact the blood bank for rescheduling or cancellation if unable to donate at the suggested time. A smaller number of donors had made their next donation appointment at their previous visit to the blood bank or by telephone. Appointments that had been made by telephone or when the donor visited the blood bank were labeled as having been made through personal contact. This category included appointments that had originated as written invitations and that subsequently had been rescheduled. All appointments with first-time donors were made through personal contact.

Files containing appointment data were assembled from the database of the computer system (ProSang, Databyrån AB, Stockholm, Sweden) with SQL queries. The primary variable reports whether or not a donor arrived on the day of his or her appointment. Eighteen explanatory variables were used. These are listed in Table 1.

Statistical modeling

The probability that a given donor arrives to his or her appointment was modeled by logistic regression. The total

TABLE 1. Explanatory variables

Variable	Definition	Minimum	Maximum	Mean
sex	1 if the donor was male; 0 if the donor was female.	0	1	0.53
personal.contact	1 if the appointment was made by personal contact, either at the time of the donor's previous visit to the blood bank, or by telephone; 0 if the donor received a letter with an invitation to donate at the specified date.	0	1	0.60
platelet.donation	1 if the donor was scheduled for a platelet donation; 0 otherwise.	0	1	0.009
first-time.donor	1 if the donor was a first-time donor; 0 otherwise.	0	1	0.05
donation.site	0 and 1 represent two different donation sites in Oslo.	0	1	0.28
donor.age	The donor's age, in years.	17	71	40
n.donations	The donor's total number of previous donations.	0	291	26
n.arrivals.2y	The number of times during the preceding 2 years that the donor had visited the blood bank.	0	30	4.5
n.deferrals.2y	The number of times during the preceding 2 years that the donor had arrived to donate, but had been deferred.	0	7	0.29
n.cancellations.2y	The number of times during the preceding 2 years that the donor had canceled an appointment to donate.	0	38	3.9
n.noshows.2y	The number of times during the preceding 2 years that the donor had failed to arrive to an appointment without canceling.	0	30	2.8
arrival.ratio.2y	Calculated as *n.arrivals.2y*/(*n.arrivals.2y* + *n.noshows.2y*). If the denominator was 0, the mean value for the data set was used.	0	1	0.66
time.since.prev.visit	The number of days between the donor's previous visit to the blood bank and the current appointment. Set to 832 if no previous visits had been recorded.	1	832	153
time.since.appt.made	The number of days between the making of an appointment and the actual appointment date.	1	218	15
day.of.year	The day number within the year.	2	365	189
day.of.week	1-5 represents Monday to Friday.	1	5	2.8
time.of.day	The time of the appointment, in hours	7.5	18.0	12.1
day.number	The day number, counting from April 1. 2001	1	971	491

number expected to arrive on a given day was then calculated by summing the individual arrival probabilities for all appointments that day.

In ordinary logistic regression (OLR),[9] the explanatory variables enter linearly into the model. The large amount of data, however, allowed us to be more ambitious, because even small departures from linearity may be detected and estimated with satisfactory precision when many observations are available. Therefore, an additive logistic regression (ALR)[10] model was fitted first. The ability to detect both linear and nonlinear relationships between explanatory variables and the outcome makes ALR well suited for an exploratory analysis. Prediction with ALR, however, is mathematically complex and requires specialized software both for estimating the model and for performing predictions. To facilitate an implementation of the prediction model within an existing blood bank computer system, the nonlinear functions found by ALR were approximated by transforming the explanatory variables, and an OLR model was formulated based on the transformed variables. The statistical models are presented in Appendix 1. All calculations were performed in computer software (S-plus, Version 6.1.2, Insightful Corp., Seattle, WA).

RESULTS

On average, 59 percent of the donors invited actually arrived to their appointment, but the day-to-day variations in donor arrival were large, ranging from 39 to 90 percent. In addition, an average of 9.45 donors per day arrived without an appointment (minimum, 1; maximum, 24; standard deviation, 4.16). This additional source of variation has not been investigated further in this study. The information available in the computer system about each donor's donation history and response to previous invitations to donate was examined, searching for items that might be of value in predicting a donor's response to an invitation to donate. Eighteen explanatory variables were selected. The choice of variables was based on previous reports on blood donor return behavior,[4-8] as well as on the accessibility of the data and our assumptions about their potential predictive value. The explanatory variables are listed in Table 1, with summary statistics.

