GRAPHICAL ELICITATION OF A PRIOR DISTRIBUTION FOR A CLINICAL TRIAL

Kathryn Chaloner,
Timothy Church,
Thomas A. Louis and
John P. Matts

University of Minnesota
Technical Report #583
September 28, 1992

Graphical elicitation of a prior distribution for a clinical trial

Kathryn Chaloner[1], Timothy Church[2], Thomas A. Louis[2]

and John P. Matts[2]

1. School of Statistics, 352 Classroom Office Building,

University of Minnesota, St Paul MN 55108, USA

2. Division of Biostatistics, University of Minnesota, Minneapolis MN 55455,USA

September 28, 1992

SUMMARY

Bayesian methods are potentially useful for the design, monitoring, and analysis of clinical trials. These methods, however, require that prior information be quantified and that the methods be robust. This paper describes a method to help quantify beliefs in the form of a prior distribution about regression coefficients in a proportional hazards regression model. The method uses dynamic graphical displays of probability distributions that can be freehand adjusted. The method was developed for, and is applied to, a randomized trial comparing prophylaxes for toxoplasmosis in a population of HIV positive individuals. Prior distributions from five AIDS experts are elicited. The experts represent a community of consumers of the results of the trial and these prior distributions can be used to try to make the monitoring and analysis of the trial robust.

1 Introduction

Bayesian approaches to clinical trials, as discussed for example by Spiegelhalter and Freedman (1988) and Freedman and Spiegelhalter (1992), require the specification of a prior distribution. In this paper, an AIDS clinical trial comparing potential prophylaxes for toxoplasmosis is used as a context to develop and implement methodology to aid in the elicitation of prior distributions. Following Kadane (1986), we require that a range of priors be identified that are representative of "the community." These will then be used to try to make monitoring and analysis robust. This paper focuses on the elicitation part of such an approach and we report on the prior distributions of five individuals. Bayesian analysis and monitoring of the trial is reported in Carlin et al (1992).

First we provide a description of the toxoplasmosis prophylaxis trial and then describe the elicitation method based on graphical input and feedback. We then summarize what we learned from eliciting the prior distributions, the results of the trial, and some of the problems in implementing a Bayesian approach.

2 The toxoplasmosis prophylaxis trial

The toxoplasmosis prophylaxis trial, conducted through the Community Programs for Clinical Research on AIDS, was a placebo controlled, modified double blind study to evaluate the effectiveness of both clindamycin and pyrimethamine on the subsequent development of toxoplasmosis encephalitis (TE). An effective prophylactic agent needed to be found, since TE was a major cause of mortality and morbidity among patients with AIDS, and had been reported to be a very common opportunistic infection and was the most common cause of intracerebral mass lesions.

All patients entered into the study had a positive titer for *Toxoplasma gondii.* Enrollment was

planned for one year with all patients to be followed for a minimum of two years so that the average period of follow-up would be two and a half years. A computerized literature search for information on toxoplasmosis and AIDS yielded very little, the exception being a paper by Grant et al (1990) in which it was estimated that 30% of the placebo patients would develop TE in two and a half years.

The study had four treatment groups: active clindamycin, placebo clindamycin, active pyrimethamine, and placebo pyrimethamine with an allocation ratio of 2:1:2:1 respectively. Randomization was set up so that allocation to clindamycin and pyrimethamine was unblinded and allocation within each group to active or placebo was double blind. The planned analysis was a semi-parametric proportional hazards regression model (Cox, 1972). Death was to be treated as a censoring event. In order to calculate sample size, based on frequentist methods for computing power, the protocol committee agreed upon values for the effects. They assumed that among fully compliant patients those on active drug would have a 50% lower rate of TE compared to the pooled placebo groups. Given the potential for toxicities, investigators felt that a large reduction in TE, perhaps as high as 50%, was necessary before these drugs could be recommended for routine use in the community. It was also predicted that 10% of the placebo patients would switch over to one of the active drugs, 25% of the patients assigned to active drug would not comply with their assignments, and 33% of the patients would be lost to follow-up due to death unrelated to treatment assignment. The planned sample size was 750 patients with 250 patients to be allocated to each active medication group and 125 patients to each placebo group. In the analysis the two placebo groups would be combined.

The first patient was randomized in September 1990. In March of 1991, the active and placebo clindamycin groups were terminated due to an excess of non-life threatening but serious toxicities

that resulted in permanent medication discontinuances. The details of the toxicities are reported in Jacobson et al (1992a). The pyrimethamine arm continued. After termination of the clindamycin arm, people on either clindamycin or placebo for clindamycin were offered rerandomization to either the pyrimethamine group or placebo group with a two to one allocation ratio.

In April 1992, the trial was terminated because the rate of TE in the pyrimethamine and placebo groups was much lower than expected. The rate in the pyrimethamine arm was actually slightly higher than in the placebo arm, but the difference was not significant and continuation of the trial was very unlikely to lead to a significant difference. Data that accrued after termination indicated an increased death rate on the pyrimethamine arm; (see Jacobson et al, 1992b).

3 The survival model and Bayes structure

The protocol specified an analysis by a proportional hazards model, with a nonparametric underlying hazard function. Our Bayesian analysis will follow this same general approach using the partial likelihood as a likelihood. This strategy can be justified using arguments given in Kalbfleisch (1978), if a Dirichlet process or a gamma process is used for the underlying hazard.

