Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

COST-EFFECTIVENESS EVALUATION OF DROTRECOGIN ALFA COMPARED TO STANDARD CARE IN THE TREATMENT OF SEVERE SEPSIS IN REAL PRACTICE

Stéphanie Payet[1], Lionel Riou França[1], Katel Le Lay[1], Robert Launois[1]

Correspondence

Stéphanie Payet

REES France – 28, rue d'Assas – 75006 Paris – France

Tel: +33 1 44 39 16 90

Fax: +33 1 44 3916 92

Cost-effectiveness evaluation of drotrecogin alfa compared to standard care in the treatment of severe sepsis in real practice

ART-5184/05

ABSTRACT

AIM To establish the cost and cost-effectiveness of drotrecogin alfa (DA) in the treatment of severe sepsis with multiple organ failure compared to standard care in daily practice.

METHODS A pre-post study was conducted between September 2002 and February 2005, before and after introduction of drotrecogin alfa (DA) onto the market. Recruitment bias was suspected, as the treatments were not randomly assigned. Patients from both phases with similar characteristics were matched by the propensity score (PS) methodology to remove this bias. Total hospitalisation costs were quantified using a regression equation involving specific tools to describe intensive care units (ICU) activity. Acquisition costs for DA were summated, assuming that all costs associated with DA were already included in the equation. Cost comparisons were conducted using the non parametric bootstrap method, as this appeared to be the most robust technique for skewed data compared to the Student t-test, Mann-Whitney U-test and logarithmic transformation of cost data. Cost-effectiveness quadrants and acceptability curves were used to assess uncertainty of the cost-effectiveness ratio.

RESULTS 1,096 patients were recruited into the study. The PS matching procedure retained 840 patients. The total hospitalisation costs were 36,717.03 € in the "before" phase and 47,870.09 € in the "after" phase. The 95% confidence interval of the difference between the two phases was [6,600.68; 15,709.45], revealing a significant increase in total costs for DA treated patients. The increase was associated not only with acquisition costs for DA, which, on average, were 6,716.92 €. ICU activity was also higher for DA treated patients for whom the Omega score was 14.45% higher and length of stay increased from 21.27 days to 24.42 days on average (p = 0.0018). The cost-effectiveness ratio was 20,278.29 € per life year gained. The probability that DA is an acceptable treatment was approximately 74% against the willingness to pay of 50,000 €.

CONCLUSION The additional costs associated with the prescription of DA cannot be related only to its price. DA treated patients undergo more technical medical procedures and have a longer length of stay. The acceptability threshold for DA is reasonably high, and could have been even higher without the lack of power of the study to accurately measure the effect of treatment.

INTRODUCTION

In 2000, the support programme for innovative and costly techniques (STIC) was launched by the Directorate for Hospitals and Organisation of Care (DHOS). This programme is located downstream from clinical research and allows establishments which receive a global funding package to be encouraged to adopt innovative high cost techniques on a national scale. The programme has three objectives: to promote patient access to high cost innovations by allocating specific credits to the health care establishments taking part in the project, to analyse the conditions of use of the new development in the management of patients and to assess this in medical and financial terms and to promote exchange and professional consensus formation between the doctors taking part in the study. The medico-economic evaluation of the development is designed on a pragmatic basis once the development has been validated by a clinical research stage and has begun to be distributed.

The PREMISS study (Intensive Care Study for Medico-Economic Evaluation of a New Development in Severe Sepsis) was launched in 2002 by two intensive care learned societies, the SFAR (French Society of Anaesthesia and Intensive Care) and the SRLF (French Speaking Intensive Care Society). It was designed to assess the observed costs and efficacy of a new adjuvant treatment, PCArh (drotrecogin alfa or DA), compared to previous conventional treatment in the management of severe septic syndromes and/or septic shock in intensive care.

Specific tools have been developed by intensive care physicians to defend the specific features of their speciality in the general measure of hospital activity. The Omega field in the catalogue of medical procedures reflects the activity of a department through a listing of 47 therapeutic actions during the entire stay, which are weighted from 1 to 10. A score is then obtained for each stay which is given by the sum of the points associated with each action being performed. The New Simplified Acute Physiology Score (SAPS II) allows the risk of death in intensive care to be estimated. This is calculated at the 24th hour of the stay from 17 clinical and biological items. Well correlated with burden of care, it has been introduced since 1st January, 1997 in the RUM (Summary of Medical Units) in order to improve PMSI (Information Systems Medicalisation Programme) performance. Finally the length of stay is also a good indicator of the cost of the stay. The equation we have chosen to measure the complete hospitalisation costs of patients suffering from severe sepsis uses these three indicators.

In order to achieve the objectives, a quasi-experimental "Before"/"After" study design was used. Inclusions incurred before and after the drug was made available in the regulatory context. The control group did not received DA but conventional treatment. All patients during the "After" phase received DA in addition to conventional treatment. The recommended dose was 24 µg/kg/h for 96 hours. Data were collected on a decentralised basis from each investigating center by an electronic record designed, created and administered by REES France, in conjunction with the study coordinators and investigators.

In contrast to a randomised clinical trial, this type of study design does not guarantee comparability of the patients recruited during the two phases, as the decision to include a patient into one or other phase does not occur as a result of a random process. Recruitment bias therefore potentially exists. One approach has been proven to reduce this bias: the propensity score (PS).[1] This involved the construction of a sample of similar patients based on most of the characteristics measured between the two phases.

The cost comparison in such a sample between the two phases is therefore less biased. As mortality data were also recorded 28 days after the start of severe sepsis, a cost-effectiveness analysis was conducted. The effectiveness criterion used was the number of life years gained.

METHODS

A costs analysis based on observational data requires specific methods. We firstly describe the methodology used to identify and exclude possible recruitment bias and then describe the cost equation chosen to calculate complete hospitalisation costs. Finally we justify the use of a non-parametric bootstrap analysis for the cost comparison in the two treatment phases and for the uncertainty's representation of the cost effectiveness ratio.

