Quality Engineering
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597292

Case Study: An Application of Logistic Regression in a Six Sigma Project in Health Care

Frank van der Meulen[a]; Thijs Vermaat[b]; Pieter Willems[c]
[a] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands [b] TNT Post, The Hague, The Netherlands [c] Virga Jesse Hospital, Hasselt, Belgium

Online publication date: 09 March 2011

To cite this Article van der Meulen, Frank , Vermaat, Thijs and Willems, Pieter(2011) 'Case Study: An Application of Logistic Regression in a Six Sigma Project in Health Care', Quality Engineering, 23: 2, 113 — 124
To link to this Article: DOI: 10.1080/08982112.2011.553761
URL: http://dx.doi.org/10.1080/08982112.2011.553761

Quality Engineering, 23:113–124, 2011

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982112.2011.553761

Taylor & Francis
Taylor & Francis Group

Case Study: An Application of Logistic Regression in a Six Sigma Project in Health Care

Frank van der Meulen[1], Thijs Vermaat[2], Pieter Willems[3]

[1]Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands
[2]TNT Post, The Hague, The Netherlands
[3]Virga Jesse Hospital, Hasselt, Belgium

ABSTRACT Health care today is facing serious problems: quality of care does not meet patients' needs and costs are exploding. In the cardiology department of the Virga Jesse Hospital in Belgium, discharged patients are advised to participate in a rehabilitation program. However, many of the discharged patients do not join the program, and others quit before being declared cured (a so-called dropout). An improvement project was started that aims to increase revenues by either attracting more patients to the rehabilitation program or reducing the fraction of dropouts.

A large data set with 516 treated patients was available. We model the probability that a patient joins the program as a function of various numerical and categorical influence factors. First an exploratory data analysis is performed, using bar charts and box plots. This is followed by a more formal statistical analysis using logistic regression.

The logistic regression model reveals the important influence factors. The probability of joining the program depends on whether a patient has a car at his or her disposal and the distance from a patient's home to the hospital. As a solution, various measures to stimulate carpooling were implemented. Prior to the implementation, a cost–benefit analysis was conducted using the fitted regression model.

KEYWORDS DMAIC, generalized additive model, logistic regression

PROCESS DESCRIPTION

All over the world health care is facing serious issues. Costs are increasing and the quality of care consistently fails to meet expectations (cf. institute of Medicine 2001). Quality improvement is therefore a major strategic issue in health care organizations and improvements have to be implemented to reduce costs and increase quality. The Six Sigma program is an effective management methodology, developed in industry and also adopted in health care; see Barry (2002) and Bisgaard (2009). Six Sigma improvement projects are executed by a fixed step-by-step approach, the DMAIC

Address correspondence to Dr. Frank van der Meulen, Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands. E-mail: f.h.vandermeulen@tudelft.nl

roadmap. It encompasses five phases: the define, measure, analyze, improve, and control phases. Projects are executed by project leaders. The DMAIC roadmap assists them in organizing their findings in a structured manner. For a description of these phases, see De Mast et al. (2006) and Breyfogle (2003).

In 2005 the Virga Jesse Hospital in Hasselt, Belgium, decided to use the Six Sigma method to improve their processes. In this article we will explore a project on the retention of heart rehabilitation patients. Its aim was to attract more patients in the rehabilitation program or reduce the fraction of patients who drop out during the program. A succesfull project will increase the hopital's revenues and will be beneficial to patients' health as well. We will discuss the project, focusing on the analyze and improve phases.

After cardiac surgery, patients with heart disease are treated in the cardiology nursing department. When a patient's condition is stable, he or she is discharged from the department and goes home. For safety reasons, patients are advised to join the hospital's cardiac rehabilitation program. In addition to psychological support and advice on a healthier diet and a less stressful lifestyle, patients in this program participate in physical training under full supervision of a physical therapist. Patients visit the rehabilitation center two or three times a week for a 2-h session, with a maximum of 45 sessions.

Many patients treated at the nursing department do not enroll in the rehabilitation program after discharge. Moreover, many patients who do start the program leave halfway through, before being physically fit. The latter is called a *dropout patient*. In both cases the hospital loses revenues; every visit is charged individually.

DATA COLLECTION

The measure phase starts with the definition of the internal critical to quality characteristics (CTQs). In this project the strategic focal point is the increase of revenue, which links directly to the following CTQs:

- *CTQ1*: the number of patients who participate in the rehabilitation program every month
- *CTQ2*: the number of sessions per participant

To measure the number of participants and sessions each month, one simply looks at the number of invoices. To assess whether this measurement procedure is valid, a comparison between a sample of invoices and the corresponding list of participating patients from the department was made. These matched perfectly, validating the chosen measurement procedure.

A large data set with 516 treated patients was available. Of these patients, 49% participated in the rehabilitation program. For each patient we have the following data available:

- *distance* between the patient's home to the hospital in kilometers (x_1, numerical)
- *age* (x_2, numerical)
- *mobility*; whether or not the patient has a car (x_3, categorical)
- *gender* (x_4, categorical)
- place of residence
- *participation*; whether or not the patient participates in the rehabilitation program (Y, binary; $Y = yes$ if the patient shows up at least once, else $Y = \text{no}$).

