Open Access Medical Statistics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A Bayesian approach to the analysis of clinical trial data using logistic regression: example from a randomized placebo-controlled crossover trial of propranolol for migraine prevention

This article was published in the following Dove Press journal:
Open Access Medical Statistics
15 June 2013
Number of times this article has been viewed

Christopher Dardis
Yogesh Moradiya
Arnold Eggers

SUNY Downstate Medical Center, Brooklyn, NY, USA

Abstract: Bayesian methods enable the "prior" (or informative) beliefs of an audience to be combined with the results of a clinical trial to arrive at a final "posterior" belief. This example concerns previously published data from a double-blind placebo-controlled trial of propranolol to reduce the number of episodes of migraine, where subjects were crossed-over after 3 months of treatment. The informative prior range was supplied by an educated audience (members of our Faculty of Neurology) who were given review papers on propranolol in migraine prophylaxis and placebo responses in migraine trials. We used logistic regression models to try to predict those whose symptoms improved (based on treatment or on the time period under consideration; ie, the first or second 3-month period, or based on both factors considered together). The posterior was generated using the Markov–Chain Monte–Carlo methods. For the original dataset, the Bayesian posteriors tended to be more tightly defined than those with no prior or minimally informative prior beliefs, thus yielding firmer conclusions in light of the trial. When compared with a larger dataset (which was generated from the original, but was arrived at by multiplying the number of observations by 10), the influence of prior beliefs was much less marked, but the posteriors did tend to be marginally more narrowly defined. This finding is in keeping with existing work on Bayesian methods, highlighting their value in aiding interpretation of trials with a small number of observations.

Keywords: Bayesian, migraine, randomized trial, placebo, propranolol

Introduction

Bayesian methods can provide a collaborative way of bridging beliefs and a given dataset. The methods are useful where an audience has a strong (or any) prior belief about something; the data will then alter/modify those beliefs, thus generating a posterior belief (ie, a combination of the prior belief and the results from the data). If the prior belief is loosely defined, the posterior belief will be heavily weighted towards the data; if the prior belief is narrowly defined, then the data will have little impact. In addition, if there is a large amount of data, the posterior belief will be more heavily influenced by the data.

Bayes theorem was published posthumously in 1763.[1] Practical implementation has, until recently, been hindered by barriers to specialized technical knowledge and by computational limitations. However, with the development of the Bayesian inference Using Gibbs Sampling software from 1989 on, (initially for Windows; Microsoft Corporation, Redmond, WA, USA) and the release of a cross-platform

Correspondence: Christopher Dardis
Department of Neurology, Barrow Neurological Institute, 350 W. Thomas Road, Phoenix, AZ 85013, USA
Tel +602 406 6262
Fax +602 406 6260
Email christopherdardis@gmail.com

equivalent Just Another Gibbs Sampler (JAGS) in 2007, the field has become accessible up to a wider audience.[2,3] Excellent introductory textbooks on these methods are available.[4,5]

Bayesian approaches are finding increased acceptance in medicine, in fields such as dose finding, efficacy monitoring, toxicity monitoring, diagnosis/decision making, and studying pharmacokinetics/pharmacodynamics.[6] They have received approval recently in the analysis of crossover designs for clinical devices.[7] They are receiving increasing attention for the analysis of clinical trials.[8] Their value in analyses involving binomial (either/or) outcomes, such as our example, has previously been highlighted.[9]

These methods require the availability of original, "raw" data for their application. This allows for the subsequent revision of a posterior belief in the light of new data; it also allows different audiences to have different interpretations of the same data. Traditionally, those involved in clinical trials have been reluctant to disclose this raw data. This may have originally stemmed from the limitations that a paper-based system of sharing information imposed (ie, traditional journals); however, with the widespread availability of inexpensive electronic storage and rapid dissemination (ie, through the Internet), this unwillingness can no longer be so readily justified.[10] Another concern has been that "raw" data may breech the confidentiality of the patient–doctor relationship. However, over time, the balance of opinion from regulators has tended to shift in favor of sharing the original data.[11]

The clinical utility of propranolol for migraine prophylaxis is now well established, although controversy remains regarding the degree/frequency of response. The issue of placebo response varying in accordance with time period has long been recognized as potentially problematic in crossover trials. This has not generally resulted in discussion of this potential source of bias when reporting such trials, particularly in the context of migraine prophylaxis. We focused on a paper from which raw data could be readily reconstructed, allowing for the examination of these questions in more detail.

Methods

The original paper described a randomized, placebo-controlled trial of propranolol to reduce migraine frequency.[12] Nineteen participants (nine men, ten women) aged 20–60 years, all of whom "were recognized therapeutic management problems,"[11] were randomized to receive propranolol (n = 8) or placebo (n = 11) for 3 months. At the end of this period, they were "crossed-over" so that those on propranolol got placebo and vice versa. The authors had originally planned to include 25 patients, but six "failed to complete the study for reasons unrelated to the trial drug per se."[11] The dose of propranolol was 20 mg four times/day and the placebo used was mannitol. Subjects were not told that they would receive placebo for one-half of the study. Participants attended at four weekly intervals. They were allowed to use symptomatic treatment for migraines, but prophylactic use of ergotamines or methylsergide was prohibited. No explicit mention of ethical approval or consent was made, although we note that the US Food and Drug Administration's Bioresearch Monitoring Program was not introduced until 5 years after the trial data we focused on was published.[13]

The authors divided outcomes into "excellent" (nearly all headaches absent after the first week of study), "good" (more than a 50% reduction in the frequency or severity of headaches), "fair" (minimal symptomatic improvement), and "no effect." For the sake of simplicity, we looked at a dichotomous outcome – excellent or good versus fair or no effect. By a process of deduction, we were able to recast the presented data into 38 observations, each containing one subject (1 to 19), the treatment (propranolol versus placebo), and the time period under consideration (first 3 months versus second 3 months) and the response (better or not). An additional file shows this dataset. (See n19.prn in Supplementary File A). Column headers: s = Subject [1–19], x1 = on propranolol [0 = no, 1 = yes], x2 = time period [0 = first 3 months, 1 = second 3 months], y = response [0 = no effect/fair, 1 = good/excellent]).

