

Published in final edited form as:

*J Stat Theory Pract.* 2009 June 1; 3(2): 477–487. doi:10.1080/15598608.2009.10411939.

## Bayesian Regression Models for the Quality Adjusted Lifetime Data with Zero Time Duration Health States

**Kaushal K. Mishra** and

North Carolina State University, Raleigh, NC, USA. kkmishra@ncsu.edu

**Sujit K. Ghosh**

North Carolina State University, Raleigh, NC, USA. ghosh@stat.ncsu.edu

### Abstract

Clinical trial studies are often conducted in which quality of life is accessed and recorded along with other clinically measurable endpoints. Consideration of the quality of life in addition to the survival time in the statistical analysis can result in better assessment of the treatments being compared. Quality adjusted lifetime (QAL) data analysis can serve as an important tool to the medical and patient community. This article presents a Bayesian regression approach to the modeling of censored QAL data. The Bayesian hierarchical framework based on a progressive health state model with a data augmentation scheme which provides a nonzero probability to the zero time spent in any health state has been developed. Simulation studies using Markov Chain Monte Carlo (MCMC) methods were performed to validate the proposed method. A real data set was used to illustrate the application of the proposed method.

### Keywords

Bayesian Inference; Data Augmentation; Markov Chain Monte Carlo; Quality Adjusted Survival; TWiST

## 1 Introduction

In the evaluation of treatments concerning chronic diseases such as cancer, acquired immunodeficiency syndrome (AIDS), leukemia or cardiovascular diseases, extending the overall survival time of the patient may not be the only criterion of comparison. The improvement in patient's quality of life must also be regarded as an objective and must reflect in the criterion of comparison. Quality-adjusted lifetime (QAL) is a measure which combines patients' quality of life and the survival time together and provides a useful criterion for comparison across treatments.

For the QAL analysis, quality of life assessments must be made at regular intervals together with clinically measurable endpoints like overall survival, time to time disease progression etc. during the conduct of clinical trial. For the assessments the overall survival time of patients is partitioned into different time periods (health states) which differ in quality of life. Each of these health states can be assigned different utility coefficients depending of the patient's or the physician's perception of the quality of life relative to the state of perfect health. These utility coefficients typically vary between 0 and 1, where 1 represents a state of perfect health and 0 a state equivalent to death. The quality adjusted lifetime of the patient

is the sum of the time periods spent in these different health states weighted by the corresponding utility coefficients.

In most clinical trials, patients enter the study over a period of time, and QAL is not always observed for every patient due to follow-up losses and study termination. The inference on QAL thus has to be made using censored data. Censoring poses a unique problem for making inference on QAL. Different models have been proposed in the literature for analyzing QAL censored data. Cole et al. (1993) used a partitioned health state model and fitted Cox proportional hazards regression models for each transition time. In this case the mean QAL corresponding to a particular covariate value can be obtained by integrating the survival curves for that covariate value. Fine and Gelber (2001) proposed an accelerated life model for the distribution of QAL. Another method for estimating the survival distribution of QAL is using the Kaplan-Meier (KM) estimator with censoring. The estimator for the mean QAL can then be obtained by integrating the survival curve of QAL. However, it has been shown that this estimator is biased due to the informative censoring induced by the different weighting of health states (Gelber, Gelman and Goldhirsch, 1989). In another approach, Glasziou *et al.* (1990) proposed progressive state model in which patients' transition from one health state to another in a sequential manner. In their approach termed as partitioned survival analysis, a weighted sum of the area under the KM survival curves between successive transition times with the weights being the utility coefficients was shown to be a consistent estimator of mean QAL. While the bootstrap method has traditionally been used to measure the errors of such estimates, Murray and Cole (2000) derived asymptotic properties of these estimators. Zhao and Tsiatis (1997) developed a consistent non-parametric approach for estimating the survival function. They also provided a consistent estimator for its asymptotic variance. Lann and Hubbard (1999) extended their model to incorporate covariate information and allows for time dependent censoring. Klein and Moeschberger (1997) proposed a multi-state modeling strategy to jointly model the different health states. However, it is not clear how to incorporate the relative weights, or the utilities of the different health states directly into this approach. Wang and Zhao (2007) derived estimating equations for the regression parameters, regressing the mean QAL on the covariates, using the semi-parametric theory developed by Robins and Rotnitzky (1992).

Recently, Ghosh and Mukhopadhyay (2007) proposed a Bayesian approach in which they consider a joint probability model for the time spent on different health states. They treated the censored observations as missing variables, and used data augmentation technique (Tanner and Wong, 1987) integrated under Markov Chain Monte Carlo (MCMC) algorithm, to generate these observations. This method avoids the performance of high dimensional integrals in the observed likelihood by augmenting the censored observations. However, the proposed model does not take into account zero time spent in any health state by the patients. In the present work, a Bayesian model with modified data augmentation technique is being presented which provides a nonzero probability to the event of zero time spent by a patient in any particular health state. Thus, transition from one health state to another can have jumps with some health states completely skipped. A Bayesian approach provides an avenue for incorporating prior information into the analysis and can be useful if experience is available from previous studies and/or from expert opinions.