An ALR model was estimated based on the full data set. The estimated relationships between each explanatory variable and the probability of arrival are shown as black lines in Fig. 1. Each plot shows the variable's contribution to the log odds of arrival. The curves for binary variables are plotted as straight lines, although they are only defined at the end points. All curves were shifted to a function value of 0 at the left end point, and the intercept was adjusted accordingly. For explanatory variables with negative effects (decreasing curves), function values then ranged from approximately –3 to 0. For explanatory variables with positive effects, function values ranged from 0 to approximately 3. The relative importance of the variables may be compared by inspection of Fig. 1. The probability of arrival was higher for donors whose appointment had been made through personal contact than for donors who had received a written invitation only. The probability of arrival also increased with increasing donor age, with the total number of previous donations, and with the number of arrivals in the preceding 2 years. A strong negative association was found with the number of no-shows during the preceding 2 years, and with increasing time from the making of the appointment to the appointment date. The number of deferrals and cancellations during the preceding 2 years also showed an inverse relationship with the probability of arrival.

One additional aspect must be taken into account when comparing the variables in Fig. 1. On a given day, the portfolio of appointments will be a mix of individual appointments with various characteristics. Therefore, when the interest is in predicting the total number of arrivals on a specific day, the importance of most explanatory variables will be reduced. The variables *day.of.year*, *day.of.week*, and *day.number,* however, will have a similar effect on all donors scheduled for that particular day. Therefore, these variables are more important than the visual impression given by Fig. 1.

A prediction system with ALR cannot easily be implemented within an existing blood bank computer system, because specialized software would be required to perform day-to-day predictions. We therefore formulated an alternative OLR model based on transformed explanatory variables. The choice of transformations was guided by the results of the ALR analysis. Variables that showed an approximately linear relationship with the log odds of arrival were kept unchanged. Other variables, such as *donor.age*, *n.arrivals.2y*, *n.noshows.2y*, *arrival.ratio.2y*, *time.since.prev.visit,* and *time.since.appt.made*, were adequately represented by two linear segments. The variables *day.of.year*, *time.of.day,* and *day.number* were represented by third-order polynomials. The resulting curves are shown as gray lines in Fig. 1. The full table of regression coefficients, standard errors, and p values is available at http://publ.bosnes.net/.

To study prediction performance, the proportion of appointments that resulted in arrival have been examined. For comparison, two simple models were fitted in addition to the ALR and OLR models. The first simple model only had an intercept, thus taking only the average arrival rate into account. The second simple model also took the type of appointment (the variable *personal.contact*) into account. Residuals, average errors, and root mean squared errors (RMSEs) were calculated for all