*CTQ*1 is directly related to participation. In fact, the value of *CTQ*1 in a month is the sum of all patients i with $Y = yes$ in that particular month. In this sense, Y is a more informative measurement than *CTQ*1, because we can relate Y to patient characteristics. The influence of the place of residence is captured by variable x_1.

ANALYSIS AND INTERPRETATION

Over the year 2005 the first CTQ (the number of participating patients) was on average 33 patients each month, with a standard deviation of 4.9 patients each month. Based on the process capability and process knowledge, the objective of the project was to increase the average number of participants to 36. This number had been attained a number of times in the past and both cardiologists and physicians claimed that such an increase was feasible.

The second CTQ (the number of sessions) was on average 29 sessions for the patients participating in 2005. Note that the maximum number of sessions per patient in a program was 45. The objective for the second CTQ was to increase the average number of sessions to 32 for each patient. The average session's revenue was 22.82 euros. Hence, increasing both the number of participants and the number of sessions per participant will increase the total revenue by 53.000 euros ($(36 * 32 - 33 * 29)$ extra sessions per month $\times$ 12 months $\times$ 22.82 euros per session makes an extra yearly revenue of 53.422 euros). There are minimal additional costs for handling these extra sessions, because the rehabilitation area with accompanying resources has overcapacity.

The second CTQ, the number of sessions of each patient, was studied by a root cause analysis: 156 patients were asked why they left the program early. Summarizing:

- 26% of the patients were readmitted for a hospital stay,
- 16% of the patients started working again and could not combine this with the rehabilitation activity (even though the center was open late),
- 16% of the patients could not join the program due to other obligations (vacations, social obligations),
- 12% of the patients dropped out for a medical reason provided by the doctor,
- 8% of the patients had their own rehabilitation facilities.

These factors were the cause of 78% of the dropout. However, none of these causes can be influenced easily. Therefore, focus shifted to *CTQ*1.

Based on brainstorm sessions with cardiologists, physical therapists, patients, and other interested parties, the following influence factors for *CTQ*1 were raised:

- Patients should be informed of the rehabilitation program at a much earlier stage.
- information on the rehabilitation program should be much more precise and attractive.
- Cardiologists should stimulate patients to participate in and finish the rehabilitation program.
- Patients should train with a heart rate monitor (polar watch) to improve their feelings of safety.
- Patients desire a smaller exercise room and are more comfortable when not with other patients.
- Patients are not likely to show up during summer holiday.

Factors that seemed to be most important can be summarized as *patient attention factors*. These factors were very important in increasing the number of participants. As a consequence, the following improvement actions were proposed:

- Writing a better brochure on the rehabilitation program.
- Writing a letter to the cardiologists to improve their attitude toward patients: to be more polite and to communicate the possibilities of the rehabilitation program at an earlier stage.

This is a typical example of jumping to conclusions, which is often experienced in practice. Below, we will explain how statistical techniques, in particular logistic regression, show a different view toward designing improvement actions. This is a good illustration of the strength of the improve phase in Six Sigma and the usefulness of logistic regression. We now give a detailed analysis of the statistics used in the improve phase. The project supervisor convinced the project leader to complete the improve phase before proceeding with the above-mentioned actions.

Analyzing Each Factor Separately

Our first step consists of studying the relation between *Y participation* and each influence factor (denoted by x_i) separately. It is useful to screen the data in this way before using more advanced techniques.

1. The first studied factor is distance. Whether the number of kilometers affects whether the patient will join the program is normally analyzed by means of logistic regression. A first simple approach consists of making boxplots for distance vs. participation. Looking at these plots, we immediately noticed two patients with very long distances ($\geq$200 km) to the hospital compared to the other patients. These patients were closely related to one of the physicians and therefore had chosen the hospital considered here. For this reason, these patients were excluded from all further statistical analysis. The left-hand figure of Figure 1 contains boxplots of the data from which these two outliers were removed. This figure suggests that patients with a short distance to the hospital tend to participate more often in the program.

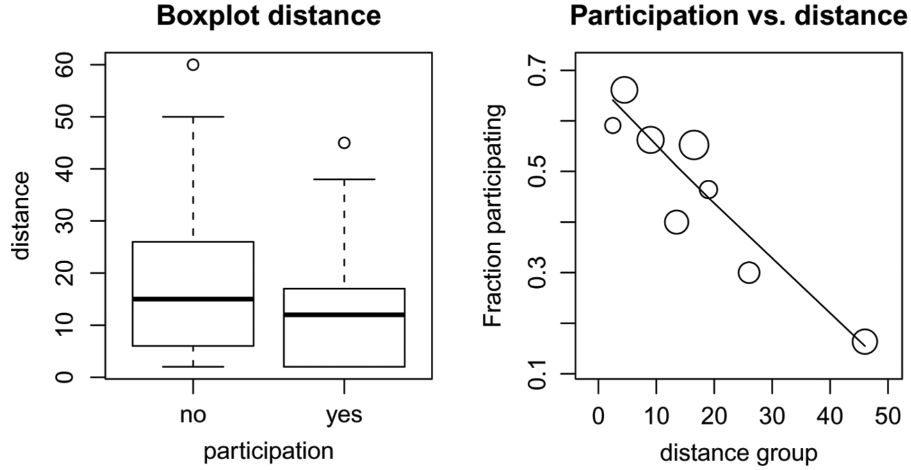

FIGURE 1 Exploratory analysis of the relation between participation and distance.