The audience that was consulted for their prior beliefs consisted of the faculty (n = 29) and residents (n = 31) in the Department of Neurology at our institution. They were presented with an online form of the survey and a paper copy was distributed at our weekly grand rounds. (See Supplementary File B). The members of the audience were also furnished with full texts of a number of review papers dealing with this subject.[14–16] In total there were eight responses; this may have reflected the group's unfamiliarity with the purpose and methods of the study, or they may have been concern about giving the "wrong" answer, as well as about the time pressures inherent in an academic hospital setting. Since the responses were anonymous, it is not possible to speculate further regarding the differences in the respondents' level of expertise in the treatment of migraine, although all had at least some experience (at least two responses came from faculty members, and two from residents in the department). It bears noting that at the time of the survey there were no members of the department

who were certified by the United Council of Neurological Subspecialties in Headache Medicine.

The first issue was to generate informative prior beliefs from the survey responses. For each question (eg, "What was the probability, do you think, that those on placebo got better?"), we took each individual survey response as indicative of a lower and upper 95% confidence interval (CI). (Results for responses indicating upper and lower CIs are shown as .prn files in Supplementary File A. In all of these files there are 8 rows, corresponding to the 8 subject's individual answers. All responses are expressed as a percentage. For Tx.prn (treatment), column headers are: x1 = upper limit for CIs on treatment, x2 = lower limit for CIs on treatment, x3 = upper limit for CIs with no treatment, x4 = lower limit for CIs on treatment. For T.prn (time period), column headers are: x1 = upper limit for CIs for first 3 months, x2 = lower limit for CIs for first 3 months, x3 = upper limit for CIs for second 3 months, x4 = lower limit for CIs for second 3 months. For TxT.prn (treatment and time considered together), column headers are: x1 = upper limit for CIs for placebo in first 3 months, x2 = lower limit for CIs for placebo in first 3 months, x3 = upper limit for CIs for propranolol in first 3 months, x4 = lower limit for CIs for propranolol in first 3 months, x5 = upper limit for CIs for placebo in second 3 months, x6 = lower limit for CIs for placebo in second 3 months, x7 = upper limit for CIs for propranolol in second 3 months, x8 = lower limit for CIs for propranolol in second 3 months.)

We took these 95% CIs as being representative of a beta distribution (ie, with range of 0 to 1), indicating probability. We were able to generate shape parameters for this distribution based on the given 95% CI. In some cases it was necessary to modify the responses as the function we were using to generate the shape parameters of the distribution (LearnBayes::beta.select) would only work if the upper and lower CIs were >4% different.[17] One responder gave the same answer for both upper and lower CIs; this was changed from 70% to 68%–72%. For the same reason, 73%–75% was changed to 72%–76%, as this was the narrowest interval we could model. Also 70%–100% was changed to 70%–99.99%, as the function could not work with a value of 100%.

Calculations are shown as Tx.R, T.R and TxT.R in Supplementary File A. These will reproduce the results in the tables and give graphical output, examples of which are shown in Figures 1–4. In each case, we began with descriptive statistics – comparing the 95% CIs from an informative prior with those from a minimally informative (Jeffreys) prior. Jeffreys prior is a beta distribution with shape parameters alpha = 0.5, beta = 0.5. A large sample from such a distribution will show a median of 50% with 95% of the values in the range 0.6%–100%. This is a well-established approach to determining a 95% CI for a proportion with minimal prior information.[18] For the informative prior, we created a mixture of the eight beta distributions representing participants' answers. We used LearnBayes::binomial.beta.mix to generate

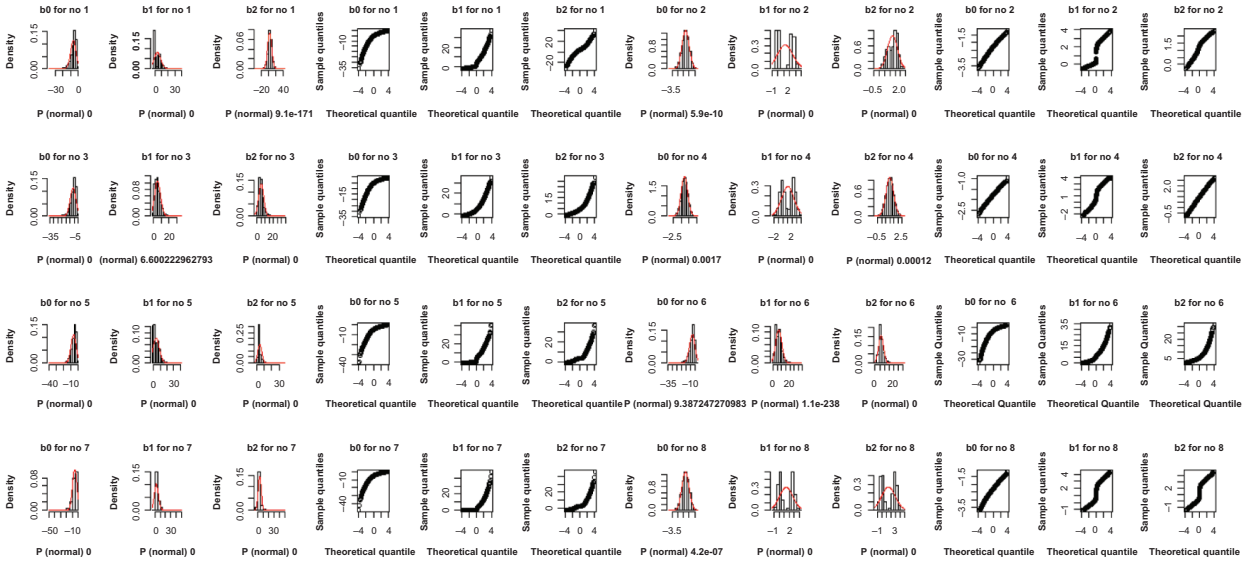

Figure 1 Distribution of priors with normal curves.

Notes: Example showing priors for coefficients (b0 (intercept), b1 (treatment) and b2 (time)) in a logistic model with two predictors, modeled as normal distributions. These coefficients were generated by sampling from beta distributions (representing participants' survey responses) then converting to the values to coefficients. The graphs to the left show the histograms of the sample, with red lines showing the closest normal approximations. The *P*-value below the graphs is the probability that the distributions are truly normal by Kolmogorov–Smirnov–Lilliefors test. Graphs to the right show quantile–quantile plots versus normal distribution (this should be linear if these are close approximations). Rows represent individual participants, two on each row. The graph was generated from TxT.R, treatment and time period considered together.