The organization of this article is as follows: In Section 2, the statistical model for the QAL data with the proposed data augmentation technique is presented. In Section 3, simulation studies are presented to assess the performance of the proposed method. In Section 4, we present a motivating example by analyzing a real dataset using the models presented in Section 2. The dataset being considered here came from clinical studies conducted by the International Breast Cancer Study Group (IBCSG) Trial V (see Section 4 for more details). Finally, in Section 5 we present some general conclusions with a summary of our findings.

## 2 The Quality Adjusted Lifetime Model

In the present article, we follow the standard notations used in the QAL literature. It is assumed that there are  $T$  treatments  $t_1, t_2, \dots, t_T$  to be compared based on their mean QAL.  $N_l, l = 1 \dots T$  is the total number of patients in each treatment. The overall survival time of the patient is partitioned into  $K$  time periods corresponding to  $K$  distinct health states denoted by  $H_1, H_2, \dots, H_K$ . Each patient passes through these health states in a sequence, however, it is possible that the patient spends zero time in some health states. The time is being measured from the start of the treatment with  $0 = T_0 \quad T_1 \quad T_2 \quad \dots \quad T_K$ . Thus, the time spent in any particular health state  $H_j$  is  $(T_j - T_{j-1})$  for  $j = 1 \dots K$ . It can be noticed that  $T_K$  is the overall survival time of the patient. Also, let the utility coefficients associated to the health states  $H_1, H_2, \dots, H_K$  be denoted by the vector  $\mathbf{q} = (q_1, q_2, \dots, q_K)$ . These utility or quality coefficients are a relative measure of the quality of life of the patient in each health state compared to the state of perfect health.

In trials involving human subjects, usually the observed survival data is censored. The time spent in one or more of the health states for some patients in each treatment group are censored. Here, we denote the censoring vector  $\mathbf{C}_l = (C_1, C_2, \dots, C_{N_l})$  for the patients having a particular treatment  $l$ . In this paper, we have assumed a progressive censoring in which if the survival time of the  $i^{\text{th}}$  patient going through  $l^{\text{th}}$  treatment is censored at the  $j^{\text{th}}$  health state  $H_j$  then it is censored for all subsequent health states  $H_{j+1}, \dots, H_K$  for that patient. Thus, the observed data for the  $i^{\text{th}}$  patient in the  $j^{\text{th}}$  health state getting a particular treatment  $l$  is  $X_{ij(l)} = \min(T_{ij(l)}, C_{i(l)})$ . Along with the observed data  $X_{ij(l)}$  a censoring indicator  $\Delta_{ij(l)} = I(T_{ij(l)} < C_{i(l)})$  is also recorded which informs the recorded data censoring status. The subscript  $l$  just the treatment indicator of the patient and will be dropped further for notation simplification. Thus, the observed dataset for every patient and every health state can be concisely represented by the ordered pair  $\{(X_{ij}, \Delta_{ij}): i = 1, 2, \dots, N_l; j = 1, 2, \dots, K\}$ . It can be observed that in the case of progressive censoring if  $j_0 = \min\{j: \Delta_{ij_0} = 0: 1 \leq j \leq K\}$  then  $\{\Delta_{ij} = 0 \text{ for all } j \geq j_0\}$ .

Given the dataset  $\{(X_{ij}, \Delta_{ij}): i = 1, 2, \dots, N_l; j = 1, 2, \dots, K\}$ , the observed likelihood is given by a product of  $(K - j_0(i) + 1)$  dimensional integral (see Ghosh and Mukhopadhyay, 2007). The maximization of such likelihood is quite complicated by analytical or numerical techniques as often these methods become either intractable or get stuck at a local optimum. Alternatively, to overcome this difficulty completely the data can be completed using some kind of data augmentation technique. In Section 2.1 we propose a data augmentation technique using Bayesian Hierarchical model to complete the data and obtain the parameter estimates from the posterior distributions.

### 2.1 Model Fitting using Bayesian Inferential Framework

In this Section we present a data augmentation technique under Bayesian framework to augment the observed data by a quantity called latent data. Under the Bayesian framework we put a prior distribution  $\pi(\theta)$  on the parameters  $\theta$  to complete the full probability model for both the data and the parameters. Statistical inferences are then based on the conditional distribution of the parameter  $\theta$  given the observed data  $\mathcal{D}$ , known as the posterior distribution given by

$$\pi(\theta|\mathcal{D}) \propto L(\theta|\mathcal{D})\pi(\theta), \quad 2.1$$

where  $L(\theta|\mathcal{D})$  is the likelihood of the augmented data. Next, we describe a data augmentation step for an arbitrary class of densities  $f_1, \dots, f_K$ .

