A practical approach for eliciting expert prior beliefs about cancer survival in phase III randomized trial

Anne Hiancea, Sylvie Chevret, Vincent Lévy

INSERM 9504, Centre d’Investigations Cliniques, Hôpital Saint Louis, Assistance Publique-Hôpitaux de Paris, Université Paris 7, France

Department de Biostatistique et Informatique Médicale, Hôpital Saint Louis, Assistance Publique-Hôpitaux de Paris, Université Paris 7, France

Accepted 14 April 2008

Abstract

Objective: To propose and compare practical approaches that allow eliciting and using expert opinions about the benefit effect on a censored endpoint, such as event-free survival (EFS), used in the planning of a clinical trial based on Bayesian methodology.

Study Design and Setting: Individual interviews of 37 experts. Bayesian normal models on the log hazard ratio (HR) of EFS were implemented. We illustrate our approach by using a trial of autologous stem cell transplantation (ASCT) vs. chemotherapy (CT) in chronic lymphocytic leukemia (CLL). We elicited experts’ prior beliefs about the difference in 3-year EFS between the two treatment arms, either roughly or throughout weights over the difference scale. Subsequently, a Bayesian synthesis of the information reported in the trial protocol with that in the experts’ prior was performed, using: (1) the postulated treatment effect based on null (skeptical) and alternative (enthusiastic) hypotheses with shared standard error; and (2) the expected difference derived from experts’ distributions.

Results: As compared with the priors based on the trial protocol data, expert priors agreed with some average from enthusiastic and skeptical information, with close standard errors.

Conclusion: This case study illustrates a rational approach to construct an expert-based prior. It should be considered as part of the design of future Bayesian trials.

Keywords: Elicitation; Prior; Bayes; Expert opinion; Chronic lymphocytic leukemia; Randomized trial

1. Introduction

Bayesian methods provide an interesting and innovative alternative to classical statistical approaches in the design, monitoring, and reporting of clinical trials. One of the advantages of a Bayesian approach is the ability to make explicit probability statements about hypotheses, for instance, the probability that the experimental treatment is superior to the standard one. Indeed, such probability statements cannot be inferred through classical approaches, although many clinicians misinterpret $P$ values and confidence intervals (CIs) as so. In contrast, Bayes’ inference directly considers uncertainty in treatment effect using a probability distribution [1,2]. This Bayesian requirement for a so-called “prior” distribution, which reflects the initial information about the treatment effect, including the investigator’s understanding of the biology of the disease, historical results for the investigational and related treatments, and preclinical results for these treatments, thus inherently subjective, has been long reported as the main drawback of such approaches. Indeed, in such a regulatory setting, it is important for sponsors and regulators to agree in advance to the prior distribution(s) that will be used [1].

For overcoming concerns about the subjective nature of prior distributions, a common approach is to assume a prior distribution that is “non-informative” [1]. In this setting, the results of the trial will carry essentially all the influence in the so-called posterior distribution. Another is to consider a variety of prior distributions in attempting to approximate the posterior distribution held by all types of readers. Spiegelhalter et al. [3] and Parmar et al. [4] proposed to use both enthusiastic and skeptical priors, based on the alternative and null hypotheses stated when computing the trial sample size, respectively. Finally, collecting and documenting subjective prior beliefs from knowledgeable experts about the potential results of
2. Materials and methods

2.1. Trial information

As an illustrative example, we choose the auto-CLL, a phase III randomized clinical trial designed to assess the benefit of ASCT-consolidation treatment following induction CT in patients with previously untreated Binet’s stage B or C CLL [13]. All patients were scheduled to receive three cycles of ChOP (cyclophosphamide, adriamycin, vincristine, prednisone) followed by three cycles of fludarabine. Once stem cell collection was completed, patients responding to CT (complete or partial response according to the National Cancer Institute (NCI) recommendations [14]) had to be randomized between a watch-and-wait policy and ASCT, with conditioning regimen consisting in total body irradiation (TBI) plus high-dose cyclophosphamide. The primary outcome was the 3-year event-free survival (EFS) rate.