In the right-hand figure of Figure 1 a more informative plot is made. We divided the range of distance into eight approximately equally sized groups. Within each group we computed the relative frequency of patients participating. Because there are ties in the distance values, not all groups were exactly the same size. The diameter of the circle for a group is proportional to the size of that particular group. To visualize a pattern among the points, we added a smoother through these points. A smoother is a nonparametric regression fit, which can be constructed by many methods. Here, we chose Friedman's "super smoother," which is implemented in the statistical software package R (function "supsmu"). Details about the construction of this smoother are of minor importance at this stage, but the interested reader may consult Friedman (1984). The R code for constructing this figure can be found on Howard Seltman's Website, http://www.stat.cmu.edu/hseltman/files/LREDA.R. From the constructed plot we clearly see that the further a patient lives from the hospital, the lower the probability that a patient will join the rehabilitation program.

2. The factor *age* can be analyzed in a similar way; see Figure 2. This factor suggests that the probability of joining the rehabilitation program decreases with age. Moreover, at approximately age 65 there seems to be a change point in the decrease of the fraction of participating patients.
3. The bar chart for mobility (left-hand picture in Figure 3) clearly indicates that the probability of joining the program is influenced by whether the patient has access to a car. Table 1 summarizes these data. The data suggest that having a car at one's disposal increases the probability

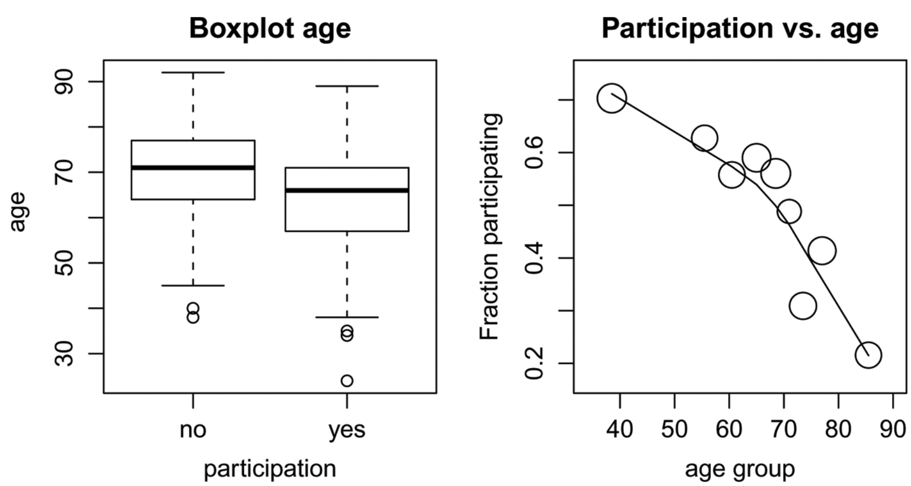

FIGURE 2 Exploratory analysis of the relation between participation and age.

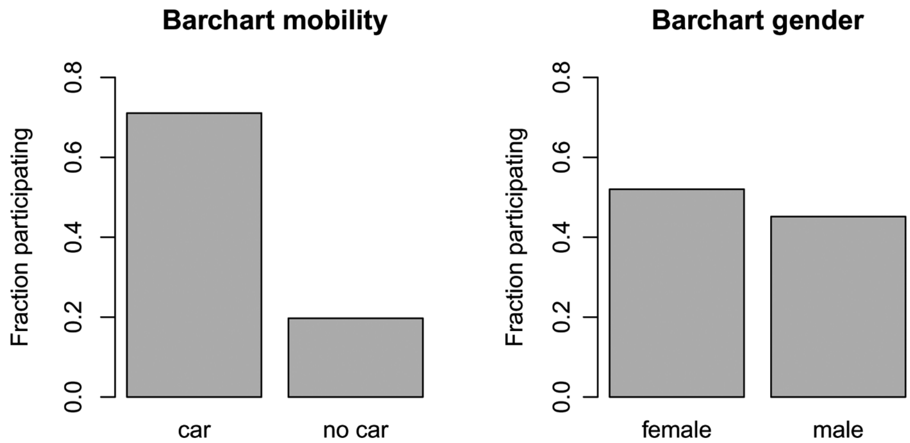

FIGURE 3 Exploratory analysis of the relations between participation and mobility and participation and gender.

for joining the program. There are missing values in the data set: for 71 patients, mobility was not registered.

4. The factor *gender* can be analyzed in a similar way as mobility. There were 13 missing values for gender in the data set. The bar chart (right-hand picture in Figure 3) indicates that this factor has a minor influence on participation. Table 2 summarizes these data.

The analysis suggests that the accessibility of the hospital has to be improved, especially for those people living far away from the hospital. Hiring a taxi service would definitely improve accessibility, though it is obvious that the costs for this service exceed the revenues of one additional session. It is of major interest to find out how much money can be invested to improve accessability of the hospital while still ensuring increased revenues. This maximal amount can be considered a break-even point. To calculate this break-even point, we need a relation between the probability that a patient will join the program and the various influence factors as an ensemble. In the next section we will show how a logistic regression model can be used to accomplish this. An introduction to logistic regression can be found in many textbooks; see, for example, McCullagh and Nelder (1989) and Myers et al. (2002).