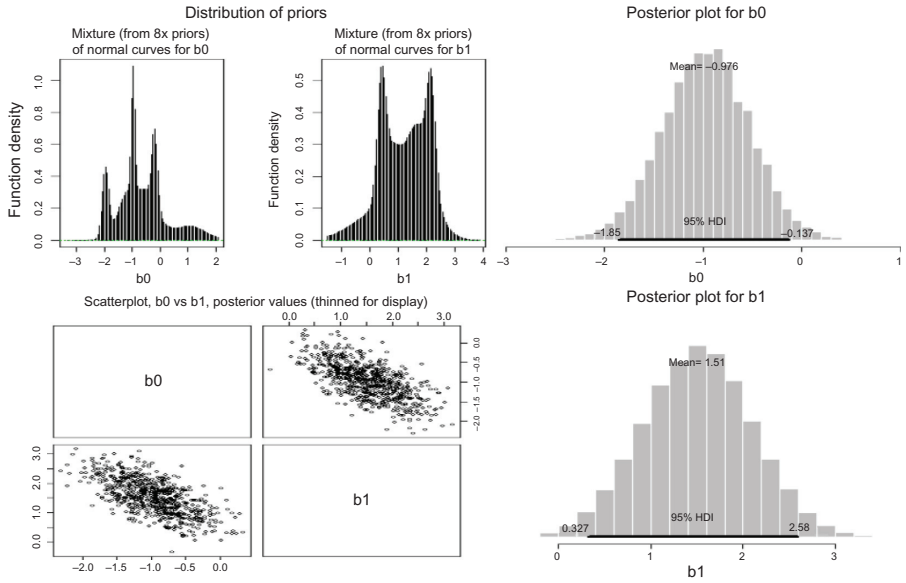

Figure 2 Example showing prior and posterior coefficients from a logistic regression model.
Notes: Upper left panel shows the normal mixture density for priors (ie, a mixture of normal curves). The right-hand panels show histograms of the posterior values of the coefficients (b0 (intercept) and b1 (time) (taken from MCMC samples, with mean and 95% HDIs). The lower left panel shows the correlation of posterior values for coefficients with a scatterplot. The graph was generated from T.R, time period considered alone.
Abbreviations: MCMC, Markov–Chain Monte–Carlo; HDI, high density intervals.

an updated beta mixture (ie, with posterior parameters and mixing probabilities) given the influence of the data (eg, the number of subjects who got better on propranolol). We then generated a sample from this mixture of betas to get a median and 95% CIs (using our function BmixP.R, shown in Supplementary File A).

For predictive statistics, given that the outcome was binary (either/or), the simplest approach to predicting it is logistic regression. We looked at logistic regression models, which would predict the likelihood of a response based on propranolol versus placebo (considered alone), time period (considered alone), and both factors considered together.

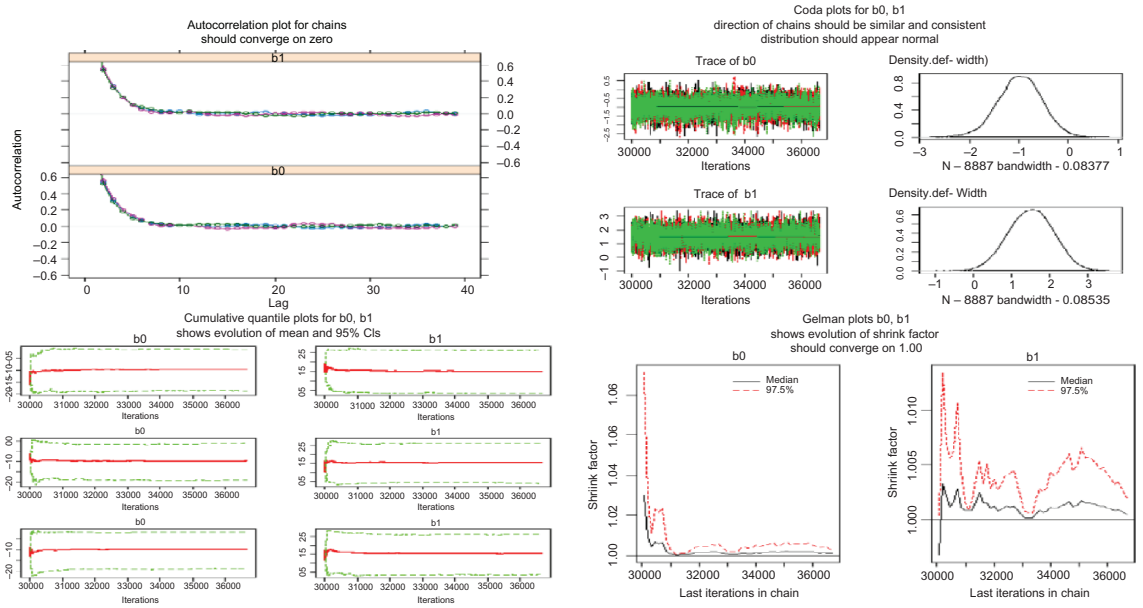

Figure 3 Example confirming the tendency of three MCMC chains to converge in a logistic regression model.
Notes: The upper left panel shows an autocorrelation plot for the coefficients (b0 (intercept) and b1 (time)). The upper right panel shows coda plots; the left side shows the trace of coefficients versus the iteration of MCMC with colors representing individual chains; the right side shows the density plot of the MCMC sample (similar to the histogram with HDIs shown in Figure 2). The lower left panel shows cumulative quantile plots with evolution coefficients and 95% CIs for three chains. The lower right panel shows Gelman plots with evolution of shrink factor converging on 1. The graph was generated from T.R, time period considered alone.
Abbreviations: MCMC, Markov–Chain Monte–Carlo; HDI, high density intervals.

