

## Distribution of the Probability of Survival Is a Strategic Issue for Randomized Trials in Critically Ill Patients

Bruno Riou, M.D., Ph.D.,\* Paul Landais, M.D., Ph.D.,† Benoît Vivien, M.D.,‡ Philippe Stell, M.D.,§ Iheb Labbene, M.D.,§ Pierre Carli, M.D.||

**Background:** Many randomized clinical trials in trauma have failed to demonstrate a significant improvement in survival rate. Using a trauma patient database, we simulated what could happen in a trial designed to improve survival rate in this setting.

**Methods:** The predicted probability of survival was assessed using the TRISS methodology in 350 severely injured trauma patients. Using this probability of survival, the authors simulated the effects of a drug that may increase the probability of survival by 10–50% and calculated the number of patients to be included in a trial, assuming  $\alpha = 0.05$  and  $\beta = 0.10$  by using the percentage of survivors or the individual probability of survival. Other distributions (Gaussian, J shape, uniform) of the probability of survival were also simulated and tested.

**Results:** The distribution of the probability of survival was bimodal with two peaks ( $< 0.10$  and  $> 0.90$ ). There were major discrepancies between the number of patients to be included when considering the percentage of survivors or the individual value of the probability of survival: 63,202 versus 2,848 if the drug increases the probability of survival by 20%. This discrepancy also occurred in other types of distribution (uniform, J shape) but to a lesser degree, whereas it was very limited in a Gaussian distribution.

**Conclusions:** The bimodal distribution of the probability of survival in trauma patients has major consequences on hypothesis testing, leading to overestimation of the power. This statistical pitfall may also occur in other critically ill patients.

TRAUMA is the leading cause of death in young people and the leading cause of lost years of life in industrialized countries.<sup>1</sup> Annual cost secondary to trauma (acute medical care, deaths, disability, lost wages, and taxes) in the United States has been estimated to exceed \$150 billion.<sup>2</sup> In the past, considerable improvement in the care of trauma patients has been accomplished by organization of prehospital care, development of regionalized trauma systems, and improvement in the assessment of

trauma patients by modern imaging techniques. More recently, randomized trials in trauma patients testing the hypothesis that a pharmacologic intervention could improve survival have appeared, indicating that traumatology now deals with the same methodologic problems as other specialties.

Many of the large prospective randomized studies recently published provided negative and disappointing results.<sup>3–6</sup> For example, although tirilazad has been demonstrated to decrease the deleterious consequences of brain trauma, it failed to decrease mortality in head trauma patients.<sup>6</sup> More recently, two large trials on the use of a hemoglobin solution (Diaspirin Cross-Linked Hemoglobin; Baxter Hemoglobin Therapeutics, Boulder, CO) in the early phase of trauma resuscitation, one in the United States and the other in Europe, were prematurely terminated. The US trial was discontinued early because of an unfavorable imbalance in between-group mortality.<sup>7</sup> The European trial was discontinued as a consequence of a failure to demonstrate evidence of benefit. The underlying reasons for early termination of these two trials have been caused, at least partially, by inappropriate study design leading to bias, inclusion of wrong patients (early deaths), and/or the impossibility of demonstrating any beneficial effect of the drug on mortality. These observations raised several issues: (1) it may not be ethically acceptable to include patients in trials that are not able to conclude anything; (2) it is possible that the development of interesting drugs in trauma patients could be prematurely stopped; (3) a very large amount of money has been spent in these negative or aborted trials, whereas resources allocated to trauma research are limited as compared with cancer or cardiovascular diseases<sup>1</sup>; (4) it seems obvious that most of the progress that could be now expected in trauma care will come from large randomized trials testing pharmacologic interventions, and, thus, methodologic pitfalls could prevent significant progress in the future.

Because two of the authors (B.R. and P.C.) were involved in the analysis of pitfalls that led to the termination of the randomized trials with the hemoglobin solution in trauma, we realized that several important methodologic problems (inclusion and exclusion criteria, the determination of the number of patients to be included, and the assumptions used for this calculation) had not been appropriately considered because the main characteristics of trauma as a disease were not completely understood. In the current study, we used a

\* Professor, Department of Anesthesiology and Critical Care, and Chairman, Department of Emergency Medicine and Surgery, ‡ Assistant Professor, § Fellow, Department of Anesthesiology and Critical Care, Centre Hospitalo-Universitaire Pitié-Salpêtrière. † Associate Professor, Department of Biostatistics, || Professor and Chairman, Department of Anesthesiology and Critical Care, Centre Hospitalo-Universitaire Necker-Enfants Malades.

Received from the Department of Anesthesiology and Critical Care and Emergency, Centre Hospitalo-Universitaire Pitié-Salpêtrière, Paris, France; and the Departments of Biostatistics and Anesthesiology and Critical Care, Centre Hospitalo-Universitaire Necker-Enfants Malades, Assistance Publique-Hôpitaux de Paris, Paris, France. Submitted for publication June 2, 1999. Accepted for publication January 9, 2001. Supported by Baxter, Hemoglobin Therapeutics, Boulder, Colorado (Drs. Riou, Vivien, and Carli). Presented in part at the annual meeting of the European Society of Anesthesiology, Amsterdam, The Netherlands, May 29–June 2, 1999.