TABLE 1 Influence of Mobility on Participation

Mobility	Number of patients	Percentage joining the program
Car	311	71
No car	132	20

TABLE 2 Influence of Gender on Participation

Mobility	Number of patients	Percentage joining the program
Female	377	52
Male	124	45

Logistic Regression Model for Modeling the Probability That a Patient Will Join the Program

In this section, we model the relation between Y and all influence factors simultaneously. In a logistic regression model, we assume that all Y_i (the response for the ith patient) are independent and identically distributed, with

$$P(Y_i = yes) = \frac{1}{1 + e^{-f(\mathbf{x}_i)}} =: p_i, \quad P(Y_i = no) = 1 - p_i,$$

or, equivalently,

$$\text{logit } (p_i) := \log\left(\frac{p_i}{1 - p_i}\right) = f(x_i).$$

Here $x_i = (1, x_{i1}, x_{i2}, x_{i3}, x_{i4})$ is the vector of predictors (including an intercept) for patient i and f is a function that has yet to be estimated. We use dummy variables in that $x_{i3} = 1$ if mobility = "no car" and zero otherwise. Similarly, $x_{i4} = 1$ if gender = "male" and zero otherwise. A generalized linear model without interaction terms assumes that

$$f(x_i) = x_i^T \beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$$

for a certain parameter vector β. It is not directly clear that such a simplified form for f is suitable for modeling the data. However, to estimate a relationship, we need to make some assumptions on the form of f. From Figure 1 it appears that the probability of participating decreases linearly with distance. If we redraw this figure with the values on the vertical axis replaced by their logit, we would again see a linear relationship. (This is a simple consequence of the fact that the mapping $p \mapsto \text{logit}(p) = \log[p/(1-p)]$ is almost linear on (0, 1), especially if p is bounded away from 0 and 1, as is the case in Figure 1.) Because this figure shows the same relationship, we did not include it. For Figure 2 similar comments apply. It appears that the probability of participating decreases in a slightly nonlinear way with age. Modern software for nonparametric estimation allows inclusion of a general smoothing function of age in the model. In this way, we obtain a generalized additive model, from which we can assess the linearity in a more formal way. This additive model takes the form

$$\text{logit}\,(p_i) = \beta_0 + \beta_1 x_{i1} + s(x_{i2}) + \beta_3 x_{i3} + \beta_4 x_{i4}.$$

To allow for an interaction effect between distance and mobility, we add an interaction term for these predictors. As such, we arrive at the following model:

$$\text{logit}\,(p_i) = \beta_0 + \beta_1 x_{i1} + s(x_{i2}) + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_{13} x_{i1} x_{i3}. \quad [1]$$

This model introduces an identifiability problem: β_0 and s are each only estimable to within an additive constant. The `mgcv` package can be used to fit the model. The main engine behind the estimation procedure consists of penalized smoothing splines and generalized cross-validation. For details about this package and the resulting fit we refer to Wood (2006). In this book, a simple convention to deal with the identifiability problem is discussed as well (see, e.g., p. 134).

We fit the model in the preceding display on those patients for whom there are no missing values in the covariates. Part of the output is

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.73540	0.23523	7.377	1.61e-13
distance (x1)	−0.06325	0.01243	-5.087	3.63e-07
mobilityno car (x3)	−1.64831	0.44456	−3.708	0.000209
gendermale (x4)	−0.07842	0.27663	−0.283	0.776814
distance: mobilityno car (x1:x3)	−0.01844	0.02622	−0.703	0.481989

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(age) (x2)	2.733	2.733	26.65	4.88e-06

The set of parametric coefficients can be interpreted just as for (generalized) linear models. Distance and mobility appear to be significant, whereas gender and the interaction between distance and mobility are not significant. The term `s(age)` represents the fitted smooth function of age. For now, it suffices to note that age is significant.

We refit the model without the insignificant terms. The results are as follows:

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.77640	0.21035	8.445	<2e-16
distance (x1)	−0.06735	0.01107	−6.086	1.16e-09
mobilityno car (x3)	−1.90108	0.27453	−6.925	4.37e-12

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(age) (x2)	2.649	2.649	27.29	3.17e-06

The smooth term is most easily examined by a picture. Figure 4 shows the fitted smooth function of age together with pointwise 95% confidence bounds. The value 2.649 from the output refers to the estimated degrees of freedom. This can be interpreted as saying that we need approximately three coefficients to parameterize the curve. Because the curve consists roughly of two linear pieces, this is not surprising.