Consider a single patient with the time readings at the change of health state be  $0 = T_0 < T_1 < T_2 < \dots < T_K$ . Suppose,  $U_j \sim f_j(u)$ ;  $j = 1, 2, \dots, K$ , where  $K \geq 2$ ,  $f_j(u)$  are density functions defined on the positive real line and  $U_1, \dots, U_K$  are independent random variables. The densities  $f_j(u)$  can be specified parametrically or non-parametrically. In this paper we have considered a parametric approach which involves parameters  $\theta$  such that  $f_j(u) = f_j(u|\theta)$  where  $\theta \sim \pi(\theta | \theta_0)$ . In the case of censoring the generation of  $U_j$ s also depends on the observed time readings  $X_j$  and the censoring indicator  $\Delta_j$  as given by Eq. 2.2 and Eq. 2.3 below.

$$U_1 = \begin{cases} X_1, & \Delta_1 = 1 \\ \sim f_1(u|\theta)I(X_1, \infty), & \Delta_1 = 0 \end{cases} \quad 2.2$$

$$U_j = \begin{cases} X_j, & \Delta_j = 1, X_j > X_{j-1} \\ \sim f_j(u|\theta)I(0, X_{j-1}), & \Delta_j = 1, X_j = X_{j-1} \\ \sim f_j(u|\theta), & \Delta_j = 0, (X_j < X_{j-1} \text{ or } \Delta_{j-1} = 1) \\ \sim f_j(u|\theta)I(X_j, \infty), & \Delta_j = 0, X_j > X_{j-1}, \Delta_{j-1} = 0 \end{cases} \quad 2.3$$

After obtaining the  $U_j$ s the data is completed by assigning  $T_1 = U_1$  and  $T_j = \max(T_{j-1}, U_j)$ . It can be noticed that  $T_1 < T_2 < \dots < T_K$  with probability one. Also,  $\Pr[T_j = T_{j+1}] > 0$  as  $\Pr[T_j = T_{j+1}] = \Pr[U_{j+1} \leq T_j] = F_{j+1}(T_j) > 0$  for  $T_j > 0$ . Similarly, it can also be shown that probability of any  $j$  time points being equal is greater than zero. Thus this model provides non zero probability to jump in health states from a health state to any higher health state with zero time being spent in the health states lying in between those two health states.

From the complete dataset the estimate of  $EQAL = E(QAL)$  is given by

$$\widehat{EQAL} = \frac{1}{N} \sum_{i=1}^N QAL_i, \quad 2.4$$

where  $QAL_i = \sum_{j=2}^K q_j (T_{i,j} - T_{i,j-1})$ . Also,  $U_j$ s are used as data to generate the samples of  $\theta$  from the posterior distribution for the next iteration. This process is continued until convergence is reached. It has been shown by Tanner and Wong (1987) that samples generated by the data augmentation technique converge to the targeted posterior distribution under fairly mild regularity conditions. Thus,  $\widehat{EQAL}$  converges to

$$EQAL = \sum_{j=1}^K q_j E(T_{i,j} - T_{i,j-1}).$$

The above model can easily be extended to a regression model include covariates if they are available. In fact, different treatments are being treated as categorical covariates in the above model. Also, the model can be easily modified to a non-parametric model by specifying the  $f_j(u)$  non-parametrically.

### 3 Simulation Study

In this section we present two simulated data studies to illustrate the performance of the proposed model through repeated data simulations. In the first example, the data was generated following the same process as the model. This study was performed to ensure the working of the model and its fitting technique. In the second example, the data was

generated from a different model and the above model was fitted to that data. This was performed to study how the model performs when the data is not actually coming from the model which is the case with almost all the real datasets.

### 3.1 Model Performance with Simulated Dataset I

In this example the data was generated following the same process as the model. The model was fitted to the generated data to ensure the working of the model and the fitting technique as mentioned above. A total of  $N = 400$  patients with  $K = 3$  health states (similar to the IBCSG Trial V dataset) were chosen. These three health states were defined as: Toxicity (TOX), Time without symptoms of disease or toxicity (TWiST =  $T_2 - T_1$ ) and the Relapse (REL). In this case the disease free survival time (DFS) is given by

$$DFS = TOX + TWiST. \quad 3.1$$

The expected quality of lifetime is given by

$$EQAL = q_1 E(TOX) + q_2 E(TWiST) + q_3 E(REL) \quad 3.2$$

where  $q = (q_1 \ q_2 \ q_3)^T$  is the vector of the quality coefficients. If we assume that  $EQAL = E(TWiST)$  i.e. the time spent in toxicity or relapse has the zero quality of life for the patient then  $q = (0 \ 1 \ 0)^T$  which is the assumption being used in this paper.