Specification of an informative prior distribution is difficult. Methods for specifying opinions have been developed for several different sampling models, for example the Bernoulli process (Winkler, 1967; Chaloner and Duncan, 1983) and the normal linear model (Kadane et al., 1980; Garthwaite and Dickey, 1990 and 1991). There are no specific methods that we know of, however, for specifying a prior distribution on the regression coefficients of a proportional hazards model. Methods for other models typically require the expert to specify certain properties of his or her beliefs and a distribution is chosen that has these properties, or properties very close to the ones provided

by the expert. The distribution is usually chosen from a specific parametric family, typically a conjugate family. Freedman and Spiegelhalter (1983) take a different approach and specify a prior distribution which is uniform over intervals, see also Spiegelhalter, Freedman and Parmar (1992).

We chose to take a combination of these approaches by starting with the specification of a prior distribution from a parametric family. The expert then has the option of making freehand adjustments. Specifying both the initial parametric distribution and the nonparametric distribution are done with the help of dynamic graphical plots to provide feedback within the XLISP-STAT environment of Tierney (1990).

4 Elicitation method

First we describe a method for specifying a parametric prior distribution on the two regression parameters corresponding to the two treatment effects, and then discuss freehand adjustments. Since the clindamycin arm was terminated early, due to toxicities, the method was converted to one for a single regression parameter. The univariate method is described in Section 4.3.

4.1 The bivariate version

Rather than direct assessment of the proportional hazards regression parameters the elicitation scheme is based on three potentially observable quantities, p_0, p_C, and p_P, the probabilities of experiencing the toxoplasmosis endpoint within the first two years of treatment on the placebo, clindamycin and pyrimethamine respectively. The expert can think about each probability as a proportion of a large group of patients, thus incorporating beliefs about noncompliance and switchover directly rather than having to consider a perfectly compliant patient (as in Freedman and Spiegelhalter, 1983). Using such potentially observable quantities follows the argument of

Kadane et al (1980) that observable quantities are easier to think about than regression parameters.

The baseline hazard for the placebo is assumed to be the same for each respective treatment arm. In the proportional hazards model, we denote the baseline survivor function for the combined placebo arms by $S(t)$ and the regression coefficients for the two treatments by β_P and β_C. For two-year follow-up:

$$\begin{aligned} p_0 &= 1 - S(2) \\ p_P &= 1 - S(2)^{\exp(\beta_P)} \\ p_C &= 1 - S(2)^{\exp(\beta_C)}. \end{aligned}$$

In our experience, clinicians consider the efficacy of a treatment as a relative risk reduction conditional on the underlying placebo rate. For example, if the incidence of toxoplasmosis for the placebo group over two years is known to be 15%, then the incidence for the pyrimethamine group might be thought most likely to be 10%; whereas if it became known that the incidence on placebo was in fact 30%, the incidence on pyrimethamine might be thought most likely to be 20%. In both cases the relative risk is approximately 0.67.

Based on these considerations the elicitation scheme asks the expert for a "best guess" of the incidence on placebo, denoted by $\hat{p}_0$. The joint probability distribution of p_P and p_C is elicited conditional on $p_0 = \hat{p}_0$. To structure a parametric approach we utilize the property that for the proportional hazards model, the complementary log-log transformation of p_P and p_C has a range of the whole real line and is linear in the regression coefficients. Specifically,

$$\log\{-\log(1-p_P)\} = \beta_P + \log\{-\log(1-p_0)\}$$

$$\log\{-\log(1-p_C)\} \;=\; \beta_C + \log\{-\log(1-p_0)\}.$$

We use a parametric bivariate distribution for β_P and β_C conditional on $p_0 = \hat{p}_0$, and we assume:

$$f(\beta_P, \beta_C | p_0 = \hat{p}_0) = f(\beta_P, \beta_C). \tag{1}$$

To check assumptions, the elicitation process can be repeated using a different follow-up interval and the elicited distributions on the β's can be compared. Under the proportional hazards assumption the prior distribution should not depend on the interval. In our example, we repeat the process by asking the expert about predictions for both two and three years of follow-up and compare the distributions on the regression coefficients.

4.2 The parametric bivariate distribution

Although it seems natural to use a bivariate normal distribution for (β_P, β_C), this choice is not particularly helpful. For this distribution on the β's, the corresponding distribution of p_P and p_C can have singularities at zero and one (Meinhold and Singpurwalla, 1987). Realistic prior opinions on the probabilities are unlikely to have such singularities. We therefore take (β_P, β_C) to have a Type B bivariate extreme value distribution as given in Johnson and Kotz (1972, pp. 251–255),

$$\begin{aligned} f(x, y|m) \;=\; & e^{m(x+y)}(e^{mx} + e^{my})^{-2+1/m} \\ & \times\{m - 1 + (e^{mx} + e^{my})^{1/m}\}exp[-(e^{mx} + e^{my})^{1/m}] \end{aligned}$$

for $m \geq 1$. Note that $m = 1$ corresponds to independence of β_P and β_C (and therefore of p_P and p_C) and $m > 1$ produces positive correlation. For a distribution with location parameters μ_P and μ_C and scale parameters σ_P and σ_C the probability density function is obtained by substituting

$\sigma_P^{-1}(x - \mu_P)$ for x and $\sigma_C^{-1}(y - \mu_C)$ for y in the above expression and multiplying the result by $\sigma_P^{-1}\sigma_C^{-1}$. Marginal distributions are extreme value with scale and location parameters (μ_P, σ_P) and (μ_C, σ_C).