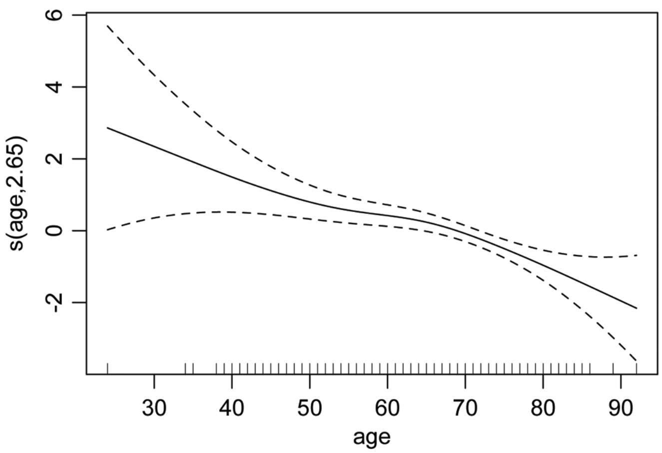

FIGURE 4 **Estimated smooth term for age.**

At this point, there are two ways to proceed. One direction consists of performing diagnostic checks for the current model and assessing its goodness-of-fit. The other direction consists of simplifying the current model by replacing the smooth function of age by a linear function, followed again by diagnostic checks and goodness-of-fit assessment. In fact, we pursued both approaches and the resulting conclusions were very similar. This is not surprising because the curve in Figure 4 is not far from linear. We favored the somewhat simpler model with linear terms and henceforth pursued the second direction in the following.

Fitting the model

$$\text{logit}\,(p_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

with the `glm` function gives

Here $\hat{p}_i$, $distance_i$, and age_i denote the fitted probability, the distance, and age of patient i, respectively.

The interpretation of the coefficient for distance (which equals -0.0675) is as follows. Suppose that at a certain distance d the odds of participating equals 2. Increasing the distance by 1 km causes the odds of participating to be multiplied by $e^{-0.0675} \approx 0.93$. So the odds at distance $d+1$ equal $2 \cdot 0.93 = 1.86$. Similarly, increasing age by 1 year causes the odds of participating to be multiplied by $e^{-0.0599} \approx 0.94$.

To visualize this fit, we fixed age at approximately its first and third quantile (60 and 75, respectively) and plotted the predicted probability to join the program against distance for both levels of mobility (see Figure 5). From this figure it is clear that the probability of joining the program decreases with distance. Moreover, irrespective of age, the probability for joining is higher for people who have a car than for those who do not.

Diagnostics

We now discuss diagnostics for the fitted model. For generalized linear models there are various types of residuals, of which the best known are the Pearson and deviance residuals. For many generalized linear models (of which logistic regression is a special case), the deviance residuals behave similarly to the residuals obtained in ordinary linear regression. Diagnostic plots, where, for instance, the residuals are plotted against covariates, can be obtained by using deviance residuals. For logistic

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	2.5%	97.5%
(Intercept)	5.77654	0.86186	6.702	2.05e-11	4.1438	7.5307
distance (x2)	−0.06752	0.01105	−6.113	9.78e-10	−0.0900	−0.0465
mobility nocar (x3)	−1.93691	0.27202	−7.121	1.07e-12	−2.4846	−1.4152
age (x2)	−0.05990	0.01189	−5.037	4.72e-07	−0.0839	−0.0372

(in the last two columns we added 95% confidence intervals). We conclude that

$$\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \begin{cases} 5.777 - 0.0675 \text{ distance}_i - 0.0599 \text{ age}_i & \text{if patient } i \text{ has a car} \\ 3.840 - 0.0675 \text{ distance}_i - 0.0599 \text{ age}_i & \text{if patient } i \text{ does not have a car} \end{cases}$$

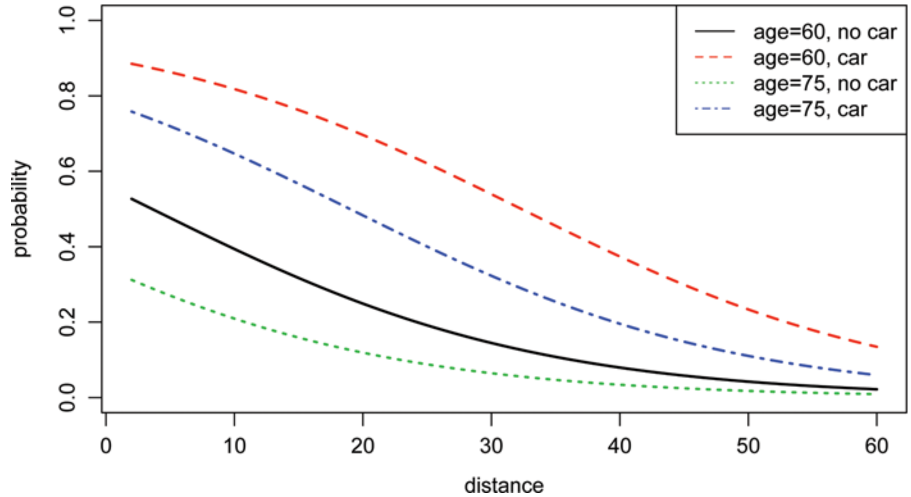

FIGURE 5 Visualization of fitted linear logistic regression model.

regression, the deviance residuals are defined by

$$D_i = 2\,\mathrm{sgn}\,(y_i - \hat{p}_i)\sqrt{y_i \log\left(\frac{y_i}{\hat{p}_i}\right) + (1-y_i)\log\left(\frac{1-y_i}{1-\hat{p}_i}\right)}.$$

From this definition one can easily see that because of the extreme discreteness of binary data, diagnostics based on these residuals are cumbersome (see, e.g., chapter 12 in McCullagh and Nelder 1989). For example, if we plot the (deviance) residuals D_i against the fitted values $\hat{p}_i$ (which is often done in ordinary regression), we always see two separate curves, one corresponding to the patients who participate and one corresponding to the patients who do not.