Address reprint requests to Dr. Riou: Département d'Anesthésie-réanimation, CHU Pitié-Salpêtrière, 47 Boulevard de l'Hôpital, 75651 Paris Cedex 13, France. Address electronic mail to: bruno.riou@psl.ap-hop-paris.fr. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

database of severely traumatized patients to simulate what could happen in a randomized trial intended to demonstrate an improvement in the survival rate. We were particularly interested in studying the distribution of the probability of survival (Ps) in trauma patients and its consequences on hypothesis testing. Our main objective was to provide information that could improve the efficiency of randomized trials in trauma and/or facilitate strategic decision-making for caregivers, health authorities, and the pharmaceutical industry. However, we also suggest that the methodologic issues raised in our study could also explain some failures in randomized trials in other domains, such as therapeutics evaluation in critically ill patients.

## Patients and Methods

### *Trauma Population and Probability of Survival*

This part of the study was designed to obtain core information on the distribution of the *a priori* Ps in severely injured trauma patients. The purpose was to provide evidence that a population of trauma patients is characterized by a marked heterogeneity in the prognosis. Indeed, we made the hypothesis that, when we determine the number of patients to be included in a randomized trial using the percentage of survivors in the population, we did not take into account the characteristics of the distribution of the Ps, mainly the marked heterogeneity in Ps.

The case histories of 350 consecutive blunt trauma patients were analyzed. All were cared for by a mobile intensive care unit (Paris SAMU system), and the severity of trauma was considered as high enough by the prehospital team to warrant direct admission of these patients to the emergency department of a level I trauma center. The SAMU system has been described elsewhere.<sup>8</sup> The on-scene triage was based on the clinical assessment of the trauma patient.<sup>9</sup> For each patient, the following data were recorded during the prehospital phase: age, sex, initial systolic arterial pressure (SAP), heart rate, respiratory rate, Glasgow Coma Score, occurrence of prehospital cardiac arrest, and SAP and heart rate at the arrival in the hospital. The following scores were determined: Abbreviated Injury Scale,<sup>10</sup> Injury Severity Score,<sup>11</sup> and Revised Trauma Score (RTS).<sup>12</sup> The Ps was then calculated using the TRISS methodology.<sup>13,14</sup> Thus, in the following text, Ps is the TRISS score when applied to trauma patients.

The observed survival (*i.e.*, survival until hospital discharge) was compared with the expected survival. The expected percentage of survivors was calculated as follows:

$$S\% = \sum Ps \cdot 100 \cdot n^{-1}$$

where n is the number of patients. As previously described, to compare our population with that of the

Major Trauma Outcome Study (MTOS),<sup>14</sup> we calculated the M, W, and Z scores. An M value less than 0.88 indicates a disparity in the severity match between the study group and the MTOS group.<sup>15</sup> The W score is the number of survivors more or less than would be expected from the MTOS prediction for 100 patients.<sup>15</sup> A Z score between  $-1.96$  and  $+1.96$  indicates no significant difference ( $P > 0.05$ ) between the actual number of survivors and that expected. Because our population was selected by the SAMU system and was expected to include more severe trauma patients with a low Ps, we also calculated the standardized W score, which represents the W score that would have been observed if the case mix of severity was identical to that of the MTOS group.<sup>15</sup> Accordingly, a standardized Z score was calculated. A Z score between  $-1.96$  and  $+1.96$  indicates no significant difference ( $P > 0.05$ ) between the number of survivors and that expected, if the case mix severity was identical to that of the MTOS.<sup>15</sup>

### *Simulation of a Therapeutic Trial*

In the second part of the study, we used our trauma population with the individual values of *a priori* Ps and its characteristic bimodal distribution of Ps to simulate a trial. This simulation does not take into account the fact that the number of survivors may differ from the expected number of survivors.

We tested the effects of a theoretical treatment that could be able to increase Ps by 10, 20, 30, or 50%. Because Ps must be between 0 and 1, we applied the following rule: for example, if a drug is thought to increase Ps by 10%, when Ps was below 0.50, it was increased by 10% of Ps, and when Ps was greater or equal to 0.50, it was increased by 10% of  $1 - Ps$ .

To calculate the number of patients to include in a randomized trial comparing survival in two groups (control *vs.* treated groups), four numbers are required: the percentage of survivor in the control group, the expected percentage of survivor in the treated group, and the values of type I error ( $\alpha$ ) and type II error ( $\beta$ ). We calculated this number of patients to include in a trial, assuming  $\alpha = 0.05$  and  $\beta = 0.10$ . We first calculated this number using the global percentage of survivors in our population, as usually performed in trials. For example, if a drug is thought to increase Ps by 10% and if the percentage of survivors was 75.0% in the control group, we expected that the percentage of survivors should be 77.5% in the treated group. Second, we calculated the number of patients to include in a trial by using the individual Ps. For this purpose, we applied the rule explained above to all individual Ps values, then we summarized all of these modified Ps values to obtain the percentage of survivors in the treated group, as described above. Nevertheless, a moderate discrepancy between the number of patients to be included using the global percentage of survivors or the individual Ps was

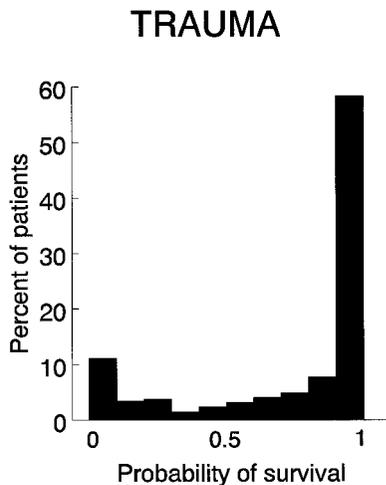


Fig. 1. Histograms showing the distribution of the probability of survival in the studied group of trauma patients (N = 350).

expected (mathematically, multiplying the mean Ps by x percent is not equivalent to taking the mean of the multiplication of individual Ps by x percent, except in the case where all patients have exactly the same Ps). Therefore, we also randomly generated a population of 1,000 theoretical patients whose distribution of the Ps was Gaussian and whose mean percentage of survivors was 75.0%, as observed in our trauma population.