When planning the trial in 2000, an absolute difference of 40% in 3-year EFS was expected with ASCT over CT (CLL 90 trial [15]). This led to a recommended sample size of 80 responders (2 × 40), with an 80% power and two-sided significance level of 5%. Owing to the 40% expected prevalence of response, a total of 200 required CLL patients had to be enrolled.

2.1.1. Elicitation of prior opinion

Thirty-seven French hematologist experts, all involved in the treatment of patients with CLL, were interviewed. The demographic characteristics of the experts are detailed in Table 1. Interviews took place either during the French Haematology Society’s (SFH) annual meeting in Paris in March 2006, or by e-mail, followed by telephone calls. The investigators’ prior opinions were elicited by the use of a questionnaire.

To make the elicitation as easy as possible for subject-matter experts to tell us what they believe, while reducing how much they need to know about probability theory to do so, we decided to ask predictive rather than structural questions, following the argument of Kadane and Wolfson [16]. Thus, experts were asked to assess only observable quantity conditionally on the treatment group, which is the probability of experiencing the endpoint within the first 3 years of ASCT as consolidation treatment and CT alone, respectively. Subsequently, the experts were asked to give their own weights of belief over the difference scale, involving, in some sense, the assessment of individual probabilities. The questionnaire also aimed to capture the investigators’ range of equivalence, which is the interval of difference in 3-year EFS, within which, the investigator feels that the two treatments are equivalent. In addition, they were also asked for their opinion on the 3-year EFS

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
</table>

Main characteristics of the 37 experts

<table>
<thead>
<tr>
<th>Variables</th>
<th>Experts (N = 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>N = 24 (64.8%)</td>
</tr>
<tr>
<td>Female</td>
<td>N = 13 (35.1%)</td>
</tr>
<tr>
<td>Time from MD degree (years)</td>
<td>Median: 17.5</td>
</tr>
<tr>
<td></td>
<td>Range: 3–29</td>
</tr>
<tr>
<td>Institution</td>
<td></td>
</tr>
<tr>
<td>University Hospital</td>
<td>N = 34 (91.9%)</td>
</tr>
<tr>
<td>General Hospital</td>
<td>N = 2 (5.4%)</td>
</tr>
<tr>
<td>Cancer Center</td>
<td>N = 1 (2.7%)</td>
</tr>
<tr>
<td>Discipline</td>
<td></td>
</tr>
<tr>
<td>Clinician</td>
<td>N = 34 (91.9%)</td>
</tr>
<tr>
<td>Biologist</td>
<td>N = 3 (8.1%)</td>
</tr>
<tr>
<td>Function</td>
<td></td>
</tr>
<tr>
<td>Head</td>
<td>N = 7 (18.9%)</td>
</tr>
<tr>
<td>Professor</td>
<td>N = 10 (27%)</td>
</tr>
<tr>
<td>Associate professor</td>
<td>N = 14 (37.8%)</td>
</tr>
<tr>
<td>Assistant professor</td>
<td>N = 5 (13.5%)</td>
</tr>
<tr>
<td>Attending</td>
<td>N = 1 (2.7%)</td>
</tr>
<tr>
<td>CLL as first domain of interest</td>
<td>N = 15 (40.5%)</td>
</tr>
<tr>
<td>Inclusions in the auto-CLL trial</td>
<td>N = 18 (50%)</td>
</tr>
<tr>
<td>Interviewed by phone</td>
<td>N = 8 (21.6%)</td>
</tr>
</tbody>
</table>

Abbreviation: CLL, chronic lymphocytic leukemia.
for CT alone, and the number of CLL patients to whom they had given auto-graft in the preceding 12 months. Otherwise, experts were asked to consider themselves as the main investigators of this trial and to give the number of patients they would reasonably plan to include to answer the question of the superiority of ASCT over CT. Finally, they had to quantify the number of patients they would be ready to treat with ASCT, so that only one patient would benefit of a lengthening of EFS, that is, the number needed to treat (NNT).

The questionnaire was closely based on that used by both Parmar et al. [4] and Tan et al. [17]. It is given in Appendix 1. It was either hand-delivered or sent to the investigators by e-mail, by two physicians (A.H. and V.L.).

