

The logistic regression model in the statistical justification of the cost of hematopoietic stem cell transplantation

Dmitry A. Bagge, Boris I. Smirnov, Boris V. Afanasyev

R.M.Gorbacheva Memorial Institute of Children Hematology and Transplantation, and Department of Hematology, Transfusiology and Transplantology, and St.Petersburg State Medical I.Pavlov University, Russia

Abstract

The logistic regression model allows the calculation of the minimal acceptable cost of HCT, having non-random impact on HCT outcome, under which the probability of a positive outcome is 50%. There were 209 patients enrolled in the study, who received autologous HCT, and allogenic related and unrelated HCT. The non-random cost and medical parameters connected with patient status were defined before HCT and HCT outcomes. The application of reduced toxicity (RTCR) or myeloablative conditioning regimens (MCR), the presence of relapse before HCT, and the cost of drugs and blood transfusion all have an influence on the outcome, and the weight coefficients of these parameters were calculated. This allows the connection of costs and clinical parameters, and the calculation of the minimal acceptable cost of HCT.

Keywords: logistic regression model, reduced toxicity conditioning regimens (RTCR), myeloablative conditioning regimens (MCR), cost of drugs, cost of blood transfusion, weight coefficient

Introduction

The basic criterion for choosing a particular method of treatment in medicine ought to be its effectiveness concerning the disease. This approach should be used for managing the patient: choosing the most effective method despite the cost. However, in many cases the high cost of a method combined with debatable effectiveness challenges their use and development, as well as their state financing.

Therefore, the aim of the study was to find some logical evaluation procedures of the treatment cost of HCT.

Materials and methods

There were 209 patients enrolled in the study, who received autologous HCT, and allogenic related and unrelated HCT from 1998 to 2005. There were 91 patients (43.5%) aged from 1 to 66 years with allo-HCT, among them 57 from unrelated donors and 34 from related donors; 118 patients (56.5%) had auto-HCT.

Using their medical charts we performed the treatment efficacy analysis and treatment cost analysis for these aforementioned patients. Clinical effectiveness was assessed using a 2-year overall survival. Evaluation of the total cost and its components in related and unrelated allo-HCT, and assessment of allo-HCT cost under myeloablative and reduced toxicity regimens was also performed. We have independently analyzed the cost of HCT with and without complications. There was an attempt to reveal the statistically justified clinical and cost parameters that influenced

HCT effectiveness. The results were analyzed using parametric and non-parametric statistics; data were assessed with regression analysis.

Results

The result of data analysis suggests using the total cost of HCT. The expression for total HCT cost can be expressed like this:

$$C_{\text{total}} = C_{\text{exam}} + C_{\text{p/d}} * N_{\text{d}} + C_{\text{transph}} * N_{\text{d}} + C_{\text{cond}} + C_{\text{drug}} * N_{\text{d}} + (C_{\text{donor search}}) \quad (1)$$

C_{exam} – the cost of laboratory examinations performed while the patient is in the clinic. This depends on the disease and duration of in-patient period.

$C_{\text{p/d}}$ – the cost of one patient day. This value is stable and calculated by the clinic's statistics department.

N_{d} – the duration of in-patient period. This depends on disease severity and complications.

C_{transph} – the cost of transfusions per day (24 hours). This also depends on disease severity, and complications that require blood transfusions.

C_{cond} – the cost of a conditioning regimen. This value is stable.

C_{drug} – the cost of drugs per day (24 h). Dependent on disease severity.

C_{infus} – the cost of apheresis and storage of BM and PBSC. This value is stable.

$C_{\text{donor search}}$ – this value should be considered only in unrelated allo-HCT; it corresponds to the cost of a donor search in international donor registers.

Before the model's construction, which could be a function of clinical, cost and other parameters, it is reasonable to perform exploratory statistical analysis. The aim of such analysis is to reveal the statistically significant influence on the outcome of such features as: sex of patient, diagnosis, type of HCT, donor sex, source of hematopoietic stem cells (BM or PBSC), presence of relapse or progression, complications, particularly infections that require blood transfusions, conditioning regimen, age at HCT, and/or presence of GVHD. According to the character of the clinical parameters, we used analysis of continuous and category variables.

Our first step was to perform non-random difference of central tendency of continuous variables analysis (1). Table 1 presents the continuous variables that help to reveal a significant difference in groups by patient status.

Additionally, we analyzed the clinical, continuous and category variables influencing the HCT effectiveness. The statistically significant influence on HCT outcome caused:

- HCT type – autologous or allogenic (p=0.002),
- presence of relapse or progression (p=0.048),
- presence of blood transfusions complications (p=0.003),
- type of conditioning regimen: myeloablative or reduced toxicity (p=0.023).