Measurement and reduction of recruitment bias

DA has already been shown to be effective in the PROWESS[2] clinical trial and no randomisation was therefore conducted in the PREMISS trial in order that none of the patients included after the molecule had been made available on the market suffered loss of opportunity. The patients included were therefore obtained from a two level sampling process: each intensive care unit, recruited on a voluntary basis, selected a patient sample. Thus, all patients who met the inclusion criteria were not necessarily recruited in these units. This study design is therefore subject to potential recruitment bias and there is nothing *a priori* to guarantee patient comparability in the two phases.

One of the most widely used criteria to identify recruitment bias is the balance of initial features between the two phases by calculating standardised differences.[3] This is measured as followed for a quantitative variable:

$$d=\frac{\overline{x}_{After}-\overline{x}_{Before}}{\sqrt{\frac{s^2_{After}+s^2_{Before}}{2}}}$$

It therefore represents the difference between the means, ($\overline{x}$), weighted by the common standard deviation. This can be generalised in the cases of binary and qualitative variables. A standardised difference of more than 10% is an indicator of imbalance of the variables studied between the two phases and indicates the presence of recruitment bias.

Several competing methods may be used to reduce this bias: pairing, stratification or adjustment. The first two must be performed at the inclusion step. The last one is the most widely used method. It involves introducing potential confounding variables into a multivariate model. Its quality, however, depends greatly on the quality of the model: if it is too far from reality the bias from a given covariate which is adjusted may be increased. The propensity score (PS) method is an alternative to the previous ones. The PS indicates the probability that a person of given characteristics is exposed to treatment.[4] It can reduce a large number of covariates into a single composite variable which correctly summarises all of the features observed. Its distribution provides a criterion to assess comparability between populations which are exposed or not exposed to treatment.[5]

In the context of PREMISS a logistic regression model allowed the PS to be estimated. This was then used to form a sample of similar patients between the two phases by matching, i.e. pairing a patient from the "Before" phase with a patient from the "After" phase who had a similar PS. The matching algorithm used was the SAS[©] "match" macro.[6] This is described as optimal as it matches patients from two phases depending on their propensity score in order to minimise the total distance between the PS of matched patients (each distance representing the absolute value of the difference between the two PS of the matched patient couple). The sample obtained is more balanced than the initial sample in terms of the observed features.

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

Calculation of complete hospitalisation costs

Complete hospitalisation costs were estimated from the CUB-Rea[7] (College of Intensive Care Database Users) database and a multiple regression equation constructed from a micro costing study, based on 211 stays in intensive care units in 1996.[8] In addition to the RUM information relating to the PMSI (age, sex, length of stay, admission and discharge methods, diagnoses and procedures), this database includes intensive care specific indicators: the SAPS II, Omega field, McCabe score and admission type. The complete hospitalisation costs counted in the micro-costing study are broken down as follows:

✍ In the intensive care unit:

- Variable direct costs: investigations (laboratory, imaging); small materials; drugs and blood products; care staff (SRN[*], HCA[**]), calculated by time spent.

- Fixed direct costs: medical nursing staff calculated pro rata for the length of stay.

- Variable indirect costs: restaurant services; laundry; pharmacy; administration.

✍ Post intensive care: number of days, costed by the departmental tariff category.

The equation obtained[9] has a determination coefficient (R^2) of 93% and is expressed as follows:

$$CC = \beta_0 + \beta_1 * LOS + \beta_2 * LOS * 1_{DCR=1} + \beta_3 * \Omega_{TOT} + \beta_4 * (SAPS2)^2 + \beta_5 * 1_{DCR=1}$$

Where:

- CC: Total Complete Cost of the Hospital Stay (in FF_{96})
- LOS: Length of Stay in Intensive Care
- Ω_{TOT}: Total Omega Score
- SAPS2: SAPS II Score
- $1_{DCR=1}$: Variable indicating death in intensive care
- β_0 = -8,881.50; β_1 = 5,465.60; β_2 = 3,715.10; β_3 = 183.75; β_4 = 5.27; β_5 = -18,078.50

For short length of stays (less than 5 days) therefore, the cost of survivors is greater than the cost of patients who die in intensive care. Beyond that given threshold patients who die cost increasingly more as their length of stay increases.

---

[*] State Registered Nurse
[**] Health care assistant

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

Addition of DA acquisition costs and expression of costs in euros 2004

This general equation applies both to patients suffering from severe sepsis and to those suffering from other diseases but does not take account of the medical costs associated with administration of DA. The acquisition costs of DA were therefore added to the complete hospitalisation costs, assuming that all the connate costs associated with DA administration (adverse events, longer term follow up etc.) are incorporated in the equation through the Omega score, the SAPS II and the length of stay in intensive care. This is an essential assumption, as it ensures that the total cost of patients receiving care with DA is not under-estimated and a realistic assumption, as these three indicators have been designed to best represent activity in intensive care.

The year 2004 was chosen to harmonise all of the costs calculated in this study as it contains the latest inclusions. The CUB-Rea equation is expressed in French Francs 1996, therefore the INSEE[10] inflation rates were used to obtain nominal values in 2002, 2003 and 2004 for the costs expressed initially in French Francs 1996.

Table 1: Purchasing power parity in 2002, 2003 and 2004

	Value in € 2002	Value in € 2003	Value in € 2004
1 F 1996 is worth	0.16456	0.16798	0.17156

Once the costs have been expressed as nominal values for each year, the time passed must be incorporated by capitalising all costs for the year 2004. In our case we shall use a capitalisation rate, r, of 3.5%. 2004 values are obtained as follows: $V_{€02 \rightarrow €04} = V_{€02} \times (1 + r)^2$ and $V_{€03 \rightarrow €04} = V_{€03} \times (1 + r)$.