2.1.2. Constructing the priors

Because the original protocol used a fixed hazard ratio (HR) model, and because the HR represents the most commonly used scale for comparing survival between treatments, as it is specifically designed to allow for censoring and time to an event [18], we decided to use these grounds alone. Therefore, a transformation was required to convert the difference results to the log HR (LHR) scale, throughout the equation:

$$\text{LHR} = \log[\log(\text{P1})/\log(\text{P2})]$$

(1)

where P1 is the 3-year EFS rate in the experimental group (CT + ASCT), and P2 is the 3-year EFS rate in the conventional group (CT). An LHR of more than 0 (i.e., HR > 1) indicates an advantage to CT and the reverse.

Furthermore, the use of LHR and its variance allows prior normal distributions to be fitted directly [19]. We assumed LHR to be normally distributed, with $\mu_0$ and $\sigma_0$ being the prior mean and the prior standard deviation (SD), respectively. This prior is equivalent to a normal likelihood arising from a hypothetical trial of $n_0$ patients with an observed value $\mu_0$ of the treatment difference statistic LHR [1].

Several normal prior distributions were fitted, differing in terms of mean $\mu_0$ and $\sigma_0$.

First, we considered two, either enthusiastic or skeptical, priors based on the alternative and null hypotheses used when computing the trial sample size, respectively, with shared variance, as proposed by Spiegelhalter et al. [3] and Parmar et al. [4]. The enthusiastic prior was, thus, directly computed from the equation (1) mentioned earlier, on the basis of a 3-year EFS rate of P2 = 50% for the control group and P1 = 90% for the experimental group, resulting in a mean LHR of $\mu_0 = -1.884$. The variance, $\sigma_0^2$, was approximated by $4/n_0$ [19], where $n_0$ is the total number of adjusted deaths from all the studies summarized [20]. In absence of external evidence, we based this computation on the information used when planning the trial from the Schoenfeld formula [21], that is, $n_0 = 4 \left( \Phi(1 - \alpha/2) + \Phi(1 - \beta) \right) \mu_0^2 / \mu_0^2$, where $\Phi(\cdot)$ is the standard normal distribution function, and $\alpha$ and $\beta$ are the type I and the type II errors, respectively. Both type errors were fixed at 5%. Therefore, the prior variance was assumed to be fixed at $\sigma_0^2 = 4/14.65 = 0.273$.

In contrast, the skeptical prior represents the belief of individuals who are reluctant to accept that high response rates are possible; hence, it formalizes the belief that large treatment difference is unlikely, based on a 3-year EFS rate of 50% for both treatment groups, that is, P1 = P2 = 50%. Consequently, the prior skeptical mean was $\mu_0 = 0$, with the same variance as the enthusiastic prior [22]. In other words, the skeptical prior was equivalent to having performed a trial with $n_0$ patients, all of whom have died and, in which, no difference has been observed between the two arms. Otherwise, the skeptical prior distribution was such that the probability of an effect as large or larger than $\log(h1)$ was 0.013, in agreement with its skeptical nature [23].

Subsequently, we used the experts’ information achieved by the questionnaire to solicit their prior beliefs. Because two main results were available, namely individual 3-year EFS in both groups, and individual weights on the treatment difference scale, different priors were computed. The first expert prior averaged the informative prior beliefs about the treatment effect of all experts. It was built by: (1) transforming each individual answer in terms of difference on 3-year EFS in the LHR scale by using equation (1); and (2) computing mean and SD of observed distribution among the experts. Otherwise, each expert’s weights could be considered as a discrete probability distribution for the expected 3-year difference in outcomes across randomized groups, on a discrete scale of 18 predetermined values. To synthesize these different distributions into a single distribution, unweighted average of the individual probability distributions comprising it [24] were computed. First, we averaged the own full priors of each expert by combining expert weights (expert prior 2). Briefly, it consisted in: (1) plotting distributions of individual weights across LHR after having transformed differences in that scale (Fig. 1A); and (2) estimating the mean and SD of such a distribution by integration over the whole LHR scale. Finally, to better reflect the fact that individual weights corresponded to a discrete probability distribution, assuming all experts to have access to the same amount of information, we fitted a multinomial model to the data; this will be referred further as expert prior 3.