The data generation scheme is as follows:

$$\begin{aligned} U_{ij} &\sim \text{Gamma}(\alpha_j, \beta_j) \\ \alpha &= [\alpha_1 \ \alpha_2 \ \alpha_3] = [10 \ 30 \ 35] \quad \text{and} \quad \beta = [\beta_1 \ \beta_2 \ \beta_3] = [2 \ 5 \ 4] \\ T_{i1} &= U_{i1} \quad \text{and} \quad T_{ij} = \max(T_{i,j-1}, U_{ij}) \\ C_i &\sim \text{Gamma}(\alpha_o, \beta_o) \\ \alpha_o &\sim \text{Unif}[1, n \max(\alpha)] \quad \text{and} \quad \beta_o \sim \text{Unif}[1, n \max(\beta)] \\ X_{ij} &= \min(T_{ij}, C_i) \quad \text{and} \quad \Delta_{ij} = I(T_{ij} < C_i) \end{aligned} \quad 3.3$$

The value of 'n' in the above data generation scheme can be selected to adjust the percentage of censoring desired in the data. Two datasets using different values of 'n' were generated to have 10.50% (1.75% censoring in  $H_1$ , 14.75% censoring in  $H_2$  and 15.00% censoring in  $H_3$ ) and 30.17% (5.00% censoring in  $H_1$ , 42.25% censoring in  $H_2$  and 43.25% censoring in  $H_3$ ) censoring. The model was fitted to both of these datasets to observe the effect of censoring in the estimated EQAL.

The model fitting to the dataset was performed using Matlab and WinBugs. For this purpose the family of distribution for  $f_j(u|\theta)$  and the priors were:

$$\begin{aligned} f_j(u|\theta) &= \text{Gamma}(\alpha_j, \beta_j) \\ \alpha_j &\sim \text{Gamma}(0.1, 0.01) \\ \beta_j &\sim \text{Gamma}(0.1, 0.01) \end{aligned} \quad 3.4$$

It can be observed that the priors are defined on positive real line and have a high variance of 1000. The samples from the posterior distribution in every iteration were generated through MCMC technique using WinBugs.

The result of the simulation is shown in Fig.1. The EQAL obtained for both the censoring cases along with 95% posterior interval is given in Table.1. It can be clearly observed that

for the 10.50% censoring the estimated EQAL is very close to the actual EQAL but for 30.17% censoring EQAL is underestimated but the actual EQAL still lies inside the 95% posterior interval. Usually, in data with such a high censoring rate the models tend to underestimate the parameter of interest. Also, it can be observed that the posterior distribution of EQAL has very less spread and even if we underestimate the EQAL in the case of high censoring the actual difference is not that high in terms of the time units which are in months to be consistent with the IBCSG Trial V dataset. Further, it can be observed that the spread of the posterior density increases with the increase in percentage censoring.

### 3.2 Model Performance with Simulated Dataset II

In this example the data was generated from a different model keeping the same number of patients and the health states as in dataset I. The model was fitted to check the performance of the model when the data is not actually coming from the model. The data in this case was generated from the following model:

$$\begin{aligned}
 T_{i1} &\sim \text{Gamma}(\alpha_1, \beta_1) \\
 T_{ij} &= \begin{cases} \sim \text{Gamma}(\alpha_j, \beta_j) I(T_{ij} > T_{i,j-1}) & \text{if } p < p_0 \\ T_{i,j-1} & \text{if } p \geq p_0 \end{cases} \quad \text{for } j=2 \dots K \\
 \alpha &= [\alpha_1 \quad \alpha_2 \quad \alpha_3] = [10 \quad 30 \quad 35] \quad \text{and} \quad \beta = [\beta_1 \quad \beta_2 \quad \beta_3] = [2 \quad 5 \quad 4] \\
 p &\sim \text{Unif}[0, 1] \\
 C_i &\sim \text{Gamma}(\alpha_o, \beta_o) \\
 \alpha_o &\sim \text{Unif}[1, n \max(\alpha)] \quad \text{and} \quad \beta_o \sim \text{Unif}[1, n \max(\beta)] \\
 X_{ij} &= \min(T_{ij}, C_i) \quad \text{and} \quad \Delta_{ij} = I(T_{ij} < C_i)
 \end{aligned} \tag{3.5}$$

where  $p_0 = 0.85$  was selected. Three datasets with 10.42% (1.00% censoring in  $H_1$ , 13.75% censoring in  $H_2$  and 16.50% censoring in  $H_3$ ), 34.83% (2.50% censoring in  $H_1$ , 54.00% censoring in  $H_2$  and 56.00% censoring in  $H_3$ ) and 50.00% (2.50% censoring in  $H_1$ , 66.75% censoring in  $H_2$  and 80.75% censoring in  $H_3$ ) censoring were generated to study the effect of censoring.

The same model as in Eq. 3.4 was fitted to the three datasets with different censoring. The density estimate of the EQAL is shown in Fig.2. Table 2 contains the estimated EQAL along with the 95% posterior interval. In this case we observe that the actual EQAL for all the three cases lies in the estimated 95% posterior interval. But the EQAL is again underestimated with the difference being larger for higher censoring in the data. Further, we can also observe that the spread of the EQAL posterior density is similar to what was observed in the previous simulated dataset. Thus, we can observe that even when the data is not coming from the actual model being fitted the estimated EQAL is very close to the actual EQAL with a little higher bias for higher percentage censoring cases. When two different treatments are being compared based on the difference in their EQAL, the bias in the estimated difference will decrease further if we assume that the dataset for both the treatments has almost same amount of data censored.