#### Selection of a Subpopulation of Trauma Patients

In the third part of the study, because the bimodal distribution of the Ps in trauma patients (fig. 1) implies a marked heterogeneity, we tried to reduce this heterogeneity by selecting an appropriate subpopulation. We considered three groups of trauma patients: those with a very high Ps ( $Ps \geq 0.90$ ), those with a very low Ps ( $Ps \leq 0.10$ ), and those with an intermediate Ps ( $0.10 < Ps < 0.90$ ). We used all prehospital variables available (age, sex, SAP, heart rate, respiratory rate, RTS, Glasgow Coma Score, occurrence of cardiac arrest) to define the best criteria enabling to select trauma patients with a low Ps, a high Ps, and an intermediate Ps. In these three subgroups, we calculated the number of patients to be included in a trial assuming a type I error of 0.05 and a type II error of 0.10, as described above.

#### Simulation of Other Distributions

In the fourth part of the study, to assess the effects of other possible distributions of the Ps that could be encountered in nontraumatized critically ill patients, we randomly generated two other populations of 1,000 theoretical patients whose distribution of the Ps were (1) uniform and (2) J shape. Indeed, a J-shape distribution has been observed in other critically ill patients such as patients with septic shock.<sup>16</sup> Because a uniform distribution implies a survival rate of 50%, we chose a J-shape distribution with a survival rate close to 50%.

#### Statistical Analysis

Data are presented as mean  $\pm$  SD. All *P* values were two-sided. The NCSS statistical program (Release 6.0, Statistical Solution Ltd, Cork, Ireland) was used for all statistical analyses.

To calculate the number of patients to be included, we applied the Casagrande and Pike<sup>17</sup> method assuming  $\alpha = 0.05$  and  $\beta = 0.10$  and using the percentage of survivors in control conditions and the percentage of survivors expected because of the drug effect (percentages were rounded to one decimal) and based on a two-sided analysis.

Identification of trauma patients in the high- and low-Ps groups were performed separately. For each procedure, the following steps were followed: first, all criteria were analyzed using an univariate analysis (unpaired Student *t* test or Fisher exact method). Second, for noncategorical variables that indicated a significant difference between groups, a cutoff value was determined using the receiver operating characteristics curve. The cutoff value was chosen as the value associated with the maximum value of sensitivity plus specificity. Then a stepwise logistic regression (*P* value at entry = 0.05) was applied to determine the appropriate criteria, using only categorical variables. The odds ratio and its 95% confidence interval were calculated.

We used the random function of Excel 5.0 software (Microsoft Corporation, Seattle, WA) to randomly generate populations of 1,000 theoretical patients. By taking random numbers (RN) between 0 and 1, we generated a uniform distribution (mean Ps, 0.50). By taking  $RN^{1.8}$ , we generated a J-shape distribution (mean Ps, 0.41). Lastly, by taking the mean of 10 random numbers + 0.25, we generated a Gaussian distribution (mean Ps, 0.75). Indeed, according to the central limit theorem, the sampling distribution tends toward a Gaussian distribution, whatever the initial distribution.<sup>18</sup>

## Results

In this population (N = 350), mean age was  $36 \pm 16$  yr (range, 15–89 yr), and 253 (72%) patients were male. The mean Injury Severity Score was  $28 \pm 16$  (median, 29), mean RTS was  $6.16 \pm 2.27$  (median, 7.11), and the mean TRISS was  $0.750 \pm 0.341$  (median, 0.940). Figure 1 depicts the distribution of Ps values in this population and reveals a bimodal distribution. The number of survivors (n = 277, 79%) was higher than that expected (n = 262, 75%). The W score was +5.1% (Z = 12.25, *P* < 0.001). As expected, the value of M = 0.64 indicated a disparity in the severity match between the study group and the MTOS group.<sup>15</sup> This disparity was mainly caused by a lower proportion of patients with a very high Ps (TRISS > 0.95, 46.9 vs. 82.8%) and a higher proportion of patients with a low Ps (TRISS < 0.25, 16.9 vs. 3.6%).

**Table 1. Trauma Population: Discrepancies between the Number of Patients (N) To Include in a Trial ( $\alpha = 0.05, \beta = 0.10$ ) When Considering the Global Percentage of Survivors (Global) or the Individual Value (Individual) of the Probability of Survival (Ps)**

Trauma Population	Control	Expected Drug Effect			
		+10%	+20%	+30%	+50%
<b>Global</b>					
% of survivors	75.0	77.5	80.0	82.5	87.5
N	—	12,018	2,848	1,192	374
<b>Individual</b>					
% of survivors	75.0	75.5	76.0	78.0	79.9
N	—	312,400	63,202	6,730	3,080

Several hypotheses were tested regarding the expected effect of the drug on Ps (+10, +20, +30, and +50%). See text and figure 1 for explanation.