The exploratory statistical analysis performed helps to form the group of continuous and category variables that can be used as predictors for regression modeling of patient status according to clinical and cost parameters.

The construction of a regression model with dependent variables, simulating dichotomous category variables and independent variables creates the need for logistic regression. Logistic regression connects event probability (one of the events of disease outcome variables) with independent variables (predictors), the impact of which was described in the previous section. Considering that the dependent variable is measured by probability mass, and independent variables include continuous and category parameters, it is necessary to make a functional transformation of the independent variables into the interval 0–1. Such functional transformation is done by function (2).

$$P = \frac{1}{1 + e^{-z}}, \quad (2)$$

this is called the logistic, with Z parameter.

The parameter $Z=B_1X_1+B_2X_2+B_3X_3+B_4X_4+B_5X_5$ connects independent variables (predictors).

The procedure of both logistic regression construction and standard regression includes three steps:

- Creation of a logistic regression model,
- Evaluation of the significance of weight factors in the formula (B_1, B_2, B_3, B_5),
- Assessment of model stability.

By means of SPSS, using step-type variants of logistic regression to include or exclude parameters from the model and Wald criterion, we obtained the model, the parameters of which are stated in Table 2 (see overleaf).

According to formula (2), the parameter

$$Z = -6.027 \cdot 10^{-5} \cdot X_1 + 0.0095 \cdot X_2 + 3.513 \cdot 10^{-6} \cdot X_3 - 2.159 \cdot X_4 - 2.059 \cdot X_5. \quad (3)$$

Formula (3) includes:

- X1 – total cost of blood transfusions
- X2 – the cost of blood transfusions per 24 h
- X3 – the cost of drugs
- X4 – conditioning regimen
X4=0 in reduced toxicity regimen,
X4=1 in myeloablative regimen,
- X5 – relapse before HCT
X5=0 no relapse before HCT,
X5=1 presence of relapse before HCT.

Since the total cost of blood transfusions is obtained via the multiplication of the cost of blood transfusions per 24 h on the number of patient days, $X_1=X_2 \cdot N_d$, formula (3) can be expressed like this:

$$Z = -6.027 \cdot 10^{-5} \cdot X_2 \cdot N_d + 0.0095 \cdot X_2 + 3.513 \cdot 10^{-6} \cdot X_3 - 2.159 \cdot X_4 - 2.059 \cdot X_5 \quad (4)$$

As stated in Table 2, the model has non-random parameters with $p < 0.05$.

The quality of patient status prognosis could be assessed using the Table of forecast classification (Table 3, see overleaf).

The values in Table 3 characterize the power of testing (probability of status Alive prediction in present alive condition) and its specificity (probability of status Dead prediction in present dead condition). Thus the power of forecast is 93.5% and its specificity is 63.6%. The probability of a correct forecast is 81.1%.

Table 1. Statistically significant cost predictors influencing the HCT outcome

Predictor	Total cost	The cost of blood transfusions	The cost of blood transfusions per 24 h	The cost of conditioning regimen per kg	Total cost of conditioning regimen	The cost of drugs	The cost of drugs per 24 h
Statistical significance	p=0.002	p=0.01	p=0.003	p=0.05	p=0.047	p=0.009	p=0.007

Table 3. Table of forecast classification

Regression step	Parameter		Outcome (status)		Percent of correct forecasts
			Alive	Dead	
8	Outcome (status)	Alive	29	2	93.5
		Dead	8	14	63.6
	Total percent		81.1		

The stability of the model should be checked using other samples, but we currently have no such data, so practical statistics recommends repeating the analysis using only a part of data for model creation, with the other part of data acting as the model validity check. In our study we embraced this approach.

All data were randomly divided into two parts using the Bernoulli distribution. The first part (selected observations) included approximately 70% of the data, the second (unselected observations) the remaining data. The first sample was used in a logistic regression procedure for model creation, while the other was used to check its validity.

The model had the same parameters as described above, and its stability was confirmed using the Table of classification (Table 4), created independently for the two samples.

As is clear from the data in Table 4, test power and specificity between Selected observations and Unselected observations have different meanings. Such differences are caused by random variations. The absence of non-randomness for such differences can be assessed by Fischer’s test. In this case:

For observation Alive give the opportunity $p = 0.1594$, for observation Dead give the opportunity $p = 0.7081$.