2.2. Statistical analysis

Summary statistics (mean with 95% CI) were computed for each item of the questionnaire. Association between experts’ characteristics and prior beliefs regarding the difference in the 3-year EFS between the two arms were tested by using the Wilcoxon rank-sum test. They consisted of sex, time from Medical Doctor Degree (MD), CLL as the first domain of interest, participation as an investigator of the Auto-LLC trial, number of CLL patients treated with ASCT, and the number of ASCT in the previous 12 months.

3. Results

3.1 Results of the questionnaire

The questionnaire was proposed to 37 experts. Table 1 summarizes their main characteristics. About two-thirds were male (65%), and one-half had their MD degree for more than 17.5 years. Ninety-two percent worked in a university teaching hospital, and were clinicians. More than 40% declared CLL as their main domain of interest. The median number of ASCT performed during the previous 12 months was 45 (range: 0–140), with a median of CLL patients treated with ASCT of 2 (range: 0–9).

All fulfilled the information requested, except one hematologist, who declined to provide any answer. Time devoted for each interview ranged from 20’ to 45’.

Table 2 summarizes the main answers to the questionnaire. The average prior beliefs on the 3-year EFS for CT alone were estimated at 43.7% (95% CI: 38.9–48.6%). This was less optimistic than that used when planning the trial, because a 3-year EFS rate of 50% was expected in the reference group. The estimated total median sample size was 200 (range: 76–400). It was close to the initial calculation, because, in 2000, 208 patients were planned to be included in this trial. Therefore, clinicians have a good perception of the number of patients required to give a statistically significant result related to the question the trial aims to answer. The median NNT was 5 (range: 0–50).

Of note, 17 clinicians refused to estimate NNT, whereas five clinicians estimated a null NNT.

The mean reported “interesting,” “discounted,” or “minimum worthwhile” difference in the 3-year EFS rate between the two arms was 22.4% (95% CI: 17.2–27.6%). Only the inclusion in auto-CLL trial appeared to modify this finding; experts who included patients in the trial reported a mean difference of 22.11% (95% CI: 16.29–27.91%), whereas those who did not participate in the trial reported a mean difference of 16.94% (95% CI: 12.42–21.46%). Otherwise, the overall mean prior belief regarding the difference in the 3-year EFS between the two treatments was in the range of equivalence, which is 12.5% to 23.2%. Eleven of the 36 clinicians stated the range of equivalence to be 10% to 20% in favor of ASCT. Although experts mostly estimated the minimum worthwhile difference far from it.

The distributions of the probability values that experts individually assigned to the various intervals of 3-year differences in EFS was centered above 10, indicating that experts mostly expected that ASCT could improve the 3-year EFS rate compared with CT alone. Most experts expected a difference of at least 15%, which is below their demanded minimum worthwhile difference, meaning that experts are quite pessimistic with regard to the trial’s results.

3.1.1 Estimation of the prior distribution for treatment effect size based on expert data

Having obtained the information on the 3-year EFS for CT alone, and that for the difference in 3-year EFS, we transformed the data to the LHR scale for each expert, using equation (1). The resulted mean LHR among the 37 experts was −0.69, with SD = 0.494 (expert prior 1; Table 3).

The box plots of the 18 recorded weights across individual experts are displayed in Fig. 1A. The second expert prior was first fitted to the observed expert distribution corresponding to the mean’s weight of beliefs for each treatment difference given in the questionnaire, as revealed...
Table 3