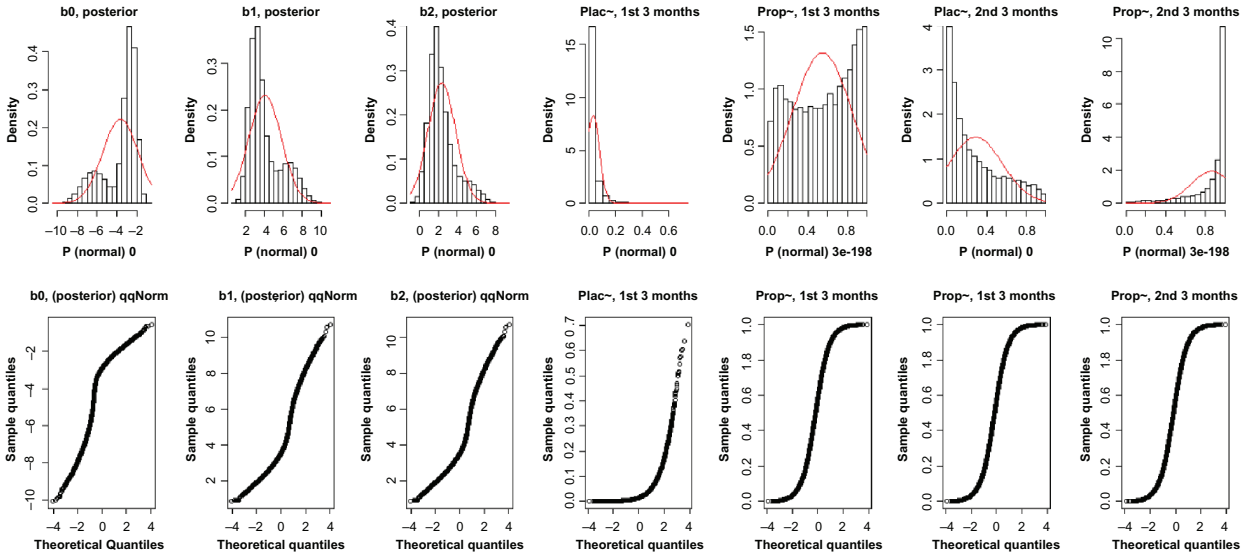

Figure 4 Posterior distributions for treatment and time period.

Notes: Example of posterior coefficients (b0 (intercept), b1 (treatment) and b2 (time)) and corresponding probabilities from a logistic regression model. The graphs above show histograms of samples, with red lines showing the closest normal approximations. The graphs to the left show the coefficients (three in this case); on the right are the probabilities derived from these coefficients (four in this case). *P*-values below the graph are the probability levels indicating that the distributions are truly normal by the Kolmogorov–Smirnov–Lilliefors test. The graphs below show quantile–quantile plots versus normal distributions (these should be linear if they are close approximations). The graph was generated from TxT.R, treatment and time period considered together.

Abbreviations: Prop~,Propranolol; Plac~, Placebo.

We recognize that a more appropriate approach would have been to use multilevel/hierarchical/mixed-effects models to account for intersubject variability; however, we found it difficult to translate equations incorporating such priors into "plain English," which would be easily understood by survey respondents. Likewise, the inclusion of an interaction term (considering treatment and time period together) was not considered further. Finally, it could be argued that the six drop-outs from the trial were "informative" (ie, nonignorable), and a more rigorous approach to accounting for this in modeling a crossover trial should have been adopted, possibly with the imputation of missing values.[19] However, our goal was to keep the questions relatively brief and simple. Although the logistic models fall short of representing these subtleties, we were guided by the principle that "the good is not the enemy of the best".

Regarding mixed-effects models, the interested reader is referred to glmeModels.R in Supplementary File A. The "random-effect" in this case is the person (the study subject), whether certain individuals were innately more predisposed to get better (or were innately more predisposed to migraine) and/or had a greater tendency to get better given particular conditions than others. With only two observations per subject, speculation on the influences is thus necessarily greatly limited. The subject, considered in isolation, was not a significant predictor of improvement. Many mixed models are possible, depending on whether a separate baseline (intercept) chance of getting better is present for each subject, and whether we allow the response (slopes) to treatment and/or to time period to vary on a per-subject basis; interactions are not considered here. Using frequentist methods, these models can be slow to calculate and tend give large standard errors for the relevant coefficients. In general, simpler models are to be preferred. Statistically, this is based primarily on the Akaike Information Criterion and also on the observation that including any random effect tends to diminish the significance of the other coefficients in the model.[20] Clinically, there is no reason to think that responses to propranolol should vary dramatically based on the subject (and even less to think that responses should vary per time period), and there is no reason to favor any of these more complex models over the simpler ones presented here. However, the issue of individual response to treatment has received little attention in the literature on migraine prophylaxis, and it may indeed be that some individuals are particularly migraine susceptible or resistant. This would be modeled as a mixed-effects logistic regression where the baseline/intercept is allowed to vary by subject and the coefficient (eg treatment) remains constant for the group. Examples of how such models may be specified in JAGS with arbitrary and minimally informative priors are given are given in glmeJAGS in Supplementary File A.

In order to fit logistic models incorporating informative priors, we began by generating informative priors for the

coefficients in a logistic regression. We started with the eight beta distributions representing participants' answers. We sampled repeatedly from each distribution and converted these values into values of coefficients in a logistic model (see formsLogReg.R in Supplementary File A). These coefficients we approximated with a normal distribution (done primarily to allow us to work with a JAGS model which incorporates a mixture of normal distributions (available in rjags as dnormix)). Larger sample sizes (eg, >1000) tended to make formal tests of normality (eg, Lilliefors) highly improbable. This is an inevitable tradeoff when generating larger sets with resampling to increase precision.[21] Also, larger samples tended to reduce the variance, although this became asymptotic with samples > 10,000. Overall, we felt that the normal distribution remained a reasonable approximation for the majority of the responses, particularly in the case of a model with one predictor. The prior distribution thus consisted of a mixture of normal distributions based on these results.

Given 38 observations, we thought it reasonable to try to estimate two predictors (which was justified as we had >10 observations/predictor). A prior based on a mixture of normal distributions appeared to be an appropriate approximation for the models with just one predictor.

A caveat with the two-predictor model was converting four levels of probability to three levels for the coefficients. In the logistic model, the predicted probabilities are constrained such that when three are known, the fourth is a given; however, we could not expect survey participants to be so internally consistent. For example, when translating from probabilities to coefficients in a logistic model with two predictors, coefficient b1 can take on two values depending on whether coefficient b2 is 0 or 1 (Table 1). At times, this created a more of a bimodal than a normal distribution (an example from the output of TxT.R to illustrate this is shown in Figure 1). We recognize that a customized distribution and sampler would have offered some improvement here.