Statistical methods for cost analysis: incorporating dispersion

As the main objective was to estimate the additional cost associated with using DA for the treatment of severe sepsis we had to use suitable tools for comparing dispersed data. Indeed, some patients incur very high costs, extending the cost distribution towards the right and making the statistical analysis more complex than for symmetrical quantitative variables. There is no widely accepted cost analysis method. Two common principles clash in this area. According to Thomson and Barber,[11] the main information which must be obtained from an economic study lies in the arithmetical mean of the per patient costs. Tests based on the distribution (Mann-Whitney test) or on the geometrical mean (logarithmic transformation of costs) should not be used.

Conversely, O'Hagan and Stevens[12] stressed the fact that the dispersed nature of the data must be taken into account using suitable methods such as logarithmic transformation in order to obtain reliable results. They have therefore demonstrated that the use of the mean may lead to incorrect results if the cost dispersion is too great. We had to do a choice. Pros and cons of most widely used tests are presented below.

The Student t test is generally used to compare the quantitative features of two groups. This is based on the hypothesis that the mean of the feature studied tends towards a normal distribution for large samples (Central Limit Theorem CLT). The distribution of cost data is generally asymmetrical and right-skewed, making this convergence difficult and casting doubt on the validity of the results.

Non-parametric tests are generally preferred when data are not normally distributed. In particular, the Mann-Whitney test may be used to compare two distributions of any type. It is based on the principle of ranking instead of values: the sum of the ranks in each group is compared. The advantage of this method is that it removes the assumption of a normal distribution. It does not, however, compare mean costs, but cost distributions. In addition the results are sensitive to variances heterogeneity in the two groups which are compared.[13,14]

Many authors recommend logarithmic transformation of costs to take account of their dispersion.[15] Although the mean of a sample is the value of interest, this is greatly influenced by extreme individual values. Logarithmic transformation often allows a normal distribution to be obtained which can then be analysed using the Student t test. The test therefore compares geometrical, and no longer arithmetic, means. An equation does exist, however, which may be used in certain hypotheses,[16] to estimate the mean from this transformation: $\mu = \exp(lm + lv/2)$ where lm and lv are the mean and variance of the logarithm of the feature compared. Briggs[17] has shown that if the log normality hypothesis is verified the estimator μ is more effective than the arithmetic mean of the sample. If not, the arithmetic mean is still a convergent estimate whereas the log-normal estimate may produce biased results.

The non-parametric bootstrap[18] method is an alternative to the above. It is based on the principle that an unknown distribution F of a quantitative factor X can be estimated from the empirical distribution function (EDF). B samples of size n (starting sample size) obtained from the EDF can be generated by drawing n individuals randomly from the initial sample with replacement on B times. It has been shown that the empirical distribution of the B mean values in each of the samples, $\bar{x}_1^*,\ldots,\bar{x}_B^*$, provides an approximation of the distribution of the true mean of X.

In a cost comparison X may be taken as the difference in cost between the two treatment groups. The mean cost in each bootstrap sample is calculated for the first group, together with the mean cost in the second group and the difference between the two mean costs. If B is sufficiently large, the mean value of the B differences converges on the true difference in costs between the two groups. The percentile method may then be used to obtain a 95% confidence interval of cost difference and to determine whether or not this difference is significant.

This latter method has two major advantages: firstly it does not make any assumptions on cost distribution and secondly it calculates directly from the arithmetic means. The bootstrap samples, however, also result in dispersed cost distributions which therefore influence the mean values, calculated as they are derived from the initial sample. This dispersion is nevertheless reduced in these samples: the extreme costs are represented less through the drawing with replacement process. The results are therefore more robust than those obtained by the parametric Student t test and the CLT still applies.

The numbers in the two phases of the study were large enough and we therefore adopted the non-parametric bootstrap method to compare costs in the two treatment phases.

Calculating the cost-effectiveness ratio and analysis of uncertainty

The PREMISS study has the secondary objectives of calculating the effectiveness and efficiency of treatment. The effectiveness criterion chosen was the life expectancy 28 days after the onset of sepsis. This data point is however not directly available: only 28 day mortality is recorded in the case report forms. The life expectancy of survivors was therefore estimated from the McCabe score:

- Patients suffering from a short term fatal disease (1 year) were allocated a life expectancy of 0.5 years;

- The life expectancy of patients suffering from a long term fatal disease (5 years) was estimated to be 3 years;

- The life expectancy of patients without fatal comorbidities was calculated from the life expectancy of the general population published in the INSEE[19] tables grouped by age and sex for the year 2003 and divided by two.[20]

The uncertainty of the calculated cost-effectiveness ratio cannot be obtained using standard methods. It is difficult to obtain the distribution of a ratio.[21] The non-parametric bootstrap again allows us to resolve this problem.

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice
ART-5184/05

By taking B bootstrap samples containing couples (cost : years of life gaigned) for each patient drawn, B ratios can be constructed and represented in the (incremental cost : incremental effectiveness) plan. This graph allows the dispersion of the ratio to be understood. From the same bootstrap samples, an acceptability curve of the new treatment can also be constructed. This curve shows the probability that the treatment is efficient according to the decision-makers' willingness to pay. For a willingness to pay of λ, this probability is equal to the proportion of bootstrap samples in which the ratio calculated is less than λ. This curve provides another measurement of uncertainty linked with the overview estimate of the cost-effectiveness ratio.[22]

RESULTS

Description of the population

509 and 587 patients were recruited into the "Before" and "After" phases respectively, between September 2002 and January 2003 for the former group and between January 2003 and February 2004 for the latter group. The major characteristics of these patients are summarised in the table below:

Table 2: Initial characteristics of PREMISS patients

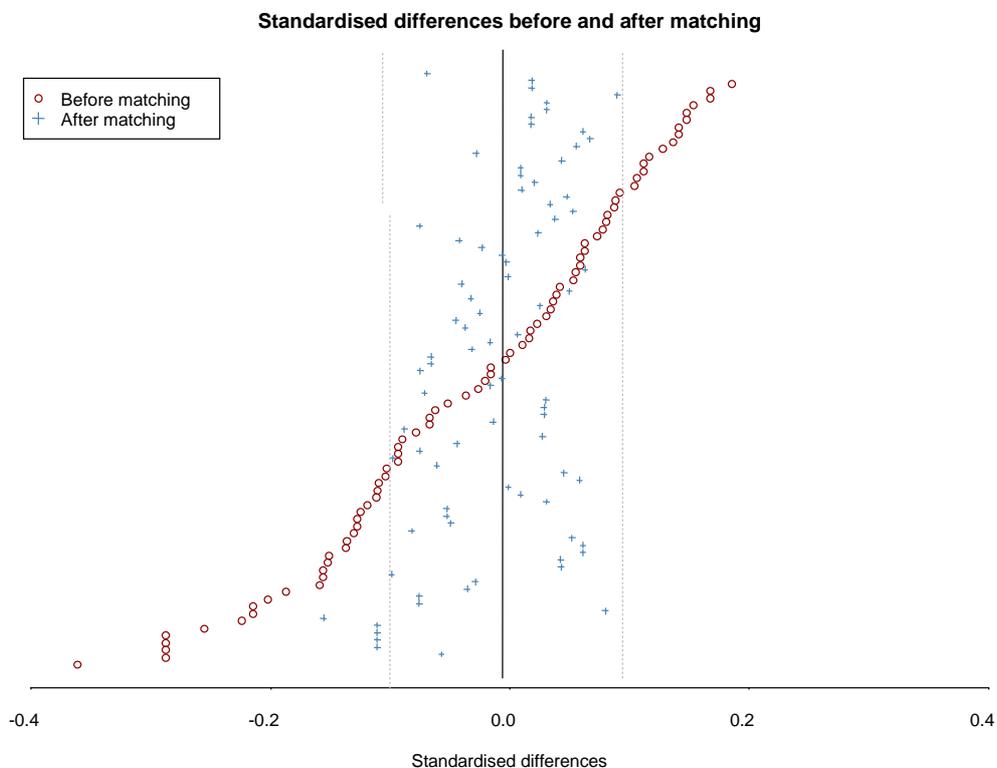| Initial characteristics	Description of sample[*]
Age	60.79 years (± 16.27)
Men	62.04%
Admission to intensive care through internal transfer	40.42%
Medical stay	71.72%
Septic shock	93.70%
SAPS II on admission to intensive care	56.56 (± 18.56)
SAPS II on entry to the study	58.15 (± 17.31)
LODS[**] on entry to the study	8.76 (± 3.40)
Respiratory insufficiency	14.23%
Cardiac insufficiency	12.86%
Pulmonary infection	46.81%
Intra-abdominal infection	24.91%
Number of organ failures	3.59 (± 0.99)

*: mean (± standard deviation) for quantitative variables: proportion for qualitative variables

**: Logistic Organ Dysfunction System

It can be seen from the magnitude of the standardised differences (not shown), that even though the patients recruited in the two phases had similar severity indices (SAPS II and LODS), they did not have the same type of severity. More patients in the "After" group had respiratory failure whereas patients in the "Before" group were more severe in neurological terms. In addition, patients recruited for DA treatment were younger and less liable to die within the year. More patients in the "Before" phase were admitted through internal transfer into the intensive care department. More of them were also suffering from endo-cardiovascular and urinary tract infections.

Matching using the propensity scores produced a sample of 840 patients, 420 in each phase. After the matching almost all of the standardised differences were below the 10% threshold (Figure 1): we therefore gained considerably in comparability between the two patient groups. The proportion of patients over 80 years old remained higher however in the "Before" phase, as did the proportion of unventilated patients. We will conduct further work on this sample in the subsequent analyses.

Figure 1: Changes in standardised differences before and after matching

11

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

Comparison of complete hospitalisation costs

The mean complete hospitalisation costs were 36,717.03 € in the "Before" phase and 47,870.09 € in the "After" phase, i.e. an additional cost produced due to DA treatment of 11,153.06 €. 10,000 bootstrap samples were generated in order to determine the significance of this additional cost. The difference in mean cost for each phase was calculated from each sample. The graph below shows a histogram of these differences for the 10,000 samples. The 95% confidence interval of the additional cost was [6,600.68 ; 15,709.45]. This interval does not contain the value zero and we can therefore conclude that the patients receiving DA treatment incurred higher costs than those treated conventionally.