<table>
<thead>
<tr>
<th>Prior Distribution</th>
<th>LHR ($\mu_0$)</th>
<th>SD ($\sigma_0$)</th>
<th>Proportion in the range of equivalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enthusiastic</td>
<td>$-1.884$</td>
<td>$0.523$</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Skeptical</td>
<td>0</td>
<td>$0.523$</td>
<td>8.81</td>
</tr>
<tr>
<td>Expert's prior 1</td>
<td>$-0.690$</td>
<td>$0.494$</td>
<td>27.29</td>
</tr>
<tr>
<td>Expert's prior 2</td>
<td>$-0.717$</td>
<td>$0.628$</td>
<td>21.76</td>
</tr>
<tr>
<td>Expert's prior 3</td>
<td>$-0.3565$</td>
<td>$0.313$</td>
<td>38.02</td>
</tr>
<tr>
<td>Expert's prior 4</td>
<td>$-0.660$</td>
<td>$0.393$</td>
<td>42.67</td>
</tr>
</tbody>
</table>

*Abbreviations:* LHR, log hazard ratio; SD, standard deviation.

a According to the minimum worthwhile difference on the 3-year EFS rate.

b According to the average of weights over the scale of expected differences in the 3-year EFS rate.

c According to the multinomial fit of weights over the scale of expected differences in the 3-year EFS rate.

d According to the minimum worthwhile difference on the 3-year EFS among experts who did not participate in the auto-CLL trial.

earlier. This reached a mean LHR of $-0.717$ with SD = 0.628. Finally, the third expert prior, resulting from fitting the multinomial model to individual weights (Fig. 1B), has a mean LHR of $-0.3565$ and SD of 0.3133. Of note, these resulting experts’ prior distributions lie in between the enthusiastic and skeptical priors derived from the protocol information (Fig. 2). This suggests some substantial uncertainty about which of the treatments would benefit the patient most, fulfilling the principle of equipoise. As expected, the experts’ priors 1 and 2 were almost superposed, whatever it was, based on the minimum worthwhile difference or on the weights on the expected difference, although the variability of the latter was increased. In contrast, expert prior 3 was closest to the skeptical prior, illustrating that elicited distributions varied from other distributions, which might also be consistent with the expert’s knowledge.

4. Discussion

Bayesian methods are being increasingly recognized as appropriate for statistical analysis of clinical trial data. Advantages of Bayesian analysis over classical analysis of clinical trials include: the ability to incorporate prior information regarding treatment effect into the analysis; the ability to make multiple unscheduled inspections of accumulated data without increasing the error rate of the study; and the ability to calculate the probability that one treatment is more effective than another.

However, the implementation of Bayesian methods requires the specification of prior distributions. A prior is often the purely subjective assessment of an experienced expert. However, subjectivity in prior distributions is explicit [3] and open to examination (and critique) by all. The final report of trial analysis could display the dependence of the posterior distribution on the different choices of informative priors, including non-informative priors as reference models to be used as standard of comparison against situations where subjective priors are used [25,26]. To formulate a prior distribution, opinion of informed experts is needed, and this process is defined by elicitation. Elicitation is largely described in literature, because it can be applied in many domains. Moreover, eliciting prior beliefs is ethically important for documenting the nature of the uncertainty or equipoise [5,17]. However, despite the existence of a broad and diverse literature in elicitation, which has provided many valuable procedures and insights, there remains very considerable scope for easy application. Regarding randomized phase III trials, Bayesian methods and elicitation process have been less described. The literature devoted to elicitation techniques used in clinical trials is based on a small number of experts ($n = 4$ [27], $n = 10$ [28], $n = 11$ [4], or $n = 14$ [17]). In this study, 37 experts were elicited, and we performed the elicitation process during an individual interview. Face-to-face elicitation proved to be useful, considering the time spent with each expert (ranging approximately from 20’ to 45’).

Elicitation of prior distributions is based on statistic interpretation, and experts are not necessarily familiar with it. Literature on elicitation recommended the training of the experts first to familiarize them with the interpretation of probabilities, but this method was difficult to perform in practice. We used a questionnaire with clinically meaningful questions and direct interaction with clinicians, although some were contacted by e-mail.
Clinicians were asked to estimate observable quantities that appeared easier to think about than regression parameters. Many inconsistencies were encountered, possibly because of misunderstandings. For instance, the fact that clinicians’ answers broadly differed from the question concerning the minimum clinically worthwhile difference and that regarding the range of equivalence revealed some elicitation issues. Parmar et al. [4] interpreted the range of equivalence as the minimum worthwhile difference, but nine clinicians (25%) answered in a very different way to the two questions. This highlights the impact of questions’ formulation and the cautious interpretation of clinicians’ answers. Otherwise, results regarding the NNT were difficult to analyze. On one hand, 17 clinicians (47%) refused to estimate this quantity and, on the other hand, five quantified the NNT to 0. No agreement between reported NNT and benefit was observed.