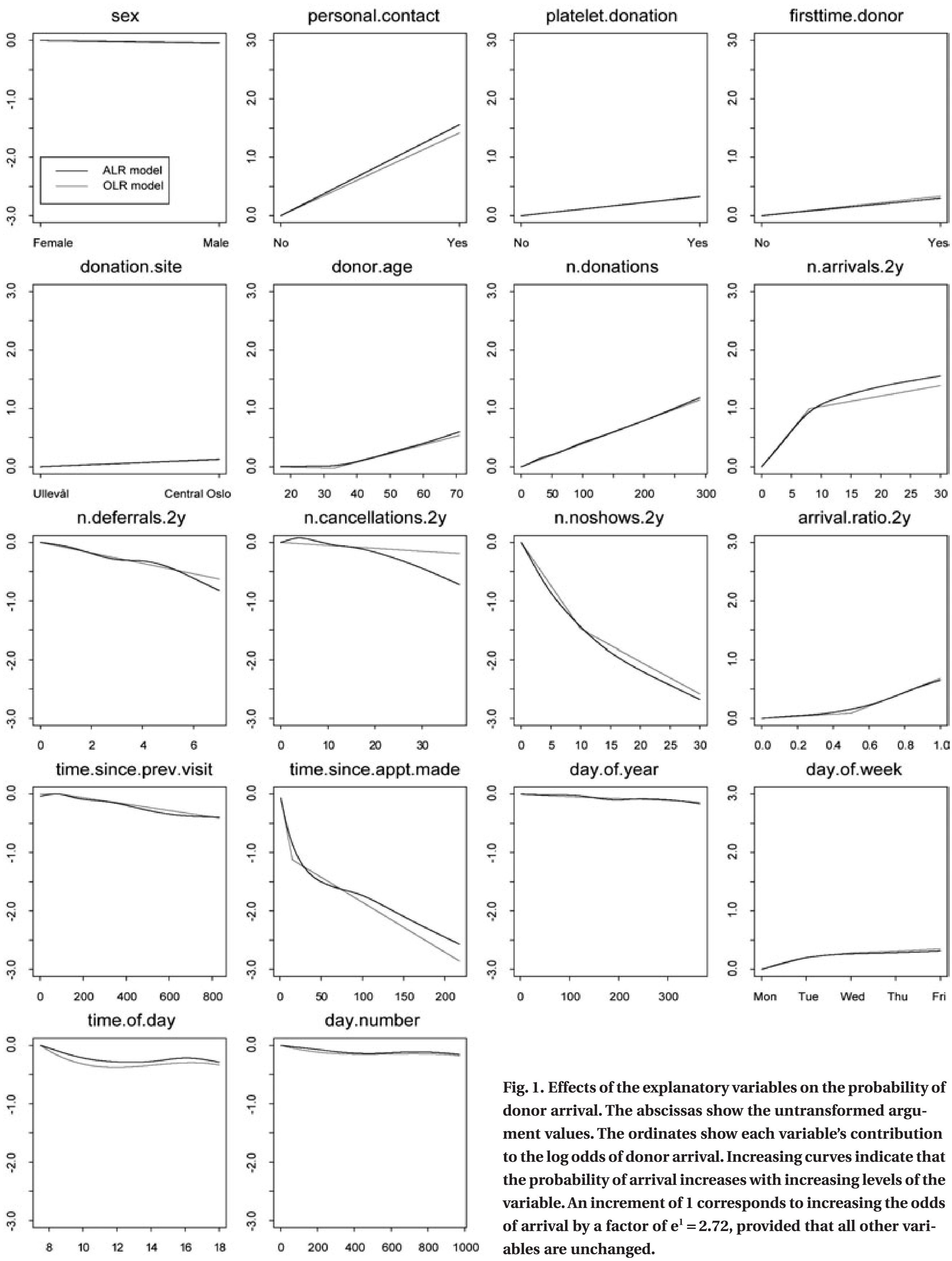

Fig. 1. Effects of the explanatory variables on the probability of donor arrival. The abscissas show the untransformed argument values. The ordinates show each variable's contribution to the log odds of donor arrival. Increasing curves indicate that the probability of arrival increases with increasing levels of the variable. An increment of 1 corresponds to increasing the odds of arrival by a factor of $e^1 = 2.72$, provided that all other variables are unchanged.

TABLE 2. Prediction errors of various models

	Full data set*		Split data set†	
Model	Average error	RMSE	Average error	RMSE
Intercept only	0.002	0.060	0.001	0.058
Intercept + *personal.contact*	0.003	0.050	–0.020	0.052
OLR	0.002	0.034	–0.004	0.033
ALR	0.003	0.034	0.000	0.032

* The full data set (April 1, 2001, to November 30, 2003) was used for both estimation and prediction.

† The data set was split, with different data for estimation (April 1, 2001, to November 30, 2002) and prediction (December 1, 2002, to November 30, 2003).

models. The two leftmost columns of Table 2 show the average error and the RMSE of the various models, when calculated from the full data set.