## 4 Analysis of Breast Cancer Data

To illustrate our proposed model we consider again the IBCSG Trial V breast cancer data. The IBCSG Trial V was a randomized clinical trial investigating two treatments in women with the node-positive breast cancer: short-duration perioperative adjuvant chemotherapy (1 month) and long-duration chemotherapy (6 or 7 months). A total of 1229 patients were randomized to receive one of the two treatments between November, 1981 and December, 1985 with 413 patients randomized to short-duration chemotherapy and 816 patients randomized to long-duration chemotherapy. The median follow-up for the study was 7



years. Six covariates were recorded from each patient upon enrollment in the trial, with included age, treatment, tumor size, tumor grades (medium or high) and number of nodes involved. During the study, along with the overall survival, the amount of time attributed to treatment toxicity and the time to disease progression were recorded for each patient. It was observed that short-duration chemotherapy was less toxic to patients but had lesser survival benefits compared to the long duration chemotherapy. However, it was not clear whether the survival advantage of the long-duration therapy outweighed the longer duration of toxicity associated with this treatment. Thus, it was of interest to learn how much patients' EQAL differed across treatments. Other covariates can also be included to extend to the present model to a regression model to study their effects on the EQAL. In this paper only the treatment effect was studied.

In the breast cancer data same three distinct health states can be conceived as mentioned in Section 3.1. Thus, the quality coefficient vector assuming  $EQAL = E(TWiST)$  is again  $q = (0 \ 1 \ 0)^T$ . The model with the same priors as in Eq. 3.4 was fitted to data of the two treatments consisting of 411 patients in short-duration therapy and 804 patients in long-duration therapy. Two patients in short-duration and 12 patients in long-duration therapy were not included in the analysis as they had inconsistent time entries as  $0 = T_0 \ T_1 \ T_2 \ T_3$  was not satisfied. Removing the patients with inconsistent time entries it was calculated that 28.87% (0% censoring in  $H_1$ , 34.55% censoring in  $H_2$  and 52.07% censoring in  $H_3$ ) and 38.18% (0% censoring in  $H_1$ , 51.37% censoring in  $H_2$  and 63.18% censoring in  $H_3$ ) of the time point values are censored for short-duration and long-duration chemotherapy respectively. The density estimates of the EQALs for both the treatments are shown in Fig.3 and the estimated EQAL along with the 95% posterior interval is given in Table.3. If we observe the estimated EQAL along with the 95% posterior interval we can conclude that there is significant difference in the EQAL between short and long duration chemotherapy in terms of extending the overall quality adjusted lifetime of the patients. The long-duration chemotherapy extends the TWiST about 18 months more than the short-duration chemotherapy. Also, as was observed in the simulation study EQAL being underestimated when the percentage censoring increases, it can be safely concluded that the difference in actual EQAL can be higher than 18 months as long-duration therapy has higher percentage of censored data compared to the short-duration therapy. If we observe the density estimates of QAL for both the treatments it becomes further clear that the spread of QAL about the mean is almost same for both the treatments. Thus, the analysis suggests that the long-duration therapy is likely to be more beneficial over the short-duration therapy even after severely discounting for the longer period of toxicity.

## 5 Conclusion

In this article, Bayesian method with possibility of spending zero time in any particular health state was developed to fit a parametric model to quality adjusted lifetime data. The proposed model can take into account jump across any number of health states with nonzero probability. The proposed method can be easily extended to a regression model if other covariates need to be considered. The Bayesian approach has additional advantage of being able to incorporate the prior information into the model by specifying the prior distribution of the model parameters appropriately. In the present case the Gamma family was used to specify the priors but other parametric families (like Lognormal, Weibull etc.) and their mixture can also be used. More, generally a complete non-parametric model can also be used for this purpose.

Extensive simulation study was performed to validate the model and the fitting technique. It was observed that the model slightly underestimated the actual EQAL in case of high percentage of data censoring. Also, even when the data is not actually coming from the

model the estimated EQAL is very close to the actual EQAL. Further, the model was used to fit the IBCSG Trial V data and it was observed that the long-duration therapy has more advantages in terms of the extending the lifetime of the patients even after severely accounting for the period of toxicity which happens to be usually shorter for short-duration therapy. This conclusion is in agreement with those published previously.