The standardized *W* score was +0.4% (*Z* score = 0.15, nonsignificant), indicating that it was not significantly different from that of the MTOS. In fact, we observed that only trauma patients with a low Ps (*i.e.*, TRISS < 0.25) experienced a higher number of survivors than that expected by MTOS (*W* = 25.7%, *Z* = 57.1, *P* < 0.001).

*Simulation of a Therapeutic Trial*

As shown in table 1, there were major discrepancies between the number of patients to be included in a trial when considering the global percentage of survivors or the individual value of Ps. As shown in table 2, this phenomenon was very limited in the theoretical population with a Gaussian distribution of Ps (fig. 2), demonstrating that it was mainly related to the marked heterogeneity of the trauma population and not to the mode of calculation (global percentage of survivors *versus* individual probability or survival).

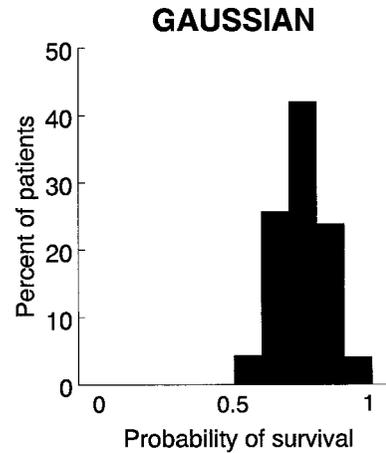
*Subpopulations of Trauma Patients*

We compared trauma patients with a low Ps ( $Ps \leq 0.10$ ) and the other trauma patients. The univariate anal-

**Table 2. Gaussian Distribution: Discrepancies between the Number of Patients (N) to include in a Trial ( $\alpha = 0.05, \beta = 0.10$ ) when Considering the Global Percentage of Survivors (Global) or the Individual Value (Individual) of the Probability of Survival (Ps) in a Theoretical Population with a Gaussian Distribution of Ps (Fig. 2) and a Mean Survival Rate of 75%**

Gaussian Distribution	Control	Expected Drug Effect			
		+10%	+20%	+30%	+50%
<b>Global</b>					
% of survivors	75.0	77.5	80.0	82.5	87.5
N	—	12,018	2,848	1,192	374
<b>Individual</b>					
% of survivors	75.0	77.4	79.9	82.4	87.3
N	—	13,066	2,972	1,228	388

Several hypotheses were tested regarding the expected effect of the drug on Ps (+10, +20, +30, and +50%).



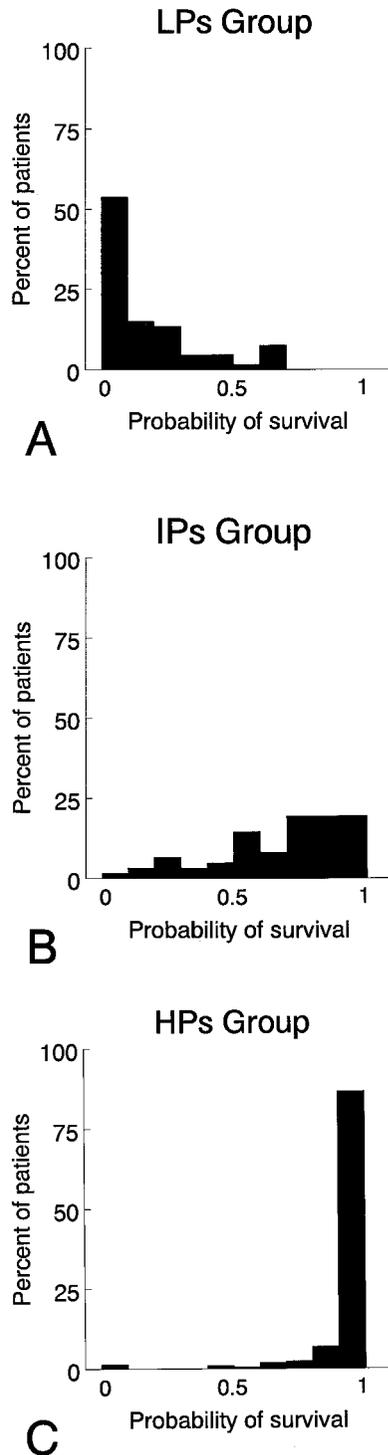
**Fig. 2. Histograms showing the distribution of the probability of survival in a population of theoretical patients (N = 1,000) generated by the computer with a Gaussian distribution (mean probability of survival, 0.75).**