The average errors were small for all models indicating that none of the models systematically over- or underestimated the arrival probability. The RMSEs of the two simple models were much larger than those of the ALR and OLR models. This implies that the ALR and OLR models have a better prediction precision than the simple models. An RMSE of 0.034 corresponds to an expected difference between the true and predicted arrival proportions of less than $1.96 \times 0.034 = 0.066$ in approximately 95 percent of the days, assuming normality.

The validity of the normal approximation was confirmed by plotting quantiles of standardized residuals against quantiles of a standardized normal distribution and observing that the dots fell roughly on a straight line (data not shown).

In the results presented so far, the same data were used for estimation and for studying the fit of the models. To study the prediction performance in a more realistic setting, the models were reestimated based on data from a 20-month interval from April 1, 2001, to November 30, 2002, and the arrival proportions the following year were predicted.

The variable *day.number* was treated with some caution. There had been a slightly decreasing trend in arrival probability during the estimation period (see *day.number* in Fig. 1). Nevertheless, we believed that it would be dangerous to predict that this trend would continue into the future. Therefore, when predicting, we set the value of *day.number* equal to the last day in the estimation period, reflecting that we had no means of knowing whether the trend would continue or turn. Figure 2 shows the true arrival proportions and the predictions based on the OLR model, with 95 percent prediction intervals.

RMSEs and average errors are shown in the two rightmost columns of Table 2 and do not differ much from those that were obtained when the same data were used for estimation and prediction. Compared to a model taking only the average arrival rate into account, the prediction intervals of the OLR model were reduced by 1 – 0.033/0.058 = 43 percent.

DISCUSSION

Voluntary, nonremunerated donation of blood is an act of altruism and solidarity. Upon that act rests the health of fellow human beings and the function of major parts of somatic health care. It seems self-evident therefore that facilities for blood donation should be managed so as to make donors perceive donation as a good experience. Measures should be sought to adapt the service to the needs of the donors and to make the transfusion service appear friendly and efficient and to minimize donors' loss of time. Prediction of donor arrival may contribute to efficient management of resources and staff at the blood center and hence to minimize waiting before donation.

We have shown that the arrival of blood donors to scheduled appointments could be predicted with logistic regression, based on data that were available from the blood bank computer system, such as information about each donor's donation history. The computer systems used at some blood collection centers do to some extent utilize information about each donor's donation history when scheduling appointments. We are not aware of any previous studies that have focused on the prediction of donor arrival to scheduled appointments, however. Flegel and coworkers[6] have developed a logistic regression model for predicting a donor's likelihood of returning for a second donation within a given time interval, to decide whether the donor's plasma should be stored for quarantine purposes.

An exploratory analysis was performed with ALR. This technique was chosen because it is well suited for detecting both linear and nonlinear relationships between the explanatory variables and the outcome. A prediction system with ALR is mathematically complex, however, and would be difficult to integrate in a blood bank computer system. Therefore, an OLR model was also formulated, based on the insights learned from the ALR model. The two models gave similar predictions.

The techniques used allowed us to calculate not only a point estimate of the proportion of donors expected to arrive, but also the variance. This enabled us to calculate confidence limits for the predictions. Knowledge of the confidence limits could be helpful in balancing the allocation of blood bank staff to the number of donors expected to arrive. Moreover, if the predictions indicated that too few donors would arrive, the blood bank could determine how many more donors would need to be contacted to be 95 percent sure that a given minimum of donors would arrive.

The OLR prediction model reduced the prediction intervals for the proportion of donors arriving on a given day by 43 percent, compared to a model that took only the