Prior distributions from the experts were not obtained before the trial onset. However, despite the name, “prior” suggests a temporal relationship; it does not refer to time, but to a situation where we assess what our evidence would have been if we had no data [3]. In reality, no significant external evidence of ASCT benefit in CLL has been reported. Of significance is the inclusion of patients in the auto-CLL trial as a significant factor influencing minimum worthwhile 3-year EFS difference ($P = 0.031$). However, the 17 experts who did not include patients in the trial were rather more skeptical than the others. Finally, little or no impact of the trial data on the beliefs of the experts has been previously reported [29].

There were conceptual main differences between the expert priors. Indeed, the first expert prior did not correspond to the belief of any one expert, but averaged the informative prior beliefs about the treatment effect of all experts. Instead, the other priors averaged the own full priors of each expert by combining expert weights. To synthesize these different weights into a single distribution, we used one of the most popular methods known as “opinion pools,” which was a weighted average of the individual probability distributions comprising it, with simple equal weighting of the experts [24]. As different techniques may produce different distributions, because the method of questioning may have some effect on the way the problem is viewed, some differences in the resulting expert priors were expected [24]. We performed these different approaches as a sensitivity analysis to allow determining whether, when the elicited distribution varied from other distributions that might also be consistent with the expert’s knowledge, the results derived from that distribution change appreciably. In reality, observed expert prior distributions 1 and 2 were very close, whereas distribution 3 was somewhat different. It can be partly explained by the fact that the expert prior 2 was derived from a normative method for combining the quantitative judgments of the experts with regard to probability distribution. Expressing opinions as probability measures seemed to lead invariably to arithmetic averaging, as previously reported in the work of Genest and Zidek [24]. Nevertheless, this allowed getting further insights in the variability of experts’ beliefs, as illustrated by larger SD of the expert prior 2 contrasting with the smaller SD of expert prior 3. Finally, “appreciable” changes could have been more obvious if differences in priors had been translated, as previously proposed [20], on a sample size scale. Indeed, because the expert prior’s LHR mean ranged from $-0.3565$ to $-0.717$, this would result in a required number of deaths (with $\alpha = \beta = 0.05$) ranging from 101 to 409, contrasting with the required 15 deaths scheduled from the alternate hypothesis.

In summary, Bayesian approaches for randomized clinical trials appear promising. The elicitation process from a small set of experts should be considered as part of the design of future trials. We have shown that choices must be made in converting expert opinion into statistical distributions. Nevertheless, as stated by Spiegelhalter et al. [3], there is no “correct” or “true” prior; instead, a “community” of priors, expressing a range of reasonable opinions that have at least consensus support to be generally accepted by a wider community, is used.

**Acknowledgment**

The authors wish to thank the 37 experts who participated in the survey (Appendix 2).

**References**

guidelines for chronic lymphocytic leukemia: revised guidelines for
Desablens B. Randomized comparison of fludarabine, CAP, and
ChOP in 938 previously untreated stage B and C chronic lymphocytic
[17] Tan SB, Chung YFA, Tai BC, Cheung YB, Machin D. Elicitation of
prior distributions for phase III randomized controlled trial of adju-
vant therapy with surgery for hepatocellular carcinoma. Control Clin
[18] Kalbfleish JD. Nonparametric Bayesian analysis of survival time
[19] Tsiatis AA. The asymptotic joint distribution of the efficient scores
test for the proportional hazards model calculated over time. Biome-
subjective information ever be used to reduce the size of randomized
[21] Schoenfeld D. Sample-size formula for the proportional-hazard re-
[22] Spiegelhalter DJ, Freedman LS, Parmar MK. Applying Bayesian
ideas in drug development and clinical trials. Stat Med
[25] Bernardo JM. Reference posterior distributions for Bayesian infer-
[26] Van Dongen S, Talloen W, Lens L. High variation in developmental
instability under non-normal developmental error: a Bayesian per-
about influence of patient characteristics on treatment effects:
3805–21.
[28] Abrams K, Ashby D, Errington D. Simple Bayesian analysis in clin-
[29] Rovers MM, van der Wilt GJ, van der Bij S, Straatman H, Ingels K,
Zielhuis GA. Bayes’ theorem: a negative example of a RCT on
23–8.
Appendix 1