However, we can still consider leverage values. A high leverage value indicates that a point is an outlier in the space spanned by the predictors. These points can potentially have a large influence on the fitted model. (Note that two obvious leverage points (due to a large distance) were excluded from the analysis at the beginning.) Figure 6 shows a plot of the covariates that appear in the fitted model. Except for patients 88 and 93, no extreme points appear in the covariate space.

A quantity that directly measures the influence of a single point on the fitted model is given by Cook's distance. For each patient, the model coefficients are calculated without the data for that patient and compared to the coefficients obtained using all data.

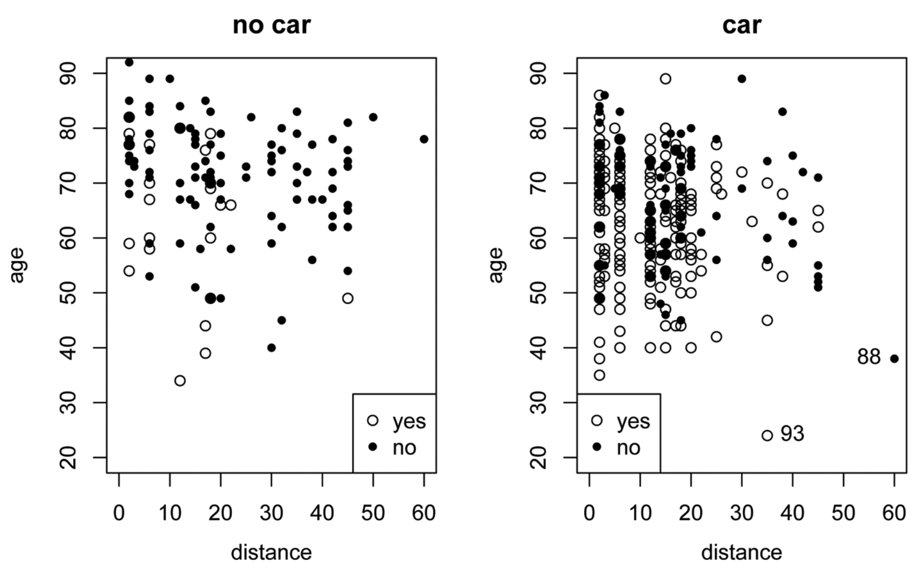

FIGURE 6 Covariate space. Black (open) circles refer to participation = no (yes).

A certain scaled distance between the two sets of coefficients is known as *Cook's distance*. A large Cook's distance (a rule of thumb is that *large* means greater than 0.5; see, e.g., Cook and Weisberg 1999) is either caused by a large residual or a high leverage value or both. To see if there are points in the data that have a relatively large influence on the fitted model, we plotted Cook's distance against the observation numbers; see Figure 7. Only patient number 170 seems to have a somewhat larger Cook's distance relative to the other patients. Because all Cook distances are smaller than 0.5, we conclude that no further analysis of influential points is necessary.

Goodness of Fit

Testing for goodness-of-fit is known to be a particularly hard problem in the case of logistic regression. If there are a limited number of different covariate patterns and replicated measurements for each covariate pattern, goodness of fit can be assessed by methods for categorical data. A typical example of such a method is Pearson's chi-square test. This approach cannot be pursued here, because both distance and age are continuous and hence for these covariates replicated measurements are not available. As a solution, grouping of the data has been advocated. The Hosmer-Lemeshow test is a well-known example of this approach; see, for example, chapter 5 of Hosmer and Lemeshow (2000). For this test, the user has to specify a number of groups G. A default choice is 10. Groups are formed by computing the 0, $1/G$, $2/G, \ldots,$ 1-quantiles of the vector of predicted probabilities (if $G=10$, these are simply the deciles). Let $O_{i,0}$ and $O_{i,1}$ denote the number of zeroes and ones respectively for the ith group. Let $E_{i,0}$ and $E_{i,1}$ denote the expected number of zeroes and ones respectively for the ith group under the fitted model. The Hosmer-Lemeshow statistic is given by

$$\sum_{i=1}^{G}\sum_{j=0}^{1}\frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}.$$

Critical values can be obtained from a χ^2 distribution with $G-2$ degrees of freedom. Observed and expected frequency counts are given in Table 3. Visual inspection of this table suggests that the data fit the model quite well. The observed value of the test statistic equals 4.70. The corresponding p value equals 0.79.