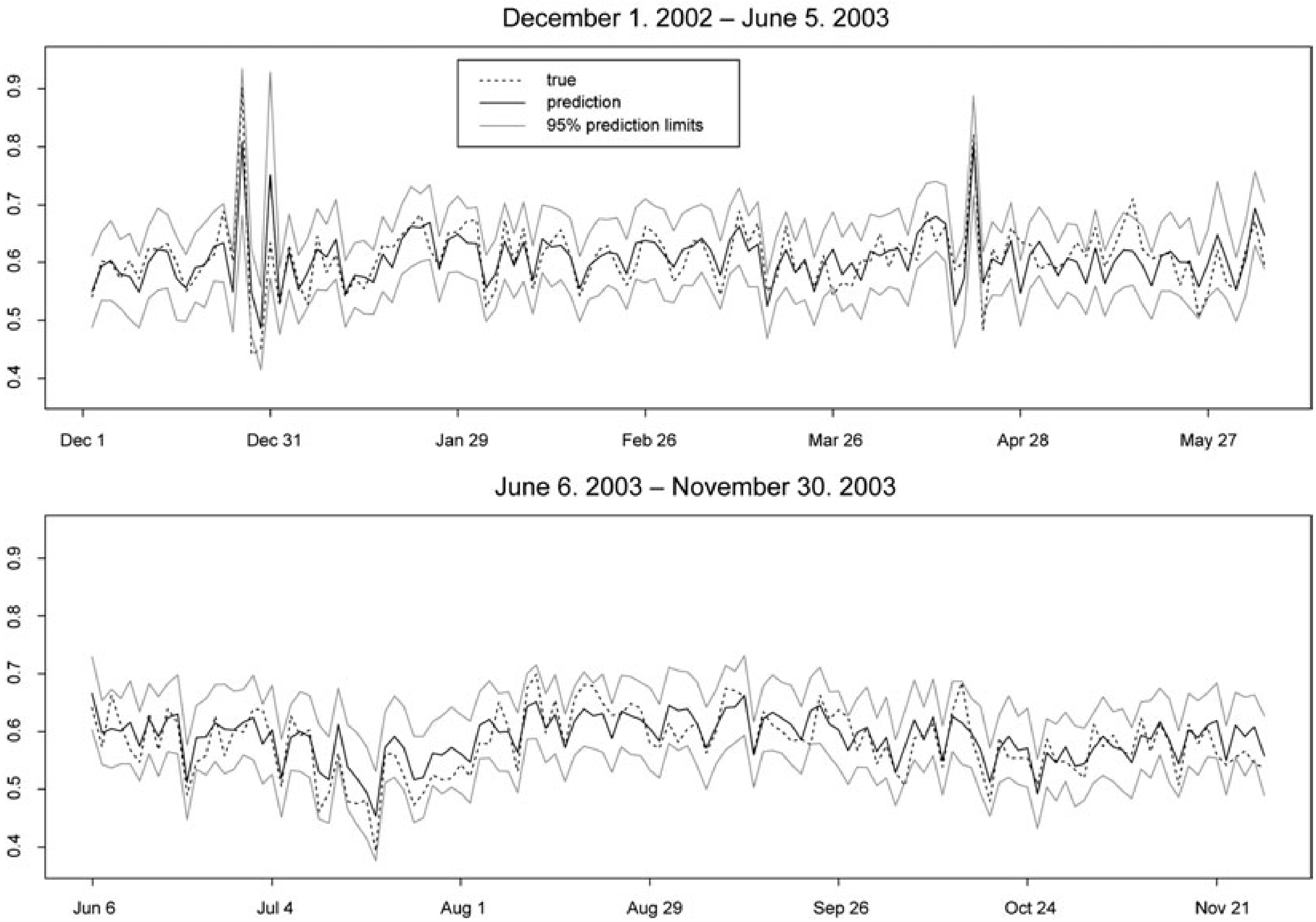

Fig. 2. True and predicted arrival proportions of the OLR model. Prediction intervals have been adjusted for overdispersion.

average arrival rate into account. Regardless of the prediction model being used, prediction intervals will depend strongly on the average arrival rate, with high arrival rates giving small prediction intervals. Therefore, an alternative approach to reducing the uncertainty in donor arrival might be attempting to increase the overall arrival rate. When using a model that takes only the average arrival rate into account, however, a 43 percent reduction of the prediction intervals would require a substantial increase in average arrival rate, from the present 59 percent to 87 percent (calculation given in Appendix 2).

Many of the explanatory variables that were used in this study are inherently correlated. This is acceptable in a model designed for prediction, as long as such intercorrelations are present to a similar extent in the data set used for estimation and the data set used for prediction. The fact that the results when splitting the data set were similar to those obtained when using the entire data set both for estimation and prediction indicates that prediction was not negatively affected by correlations between variables.