ysis showed that initial SAP ( $43 \pm 47$  vs.  $114 \pm 32$  mmHg), heart rate ( $67 \pm 60$  vs.  $98 \pm 23$  beats/min), Glasgow Coma Score ( $4 \pm 2$  vs.  $12 \pm 4$ ), and respiratory rate ( $10 \pm 11$  vs.  $20 \pm 5$ ), occurrence of prehospital cardiac arrest (51 vs. 2%) and RTS ( $1.22 \pm 1.50$  vs.  $6.74 \pm 1.51$ ) were significantly different between groups. The area under the receiver operating characteristics curve was  $0.98 \pm 0.16$  for the RTS, with a cutoff value of 4.5 (sensitivity = 97%, specificity = 90%). The logistic regression indicated that an RTS less than 4.50 was the only independent predictor of a low Ps (table 3). We also compared trauma patients with a high Ps ( $Ps \geq 0.90$ ) and the other trauma patients. The univariate analysis showed that initial SAP ( $120 \pm 25$  vs.  $88 \pm 49$  mmHg), Glasgow Coma Score ( $14 \pm 2$  vs.  $7 \pm 4$ ), and respiratory rate ( $20 \pm 5$  vs.  $17 \pm 9$ ), occurrence of prehospital cardiac arrest (0 vs. 16%) and RTS ( $7.54 \pm 0.62$  vs.  $4.23 \pm 2.34$ ) were significantly different between groups. The area under the receiver operating characteristics curve was  $0.93 \pm 0.07$  for the RTS, with a cutoff value of 6.3 (sensitivity = 94%, specificity = 82%) and  $0.91 \pm 0.07$  for the Glasgow Coma Score, with a cutoff value of 12 (sensitivity = 88%; specificity = 85%). The logistic regression indicated that an RTS greater than 6.30 and a

**Table 3. Factors Associated with a Low ( $\leq 0.10$ ) or a High ( $> 0.90$ ) Probability of Survival (Ps) in Trauma Patients (N = 350)**

Variable	<i>P</i> Value	Odds Ratio (95% CI)
Low Ps		
RTS < 4.5	0.0001	327.5 (43.4–2471.9)
High Ps		
RTS > 6.3	0.0001	3.65 (2.20–6.04)
GCS > 12	0.015	1.21 (1.04–1.40)

CI = confidence interval; RTS = revised trauma score; GCS = Glasgow coma score.



**Fig. 3.** Histograms showing the distribution of the probability of survival (Ps) in trauma patients with a (A) low Ps (LPs group,  $n = 67$ ), (B) an intermediate Ps (IPs group,  $n = 62$ ), and (C) a high Ps (HPs group,  $n = 221$ ).

Glasgow Coma Score greater than 13 were independent predictors of a high Ps (table 3).

Figure 3 shows the distribution of Ps in the low-Ps, high-Ps, and intermediate-Ps groups. Sixty-seven patients (19%) were included in the low-Ps group, 62 (18%) in the intermediate-Ps group, and 221 (63%) in the high-Ps

**Table 4.** Selected Trauma Populations: Discrepancies between the Number of Patients (N) To Include in a Trial ( $\alpha = 0.05$ ,  $\beta = 0.10$ ) When Considering the Global Percentage of Survivors (Global) or the Individual Value (Individual) of the Probability of Survival (Ps)

	Control	Expected Drug Effect			
		+10%	+20%	+30%	+50%
<b>Low Ps</b>					
Global					
% of survivors	15.5	17.0	18.6	20.1	23.2
N	—	25,158	6,054	2,816	1,050
Individual					
% of survivors	15.5	16.2	17.0	19.5	22.2
N	—	113,862	20,528	2,998	1,418
<b>Intermediate Ps</b>					
Global					
% of survivors	67.3	70.6	73.8	77.1	83.6
N	—	8,140	2,002	834	264
Individual					
% of survivors	67.3	68.6	69.9	74.4	79.2
N	—	53,864	10,798	1,348	564
<b>High Ps</b>					
Global					
% of survivors	94.8	95.3	95.8	96.4	97.4
N	—	78,316	18,424	6,652	2,170
Individual					
% of survivors	94.8	94.9	95.2	96.1	97.1
N	—	2,049,664	101,078	8,522	3,018

Study used the Selected Population of Trauma Patients with Low (LPs group), High (HPs group), or Intermediate (IPs group) Probability of Survival. Several hypotheses were tested regarding the expected effect of the drug on Ps (+10, +20, +30, and +50%). See text and figure 1 for explanation.

group. Using these selected populations, the number of patients to be included in a trial when considering the global percentage of survivors or the individual value of Ps are shown in table 4.

#### Other Distributions

Figure 4 shows the distributions (uniform or J shape) of Ps generated by the computer in 1,000 theoretical patients. As shown in table 5, there were some discrepancies between the number of patients to be included in a trial when considering the global percentage of survivors or the individual value of Ps in both groups. For a drug expected to increase Ps by 20%, the increase in the number of patients to be included was 22-fold in the bimodal distribution (table 1), 3.2-fold in the uniform distribution, 2.5-fold in the J-shape distribution (table 5), and only 1.04-fold in the Gaussian distribution (table 2).