The marginal distribution of p_P has probability density function:

$$f(p_P|\mu_P, \sigma_P, p_0 = \hat{p}_0) = \frac{1}{(1-p_P)\{-\log(1-p_P)\}\sigma_P} exp\{-exp[\frac{log\{-\log(1-p_P)\} - \mu_P}{\sigma_P}]\}. \quad (2)$$

The corresponding distribution of $[-\log(1 - p_P)]$ is Weibull and of course, $\log[-\log(1 - p_P)]$ is extreme value.

To specify the distribution on p_P and p_C the expert specifies the upper and lower quartiles for p_P and p_C. These quartiles are used to calculate initial values for μ_P, μ_C, σ_P, and σ_C. Starting with $m = 1$, i.e. independence, the expert is presented with plots of each marginal distribution and a dialog box with five sliders. The first four sliders adjust the four parameters μ_P, μ_C, σ_P and σ_C. As the values are changed, the plots change accordingly. The sliders allow the expert to adjust interactively the specified values and see the consequences directly, in terms of the marginal distributions for p_P and p_C.

The quantities p_P and p_C are likely to be dependent with a positive correlation. Our current approach does not address negative correlation. To pick an initial value of m, the expert specifies the probability that both probabilities p_P and p_C are larger than their respective marginal medians. If p_P and p_C are independent, this probability is 0.25. For the type B extreme value distribution the probability is $(\frac{1}{2})^{2^{1/m}}$ (Johnson and Kotz 1972, p. 253) and the probability must be between 0.25 and 0.50.

The parameter m can be adjusted using the fifth slider. Graphical feedback on the joint dis-

tribution is provided by a contour plot of the joint prior distribution of p_P and p_C with regions corresponding to approximate 20%, 40%, 60% and 80% density regions based on a chi-square approximation to the log likelihood. Changing the value of m does not change the marginal distributions but it does change the contour plot.

Figure 1 shows the screen of the workstation running the program to elicit a parametric approximation to a prior opinion. The figure shows the two marginals, a contour plot of the joint density and the dialog box with the five sliders.

The single expert on whom this bivariate method was tried reported that his beliefs were represented by this parametric family. Methods are being developed for extending this scheme to allow for non-parametric adjustments using more general copula combinations of marginal distributions (see eg. Shih, 1990).

4.3 The univariate version

The univariate version of the elicitation method is very similar to but more flexible than the bivariate version. The expert specifies the distribution of p_P by first specifying quartiles and then adjusting interactively within the parametric family given by Equation 2. Only two sliders are required, one for each of μ_P and σ_P. The expert subsequently adjusts the plot by freehand using a mouse.

5 Experience with prior elicitation

The elicitation process has been used with five individuals with quite different backgrounds whom we will refer to as A, B, C, D and E. A, B and C are physicians: A practices at a university AIDS clinic, B specializes in the neurological manifestations of AIDS and HIV infection, and C practices

at a community clinic. D and E are non-physicians involved in research in AIDS clinical trials: D manages studies and E is an infectious disease epidemiologist.

In eliciting the opinions we used a written script so that questions were asked in a standard form.

5.1 Bivariate elicitation

Expert A provided bivariate information before the clindamycin arm was terminated. The other individuals provided opinions after this time and only gave their beliefs about the effect of pyrimethamine.

The initial step requires a best guess value of p_0. Expert A experienced difficulty in treating death without TE as an uninformative censoring event. Expert A decided to partition the population into four groups: people who will die without experiencing TE, people who die from an initial attack of TE, people who experience TE and recover from the first attack and people who do not die and also do not experience TE. A's best guess for a large group of patients is given in Table 1 for the placebo group and the clindamycin group. The increased incidence of death from non-TE causes in the treatment group in Table 1 results from reducing the risk of TE and thus allowing other clinical events to cause death. Expert A could not agree with treating death from non-TE related causes as uninformative censoring.

In light of this we proceeded to elicit A's opinion using the combined endpoint of death or TE. A's best guess was $\hat{p}_0 = 0.69$. Conditional upon this, A specified initial values for the mode, 25th percentile and 75th percentile for the distribution of p_C: 0.60, 0.40 and 0.80 respectively. A adjusted the plot derived from these initial values ending up with $\mu_C = -0.15$ and $\sigma_C = 0.25$. We then discussed the pyrimethamine treatment effect. Expert A believed that the marginal distribution

was essentially equal to that for clindamycin.

Following the script, we explained the idea behind the joint probability that both proportions would be above their medians and A reported a joint probability of 0.45 (this corresponds to strong dependence). Figure 1 shows the distribution specified by this physician as displayed on the workstation screen. Expert A found the plots extremely helpful and suggested some modifications to the program.

5.2 Univariate elicitation

The remaining four experts specified their beliefs using the univariate version of the method for p_P. None of these individuals had the difficulty experienced by A of treating death as uninformative censoring. Each individual had a very different approach to answering the questions.

Experts B, C and E had no difficulty providing initial quartiles and a probability distribution with the process completed in about 30 minutes. Expert D took much longer, over an hour. and needed to make several side calculations and assumptions about patient attributes. Specifically, D assumed that among the patient population CD4 counts were approximately uniformly distributed below 200. Other factors that influenced D were that the treatment could have some efficacy in preventing other opportunistic infections such as *Pneumocystis carinii pneumonia* (PCP).