Routinely used packages such as SPSS and Minitab use different grouping strategies and, as a consequence, may yield different results for the same problem (see Pigeon and Heyse 1999). This illustrates the sensitivity of the Hosmer-Lemeshow test to the grouping method. Furthermore, it has been reported that the power of the Hosmer-Lemeshow test is low compared to certain competitors (see Hosmer et al. [1997], where a comprehensive comparison of goodness-of-fit tests is given). More recent work on this topic was performed by Xie et al. (2008), in which groups were obtained by cluster analysis in the covariate space. From the work by Hosmer et al. (1997) it follows that the le Cessie-van Houwelingen-Copas-Hosmer (CHCH test) unweighted sum of squares test for global goodness of fit performs quite well in simulations. Because this statistic is readily explained and also implemented in `R` in

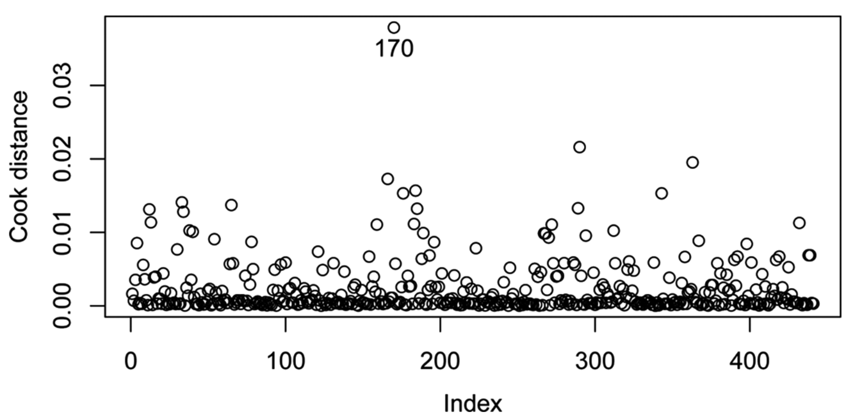

FIGURE 7 Cook's distance vs. observation number.

TABLE 3 **Cells, Observed Counts, and Expected Counts for Computing the Hosmer-Lemeshow Test Statistic**

Cell i	$E_{i,0}$	$E_{i,1}$	$O_{i,0}$	$O_{i,1}$
[0.0076, 0.119]	43.06	2.94	44	2
(0.119, 0.214]	36.05	6.95	37	6
(0.214, 0.367]	31.00	13.00	28	16
(0.367, 0.532]	23.67	20.33	24	20
(0.532, 0.66]	17.28	26.72	18	26
(0.66, 0.729]	13.47	30.53	11	33
(0.729, 0.779]	10.72	33.28	9	35
(0.779, 0.821]	9.29	37.71	13	34
(0.821, 0.872]	6.46	34.54	6	35
(0.872, 0.972]	3.99	40.01	5	39

the Design library (using the command lrm.residuals), we will also assess the fit of our model using this test statistic. The expected response for the ith patient equals $p_i = p_i(\beta)$. The fitted value for this patient equals its estimated expected response, which is therefore given by $\hat{p}_i := p_i(\hat{\beta})$. The CHCH test is a studentized version of

$$T = \sum_{i=1}^{n} (Y_i - \hat{p}_i)^2.$$

For large data sets critical values can be obtained from the standard normal distribution. Because the data considered here contain over 400 patients, the work by Hosmer et al. (1997) suggested that the test should have about 90% power to detect moderate departures from linearity.

Applying the test to our data and model gives

Sum of sq. errors	Exp. value|H0	S	Z	P
72.98	72.75	0.54	0.42	0.67

From left to right, the output gives the observed value for T, an approximation of its expectation and standard deviation under the null hypothesis, the value of the studentized test statistic (Z), and the corresponding p-value (P). Again, there is no reason to doubt the model's fit.

Improvement Actions Based on a Break-Even-Point Analysis

The factors age and distance are nuisance factors: they cannot be controlled. Mobility, on the other hand, can be influenced. Past data show that every month approximately 15 patients discharged from the nursing department do not have a car. For these 15 patients we aim to increase the probability that they join the program by improving transport to the hospital. A major question is how much can be invested to acquire these patients.

Let

$$\delta(\text{distance, age}) = \hat{p}(\text{distance, age, mobility} = \text{car}) - \hat{p}(\text{distance, age, mobility} = \text{no car}). \quad [2]$$

Figure 8 shows a contour plot of δ.

From this plot we can see for which patients investing in mobility pays off the most. These are the patients with age and distance that fall within the white-colored area. However, for ethical reasons, the hospital decided not to discriminate between patient characteristics.

Therefore, we calculated the *average* value of δ for all patients without a car in the hospital using [2] and the fitted model. This average probability equals 0.35. Hence, the maximum amount that can be invested to ensure transport for each patient equals 0.35 times 29 sessions on average times 22.83 euros per session = 232 euros. Because there are approximately 15 patients a month without a

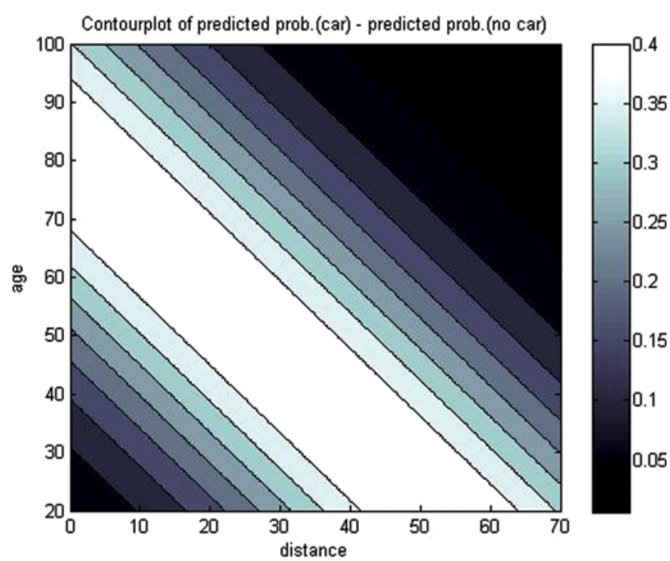

FIGURE 8 **Contour plot: Probability of joining the program for a patient with a car minus the probability of joining the program for a patient without a car, depending on distance and age.**

car, we can invest $15 \times 232 = 3{,}480$ euros a month to ensure transport.