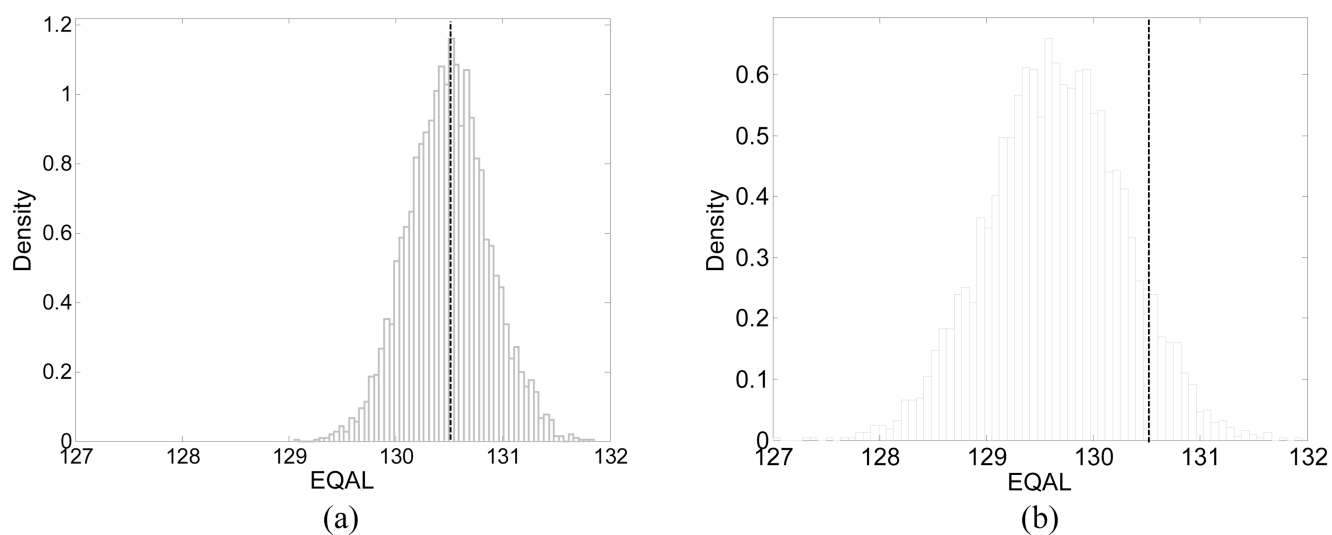
## Acknowledgments

We thanks Spiegelhalter, et al of MRC Biostatistics unit, Institute of Public Health, Cambridge, UK for providing the WinBugs program which was used to perform MCMC sampling. We also thank Professor Richard Gelber and Dr. Bernard F. Cole of the IBCSG for providing the IBCSG Trial V dataset used as an illustration. The authors would also like to thank Kevin Murphy and Maryam Mahdavian for the MATBUGS code which provides an interface between Matlab and WinBugs. Finally, the authors would like to thank the editor and two anonymous referees for providing valuable suggestions which lead to an improved version of an earlier manuscript.

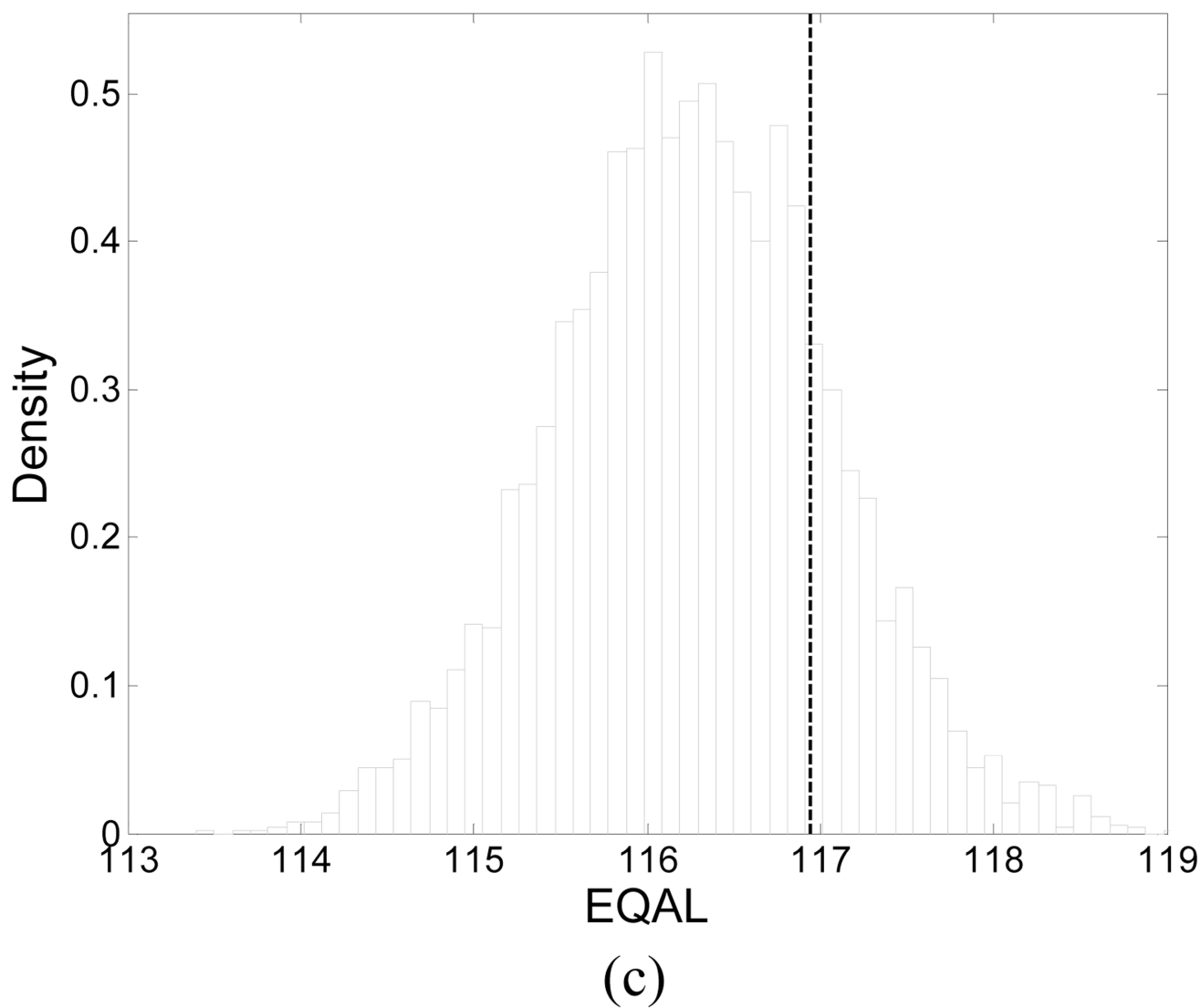
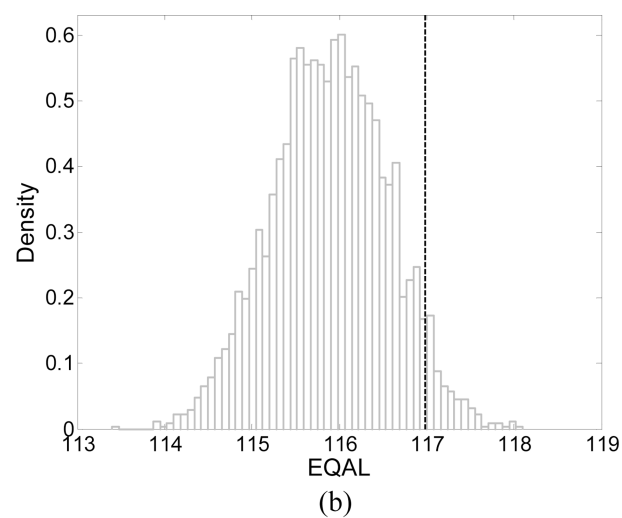
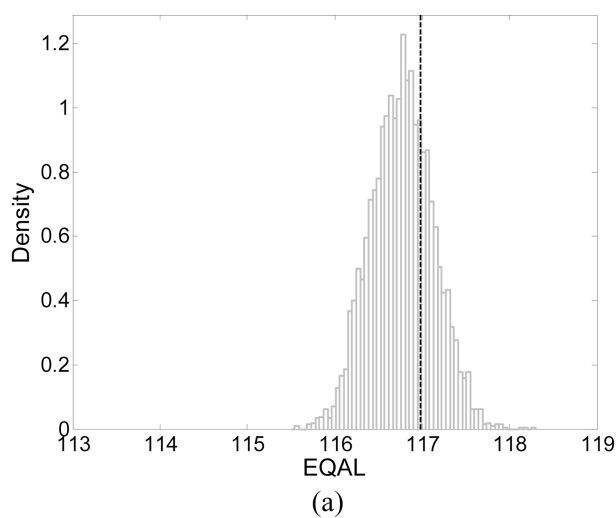
## References