Figure 2: Histogram of mean differences in costs between the "After" phase and the "Before" phase on the 10,000 bootstrap samples
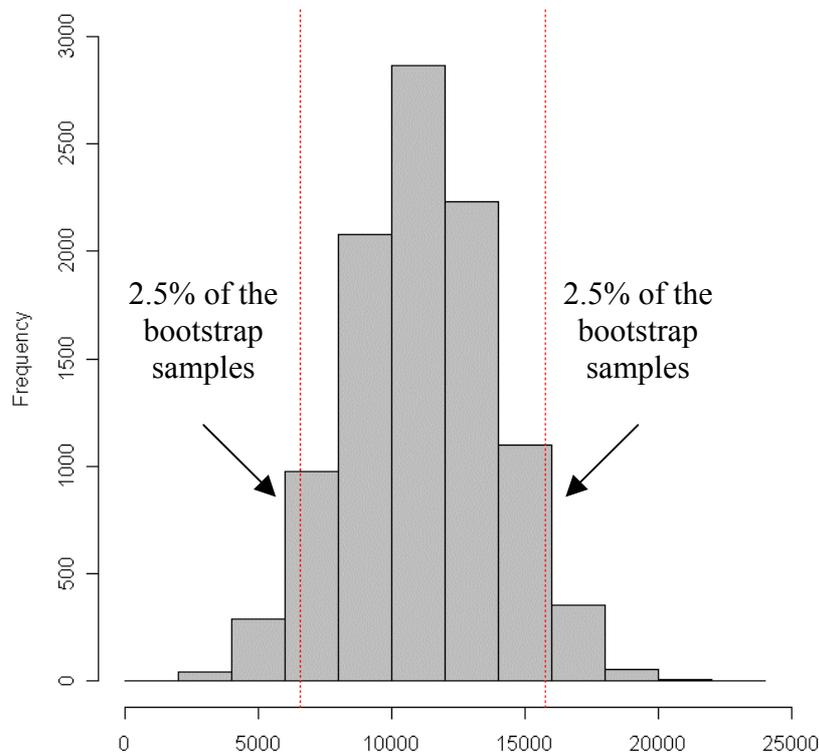
The additional cost is due to several reasons. Firstly the acquisition cost of DA reached 6,716.92 € in average, and therefore does not explain all of the additional cost found. Additional costs associated with the use of DA are incurred reflected by the increase in the Omega score and in the length of stay in intensive care. The mean Omega score rose from 372.84 in the "Before" phase to 426.60 in the "After" phase, i.e. a 14.45% increase in activity in the intensive care units (p = 0.0010, Mann-Whitney test).

Cost-effectiveness evaluation of drotrecogin alfa compared to standard care in the treatment of severe sepsis in real practice

ART-5184/05

The length of stay in intensive care was also significantly longer in patients included in the "After" phase (p = 0.0018, Mann-Whitney test). The patients spent an average of 24.42 days in the intensive care department compared to 21.27 days for patients recruited in the "Before" phase. These two indicators, which are positively associated with complete hospitalisation costs in the CUB-Rea equation, explain part of the additional cost found in the "After" phase.

Construction of the effectiveness criterion: life expectancy at 28 days

There were fewer deaths 28 days after the onset of severe sepsis in the "After" phase. The absolute difference in deaths compared to the "Before" phase was 3.29%. This difference was not, however, significant (p = 34.54%, Fisher's exact test).

Table 3: Status on D28 and recruitment phase – matched sample

	Alive	Dead	Total
"Before" phase	260	155(37.35%)	415
"After" phase	271	140(34.06%)	411
Total	531	295(35.71%)	826

The life expectancy of survivors was calculated from the age and sex distributions by comorbidity. The mean life expectancy of survivors who received DA was 10.13 years (± 6.82) compared to 9.79 years (± 6.86) for patients treated conventionally.

Combining the survival rate with the life expectancy of survivors produced a mean life expectancy of 6.68 years (± 7.33) for patients in the "After" phase and 6.13 years (± 7.20) for patients in the "Before" phase. This difference (0.55 years) was also not significant (p = 2.33%, Mann-Whitney test).

Cost-effectiveness ratio and assessment of uncertainty

Although the effectiveness criterion was not significantly different between the two phases, it was entirely legitimate to perform a cost-effectiveness analysis. Firstly, it is well known that failure of an effect to be significant is not proof of lack of effect (" The absence of evidence is not evidence of absence"). In our case, the study numbers are too small to obtain a significant fall in mortality through the use of DA. We concluded from the death rates reported in the PROWESS trial that 600 patients would have been needed per group (instead of 420) in order to find a significant difference in mortality between the two phases.

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

In addition, the purpose of a cost-effectiveness analysis is to estimate one parameter, the incremental ratio, with an appropriate representation of uncertainty.[23] It is therefore the combined density of the difference in cost and difference in effectiveness which must be analysed and not each density separately.

For a cost difference of 11,153.06 € and a difference in life expectancy of 0.55 years, the cost-effectiveness ratio $R = \dfrac{11153.06}{0.55} = 20278.29$ € per life year gained. The uncertainty of this ratio was measured in two different ways: by analysing the scatterplot formed by 10,000 bootstrap replicates of the ratio and by constructing the acceptability curve for DA. The 10,000 couples (Δcost-Δeffectiveness) obtained from the bootstrap samples are shown in figure 3.

Figure 3: 10,000 couples (incremental cost; incremental effectiveness) obtained from the bootstrap samples
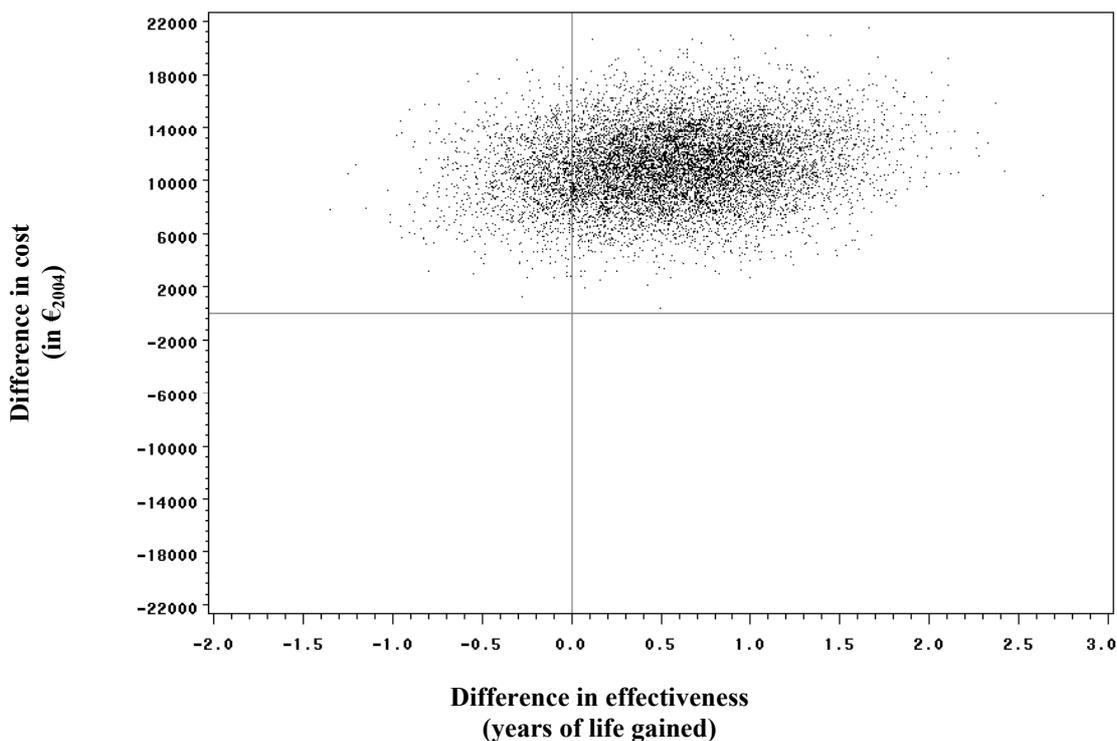