We should, however, be cautious in the interpretation of the relative importance of variables that are correlated. If one variable were to be omitted from the model, the importance of correlated variables would increase.

Previous reports have shown that various aspects of each donor's donation history contribute to explaining variations in donor return. It is reasonable to expect that factors that influence the probability of a donor returning for subsequent donations would also be of value in predicting whether a specific donor will arrive to a scheduled appointment. We found that the donor's total number of donations and the number of arrivals during the preceding 2 years both corresponded to an increased probability of arrival. This is consistent with the data of James and Matthews[5] who found that the proportion of returning donors increased with the donors' number of previous donation attempts. It is also consistent with the prediction model of Flegel and colleagues,[6] which used a score function based on the donor's response to previous opportunities to donate as the main explanatory variable. A high number of donations during the preceding 2 years would be expected to result in a higher value of this score function. It was also observed that the probability of arrival increased with increasing donor age. This is consistent with the data of Ownby and coworkers[4] who found that older donors were more likely to return for a second donation. A donor's age and total number of donations are clearly correlated, as are the total number of donations and the number of arrivals during the preceding 2 years. Because of these intercorrelations, we cannot determine which of these variables is the most important. Nevertheless, the results show that donors who have established a pattern of repeat donation are more likely to respond to future invitations to donate. This emphasizes the importance of keeping a continuous focus on donor retention.

The number of times during the preceding 2 years that a donor had failed to arrive to an appointment was a strong negative predictor of arrival. Identifying such lapsed donors at an early stage, and establishing a program for motivating them to resume donating, is an important element of donor retention. Recruiting lapsed donors is desirable also because they have lower rates of transfusion-transmissible viral infections than first-time donors.[11]

Ownby and associates[4] found that individuals who became regular donors tended to have a short interval between their first and second donations. Because a large proportion of our donors were active before the present computer system was introduced, this information was not accessible in our database. It was found, however, that a short interval between the previous donation and the current appointment corresponded to a higher probability of arrival.

It has previously been observed that short-time temporary deferral has a negative impact on donor return.[7,8] This observation was confirmed by our data. The negative impact of temporary deferral presents an important challenge to blood center professionals. Piliavin[7] recommended informing new donors about the possibility of future deferrals and spending some time when deferring a donor to deal with the donor's emotional response and arrange for a new donation appointment. New donors should also be informed about common reasons for short-time temporary deferrals, to avoid coming in vain. Training of staff, to ensure that donors are not deferred without good reason, is also important.

Two means of communication with the donors were used—personal contact and written invitations to donate. As would be expected, appointments that had been made through personal contact were more likely to result in arrival than unconfirmed written invitations to donate. The label "personal contact" was assigned to three distinct categories of appointments—those that where made while donating and those that were made by telephone, on the initiative of either the donor or the blood center. It is likely that the arrival rates for these categories are different. If more detailed information about the appointment making had been available, it could have been incorporated in the regression model. The model could also be extended for other modes of contact, such as e-mail or SMS text messaging. An advantage of use of a statistical model of donor arrival is that the effect of introducing new modes of contact could be easily monitored by comparing the corresponding regression coefficients.

A long interval between the making of an appointment and the appointment date resulted in a reduced probability of arrival. These missed appointments were probably forgotten. Signing up donors for their next appointment before they leave the blood bank has been advocated as a means of donor retention.[2] Our data suggest that the use of written or electronic reminders a few days in advance would probably improve donor arrival to such appointments. It should be kept in mind, however, that some donors react negatively to donation reminders.[12] Reminders should therefore only be sent after obtaining the donor's permission.