The values of $\hat{p}_0$ for B, C, D and E were 0.20, 0.75, 0.18 and 0.65 respectively. The four distributions elicited are given in Figure 2, along with the expert A's marginal distribution for p_P.

The elicited distributions were graphs of a density function, corresponding to a continuous distribution. Each graph was stored using a 32 point representation with uniformly spaced points on the x-axis. This was converted to a 31 point discrete approximation with equal probability at each point, not necessarily equally spaced. The approximation is based on a linearly interpolation

of the 32 point representation.

5.3 Distributions on the regression coefficient

The distributions on the regression coefficient, β_P, were calculated from the discrete approximation. The distributions on β_P for all five experts are given in Figure 3 as probability histograms. Recall that the distribution for A corresponds to a different endpoint than that considered by the other four.

5.4 Checking assumptions

In the elicitation process we assumed that the proportional hazards model holds, that the prior distribution on the regression coefficients are independent of that on the underlying hazard and that the independence assumption in Equation 1 holds. To check the assumptions the process was repeated for three experts by asking about three year in addition to two year predictions. The elicited distributions on β_P for C, D, and E are shown in Figure 4. Clearly the assumptions do not hold as the distributions are not sufficiently close. Further research is necessary to address this issue.

6 Discussion

6.1 Lesons learned about the elicitation process

Through this experiment in elicitation we learned that:

1. The dynamic graphical displays did help the experts to visualize probability distributions and they did provide useful instant feedback.

2. A written script was developed, and improved upon after the first session, to give clarity and consistency to the questions asked. Having a clear well defined outline questions to ask of the expert was extremely important.

3. Experts are very different both in terms of opinions and how they parse the problem. What was easy for one expert was hard for another. Developing a single method for elicitation that is good for everybody will be difficult, if not impossible.

4. Even though the process only requires specification of a best guess value for p_0 experts wanted to report their uncertainty about it.

5. Extreme percentiles may be better than quartiles. Several experts started from "95% intervals" when asked for their quartiles. The 2.5th and 97.5th percentiles may therefore be easier for them to think about, as approximate bounds, than the 25th and 75th.

6. A satisfactory endpoint is essential but may be difficult to decide upon without exploring the data.

7. Some check of the proportional hazards and other assumptions is required. The distributions obtained from asking about three year predictions were different from those obtained from two year predictions.

In summary this elicitation process is not perfect, but it is a first attempt to develop a method.

6.2 Lessons learned about the toxoplasmosis trial

At the completion of the trial we learned:

1. None of the experts predicted the very low rate of TE experienced in the trial. Out of the 396 people in the trial only 12 out of 264 on pyrimethamine got TE and only 4 out of 132 on placebo got TE. The low rate is believed to be partly due to the prophylaxis treatment for PCP. All patients were taking some drug as PCP prophylaxis throughout the trial. Among patients receiving the drug Bactrim, TE was extremely rare and so Bactrim may also have a prophylactic effect for TE in addition to PCP. During the time the trial was running, other studies indicated that Bactrim may be an effective PCP prophylaxis and many patients switched to Bactrim during the trial.

2. All five experts put high probability on a large beneficial effect of the treatment. They were wrong and the rate of TE was slightly higher in the pyrimethamine group than in the placebo group.

3. In the pyrimethamine group 46 out of 264 people died compared to only 13 out of 132 in the placebo group. The death rate was higher in the pyrimethamine group. The data indicated that death was informative and the analysis should include death in the endpoint. The assumption made by the protocol team at the design stage of the trial that there would be no difference in the non TE death rates in the different arms, was wrong. Four of the five prior distributions were elicited without death in the endpoint.

4. Using death or TE as an endpoint the proportional hazards model did not provide a good fit to the data. The hazard for death in the pyrimethamine group increased over follow-up compared to the placebo group.

Bactrim becoming the choice for standard prophylaxis for PCP during the trial illustrates that standard care for AIDS changes rapidly and might well have changed the outcome of the trial.

The changing nature of standard care and the large number of drugs many AIDS patients take introduces new challenges to the statistical analysis of an AIDS clinical trial: this is discussed in detail in Ellenberg, Finkelstein and Schoenfeld (1992). These challenges are present in either a Bayesian or frequentist approach.

Knowledge about AIDS is changing rapidly. Information is accrued during a trial from sources independent of the trial. Experts will accrue information from their clinical practice and elsewhere during a trial. In addition, other trials may yield information: trials for PCP prophylaxis and some French trials for TE prophylaxis were concurrent with our trial. Information accrued concurrently could be incorporated into either the likelihood or the prior distributions.

In order to specify a prior distribution the model that will be used for analysis must be specified. The data may indicate an alternative model. Perhaps one prior distribution and one model could be used at the design stage and, subsequently, when the statistician decides on a model of choice for an analysis, either interim or the final, the statistician could return to the expert, who remains blinded to the data, and elicit a prior distribution for that model. The prior specification of a model is also required for a correct frequentist analysis.

This excercise has illustrated some of the practical problems in implementing a Bayesian approach to a clinical trial.