The couples (Δcost; Δefficacy) covered two quadrants of the (cost-effectiveness) plan. In the North-East quadrant treatment of severe sepsis with DA is both more effective and more expensive that conventional treatment. Conversely, in the North-West quadrant, DA is still more expensive but it is now less effective than conventional treatment. The number of samples in each quadrant is shown in table 4.

Cost-effectiveness evaluation of drotrecogin alfa compared to standard care in the treatment of severe sepsis in real practice

ART-5184/05

Table 4: Distribution of bootstrap samples by positivity of the ratio

	R < 0 (North-West)		R > 0 (North-East)	
	Number of bootstrap samples	Frequency	Number of bootstrap samples	Frequency
	1408	14.08%	8592	85.92%

The acceptability curve provided another way of interpreting uncertainty of the ratio. For each willingness to pay λ, it provided the probability that DA was acceptable (i.e. the probability that the ratio was below the willingness to pay).

Figure 4: Acceptability curve for DA
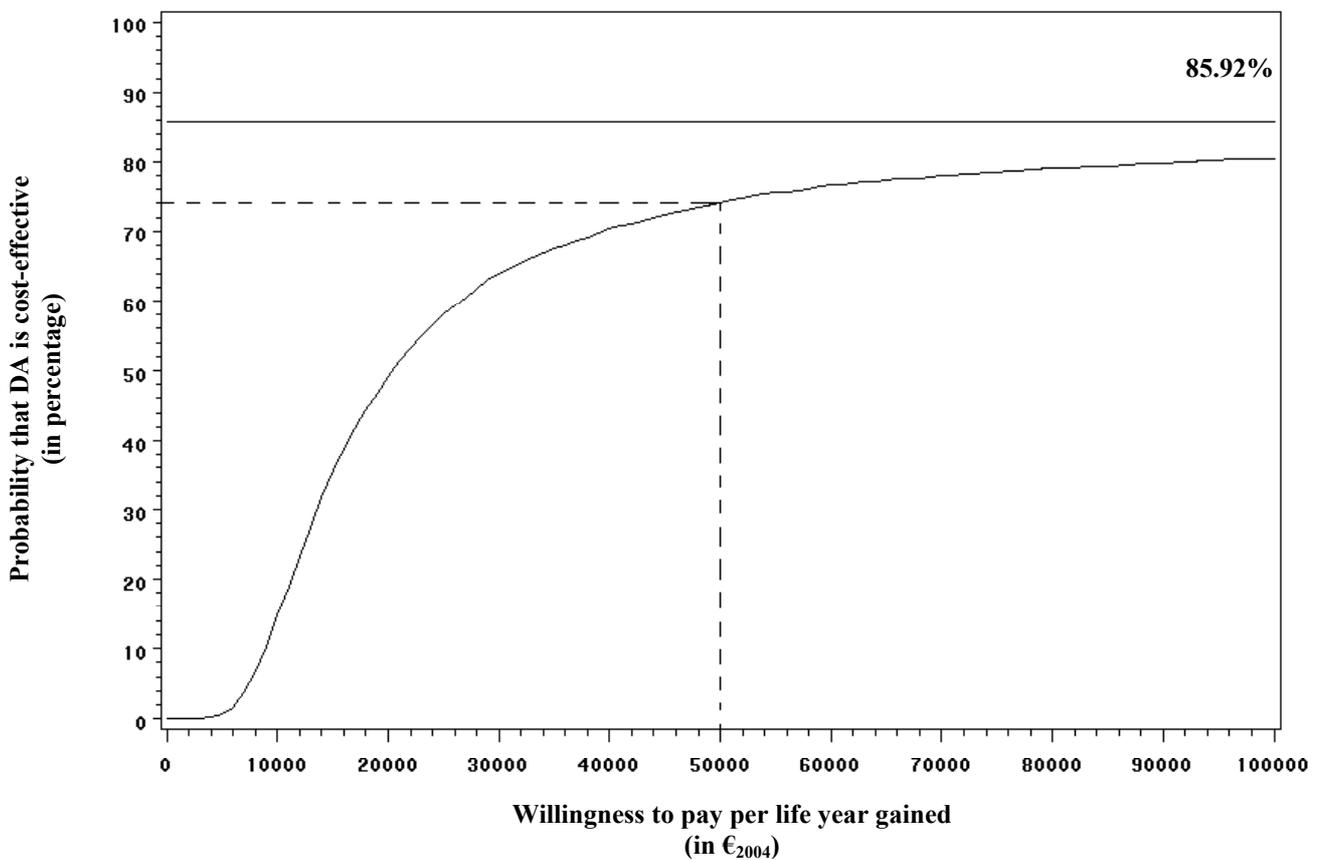
The asymptote of the acceptability curve was not equal to 1. This was due simply to the fact that bootstrap samples existed in which DA was less effective than conventional treatment. The asymptote was thus equal to the proportion of bootstrap samples for which the number of life years gained was greater in the "After" phase than in the "Before" phase (cf. table 4).