#### Discussion

The current study demonstrated that the bimodal distribution of the Ps in trauma patients has major consequences on the main assumptions of hypotheses testing. Obviously, the power of randomized trials intended to demonstrate an improvement in survival was markedly overestimated. This problem also occurred in other

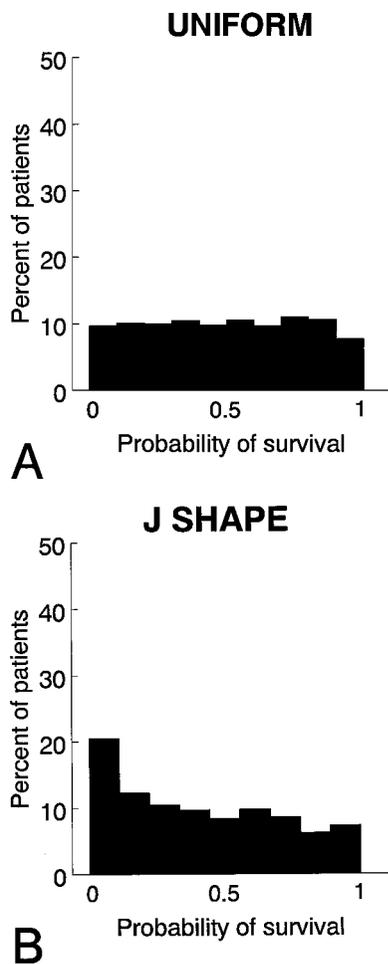


Fig. 4. Histograms showing the distribution of the probability of survival in two populations of theoretical patients (N = 1,000) generated by the computer with a uniform (A; mean probability of survival, 0.50) and a J-shape (B; mean probability of survival, 0.41) distribution.

types of distribution (uniform, J shape), although to a lesser degree, when compared with a bimodal distribution such as the one observed in trauma patients.

Our trauma population slightly differed from the MTOS population because we included more severe trauma patients. This discrepancy was expected, because only trauma patients who underwent severe trauma and/or were suspected to have severe trauma lesions were directly admitted in our unit. The senior physician in the SAMU system was responsible for this on-scene triage.<sup>8</sup> The TRISS methodology has been widely used to assess Ps in trauma patients. The TRISS score has been criticized and other severity scores have been proposed, such as ASCOT or International Classification of Disease related scores.<sup>19,20</sup> Three main limitations of the TRISS have been documented.<sup>21</sup> First, there is a lack of homogeneity within the patient subcategory of penetrating injuries (*i.e.*, gun shot *vs.* stab wounds). Second, there is an inability of the TRISS to predict the survival rate of patients suffering low falls. Third, the TRISS is unable to

Table 5. Other Distributions: Discrepancies between the Number of Patients (N) To Include in a Trial ( $\alpha = 0.05$ ,  $\beta = 0.10$ ) When Considering the Global Percentage of Survivors (Global) or the Individual Value (Individual) of the Probability of Survival (Ps)

	Control	Expected Drug Effect			
		+10%	+20%	+30%	+50%
Uniform distribution					
Global					
% of survivors	49.7	54.7	59.6	64.6	74.5
N	—	4,112	1,018	432	140
Individual					
% of survivors	49.7	52.2	54.8	57.3	62.4
N	—	16,646	3,216	1,424	688
J-shape distribution					
Global					
% of survivors	40.7	44.8	48.8	52.9	61.0
N	—	6,018	1,530	666	230
Individual					
% of survivors	40.7	43.0	45.3	47.5	52.1
N	—	19,162	3,892	1,770	792

Study used computer-generated populations of patients (n = 1,000) with various distributions of Ps corresponding to figure 5. Several hypotheses were tested regarding the expected effect of the drug on Ps (+10, +20, +30, and +50%).

account for multiple severe injuries to a single part of the body. In our study, only blunt trauma patients were included. Patients suffering low falls were not included. Thus, the first two limitations did not seem to apply. However, patients with multiple injuries to a single part of the body did occur, so this limitation could have occurred in our study. This being said, the TRISS score has been widely used not only in the United States but also in Canada<sup>22</sup> and in European countries<sup>23,24</sup> and should be presently considered as the most reliable and validated method to predict survival rate in trauma patients.

By calculating the number of patients to include in a simulated randomized trial, we observed marked differences when the hypotheses were based on the percentage of survivors in the population or on the individual Ps (table 1). The bimodal distribution of Ps in our population obviously explains this discrepancy. In fact, when considering the percentage of survivors in the population, we imply that most patients had nearly the same prognosis, which is thought to be reflected by the observed percentage of survivors. Mathematically, multiplying the mean Ps (*i.e.*, the percentage of survivors) by x percent (the expected drug effect) is not equivalent to taking the mean of the multiplication of individual Ps by x percent, except in the case where all patients have exactly the same Ps. As expected, when using a Gaussian distribution of Ps in a theoretical population (fig. 2), the discrepancy between the two modes of calculation of the number of patients was very moderate (table 2)

compared with that observed in the bimodal distribution (table 1). Indeed, in a bimodal distribution, the mean Ps (equivalent to the percentage of survivors) does not appropriately describe the distribution of Ps in this population: the mean Ps (0.75) was very different from the median Ps (0.94). Although the bimodal distribution of Ps in trauma patients has been previously described,<sup>14,15</sup> we outline that the consequences on hypothesis testing were not, to date, appropriately drawn.

In an attempt to reduce the heterogeneity of our trauma population, we identified simple prehospital variables (Glasgow Coma Score and RTS), which were appropriately able to identify trauma patients with an intermediate Ps (table 3 and fig. 3B). The aim of this selection was to identify either patients with a high or low Ps. Using these selection criteria, the discrepancy between the two methods to calculate the number of patients to be included in a study was markedly reduced, especially in patients with an intermediate Ps (table 4). Nevertheless, it should be pointed out that the number of patients to include remains very high because a pharmacologic intervention that could potentially increase Ps by 50% is not very realistic. Because patients from the intermediate-Ps group represented only 18% of trauma patients, it suggests that a multicenter design is probably relevant.