We thus generated mixtures of normal distributions (with equally weighted probabilities) and used these as prior distributions for the logistic models. Using a mixture distribution does introduce a lack of consistency that could be overcome by generating posterior probabilities for each individual in isolation. Our goal was to model a consensus opinion, bearing in mind the loss of certainty that this entails.

We next "factored in" the data to generate a posterior distribution for each proportion. This was done using JAGS to generate Markov–Chain Monte–Carlo samples. An example of the graphical output showing the priors and posteriors from the time period (T.R) is shown in Figure 2. An example of diagnostic plots to ensure that the chains tended to converge from the same file is shown in Figure 3.

Finally, we looked at the resulting Markov–Chain Monte–Carlo values for the coefficients and converted these back into probabilities (see formsLogReg.R in Supplementary File A for formulas). An example of output from TxT.R showing the posteriors for the coefficients and for probabilities is shown in Figure 4. We compared these results with those resulting from no prior beliefs and with a weakly informative prior (Cauchy) distribution. To model no beliefs, we took 95% CIs for the coefficients in the regression model and converted these into the relevant probabilities. The Cauchy distribution has been suggested as optimal for logistic regression, although alternatives have been suggested.[22] Its value is most evident where complete separation occurs, as occurred in our example when considering both factors together (ie, zero/null values occur in cross-tables of factor levels by outcome, meaning that in some cases the predictor can perfectly predict the outcome). We followed the suggestion of Gelman et al and used a Cauchy distribution with the parameters location = 0, scale = 2.5.[23]

For each of the above steps, we also simulated ten times the original dataset to demonstrate the influence that a larger amount of data would have on the posterior beliefs.

Table 1 Relationship between four probabilities/outcomes and three coefficients

Probability	Coefficients
First 3 months on placebo	b0
First 3 months on propranolol	b0 + b1
Second 3 months on placebo	b0 + b2
Second 3 months on propranolol	b0 + b1 + b2

Note: This shows that one coefficient (eg, b1) has a different interpretation depending on the value of another (eg, b2).

Results

Survey responses

There are some notable inconsistencies in the survey responses. Considering each factor in isolation, the most striking is that respondents gave a higher likelihood of improvement in both time periods than when propranolol was considered alone (60.3% [50.2%–70.1%] for propranolol alone versus 73.3% [52.1%–91.4%] for first 3 months (not considering effect of propranolol) or 78.3% [53.8%–97.0%] for 2nd 3 months (not considering effect of propranolol); 95% CIs; see Table 2). It should also be noted that the beliefs regarding time period alone are rather similar and ill defined: 73.3% [52.1%–91.4%], 78.3% [53.8%–97.0%]. By contrast,

Table 2 Descriptive statistics for proportions

Factors	Prior	Data	Posterior
Treatment			
Placebo	Jeffreys	10.5	10.5 (2.25–29.7)
Placebo	14.5 (9.8–19.7)	10.5	14.2 (9.7–19.1)
Propranolol	Jeffreys	78.9	78.9 (57.4–92.4)
Propranolol	60.3 (50.2–70.1)	78.9	63.5 (54.5–72.4)
With 10× the number of observations			
Placebo	Jeffreys	10.5	10.5 (6.8–15.5)
Placebo	14.5 (9.8–19.7)	10.5	12.5 (9.3–16.0)
Propranolol	Jeffreys	78.9	78.9 (72.7–84.2)
Propranolol	60.3 (50.2–70.1)	78.9	72.9 (67.7–78.0)
Time period			
First 3 months	Jeffreys	26.3	26.3 (10.8–48.4)
First 3 months	73.3 (52.1–91.4)	26.3	48.5 (32.7–64.4)
Second 3 months	Jeffreys	63.2	63.2 (40.9–81.8)
Second 3 months	78.3 (53.8–97.0)	63.2	68.8 (52.1–84.0)
With 10× the number of observations			
First 3 months	Jeffreys	26.3	26.3 (20.4–32.9)
First 3 months	73.3 (52.1–91.4)	26.3	30.2 (24.0–36.5)
Second 3 months	Jeffreys	63.2	63.2 (56.1–69.8)
Second 3 months	78.3 (53.8–97.0)	63.2	64.0 (57.3–70.5)
Treatment and time period			
First 3 months on placebo	Jeffreys	0	0 (0–20.0)
First 3 months on placebo	9.2 (4.7–14.5)	0	8.5 (4.3–13.4)
First 3 months on propranolol	Jeffreys	62.5	62.5 (29.5–88.1)
First 3 months on propranolol	60.3 (50.2–70.1)	62.5	60.5 (50.8–69.9)
Second 3 months on placebo	Jeffreys	25	25 (5.6–59.2)
Second 3 months on placebo	63.1 (50.4–75.3)	25	58.3 (46.2–70.1)
Second 3 months on propranolol	Jeffreys	90.9	90.9 (64.7–99.0)
Second 3 months on propranolol	63.1 (50.4–75.3)	90.9	67.3 (56.4–78.6)
With 10× the number of observations			
First 3 months on placebo	Jeffreys	0	0 (0–2.3)
First 3 months on placebo	9.2 (4.7–14.5)	0	4.9 (2.5–7.9)
First 3 months on propranolol	Jeffreys	62.5	62.5 (51.6–72.5)
First 3 months on propranolol	60.3 (50.2–70.1)	62.5	61.3 (54.0–68.5)
Second 3 months on placebo	Jeffreys	25	25 (16.5–35.3)
Second 3 months on placebo	63.1 (50.4–75.3)	25	40.6 (32.5–48.9)
Second 3 months on propranolol	Jeffreys	90.9	90.9 (84.5–95.2)
Second 3 months on propranolol	63.1 (50.4–75.3)	90.9	81.6 (75.5–87.2)

Notes: Shows prior beliefs (Jeffreys [uniformed] and informative by survey), results from data and resulting posteriors. Values are the percentages of subjects improved (95% CIs).
Abbreviation: CI, confidence interval.

percentages for treatment alone show clear separation and are more narrowly defined: 14.5% [9.8%–19.7%], 60.3% [50.2%–70.1%].

There is some consistency when considering both factors together, although it is striking that the chance of improvement is thought to be very similar among three groups: those receiving propranolol (either time period) and those receiving placebo in the second 3 months (60.3% [50.2%–70.1%], 63.1% [50.4%–75.3%], 63.1% [50.4%–75.3%]; the latter two values are identical only when rounded to one significant figure). There is consensus surrounding the fact that those receiving placebo in the first 3 months had significantly worse outcomes than any of the other groups (9.2% [4.7%–14.5%]).