Determination of an acceptability threshold for an intervention is still subject to controversy. NICE (National Institute for Clinical Excellence) is understood to use a threshold of between £20,000 and £30,000 GBP (between 29 000 and 44 000 €) per QALY, although its decisions are based on more complex rules.[24] Another figure widely used comes from the United States:[25] 50,000 € per life year gained. The probability that DA is acceptable for this willingness to pay is 74.12%.

DISCUSSION

Stengths and weaknesses of observational studies

The choice of "Before"/"After" design of the PREMISS study was influenced by the fact that patient randomisation was unethical, as the intervention being assessed had already been deemed to be effective from the PROWESS[2] clinical trial. In addition, as the aim of the study was principally to estimate the costs of patients for whom DA was indicated in real practice, it was essential that the study interfered as little as possible with intensive care physicians' practices.[26] The internal validity (the ability of the study to provide reproducible results in other studies) was therefore given preference over internal validity (the ability of the study to provide results which truly reflect the variables measured): rather than reproducing the results of PROWESS we decided to ensure that these could be generalised in the setting of everyday practice for French hospital intensive care physicians.

The price that must be paid to meet these objectives is considerable. Comparing patients from the two phases revealed significant recruitment bias. Whereas the intensive care physicians included more than 500 patients in approximately 5 months into the "Before" phase, it took 20 months to recruit the same number in the "After" phase. This difference in inclusion rate could be explained by lack of motivation on the part of the investigators although it is more likely to be due to a selection effect of patients who were to receive DA. The patients included in the "After" phase were therefore younger and had fewer comorbidities although the SAPS II and LODS scores were not significantly different between the two treatment phases.

The propensity score matching

There are many methods used to measure the effect of a treatment in observational studies when it is not possible to envisage a clinical trial. The main aim of these is to take account of bias due to non-random allocation of treatments to the people who are included.[27]

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

The method we chose was the propensity score method: in order to minimise recruitment differences, a sub-sample of patients who were comparable between the two phases was constructed by matching based on the PS. Whilst this method allows less biased comparisons to be made between the two phases it also results in a significant number of patients being excluded (the sample size was reduced by 23%) which influences the accuracy of the estimates. In addition, only bias with respect to the measured variables are liable to be corrected by this method. It is likely that other, unmeasured, variables were different between the two phases and may have interfere with the results.[27]

This problem may be resolved with an alternative, instrumental variable method: this involves finding a variable (or instrument) which is associated with allocation of treatment (the phase) but not with survival or costs (outputs). As this is a very strong hypothesis, it is difficult to find a suitable variable. We did not have any satisfactory candidates in the PREMISS study. Furthermore, 46 covariates were used to calculate the propensity scores. Most were prognostic for severity of the patients and it is thus unlikely that other covariates which were important to the distribution of patients within the two phases were omitted.

As we see in Figure 1, the sample matched for propensity scores is balanced for almost all of the initial characteristics. We found residual bias, however, for two of these: there were more people over 80 years old in the "Before" phase than in the "After" phase, whereas more patients in the "After" phase were being ventilated at the time they were entered into the study. The method of matching by propensity scores therefore helped to reduce, but did not completely remove, bias.

The residual bias, particularly with respect to age, may have influenced the results of the cost-effectiveness analysis. Life expectancy was based amongst other things on the mean age of the survivors not suffering from any fatal comorbidities. Patients in the "After" phase were slightly younger and their mean life expectancy was therefore higher. The incremental effectiveness was therefore greater than it would have been if all bias had been removed. However, we should note that even if age is associated with a standardised difference of more than 10%, their was no significant difference in mean age. We considered this residual bias was negligible.

Use of a bayesian context to interpret the acceptability curve

The acceptability curves show the probability that the new treatment is cost-effective in terms of the decision-makers' willingness to pay and provide information about the uncertainty of the ratio calculated.[28]

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

However, many authors stress that such a probability based interpretation may only be made in a bayesian context,[29,30,31] in which the parameters are random variables. In a frequentist context, as was used in the PREMISS study, the parameters are fixed quantities: the treatment is either cost-effective or it is not and no probability can be linked to this outcome. However, the method we used to construct the acceptability curve is valid regardless of context (non-parametric bootstrap[32]) and the bayesian context provides an intuitive interpretation of uncertainty of the calculated cost-effectiveness ratio for a decision maker: depending on the amount which the society is willing to pay to handle the health problem in question and the risk it deems tolerable in terms of the uncertainty of the cost-effective nature of the new treatment, the decision-maker may chose whether or not to agree to fund the treatment. We therefore decided to retain this interpretation in order to improve understanding of our results even though it introduces some heterogeneity into them.

Apart from this, synthesizing the evidence on the clinical and economical impacts of DA in the treatment of severe sepsis from disparate sources would have some advantages. Instead of only considering observational data from PREMISS it would be possible to combine both observational data and PROWESS clinical trial results in a bayesian model. This would increase the accuracy of PREMISS parameter estimates. The results would then be expressed in the form of posterior distributions of the parameters, summarising all of the available information. The acceptability curve in terms of probability of efficiency could then be interpreted in its true sense.

CONCLUSION

The major observation of the PREMISS study is on the financial impact of DA. Whilst adopting this treatment carries a cost, this is not restricted to the acquisition cost of the drug. The mean acquisition cost can be estimated as being slightly more than 6,700 €. The total hospitalisation costs estimated from regression (DA included) ultimately rise to close to 11,150 € in the "After" phase. The difference is explained by the change in health care organisation. Patients in the "After" phase spend 3 days more in intensive care. The burden of care measured by the Omega field is also significantly greater in the "After" phase.