Trauma is unique as a disease. Indeed, most deaths occur early (80% within 24 h),<sup>1</sup> and a precise measure of Ps requires a complete assessment of trauma lesions that usually cannot be obtained in the early phase. In contrast, it is usually thought that a pharmacologic intervention expected to improve survival rate should be administered as soon as possible, *i.e.*, before accurate determination of Ps. These characteristics contrast dramatically with cancer, where a precise assessment of the prognosis can be performed before randomization, enabling us to compare groups with relatively homogeneous prognoses. A randomized trial testing a new chemotherapy that would have included patients with skin epithelial cancer as well as those with metastatic carcinoma of the pancreas would be considered as a nonsense. However, this situation is suspected to have occurred in some of the randomized trials performed in trauma. Heterogeneity is the cornerstone of the problem. It should be pointed out that even in cancer research where prognostic factors have been extensively used for stratification, heterogeneity is now recognized as a factor that leads to a substantial loss of power.<sup>25,26</sup>

Our study may have equal consequences for clinical researchers, health authorities, or pharmaceutical companies. First, our study indicates that the number of patients should be carefully determined in a randomized trial in trauma and should consider the characteristics of the distribution of Ps. When the number of patients to be included is definitely nonrealistic, alternative approaches should probably be considered. Using a three-

variable (predicted risk of mortality, drug treatment, and predicted risk of mortality–drug treatment interaction) log-normal regression model, Knaus *et al.*<sup>16</sup> were able to demonstrate the efficacy of an anticytokine therapy in sepsis. Second, depending on the nature of the problem, information provided in table 4 may facilitate decision-making in different situations. Care providers or funders may be interested in conducting a trial in patients with a high Ps because most of trauma patients belong to this group, even if the size of the trial is large. Pharmaceutical companies may be rather interested in conducting a trial in patients with an intermediate Ps because the size of the trial should be the lower and because there are numerous arguments to suggest that this population should benefit most from a pharmaceutical intervention (resuscitation fluid, immunomodulation, *etc.*). Prehospital or emergency physicians may be rather interested in conducting a trial in patients with a low Ps because their survival is critically dependent on the efficiency of the organization of the emergency systems. Third, because of the bimodal distribution of Ps and its consequences on the power of a study, the mortality rate may be an unattainable end point in many situations. Regulatory agencies should include this information when discussing clinically relevant end points with pharmaceutical companies. In such situations, other end points should be explored to promote progress in drug therapy. For example, the possibility of using a composite end point (mortality, sepsis, renal failure, acute respiratory distress syndrome, and multiple organ failure) could be considered.

We suggest that the problem identified in the current study might not only occur in trauma trials but also in other disorders affecting critically ill patients, such as acute respiratory distress syndrome or sepsis. Indeed, numerous large randomized trials conducted in the critical care setting in the past decade failed to provide positive results. For example, randomized trials have failed to demonstrate any significant benefit of nitric oxide in acute respiratory distress syndrome, whereas it is known to markedly improve gas exchange. This topic should probably be reconsidered in the light of the current findings.<sup>27–29</sup> Establishing accurate predictors of mortality for these diseases<sup>30</sup> and verifying the nature of the distribution of Ps should be a high priority before undergoing randomized trials. Indeed, although the influence of the distribution of Ps is less important in other types of distribution (table 5) than in a bimodal distribution, it clearly remains clinically relevant. Thus, our study might have important consequences for the design of randomized trials in critical care in the future.

The authors thank Dr. David Baker, M.D., F.R.C.A. (Hôpital Necker-Enfants Malades, Paris, France), for reviewing the manuscript.