- Cole BF, Gelber RD, Goldhirsch A. Cox regression models for quality adjusted survival analysis. *Statistics in Medicine*. 1993; 12:975–987. [PubMed: 8337554]
- Fine JP, Gelber RD. Joint regression analysis of survival and quality-adjusted survival. *Biometrics*. 2001; 57:376–382. [PubMed: 11414559]
- Gelber RD, Gelman RS, Goldhirsch A. A quality of life oriented endpoint for comparing therapies. *Biometrics*. 1989; 45:781–795. [PubMed: 2790121]
- Ghosh SK, Mukhopadhyay P. Bayesian analysis of quality adjusted lifetime (QAL) data. *Journal of Statistical Theory and Practice*. 2007; 1(2):233–251.
- Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Statistics in Medicine*. 1990; 9:1259–1276. [PubMed: 2277877]
- Klein, JP.; Moeschberger, ML. *Survival Analysis*. Springer; 1997.
- Laan van der MJ, Hubbard A. Locally efficient estimation of the quality-adjusted lifetime distribution with right-censored data and covariates. *Biometrics*. 1999; 55:530–536. [PubMed: 11318210]
- Murray S, Cole BF. Variance and sample size calculations in quality-of-life-adjusted survival analysis (Q-TWiST). *Biometrics*. 2000; 56:173–182. [PubMed: 10783793]
- Robins, JM.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell, N.; Dietz, K.; Farewell, V., editors. *AIDS Epidemiology-Methodological Issues*. Boston, MA: Birkhuser; 1992. p. 297-331.
- Tanner M, Wong W. The calculation of posterior distributions by data augmentation (with discussion). *Journal of American Statistical Association*. 1987; 82:528–550.
- Wang H, Zhao H. Regression analysis of mean quality-adjusted lifetime with censored data. *Biostatistics*. 2007; 8(2):368–382. [PubMed: 16877496]
- Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*. 1997; 84:339–348.