Effect of priors on descriptive statistics

Looking at one factor in isolation, when comparing the posteriors for the proportions with uninformative priors, we can see that even where the survey response differed greatly from the data, the resulting posteriors were significantly narrower than those with a minimally informative prior. The best example is that of the first 3 months, where the survey indicated a much higher chance of improvement (73.3% [52.1%–91.4%]) than the data (26.3%), yet the posterior was more tightly defined

Table 3 Predictive statistics from logistic regression models with one predictive variable

Factors	Prior	Data	Posterior
Treatment			
Placebo	None	10.5	10.5 (0.82–24.5)
Placebo	Cauchy	10.5	12.8 (2.5–31.0)
Placebo	14.5 (9.8–19.7)	10.5	12.6 (3.5–24.1)
Propranolol	None	78.9	78.9 (31.1–99.4)
Propranolol	Cauchy	78.9	77.2 (36.8–98.3)
Propranolol	60.3 (50.2–70.1)	78.9	77.1 (43.1–97.9)
With 10× the number of observations			
Placebo	None	10.5	10.5 (6.3–14.9)
Placebo	Cauchy	10.5	10.8 (6.8–15.3)
Placebo	14.5 (9.8–19.7)	10.5	10.8 (6.8–15.2)
Propranolol	None	78.9	78.9 (65.6–89.8)
Propranolol	Cauchy	78.9	78.8 (65.6–89.5)
Propranolol	60.3 (50.2–70.1)	78.9	78.8 (65.7–89.3)
Time period			
First 3 months	None	26.3	10.5 (8.4–45.2)
First 3 months	Cauchy	26.3	28.2 (11.4–48.3)
First 3 months	73.3 (52.1–91.4)	26.3	27.6 (12.1–43.9)
Second 3 months	None	63.2	78.9 (26.2–93.1)
Second 3 months	Cauchy	63.2	61.6 (26.8–90.9)
Second 3 months	78.3 (53.8–97.0)	63.2	63.4 (31.4–88.9)
With 10× the number of observations			
First 3 months	None	26.3	26.3 (20.1–32.6)
First 3 months	Cauchy	26.3	26.5 (20.5–32.9)
First 3 months	73.3 (52.1–91.4)	26.3	26.4 (20.5–32.7)
Second 3 months	None	63.2	63.2 (50.6–75.4)
Second 3 months	Cauchy	63.2	63.0 (50.3–74.8)
Second 3 months	78.3 (53.8–97.0)	63.2	63.1 (50.4–74.8)

Notes: Shows prior beliefs (uninformed, minimally informed [Cauchy], and informed by survey), results from data, and resulting posteriors. Values are the percentages of subjects improved (95% CIs).
Abbreviation: CI, confidence interval.

than that derived with a Jeffreys prior. When 10× original data was considered, the minimally informative priors again tended to be tighter than those including survey responses when considering treatment alone, although this effect was much less marked when considering time period. The same holds true when considering both factors together.

Effect of priors on logistic models

Looking at treatment as a predictor, the posteriors were, as expected, somewhat improved by the addition of an uninformative (Cauchy) prior and improved greatly with the addition of survey priors. The same is broadly true when using time period as a predictor. Even though the survey priors are somewhat at variance with the data, their influence leads to a narrower posterior than the other methods. Considering 10× the number of observations, all of the posteriors were very similar, indicating that this larger volume of data tended to greatly outweigh the influence of the priors.

With both factors considered together, the model with no priors performed poorly, with very wide 95% CIs (eg, first 3 months on propranolol, 58.9% [5.9%–99.6%]). No one on placebo in the first 3 months improved, thus allowing for a perfect prediction of outcomes based on this factor. The Cauchy prior overcomes this difficulty, thereby proving its worth with much narrower posteriors. Except for predicting the "first 3 months on placebo" group, it did better than the survey data as a prior in this regard. This appears primarily due to the discrepancy between the responses and the data in some cases. The Cauchy prior also performed best with 10× number of observations, although here the differences were much less marked.

Discussion

It is already well known mathematically that in small samples, Bayesian conclusions will be more definite than frequentist ones (as long as the prior does not conflict with the data). Perhaps the best example from our study comes from the logistic model that predicts response to propranolol, where the mean (95% CI) narrowed from 78.9% [31.1%–99.4%] to 77.1% [43.1%–97.9%]. This improvement is more striking when considered in terms of odds ratios, since percentage

Table 4 Predictive statistics from logistic regression model with two predictive variables

Factors	Prior	Data	Posterior
Treatment and time period			
First 3 months on placebo	None	0	2.6 (0–11.3)
First 3 months on placebo	Cauchy	0	2.6 (0.3–20.9)
First 3 months on placebo	9.2 (4.7–14.5)	0	2.1 (0–13.0)
First 3 months on propranolol	None	62.5	58.9 (5.9–99.6)
First 3 months on propranolol	Cauchy	62.5	58.9 (14.0–97.5)
First 3 months on propranolol	60.3 (50.2–70.1)	62.5	61.7 (1.9–99.8)
Second 3 months on placebo	None	25	21.4 (0–83.7)
Second 3 months on placebo	Cauchy	25	21.4 (0.6–71.8)
Second 3 months on placebo	63.1 (50.4–75.3)	25	19.2 (0–93.7)
Second 3 months on propranolol	None	90.9	93.5 (36.1–100)
Second 3 months on propranolol	Cauchy	90.9	93.5 (38.0–100)
Second 3 months on propranolol	63.1 (50.4–75.3)	90.9	96.2 (12.4–100)
With 10× the number of observations			
First 3 months on placebo	None	0	2.6 (1.0–4.8)
First 3 months on placebo	Cauchy	0	2.6 (1.2–5.3)
First 3 months on placebo	9.2 (4.7–14.5)	0	3.2 (0.9–5.6)
First 3 months on propranolol	None	62.5	58.9 (33.3–81.1)
First 3 months on propranolol	Cauchy	62.5	58.9 (33.9–79.7)
First 3 months on propranolol	60.3 (50.2–70.1)	62.5	61.2 (33.2–84.0)
Second 3 months on placebo	None	25	21.4 (6.0–40.5)
Second 3 months on placebo	Cauchy	25	21.4 (7.0–40.0)
Second 3 months on placebo	63.1 (50.4–75.3)	25	21.7 (6.4–42.7)
Second 3 months on propranolol	None	90.9	93.5 (83.3–98.9)
Second 3 months on propranolol	Cauchy	90.9	93.5 (82.4–98.9)
Second 3 months on propranolol	63.1 (50.4–75.3)	90.9	93.1 (81.1–99.0)