6.3 The need for an elicitation method

Even if prior distributions are not accepted for analysis and monitoring they are clearly useful in the design stage to decide on a sample size, to decide whether or not the trial is feasible and to decide on the ethics of the trial.

In the design of this trial the protocol team used their best guesses for the magnitude of the

rates and effects to compute a sample size. The explicit incorporation of prior uncertainty into the design process would be beneficial. In addition, at the design stage, if a physician believed that an arm of a clinical trial was harmful to a patient then it would be unethical for that physician to enroll patients in a trial involving the possible use of that treatment. In this trial the beneficial effect of the prophylaxis must be large enough to outweigh the potential toxicity and side effects of the drug before it would be used routinely. The quantification of prior opinion is required for an individual to decide whether it is ethical to participate in the trial: either as a physician or as a patient.

Prior distributions are also required for Bayesian monitoring of a trial. If a treatment is to be widely used, the results of a clinical trial must be convincing to many different people: in particular to the physicians prescribing it and to the people receiving the treatment. A method for stopping a trial could be to stop when a majority of this community of interested parties would agree on the practical implications of the results. It is therefore appropriate to elicit prior opinions from as many such people as possible in order to determine when to stop the trial.

Such Bayesian monitoring of this trial using such a method is discussed in Carlin et al (1992); see also Kadane (1986), Louis (1992), Carlin and Louis (1992).

6.4 Discussion

Bayesian methods are becoming increasingly recognized as appropriate for biostatistical problems and clinical trials, see for example Breslow (1990). The implementation of Bayesian methods requires the specification of prior distributions. Prior distributions are hard to specify, however, and little guidance is available in the literature for elicitation in clinical trials.

Acknowledgements

We are grateful to Bradley Carlin for valuable input and to Keith Henry, Carlton Hogan, Sandra Melnick, Frank Rhame and Scott Strickland for their opinions. This research was supported in part by National Institute of Allergy and Infectious Diseases Contract NO1-AI-05073.

The program and the script are available from the first author (e-mail: kathryn@stat.umn.edu).

References

BRESLOW, N. (1990). Biostatistics and Bayes. *Statistical Science* 5, pp. 269–298.

CARLIN B., CHALONER, K. CHURCH, T. LOUIS, T.A. AND MATTS, J.P. (1992). Bayesian approaches for monitoring clinical trials wity an application to Toxoplasmic encephalitis prophylaxis. Manuscript.

CHALONER, K. M. AND DUNCAN, G. T. (1983). Assessment of a beta prior distribution: PM elicitation. *Statistician* **32**, pp. 174–180.

COX, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* bf 34, 187-202.

ELLENBERG, S.S., FINKELSTEIN, D.M. AND SCHOENFELD, D.A. (1992). Statistical issues arising in AIDS clinical trials. *Journal of the American Statistical Association* **87**, pp. 562–569.

FREEDMAN, L. S. AND SPIEGELHALTER, D. J. (1983). The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *Statistician* **32**, pp. 153–160.

FREEDMAN, L.S. AND SPIEGELHALTER, D.J. (1992). Application of Bayesian statistics to decision making during a clinical trial. *Statistics in Medicine* **11**, pp. 23–35.

GARTHWAITE, P. H. AND DICKEY, J. M. (1990). Quantifying expert opinion in linear regression models. *Journal of the Royal Statistical Society, Series B,* **50**, pp. 462–474.

GARTHWAITE, P. H. AND DICKEY, J. M. (1991). An elicitation method for multiple linear regression models. *Journal of Behavioral Decision Making* **4**, pp. 17–31.

GRANT, I. H., GOLD, J. W. M., ROSENBLUM, M., NIEDZWIECKI, D. AND ARMSTRONG, D. (1990). *Toxoplasma gondii* serology in HIV-infected patients: The development of central nervous system toxoplasmosis in AIDS. *AIDS* **4**, pp. 519–521.

JOHNSON, N. L. AND KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* New York: Wiley.

JACOBSON, M.A., BESCH, C.L., CHILD, C.C., HAFNER, R., MATTS, J.P., MUTH, K., WENTWORTH, D.N., DEYTON, L., AND CPCRA (1992a). Toxicity of clindamycin as prophylaxis for AIDS-associated toxoplasmic encephalitis. *Lancet* **339**, pp. 333–334.

JACOBSON, M.A., BESCH, C.L., CHILD, C.C., HAFNER, R., MATTS, J.P., MUTH, K., WENTWORTH, D.N., NEATON, J., DEYTON, L., AND CPCRA (1992b). Prophylaxis with pyrimethamine for toxoplasmic encephalitis in patients with advanced HIV disease: results of a randomized trial. *Manuscript in submission.*

KADANE, J. B. (1986). Progress toward a more ethical method for clinical trials. *The Journal of Medicine and Philosophy* **11**, pp. 385–404.

KADANE, J. B., DICKEY, J. M., WINKLER, R. L., SMITH, W. S. AND PETERS, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association* **75**, pp. 845–854.

KALBFLEISCH, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B,* **40**, pp. 214–221.

LOUIS, T.A. (1992). Comment on a paper by Ellenberg, Finkelstein and Schoenfeld. *Journal of the American Statistical Association* **87**, pp. 576–579.

LOUIS, T.A. AND CARLIN, B. (1992). Robust Bayesian approaches for monitoring clinical trials. Manuscript.