## References

1. Capan LM, Miller SM, Turndorf H: Trauma overview, Trauma Anesthesia and Intensive Care. Edited by Capan LM, Miller SM, Turndorf H. Philadelphia, Lippincott, 1991, pp 3-28
2. Elliott DC, Rodriguez A: Cost effectiveness in trauma care. *Surg Clin North Am* 1996; 76:47-62
3. Hammond JM, Potgieter PD, Saunders GL: Selective decontamination of the digestive tract in multiple trauma patients-is there a role? Results of a prospective, double-blind randomized trial. *Crit Care Med* 1994; 22:33-9
4. Dries DJ, Jurkovich GJ, Maier RV, Clemmer TP, Struve SN, Weigelt JA, Stanford GG, Herr DL, Champion HR, Lewis FR: Effect of interferon-gamma on infection-related death in patients with severe injuries: A randomized, double blind, placebo-controlled trial. *Arch Surg* 1994; 129:1031-41
5. Mattox KL, Maningas PA, Moore EE, Mateer JR, Marx JA, Aprahamian C, Burch JM, Pepe PE: Prehospital hypertonic saline/dextran infusion for post-traumatic hypotension: The USA Multicenter Trial. *Ann Surg* 1991; 213:482-91
6. Marshall LF, Maas AI, Marshall SB, Bricolo A, Fearnside M, Iannotti F, Klauber MR, Lagarrigue J, Lobato R, Persson L, Pickard JD, Piek J, Servadei F, Wellis GN, Morris GF, Means ED, Musch B: A multicenter trial on the efficacy of using tirilazad mesylate in cases of head injury. *J Neurosurg* 1998; 89:519-25
7. Sloan EP, Koenigsberg M, Gens D, Cipolle M, Runge J, Mallory MN, Rodman G, for the DCLHb Traumatic Hemorrhagic Shock Study Group: Diaspirin cross-linked hemoglobin (DCLHb) in the treatment of severe traumatic hemorrhagic shock: A randomized controlled efficacy trial. *JAMA* 1999; 282:1857-64
8. Spaulding CM, Joly LM, Rosenberg A, Monchi M, Weber SN, Dhainaut JF, Carli P: Immediate coronary angiography in survivors of out-of-hospital cardiac arrest. *N Engl J Med* 1997; 336:1629-33
9. Alexander RH, Proctor HJ: Advanced Trauma Life Support Course for Physicians: Resource Document 2. Prehospital Triage Criteria. Chicago, American College of Surgeons, 1993, pp 317-8
10. Association for the Advancement of Automotive Medicine. The Abbreviated Injury Scale. Des Plaines, Association for the Advancement of Automotive Medicine, 1990
11. Baker SP, O'Neill B, Haddon W, Long WB: The Injury Severity Score: A method for describing patients with multiple injuries and evaluating emergency care. *J Trauma* 1974; 14:187-96
12. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME: A revision of the Trauma Score. *J Trauma* 1989; 29:623-9
13. Boyd CR, Tolson MA, Copes WS: Evaluating trauma care: The TRISS method. *J Trauma* 1987; 27:370-8
14. Champion HR, Copes WS, Sacco WJ, Lawnick MM, Keast SL, Bain LW, Flanagan ME, Frey CF: The Major Trauma Outcome Study: Establishing national norms for trauma care. *J Trauma* 1990; 30:1356-65
15. Hollis S, Yates DW, Woodford M, Foster P: Standardized comparison of performance indicators in trauma: A new approach to case-mix variation. *J Trauma* 1995; 38:763-6
16. Knaus WA, Harrell FE, LaBresque JF, Wagner DP, Pribble JP, Draper EA, Fisher CJ, Soll L: Use of predicted risk of mortality to evaluate the efficacy of anticytokine therapy in sepsis. *Crit Care Med* 1996; 24:46-56
17. Casagrande JT, Pike MC: An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* 1978; 34:483-6
18. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S: Basic statistics for clinicians. 1. Hypothesis testing. *Can Med Assoc* 1995; 152:27-32
19. Champion HR, Copes WS, Sacco WJ, Frey CF, Holcroft JW, Hoyt DB, Weigelt JA: Improved predictions from severity characterization of trauma (AS-COT) over Trauma and Injury Severity Score (TRISS): Results of an independent evaluation. *J Trauma* 1996; 40:42-8
20. Rutledge R, Osler T, Emery S, Kromhout-Schiro S: The end of the Injury Severity Score (ISS) and the Trauma and Injury Severity Score (TRISS): ICISS, an International Classification of Diseases, ninth revision-based prediction tool, outperforms both ISS and TRISS as predictors of trauma patient survival, hospital charges, and hospital length of stay. *J Trauma* 1998; 44:41-9
21. Cayten CG, Stahl WM, Murphy JG, Agarwal N, Byrne DW: Limitations of the TRISS method for interhospital comparisons: A multihospital study. *J Trauma* 1991; 31:471-81
22. Guirguis EM, Hong C, Liu D, Watters JM, Baillie F, McIntyre RW: Trauma outcome analysis of two Canadian centres using the TRISS method. *J Trauma* 1990; 30:426-9
23. Ruchholtz S, Nast-Kolb D, Waydhas C, Lefering R: The trauma register of the Polytrauma Committee of the German Society of Trauma Surgery as the basis for quality management in the management of severely injured patients. *Langenbecks Arch Chir Suppl Kongressbd* 1997; 114:1265-7
24. Corbanese U, Possamai C, Casagrande L, Bordino P: Evaluation of trauma care: Validation of the TRISS method in an Italian ICU. *Intensive Care Med* 1996; 22:941-6
25. Aalen OO: Heterogeneity in survival analysis. *Stat Med* 1988; 7:1121-37
26. Azawa K, Nakamura T, Palesch Y: Power of logrank test an Cox regression model in clinical trials with heterogeneous samples. *Stat Med* 1997; 16:583-97
27. Dellinger RP, Zimmerman JL, Taylor RW, Straube RC, Hauser DL, Criner GJ, Davis K, Hyers TM, Papadakos P: Effects of inhaled nitric oxide in patients with acute respiratory distress syndrome: Results of a randomized phase II trial. *Crit Care Med* 1998; 26:15-23
28. Troncy E, Collet JP, Shapiro S, Guimond JG, Blair L, Ducruet T, Francoeur M, Charbonneau M, Blaise G: Inhaled nitric oxide in acute respiratory distress syndrome: A pilot randomized study. *Am J Respir Crit Care Med* 1998; 157:1483-8
29. Lundin S, Mang H, Smithies M, Stenqvist O, Frostell C, and the European Study Group of Inhaled Nitric Oxide: Inhalation of nitric oxide in acute lung injury: Results of a European multicentre study. *Intensive Care Med* 1999; 25:911-9
30. Monchi M, Bellenfant F, Cariou A, Joly LM, Thebert D, Laurent I, Dhainaut JF, Brunet F: Early predictive factors of survival in the acute respiratory distress syndrome. *Am J Respir Crit Care Med* 1998; 158:1076-81