Notes: Shows prior beliefs (uninformed, minimally informed [Cauchy], and informed by survey), results from the data, and resulting posteriors. Values are the percentages of subjects improved (95% CIs).
Abbreviation: CI, confidence interval.

differences are more significant when occurring close to 0 or 100% (3.74 [0.45–165.7] to 3.36 [0.76–46.6]). This pattern was true across the board – the posteriors generated with informative priors were more tightly defined. The improvements were most striking in the models with one predictor. In models with two predictors, the usefulness of the survey priors was generally less than that a minimally informative (Cauchy) prior. When compared with the larger artificial dataset (generated from the original, but multiplying the number of observations by ten), the influence of prior beliefs in narrowing the posterior was still present, but overall much less marked. It appears that even larger datasets would tend to obscure this difference completely. This is in keeping with existing work on Bayesian methods, underscoring their value in the analysis of trials with a smaller number of observations.[24]

The randomized placebo-controlled crossover trial is an established method for assessing efficacy, particularly for chronic conditions that are not expected to fluctuate greatly over time. The crossover design has been particularly advocated in preference to the parallel-group design for adolescents affected by migraines.[25] The issue of time period being important has not received much attention, and we were struck by the change in response from the first 3-month period to second 3-month period. This was complicated somewhat by more subjects receiving propranolol in the second 3-month period; however, it remained a valuable predictor in the two-factor logistic model, although it was less important than treatment or intercept ($P < 0.05$ by Wald statistic with significantly reduced deviance when included; $P < 0.5$ by Chi-square). It appears that the duration of participation in a clinical trial may (at least when considering only those subjects completing the trial) increase the likelihood of a favorable outcome. There is the possibility that there were carry-over effects for the group receiving propranolol in the first period. This need for an adequate washout period at the time of crossing over is now acknowledged, and has become standard in more recent trials.[26]

The authors of the original study concluded that "propranolol prophylaxis is a safe and effective therapy for migraine prophylaxis."[11] Our study adds little to this. The authors note that 15 of 19 (78.9%) patients responded better to propranolol than placebo, and 2 of 19 (10.5%) responded to both. Our study adds some refinement to the above. When using descriptive statistics, as the original authors did, we

have shown the wide range of the 95% CI for their estimate of 78.9% (ie, 57.4%–92.4%). Our audience was skeptical of this high response rate and opted for a more conservative estimate of 63.5% [54.5%–72.4%]. Using predictive statistics (logistic regression), which the original authors did not do, the estimates are even more conservative: 78.9% [31.1%–99.4%] with no informative prior; and 77.1% [43.1%–97.9%] with the priors from the survey. This conservatism would have been tempered by having ten times the original data, although the results would never have been quite as optimistic as in the original paper.

Our small dataset is in broad agreement with a Cochrane meta-analysis on the subject; if anything, our subjects showed a greater tendency to respond than the results from this paper (7.5× more likely to improve versus 1.9 × [95% CI 1.6–2.3] in the meta-analysis).[15] The placebo response rate at around 10% was broadly similar to that in other crossover studies presented in a meta-analysis of placebo response in migraine.[13]

Ours is far from the first application of Bayesian methods to conduct a retrospective analysis of trial data. Their value in aiding in the interpretation of large trials (n = 30,000) where a difference may assume statistical but not clinical (practical) significance has already been highlighted.[27] Their utility in iteratively updating expert opinion in the effectiveness of a treatment in light of trial data has also been shown.[28] They have also been used to help determine the number of subjects required for a clinical trial to show evidence of efficacy; in this case, a small trial (n = 30) stopped at the interim analysis to inform a larger Phase III study.[29] We note that the latter authors combined survey data to generate an informative prior rather than modeling it as a mixture distribution, as in our case. As far as we are aware, ours is the first to do so using a prior that is a mixture distribution, and which we feel more accurately reflects diversity of opinion. An alternative approach would have been to pool the data and generate one prior; however, using a mixture of priors allows for weighting to varying by opinion, which could be used to model a greater degree of expertise of one respondent. Although we gave equal weight to the responses of each participant, a refinement would be to use a measure of experience such as the number of similar cases of frequent migraines treated over the preceding 10 years.

One major shortcoming in our study was the low rate of participation with only eight of 60 (13%) of those surveyed responding. While there is no agreed minimum response rate for surveys, we acknowledge that ours was exceptionally low.[30] The survey was circulated as an online link as well as in paper form (on two occasions) to all concerned. Weekly grand rounds were the only regular meeting for all members of the department. Trying to explain an additional task following a 50-minute lecture (even though surveys could be returned at any time) may have proved too onerous an obligation, particularly as many members would have been under time pressures to attend to other responsibilities. In retrospect, a mailed survey would perhaps have increased response rates further. The low rate of response may also reflect respondents' unfamiliarity with these methods, which are far from having gained widespread acceptance. As to whether this affects the validity of the conclusions, we would have liked to receive more responses and we leave it to the interested reader to consider repeating the exercise with a larger sample.

Another caveat in the design is asking for prior beliefs when the results of the study are already known. Strictly speaking, this goes against Bayesian first principles. Those surveyed were presented with results from the study rather than given an "imaginary" study with no outcomes available. This was necessary in that a motivated participant would readily have been able to discover the original study based on the description of its design, particularly as it was one of the early landmark papers in its area. In our case the "priors" should be considered as "informative" in that participants were asked to bring their experience in treating migraine and their knowledge of relevant literature to bear on the results from the trial. The only way to avoid this difficulty would be to have the priors supplied before the results of the study were made public, which is impossible in retrospective analyses of this kind.

Some controversy remains regarding the validity of Bayesian methods.[31,32] While critics argue that they may impart a false sense of certainty to studies which are poorly designed or have insufficient numbers from which to draw conclusions, those in favor tend to highlight the value of sharing raw data and sequentially revising beliefs in light of further data or changes in opinion. A particular objection to our methods may be that the Bayesian paradigm cannot be applied in situations where participants are already aware of results of the trial in question (ie, "prior" beliefs cannot be modeled retrospectively). Our priority was rather to supplement the results of a trial with the expertise of the respondents, in effect asking how close the trial results were to the reality in question.