MEINHOLD, R.J. AND SINGPURWALLA, N.D. (1987). A Kalman-filter smoothing approach for extrapolations in certain dose-response, damage-assessment and accelerated-life-testing studies. *The American Statistician* **41**, pp. 101–106.

SHIH, J.H. (1990). Models and analysis for multivariate failure time data. University of Minnesota, Division of Biostatistics, Research Report.

SPIEGELHALTER, D. J. AND FREEDMAN, L. S. (1988). Bayesian approaches to clinical trials. *Bayesian Statistics 3*, pp. 453–477, Oxford University Press.

SPIEGELHALTER, D. J., FREEDMAN, L. S. AND PARMAR (1992). Bayesian approaches to randomsed trials. Manuscript.

TIERNEY, L. (1990). *LISP-STAT—An Object-Oriented Environment for Statistical Computing and Dynamic Graphics.* New York: John Wiley.

WINKLER, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association* **62**, pp. 1105–1120.

	die without TE	die from TE	recover from first TE	alive and no TE
placebo	51%	9%	9%	31%
clindamycin	54%	3%	3%	40%

TABLE 1: Expert A's best guesses.

List of figures

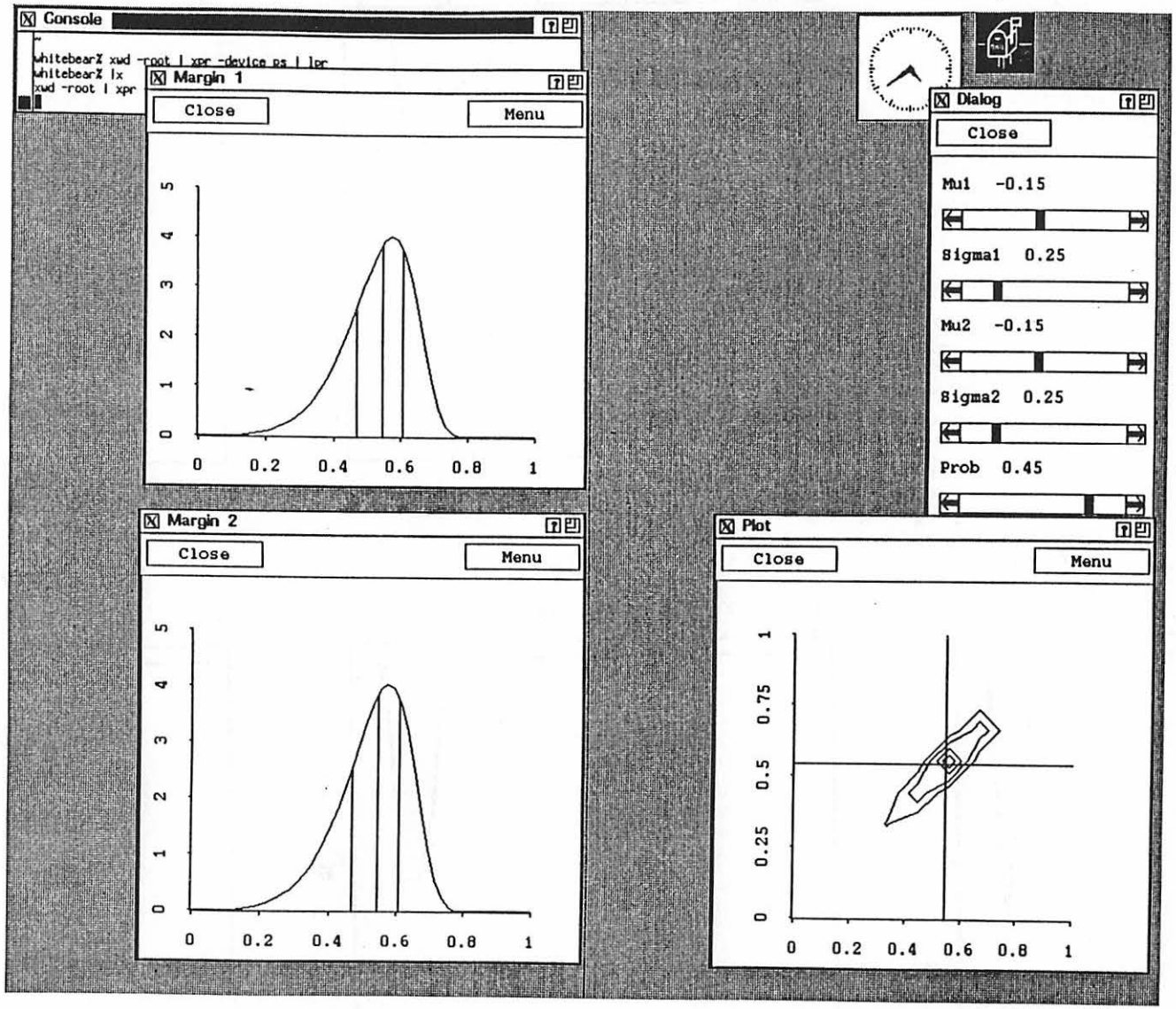

Fig 1: the screen for the bivariate elicitation for expert A.