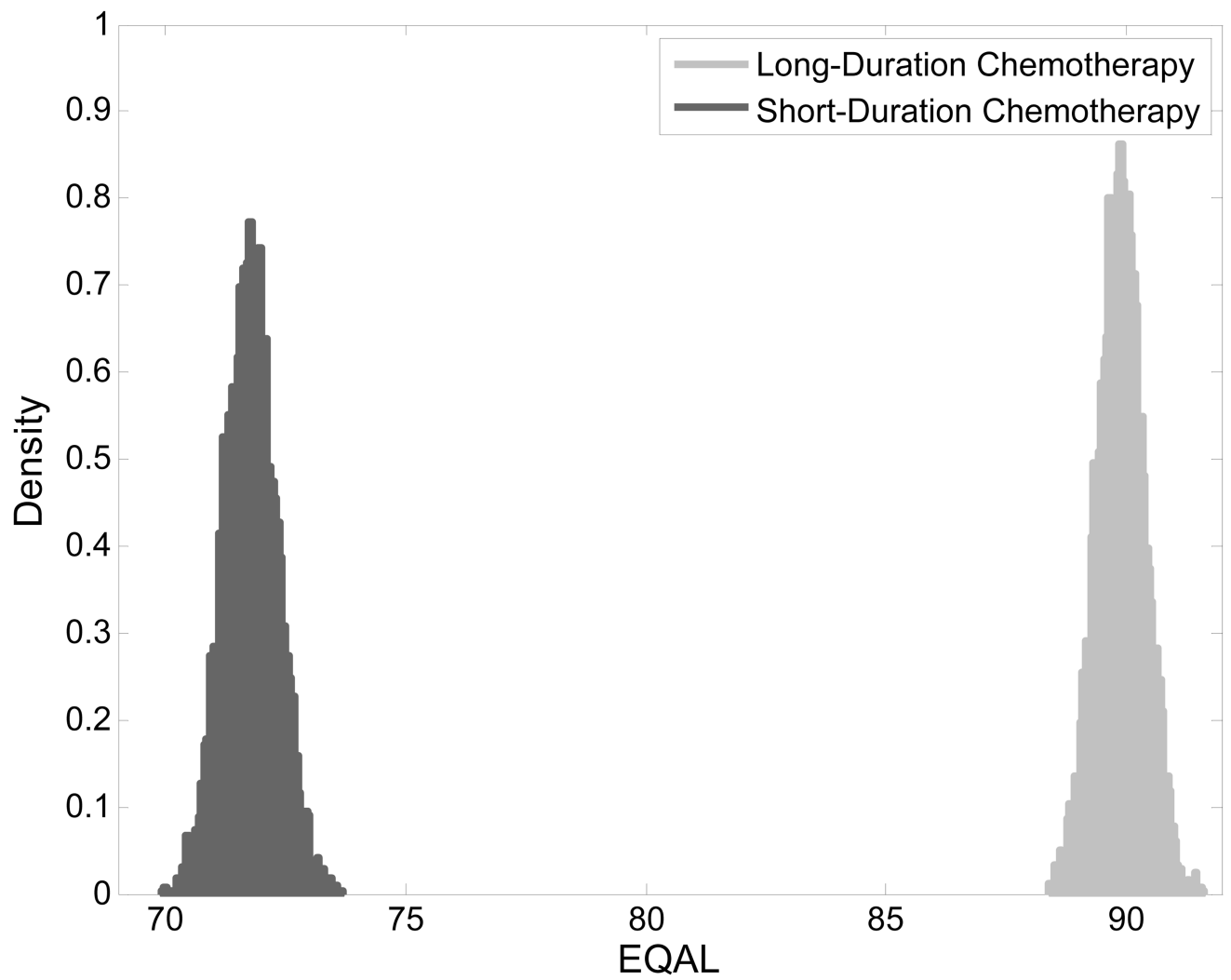


**Figure 1.**  
The posterior density of EQAL for the simulated dataset with (a) 10.50% censoring (b) 30.17% censoring. The dashed line indicates the actual EQAL.



**Figure 2.**

The posterior density of EQAL for the simulated dataset with (a) 10.42% censoring (b) 34.83% (c) 50.00% censoring. The dashed line indicates the actual EQAL.



**Figure 3.**  
The posterior density of EQAL for short and long duration chemotherapy for the IBCSG Trial V dataset.

**Table 1**

Posterior summary of simulated dataset I.

% Censoring	Actual EQAL	Estimated EQAL	2.5%ile	97.5%ile
10.50	130.52	130.49	129.76	131.26
30.17	130.52	129.65	128.42	130.87

**Table 2**

Posterior summary of simulated dataset II.

% Censoring	Actual EQAL	Estimated EQAL	2.5%ile	97.5%ile
10.42	116.95	116.77	116.09	117.48
34.83	116.95	115.89	114.60	117.15
50.00	116.95	116.26	114.68	117.83



**Table 3**

The estimated EQAL for Short-duration and Long-duration Chemotherapy for IBCSG Trial V dataset.

Treatment	Estimated EQAL	2.5%ile	97.5%ile
Short-duration Therapy	71.77	70.71	72.87
Long-duration Therapy	89.90	88.93	90.87