Methods such as those illustrated here may find an application in rapidly generating an expert consensus. Conducting "live" audience surveys is becoming increasingly straightforward.[32,33] Combining these with trial data could give an audience immediate feedback on the combination of their priors and the data. This would appear particularly valuable for bodies drafting guidelines based on the interpretation of

trials. We hope to facilitate the use and discussion of such methods by making our code freely available. We also feel that this analysis illustrates the value to working with and sharing raw data. New techniques and insights may become available long after a study is performed, and these may allow us to look at the results again from a new perspective.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Bayes T, Price R. An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. Philosophical Transactions of the Royal Society of London. 53:370–418.
2. Lunn D, Thomas A, Best, N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*. 2000;(10):325–337.
3. Plummer M. *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. In: Hornik K, Leisch F, Zeileis A, editors. Proceedings of the 3rd International Workshop on Distributed Statistical Computing; March 20–22, 2003; Vienna, Austria. Available from: http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/Plummer.pdf. Accessed January 17, 2013.
4. Kruschke JK. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*, 1st ed. Burlington, MA: Academic Press; 2010.
5. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 1st ed. New York, NY: Cambridge University Press; 2006.
6. Lee JJ, Chu CT. Bayesian clinical trials in action. *Stat Med*. 2012;31(25):2955–2972.
7. Fu H, Qu Y, Zhu B, Huster W. A Bayesian approach to the statistical analysis of device preference studies. *Pharm Stat*. 2012;11(2):149–156.
8. Pressman AR, Avins AL, Hubbard A, Satariano WA. A comparison of two worlds: How does Bayes hold up to the status quo for the analysis of clinical trials? *Contemp Clin Trials*. 2011;32(4):561–568.
9. Zaslavsky BG. Bayesian versus frequentist hypotheses testing in clinical trials with dichotomous and countable outcomes. *J Biopharm Stat*. 2010;20(5):985–997.
10. Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. *Trials*. 2011;12:249.
11. Eichler HG, Abadie E, Breckenridge A, Leufkens H, Rasi G. Open clinical trial data for all? A view from regulators. *PLoS Med*. 2012;9(4):1001202.
12. Weber RB, Reinmuth OM. The treatment of migraine with propranolol. *Neurology*. 1972;22(4):366–369.
13. Lisook A. *The History of FDA's Bioresearch Monitoring Program*. Silver Spring, MD: US Food and Drug Administration. Available from: http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/UCM164151.pdf. Accessed November 28, 2012.
14. Macedo A, Baños JE, Farré M. Placebo response in the prophylaxis of migraine: a meta-analysis. *Eur J Pain*. 2008;12(1):68–75.
15. Autret A, Valade D, Debiais S. Placebo and other psychological interactions in headache treatment. *J Headache Pain*. 2012;13(3):191–198.
16. Linde K, Rossnagel K. Propranolol for migraine prophylaxis. *Cochrane Database Syst Rev*. 2004;2:CD003225.
17. Albert J. LearnBayes: functions for learning Bayesian inference. R package version 2.12 [homepage on the Internet]. 2011. Available from: http://CRAN.R-project.org/package=LearnBayes. Accessed January 18, 2013.
18. Agresti A, Min Y. Frequentist performance of Bayesian confidence intervals for comparing proportions in 2 × 2 contingency tables. *Biometrics*. 2005;61(2):515–523.
19. Hu C, Szapary PO, Yeilding N, Zhou H. Informative dropout modeling of longitudinal ordered categorical data and model validation: application to exposure-response modeling of physician's global assessment score for ustekinumab in patients with psoriasis. *J Pharmacokinet Pharmacodyn*. 2011;38(2):237–260.
20. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974;19(6):716–723.
21. Cembrowski GS, Westgard JO, Conover WJ, Toren EC Jr. Statistical analysis of method comparison data. Testing normality. *Am J Clin Pathol*. 1979;72(1):21–26.
22. Chen MH, Ibrahim JG. Conjugate priors for generalized linear models. *Statistica Sinica*. 2003;13:461–476.
23. Gelman A, Jakulin A, Pittau MG, Su Y. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2(4):1360–1383.
24. Evans CH Jr, Ildstad ST, editors. *Small Clinical Trials: Issues and Challenges*. Washington, DC: The National Academy Press; 2001.
25. Evers S, Marziniak M, Frese A, Gralow I. Placebo efficacy in childhood and adolescence migraine: an analysis of double-blind and placebo-controlled studies. *Cephalalgia*. 2009;29(4):436–444.
26. Bruijn J, Duivenvoorden H, Passchier J, Locher H, Dijkstra N, Arts WF. Medium-dose riboflavin as a prophylactic agent in children with migraine: a preliminary placebo-controlled, randomised, double-blind, cross-over trial. *Cephalalgia*. 2010;30(12):1426–1434.
27. Brophy JM, Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA*. 1995;273(11):871–875.
28. Miksad RA, Gönen M, Lynch TJ, Roberts TG Jr. Interpreting trial results in light of conflicting evidence: a Bayesian analysis of adjuvant chemotherapy for non-small-cell lung cancer. *J Clin Oncol*. 2009;27(13):2245–2252.
29. Tan SB, Chung YF, Tai BC, Cheung YB, Machin D. Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma. *Control Clin Trials*. 2003;24(2):110–121.
30. Fowler FJ. *Survey Research Methods*, 3rd ed. Thousand Oaks, CA: Sage Publications; 2002.
31. Berry DA. Clinical trials: is the Bayesian approach ready for prime time? Yes! *Stroke*. 2005;36(7):1621–1622.
32. Howard G, Coffey CS, Cutter GR. Is Bayesian analysis ready for use in phase III randomized clinical trials? Beware the sound of the sirens. *Stroke*. 2005;36(7):1622–1623.
33. Poll Everywhere. Instant audience feedback [homepage on the Internet]. Chicago, IL: Poll Everywhere. Available from: http://www.polleverywhere.com. Accessed November 28, 2012.

Supplementary materials

Supplementary File A

http://www.dovepress.com/cr_data/supplementary_file_40540.pdf

Supplementary File B

http://www.dovepress.com/cr_data/supplementary_survey_40540.pdf