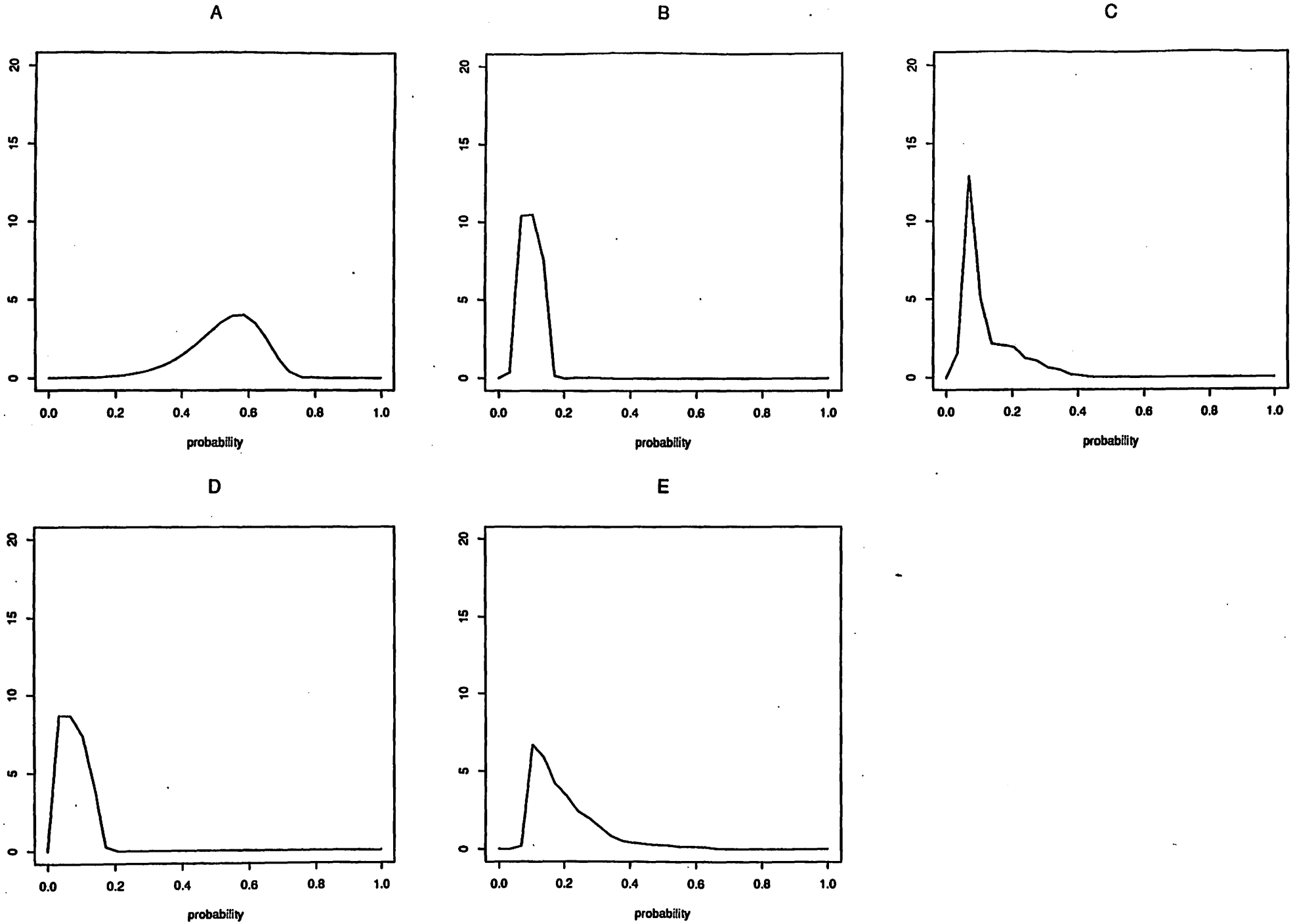

Fig 2: the prior distributions on the probabilities

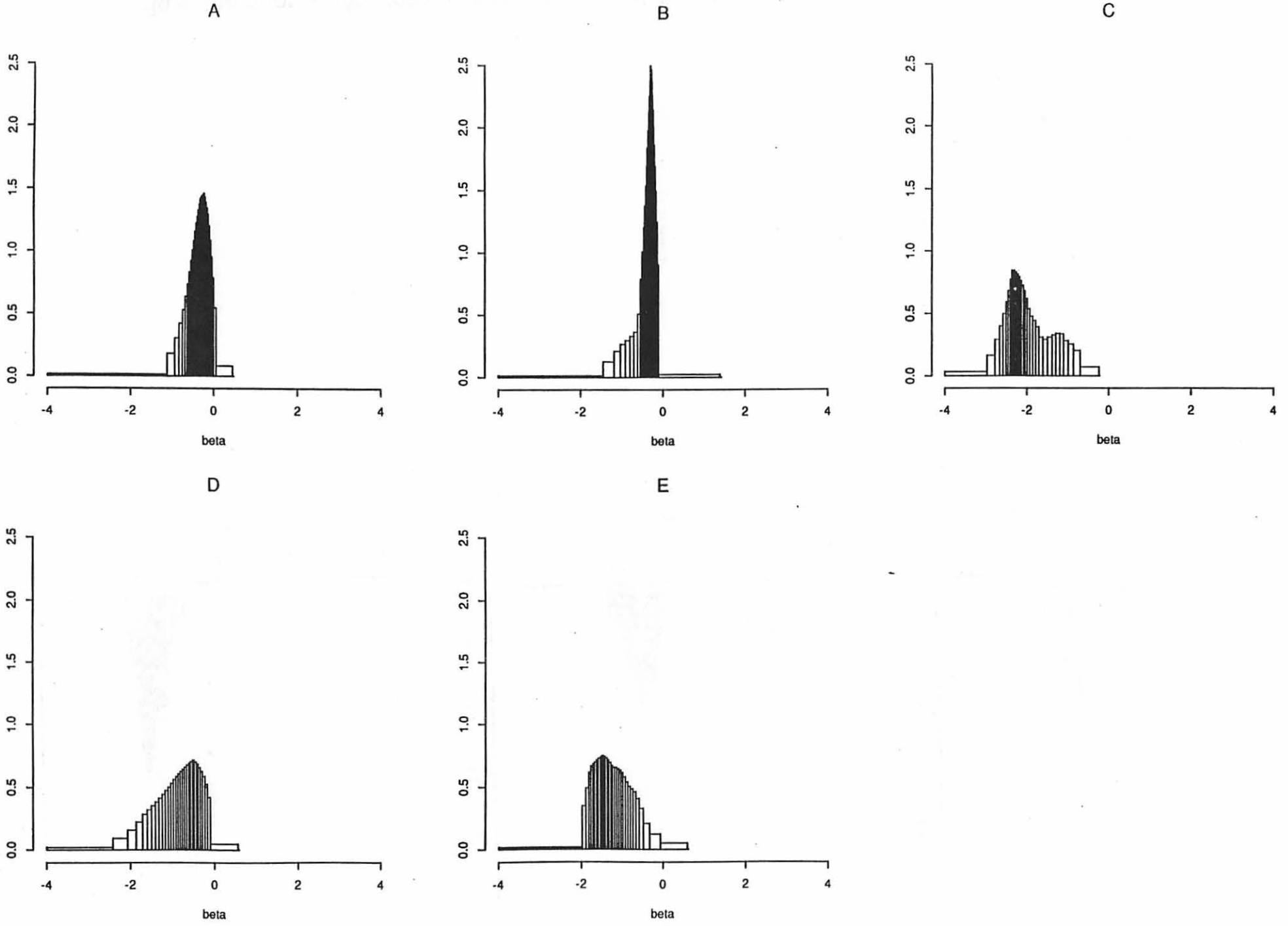

Fig 3: prior distributions on the regression