An overview of the economic and clinical results of the PREMISS study from a cost-effectiveness evaluation shows a ratio in the region of 20,300 € per life year gained. The probability of DA being cost effective at a willingness to pay threshold of 50,000 € per life year gained is 74%.

Cost-effectiveness evaluation of drotrecogin alfa compared to
standard care in the treatment of severe sepsis in real practice

ART-5184/05

These are more pessimistic results than those estimated by modelling in France,[9] which found an incremental cost-effectiveness ratio of approximately 12,000 € per life year gained with an 85% probability that treatment would be acceptable at the threshold of 50,000 $ per QALY. These differences may be explained by the fact that the DA evaluation models were all based on the PROWESS study which allowed the effect of treatment to be estimated with greater power and on more patients.

BIBLIOGRAPHY

[1] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

[2] Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, Steingrub JS, Garber GE, Helterbrand JD, Ely EW, Fisher CJ Jr; Recombinant human protein C Worldwide Evaluation in Severe Sepsis (PROWESS) study group. Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med. 2001 Mar 8;344(10):699-709.

[3] Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol. 2001 Apr;54(4):387-98.

[4] D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomised control group. Stat Med. 1998 Oct 15;17(19):2265-81.

[5] Yanovitsky I, Zanutto E, Hornik R. Estimating causal effects of public health education campaigns using propensity score methodology. Evaluation and program planning 2005;28:209-220

[6] Mayo Clinic College of Medicine. http://mayoresearch.mayo.edu/mayo/research/biostat/upload/match.sas

[7] Aegerter P, Auvert B, Buonamico G, Sznadjer M, Beauchet A, Guidet B, Le Gall JR, Cub-Réa. Organisation and quality control of a clinical database on intensive care medicine in central and suburban Paris. Rev Epidemiol Sante Publique 1998 ;46(3):226-37.

[8] Sznajder M, Aegerter P, Launois R, Merliere Y, Guidet B, CubRea. A cost-effectiveness analysis of stays in intensive care units. Intensive Care Med. 2001 Jan;27(1):146-53.

[9] Riou França L, Launois R,. Guidet B, Le Lay K, Aegerter P, Bouhassira M, Meshaka P Cost-effectiveness of drotrecogin alfa (activated) in the treatment of severe sepsis in France. Submitted for publication.

[10] Site de l'INSEE (National Institut for Statistics and Economic Studies). http://www.insee.fr/fr/indicateur/achatfranc.htm.

[11] SG Thomson, JA Barber. How should cost data in pragmatic randomised trials be analysed ? British Medical Journal 320, 1197-1200 (2000).

[12] A O'Hagan, JW Stevens. Assessing and Comparing Costs: How Robust are the bootstrap and the methods based on asymptotic normality ? Health Economics 12, 33-49 (2003).

[13] DW Zimmerman. Statistical Significance Levels of Nonparametric Tests Biased by Heterogeneous Variances of Treatment Groups. Journal of General Psychology, Oct 2000.

Cost-effectiveness evaluation of drotrecogin alfa compared to standard care in the treatment of severe sepsis in real practice

ART-5184/05

[14] A Desgagné, AM Castilloux, JF Angers, J LeLorier. The Use of the Bootstrap Statistical Method for the Pharmacoeconomic Cost Analysis of Skewed Data. Pharmacoeconomics 1998. 13, 487-497.

[15] XH Zhou, CA Melfi, SL Hui. Methods for comparison of cost data. Ann Intern Med 1997; 127(8 Part 2): 752-756.

[16] Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ. 2001 Jul;20(4):461-94.

[17] A Briggs, R Nixon, S Dixon, S Thomson. Parametric modelling of cost data: some simulation evidence. Health Economics (in press).

[18] Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman and Hall 1993.

[19] Site de l'Institut National des Etudes Démographiques (INED). http://www.ined.fr/population-en-chiffres/france/index.html.

[20] AA Quartin, RM Schrein, DH Kett, PN Peduzzi. Magnitude and duration of the effect of sepsis on survival. Department of Veterans Affairs Systemic Sepsis Co-operative Studies Group. JAMA 1997 Apr 2;277(13):1058-63

[21] Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ 2002;11:415-430.

[22] BA Vant Hout, MJ Al, GS Gordon, FFH Rutten. Costs, effects and C/E-ratios alongside a clinical trial. Health Economics 3, 309-319 (1994)

[23] AH Briggs, BJ O'Brien. The death of cost-minimisation analysis ? Health Econ 2001;10:179-184

[24] N Devlin, D Parkin. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ. 2004 May;13(5):437-52

[25] Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. Med Decis Making. 2000 Jul-Sep;20(3):332-42.

[26] A Hartz, JL Marsh. Methodologic Issues in Observational Studies. Clinical Orthopaedics and Related Research, 2003

[27] OH Klungel, EP Martens, BM Psaty, DE Grobbee, SD Sullivan, BH Stricker, HGM Leufkens, A de Boer. Methods to assess intended effects of drug treatment in observational studies are reviewed. Journal of Clinical Epidemiology 57 (2004) 1223-1231

[28] HA Glick, AH Briggs, D Polsky. Quantifying stochastic uncertainty and presenting results of cost-effectiveness analyses. Expert Rev. Pharmacoeconomics Outcomes Res. 1(1), 2001.

[29] A O'Hagan, JW Stevens, J Montmartin. Inference for the Cost-Effectiveness Acceptability Curve and Cost-Effectiveness Ratio. Pharmacoeconomics 2000 Apr. 17(4) 339-349.

[30] AH Briggs. A bayesian approach to stochastic cost-effectiveness analysis. Health Economics 8(3), 257-261 (1999)

[31] DF Heitjan, AJ Moskowitz, W Whang. Bayesian estimation of cost-effectiveness ratios from clinical trials. Health Economics 8, 191-201 (1999)

[32] M Löthgren, N Zethraeus. Definition, Interpretation and calculation of cost-effectiveness acceptability curves. Health Economics 9, 623-630 (2000)