Questionnaire submitted to the panel of 37 experts to elicit their opinion (translated from French)

The aim of this questionnaire is to ask your opinion regarding a randomized phase III trial assessing the benefit over conventional chemotherapy of intensification with autologous stem cell transplantation as the first line treatment in patients with B or C Chronic Lymphocytic Leukemia (Principal investigator: Laurent Sutton).

Question 1

For chemotherapy alone, what Event Free Survival (EFS) rate do you expect?

- % at 2 year
- % at 3 year
- % at 5 year

Question 2

Please try to quantify the interesting or discounted or minimum worthwhile difference of EFS, survival, and median survival.

EFS difference rate compared to chemotherapy alone of:

- % at 2 year
- % at 3 year
- % at 5 year

Survival difference rate compared to chemotherapy alone of:

- % at 2 year
- % at 3 year
- % at 5 year

Median difference:

- months in terms of survival
- months in terms of EFS

Question 3

We are interested in your estimation of the difference in the 3-year Event Free Survival rate that you expect between chemotherapy alone and intensification with autologous stem cell transplantation.

Please enter your weights of belief in each of the possible intervals of difference shown in the table below.

The stronger you believe that the difference will truly lie in a given interval, the greater should be your weight for that interval. If you believe it is impossible that the difference lies in a given interval your weight should be zero. Your weights should add up to 100.

Example: This is the example of a clinician who believes the most plausible situation is that intensification with autologous stem cell transplantation is better than chemotherapy alone by 10–15%. However, he is prepared to concede that it may do better than this. He even accepts the possibility that chemotherapy may harm 0–5%. As a consequence
he weighs his beliefs as 35 for the most likely outcome, 5 for the unfavourable outcome, and divides the remaining units over the range 0–5, 5–10, 15–20, equally in blocks of 20.

NB: 0–5% advantage to intensification with autologous stem cell transplantation means that the 3-years EFS rate for this treatment will be between 0% and 5% higher than that for chemotherapy alone.

**Question 4**

We are also interested in what importance you place on the extra difficulty and possible extra toxicity of autologous stem cell transplantation procedure.

Please indicate in the table below the absolute advantage in 3-year EFS rate above which you would routinely switch to intensification with autologous stem cell transplantation rather than just chemotherapy alone?

Please also indicate the rate below which you would definitely retain the use of chemotherapy alone.

Example: This is the opinion of a clinician who feels he would definitely use the intensification with autologous stem cell transplantation if it gives a 20% increase in the 3-year EFS rate. He would definitively retain the use of chemotherapy alone if the intensification with autologous stem cell transplantation resulted in less than 10% improvement.

Between these limits (i.e., 10%–20%) he feels that the treatment are roughly equivalent:

- Absolute advantage in 3-year EFS rate *above* which you use intensification with stem cell transplantation 20%.
- Absolute advantage in 3-year EFS rate *below* which you use chemotherapy alone as compared to intensification with autologous stem cell transplantation 10%.

**Question 5**

You are the principal investigator of a randomized phase III trial assessing intensification with autologous stem cell transplantation versus chemotherapy alone and EFS in the primary outcome.

How many patients would you reasonably plan to include in this trial?

**Question 6**

In the same context of a randomized trial, how many patients are you ready to autograft so that one patient would benefit for a lengthening of survival and EFS?

**Question 7**

How many patients did you autograft in the last 12 months?

- |___| total
- |___| LLC

**Appendix 2**

**List of the 37 experts who answered the questionnaire**