coefficient from 2 year probabilities

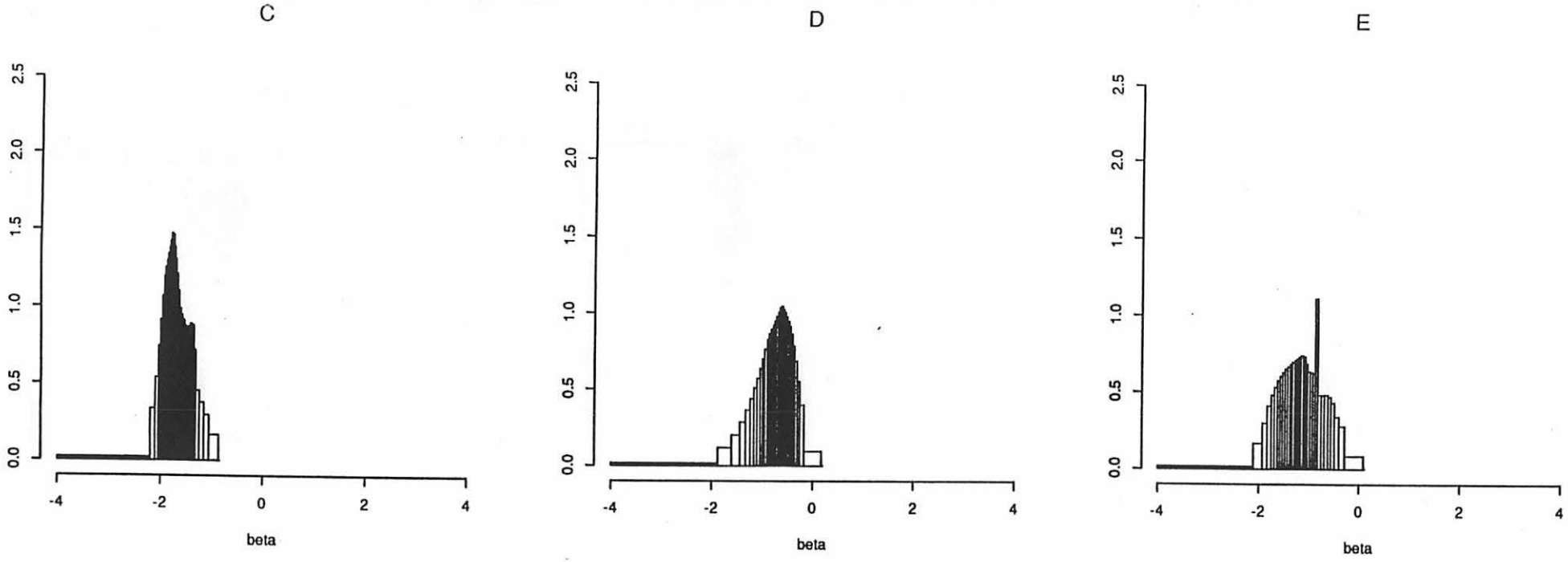

Fig 4: the prior distributions on regression coefficient from 3 year probabilities for C, D and E