A limitation of the model presented in this study stems from the fact that the data were collected on the day before the donation day. Therefore, the model can only give accurate predictions within a narrow window of time. Many donors cancel or reschedule their appointments only a few days before the donation day. To increase its usefulness as a planning tool, the prediction system could be extended by adding a model for predicting cancellation. Many of the variables that were used for predicting donor arrival may also be useful for predicting cancellation. In addition, the remaining time until the appointment day would clearly be of great importance and would have to be included as an explanatory variable. A logistic regression model for calculating the cancellation probability could be fitted, and cancellation probabilities at a given point in time could be calculated for each scheduled donor. The adjusted probability of arrival could then be calculated by multiplying the arrival probability as calculated by the original regression model with the probability of noncancellation. Valid arrival probabilities could then be obtained several days in advance.

This study was performed in a blood collection center where fixed appointments had been used for decades. Nevertheless, some donors came without appointments. A practical implementation of the prediction system would therefore need to include a submodel for the walk-in donors. For a small number of walk-in donors, increasing the arrival estimates by the average number of walk-in donors and adjusting the variance estimates would be sufficient.

Although the prediction model that was presented was designed for a scheduling system that used fixed appointments only, it may be possible to use similar techniques for predicting the response to notifications about open donation sessions that target well-defined subsets of the donor pool. Many of the variables that have been used, which measure the response to a donation appointment, could be replaced by analogous variables, which measure the response to a donation opportunity. The prediction model of Flegel and coworkers[6] was designed for walk-in mobile donation sessions, and the score function that was used for predicting the probability of return within a given time interval might also be useful in predicting a donor's probability of arrival at a given donation session.

In conclusion, it has been shown that statistical modeling with logistic regression can provide useful estimates of blood donor arrival. An integration of such prediction techniques into blood bank scheduling systems could allow for better planning of donation sessions.

ACKNOWLEDGMENTS

We thank the developers of ProSang, Databyrån AB, for providing technical information about the database structure, and for their assistance in making relevant data accessible to SQL queries.

REFERENCES

1. Misje A. Altruisme bak hver dråpe? [Altruism in every droplet?] [Internet]. Masters thesis in sociology. Bergen: University of Bergen; 2001. Available from: http://www.ub.uib.no/elpub/2001/h/710002/Hovedoppgave.pdf
2. Simon TL. Where have all the donors gone? A personal reflection on the crisis in America's volunteer blood program. Transfusion 2003;43:273-9.
3. Michaels JD, Brennan JE, Golden BL, et al. A simulation study of donor scheduling systems for the American Red-Cross. Computers Operational Res 1993;20:213.
4. Ownby HE, Kong F, Watanabe K, et al. Analysis of donor return behavior. Transfusion 1999;39:1128-35.
5. James RC, Matthews DE. Analysis of blood donor return behaviour using survival regression methods. Transfusion Med 1996;6:21-30.
6. Flegel WA, Besenfelder W, Wagner FF. Predicting a donor's likelihood of donating within a preselected time interval. Transfus Med 2000;10:181-92.
7. Piliavin JA. Temporary deferral and donor return. Transfusion 1987;27:199-200.
8. Halperin D, Baetens J, Newman B. The effect of short-term, temporary deferral on future blood donation. Transfusion 1998;38:181-3.
9. McCullagh P, Nelder JA. Generalized linear models. London: Chapman & Hall; 1989.
10. Hastie T, Tibshirani R. Generalized additive models. London: Chapman & Hall; 1990.
11. Schreiber GB, Glynn SA, Damesyn MA, et al. Lapsed donors: an untapped resource. Transfusion 2003;43:17-24.
12. Glynn SA, Kleinman SH, Schreiber GB, et al. Motivations to donate blood: demographic comparisons. Transfusion 2002;42:216-25.

APPENDIX 1

Statistical modeling

Let $y_{i,t}$ indicate whether the ith potential donor on Day t actually arrives and $p_{i,t}$ be the corresponding probability of arrival. The number of donors that arrive on Day t will then be given by $Y_t = \sum_{i=1}^{A_t} y_{i,t}$, where A_t is the total number of appointments scheduled for Day t. The expected number of arrivals is then given by

$$E(Y_t) = \mu_t = \sum_{i=1}^{A_t} p_{i,t}.$$

Because $y_{i,t}$ is binary, its variance is $p_{i,t}(1 - p_{i,t})$. If the variables $y_{i,t}$ were independent, the variance of Y_t would be

$$\sigma_{0t}^2 = \sum_{i=1}^{A_t} p_{i,t}(1 - p_{i,t}).$$

There is good reason to believe, however, that donors will not act independently, because factors such as the weather will have a similar influence on all donors scheduled for a particular day. To compensate for this effect, called overdispersion, the variance of Y_t is modeled as

$$\mathrm{Var}(Y_t) = \sigma_t^2 = \Phi\sigma_{0t}^2,$$

where $\Phi > 1$.