Table 4. Table of classification of forecasted data

	Selected observations			Unselected observations		
	Outcome		Percent of correct forecasts	Outcome		Percent of correct forecasts
	Alive	Dead		Alive	Dead	
Alive	23	2	92	4	2	66.7
Dead	5	13	72.2	1	3	75
Total percent	83.7			70		

Table 2. The variables in HCT cost formula

Regression step	Predictor	Weight coefficient B	SE	Wald criterion	Statistical significance
8	The cost of blood transfusions	$-6.027 \cdot 10^{-5}$	<0.001	2.95	0.086
	The cost of blood transfusions per 24 h	0.0095	0.004	5.98	0.014
	The cost of drugs	$3.513 \cdot 10^{-6}$	<0.001	4.038	0.044
	Conditioning regimen	-2.159	0.752	8.243	0.004
	Presence of relapse	-2.059	0.862	5.703	0.017

Both opportunities show the absence of differences in the columns of classification table, i.e., the absence of differences in diagnostics of the two samples, the first of which was used for the model construction, and the second, which was used as a control.

A number of parameters were revealed during the aforementioned analysis, which are either related or independent from the disease outcome.

If the parameters are not connected with patient status, the central tendencies, obtained during common analysis of study predictors [see formula (1)], should be used for their characteristic. Considering abnormality of studied random values distribution, the median could be considered a central tendency.

The value of the parameters connected with patient status should be obtained by means of a regression model. The logistic function (2) possesses the value 0.5 under $Z=0$. This value corresponds with the situation when patient status is determined with probability 0.5. To obtain a more strict forecast condition $Z>0$ or $Z<0$ according to the target status value should be met.

We then set the value $Z=0$ to search for minimal expenses for HCT. Then from formula (4) we get:

$$-6.027 \cdot 10^{-5} \cdot X_2 \cdot N_{\text{cyt}} + 0.0095 \cdot X_2 + 3,513 \cdot 10^{-6} \cdot X_3 = -2.159 \cdot X_4 - 2.059 \cdot X_5 \quad (5)$$

On the left side of an equation are the cost parameters, and on the right side are the clinical parameters.

For different treatment variants (k), which are fed into the right side of equation (5), we could calculate weighting values, given by equation:

$$k = 2.159 \cdot X_4 + 2.052 \cdot X_5.$$

Estimated values for k are shown in Table 5.

Using the data from Table 5 for k and setting the median for X_2 or X_3 we can calculate the minimal expenses, which depend on

Table 5. The variants of k parameter calculation

	X_4	X_5	k
	0(MCR), 1(RTCR)	0(Remission), 1(Relapse)	
1	0	0	0
2	0	1	2.159
3	1	0	2.059
4	1	1	4.218

patient status, and which patients will be alive with probability 0.5 (i.e., 50%).

There could be 2 variants of calculation:

1. We set median for X_2 (the median of cost of blood transfusions per 24 h under known clinical parameters), and the second parameter is calculated like this:

$$X_2 = \frac{-k + 6.027 * 10^{-5} * X_2 * N_6 - 0.0095 * X_2}{3.513 * 10^{-6}}$$

2. We set median for X_3 (i.e., the median for total drug cost), and the second parameter is calculated like this:

$$X_3 = \frac{-k - 3.513 * 10^{-6} * X_3}{-6.027 * 10^{-5} * N_6 + 0.0095}$$

In summary, a model for calculating minimal acceptable cost of predictors was created, having a non-random impact on HCT outcome, under which the probability of positive outcome is 50%. The calculation of minimal total cost of HCT, under which there is 50% survival, is possible by data substitution on formula (1).

Discussion

The transplantation of hematopoietic stem cells is one of the high-technology treatment methods, thus it is rather expensive due to demand for contribution to international directives (GMP, EBMT).

According to the literature, the cost of allo-HCT can vary from US \$100,000 to \$250,000 (Armitage J.O. et al., 1984; Welch H.G., Larson E.B., 1989; Beard M.E. et al., 1991; Dufour T. et al., 1992; Barr R. et al., 1996; Faucher C. et al., 1998); cost differences are caused by local features in different countries, considering economic factors, labor costs, drug costs, etc. M. van Agthoven et al. (2002) reviewed the results of allo-HCT in patients with acute leukemia (ALL and AML) for 2 years. In patients who survived, the cost of allo-BMT from HLA-matched related donors was approximately EUR 103.509, and the cost of allo-PBSCT was EUR 105.906. The cost of allo-BMT from HLA-matched unrelated donor was approximately EUR 173.587, where 1/3 of this sum was spent on a donor search.

According to the literature, the main components of clinical expenses in HCT are outgoings on drugs (38.9%);

33.7% of clinical expenses are due to the cost of patient days; 7.5% is for blood transfusions; 5.8% for laboratory examinations; 5.6% for microbiological examinations; 1.4% for radiology, and 1.9% are other expenses (Kline R.M. et al., 1998).

In our study the aforementioned components of HCT costs were analyzed on statistically significant influences on death rate after HCT and were included in a suggested model of cost assessment, which provide 50% survival.

Despite of the importance of the studied problem and limiting role of high cost of HCT in its routine use in clinical practice in some countries, at the moment there is no method for analysis that can definitely justify its cost, and moreover there are no approaches to predict the influence of expenses on disease outcome.