Based on this economic model, feasible improvement actions can be evaluated. Because a taxi service for single patients turned out to be too expensive; the project leader came up with a carpooling procedure to couple patients with a car to patients without a car. After a pilot phase of the carpooling procedure, a number of patients started using this service. Most of these patients explained that they would not have joined the program if this service were not available.

Control Phase

In the last phase of the Six Sigma project the CTQs were monitored. Using a dashboard, the department can see the number of participating patients and the number of sessions per patient for each month. The number of included patients increased from 33 to 45 (far above the objective in the analyze phase). The number of sessions remained constant, at 29 on average. The project was handed over to the department. Additional revenues turned out to be circa 96,000 euros each year. Note that all initial improvement ideas were abandoned.

CONCLUSIONS

This case study describes the success of the Six Sigma methodology in a hospital for a specific project. Verifying ideas before developing improvement actions is an important aspect in this methodology. Often people's ideas on processes are incorrect, but improvement actions based on these are still being implemented. These actions cause frustrated employees, may not be cost effective, and in the end do not solve the problem. Within Six Sigma it is obligatory to verify the ideas by data analysis or experiments. At first this may seem like a loss of time, energy, and resources, but it is, much less than what is lost when implementing the wrong improvements.

Some people argue that statistics do not solve major problems in health care; Lean principles are much more effective is the credo. This case study shows that statistical analysis can be very useful. Even when a somewhat more advanced technique like logistic regression modeling is required, exploratory graphics such as boxplots and bar charts point the direction toward a valuable solution. The usefulness of the logistic regression is demonstrated by the resulting economic model, which can assist in making deliberated improvement actions.

ABOUT THE AUTHORS

Dr. F.H. van der Meulen is assistant professor at Delft Institute of Applied Mathematics (Delft University of Technology). His present work focuses on (Bayesian) computational statistics and Bayesian nonparametrics. He earned a Ph.D. in mathematical statistics from the Free University of Amsterdam. Part of this work was carried out during a two-year period at the Institute for Business and Industrial Statistics at the University of Amsterdam.

Dr. M.B. Vermaat is Deputy Manager of the Sorting Centre Rotterdam at TNT Post in the Netherlands. Before, he was Programme Manager Lean Six Sigma at TNT Post. He is certified Master Black Belt. He has worked for the Institute for Business and Industrial Statistics at the University of Amsterdam, where he earned his Ph.D. in Mathematics in 2006.

P. Willems is corporate performance manager at Jessa Hospital in Hasselt, Belgium. He coordinates the BPM and Lean Six Sigma program. He received his Master Black Belt degree at the Institute for Business and Industrial Statistics at the University of Amsterdam and his Executive Master in Business Process Management at the Vlerick Leuven Ghent Management School.

REFERENCES

Barry, R., Murcko, A. C., Brubaker, C. E. (2002). *The Six Sigma Book for Healthcare*. Chicago: Health Administration Press.

Bisgaard, S. (2009). *Solutions to the Healthcare Quality Crisis: Cases and Examples of Lean Six Sigma in Healthcare*. ASQ Quality Press.

Breyfogle, F. W. (2003). *Implementing Six Sigma-Smarter Solutions Using Statistical Methods*, 2nd ed. New York: Wiley.

Cook, R. D., Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.

Friedman, J. H. (1984). *A variable span scatterplot smoother*. Technical Report No. 5. Laboratory for Computational Statistics, Stanford University, Palo Alto.

Hosmer, D. W., Hosmer, T., Le Cessie, S., Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, 16:965–980.

Hosmer, D. W., Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley Series in Probability and Statistics. New York: Wiley.

Ihaka, R., Gentleman, R. (1996). *R*: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.
Institute of Medicine. (2001). *Crossing the Quality Chasm; a New Health System for the 21st Century*. Washington, DC: National Academy Press.
de Koning, H., de Mast, J. (2007). The CTQ flowdown as a conceptual model of project objectives. *Quality Management Journal*, 14(2): 19–28.
de Mast, J., Does, R. J. M. M., de Koning, H. (2006). *Lean Six Sigma for Service and Healthcare*. Alphen aan den Rijn, The Netherlands: Beaumont Quality Publications.
McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.
Myers, R. H., Montgomery, D. C., Vining, G. G. (2002). *Generalized Linear Models: with Applications in Engineering and the Sciences*. New York: John Wiley & Sons.
Pigeon, J. G., Heyse, J. F. (1999). A cautionary note about assessing the fit of logistic regression models. *Journal of Applied Statistics*, 26(7):847–853.
Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. New York: Chapman & Hall.
Xie, X. J., Pendergast, J., Clarke, W. (2008). Increasing the power: A practical approach to goodness-of-fit test for logistic regression models with continuous predictors. *Computational Statistics & Data Analysis*, 52:2703–2713.