When the number of daily appointments A_t is large, Y_t will be approximately normally distributed. A 95 percent prediction interval will then be given by $\mu_t \pm 1.96\sigma_t^2$. In practice, μ_t and σ_t^2 (including Φ) must be estimated from historical data.

The individual arrival probability $p_{i,t}$ was modeled by logistic regression,

$$p_{i,t} = \exp(\eta(\mathbf{x}_{i,t}))/(1 + \exp(\eta(\mathbf{x}_{i,t}))),$$

where $\eta(\mathbf{x}_{i,t})$ is a function of the p-dimensional vector of explanatory variables $\mathbf{x}_{i,t} = (x_{1,i,t}, \ldots x_{p,i,t})$. In the initial analysis, ALR[10] was used, that is,

$$\eta(\mathbf{x}) = \beta_0 + \sum_{j=1}^{p} s_j(x_j),$$

where s_j, $j = 1, \ldots, p$ are unknown, but smooth functions.

An alternative OLR[9] model was also formulated. The p explanatory variables $\mathbf{x} = (x_1, \ldots, x_p)$ were transformed into $p' > p$ new variables $\mathbf{z} = (z_1, \ldots, z_{p'})$, by replacing variables which were found to have a nonlinear relationship with the outcome by either third-order polynomials or two linear segments. A linear predictor $\eta(\mathbf{z})'$ that approximated $\eta(\mathbf{x})$ was then used,

$$\eta(\mathbf{z})' = \beta_0 + \sum_{j=1}^{p'} \beta_j z_j,$$

where the β's are regression coefficients. Maximum likelihood estimates of the regression coefficients were obtained, and corresponding estimates $\hat{\eta}'$, $\hat{p}_{i,t}$, and $\hat{\mu}_t$ were calculated. To calculate $\hat{\sigma}_t^2$, an estimate of the overdispersion factor Φ is needed. Because $\Phi = \mathrm{Var}(Y_t)/\sigma^2_{0t}$, an estimate for Φ is given by

$$\hat{\Phi} = (1/T)\sum_{t=1}^{T} \frac{(Y_t - \hat{\mu}_t)^2}{\hat{\sigma}^2_{0t}},$$

where T is the total number of days in the data set. The resulting estimate of the overdispersion factor was 1.59.

When evaluating the predictions, the proportion of appointments that resulted in arrival $P_t = Y_t/A_t$ and its estimate $\hat{P}_t = \hat{Y}_t/A_t$ were examined. Residuals $e_t = P_t - \hat{P}_t$, average errors $\bar{e} = 1/T\sum e_t$ and $\mathrm{RMSE} = \sqrt{1/T\sum e_t^2}$ were calculated.

APPENDIX 2

The increase in arrival rate that would be required to achieve a 43 percent reduction of prediction intervals can be calculated by assuming that the ratio of the standard deviations of the expected number of donors per day under the improved and present conditions is equal to 1 – 0.43 = 0.57. Let Φ, A, and p represent the overdispersion factor, number of appointments and arrival rate under the improved (subscript 1) and present (subscript 0) conditions. The ratio is then given by

$$\sqrt{\Phi_1 A_1 p_1 (1-p_1)} \div \sqrt{\Phi_0 A_0 p_0 (1-p_0)} = 0.57.$$

If we require an unchanged number of expected arrivals $A_1 p_1 = A_0 p_0$, and further assume that $\Phi_1 = \Phi_0$, then substituting $p_0 = 0.59$ implies $p_1 = 1 - 0.57^2(1 - 0.59) = 0.87$.