Considering these facts the suggested method for statistically justified assessment of HCT cost, which helps to connect clinical parameters influencing treatment cost, as well as forecast minimal accessible cost of HCT, in which 50% is achieved, could be used for evaluation of necessary financing for this treatment method.

References

1. Dubno P.U. Using SPSS for treatment of statistical data. Moscow: LLC Publishing house AST: NT Press, 2004, 221 p. (In Russian)
2. Nasledov A.D. Computer analysis of data in psychology and social sciences. St. Petersburg: Piter, 2005, 416 p. (In Russian)
3. Robert H. Fletcher, Suzanne W. Fletcher, E. Wagner. Clinical Epidemiology: The Essentials. Moscow: Media Sphera, 1998, 352 p. (In Russian)
4. Armitage J.O., Klassen L.W., Burns C.P. et al. A comparison of bone marrow transplantation with maintenance chemotherapy for patients with acute nonlymphoblastic leukemia in first complete remission. *Am. J. Clin. Oncol.* 1984,7(3):273-278.
5. Barr R., Furlong W., Henwood J. et al. Economic evaluation of allogeneic bone marrow transplantation: a rudimentary model to generate estimates for the timely formulation of clinical policy. *J. Clin. Oncol.* 1996,14(5):1413-1420.
6. Beard M.E., Inder A.B., Allen J.R. et al. The costs and benefits of bone marrow transplantation. *N.Z. Med. J.* 1991,104(916):303-305.
7. Dufour T., Saux M.C., Terraza B. et al. Comparative cost of allogeneic or autologous bone marrow transplantation and chemotherapy in patients with acute myeloid leukaemia in first remission. *Bone Marrow Transplant.* 1992,10(4):323-329.
8. Faucher C., Fortanier C., Viens P. et al. Clinical and economic comparison of lenograstim-primed blood cells (BC) and bone marrow (BM) allogeneic transplantation. *Bone Marrow*

Transplant. 1998;21(Suppl.3):S92-S98.

9. Kline R.M., Meiman M., Tarantino M.D. et al. A detailed analysis of charges for hematopoietic stem cell transplantation at a children's hospital. Bone Marrow Transplant. 1998 Jan;21(2):195-203.

10. van Agthoven M, Groot MT, Verdonck LF, et al. Cost analysis of HLA-identical sibling and voluntary unrelated allogeneic bone marrow and peripheral blood stem cell transplantation in

adults with acute myelocytic leukaemia or acute lymphoblastic leukaemia. Bone Marrow Transplant. 2002 Aug;30(4):243-51.

11. Welch H.G., Larson E.B. Cost effectiveness of bone marrow transplantation in acute nonlymphocytic leukemia. N. Engl. J. Med. 1989;321(12):807-812.

© The Authors. This article is provided under the following license: Creative Commons Attribution-Noncommercial 2.0 Germany, <http://creativecommons.org/licenses/by-nc/2.0/de/deed.en>

Ссылка: Клеточная терапия и трансплантация, том 1, номер 1, 28 мая 2008, с. 56,
doi: 10.3205/ctt2008-05-28-003-ru-a

Модель логистической регрессии в статистическом обосновании стоимости трансплантации гемопоэтических стволовых клеток

Д.А. Багге, Б.И. Смирнов, Б.В. Афанасьев

Резюме

Модель логистической регрессии позволяет рассчитать минимально допустимую стоимость ТГСК, при которой возможно достижение 50% положительного эффекта. В исследование вошли 209 пациентов перенесших аутологичную, аллогенную родственную и неродственную ТГСК. Было определено неслучайное влияние на исход ТГСК параметров стоимости и различных медицинских факторов, в том числе возраста и пола пациента, типа донора (родственный/неродственный), источника стволовых клеток (из костного мозга или периферической крови).

В стоимость процедуры ТГСК включались расходы на пребывание больного в стационаре (число койко-дней), стоимость лабораторного обследования, расходы на кондиционирующую терапию, стоимость лекарственных препаратов в посттрансплантационном периоде, расходы на аферез и заготовку костного мозга или периферических стволовых клеток, а также стоимость поиска донора (при неродственных ТГСК). Применение режимов кондиционирования со сниженной интенсивностью дозы и миелоаблативных режимов, наличие или отсутствие рецидива, а также стоимость медикаментов и трансфузиологического пособия статистически значимо влияли на исход ТГСК, что позволило рассчитать весовые коэффициенты и связать показатели стоимости и клинические параметры. Это дает возможность рассчитать минимально допустимую стоимость ТГСК.

Ключевые слова: гемопоэтические стволовые клетки, трансплантация, стоимость, статистика, регрессионная модель