London School of Hygiene & Tropical Medicine

Cover sheet for work submitted towards Assessment 2017/18

Please attach a copy of this sheet to *each* piece of work submitted.

LSHTM Candidate Number:	109963
Module Title:	Survival Analysis and Bayesian Statistics
Module Number:	2463
Module Organiser:	Alexina Mason
Timetable slot:	C1 C2 D1 **D2** E
Date Submitted:	22nd March, 2018
Type of Assessment:	Assignment

Please remember that you have signed a declaration stating that you have read and understood the School's definition of plagiarism and cheating given in the Taught Course Student's Handbook. This declares that any assignment submitted by you during your registration as a student of the London School of Hygiene & Tropical Medicine is your own work, and that you have acknowledged all results and quotations from the published or unpublished work of other people. You should consult your tutor or Module Organiser if you are in any doubt about what is permissible.

==========

This section to be completed by the Module Administrator:

NAME:……

MSc. ……..

Bayesian statistics: Assignment - March 2018

This report was prepared for the Data Monitoring Committee (DMC) to understand if there is enough evidence to stop a randomised clinical trial on resuscitation techniques in patients with cardiac arrest.

Introduction

A two-arm un-blinded randomised clinical trial (RCT) aimed to test the hypothesis that one resuscitation technique is better at cardiac arrest survival than the other. In addition to the RCT interim results, the analysis incorporated a case control and a cross-sectional study that were found on the association between the two resuscitation techniques and death. A Bayesian approach was selected over a frequentist:

- to enable the analysis of small numbers or missing values; and
- to incorporate external information from the case control and cross-sectional study in the analysis; and
- to allow for an intuitive interpretation of results by directly answering questions with probabilities.

Methods

OpenBUGS, a Gibbs sampling software that uses Markov chain Monte Carlo (MCMC) methods assessed the odds ratios (OR) of dying with technique B compared to dying with technique A along with the posterior probability that the OR favours technique A. Data used to fit four different models can be seen in table 1, for further details on mathematical equations, how models were ran and checked refer to the technical appendix.

Model 1: information from the randomised clinical trial only

In the RCT, the number of patients undergoing each technique was fixed and the risk of death was left to randomness. The model preserved this randomness by using a binomial distribution on the risk of death for each technique along with sceptical priors to mimic the current knowledge that the effects of each technique were unclear. Priors for this model consisted of a beta distribution on the risk of death and a normal distribution for the log(OR) of death comparing technique B to A. Assumptions were:

- constant risk of death for all patients; and
- the outcome of one patient is independent from the outcome of another

Table 1. Summary of data and study results

Randomised clinical trial

Technique	Outcome: death	Outcome: survive	Total
A	0	5	5
B	3	2	5
Total	3	7	10

Case control study

Technique	Outcome: death	Outcome: survive	Total
A	10	12	22
B	20	8	28
Total	30	20	50

Cross-sectional study

Technique	Outcome: death	Outcome: survive	Total
A	10	15	25
B	25	20	45
Total	35	35	70

Model 2: information from the case control study only

The case control study identified cases (patients that died from either technique A or B) and a group of suitable controls (patients that survived resuscitation from either technique A or B). In this design, the outcome was fixed while the random element was the probability of receiving resuscitation technique B. In the model, the probability of receiving technique B followed a binomial distribution and a normal distribution was applied to the log(OR). Vague normal priors ensured no extra information was added into the model. Assumptions for this model are:

- constant probability of receiving resuscitation technique B; and
- the probability of receiving technique B for one patient is independent from the probability of receiving technique B for another patient.

Model 3: information from the cross-sectional study only

The cross-sectional study contained information on technique (A or B) and the outcome of death. In this design, the total sample size was the only fixed part and all other characteristics were random. This model used a multinomial distribution to mimic the fixed total sample size with a vague Dirichlet prior on the multinomial vector of probabilities to give equal probability to all values. The models assumptions were:

- individuals in each cell are independent and randomly drawn from the total sample size; and
- the probability of death, given the type of technique is not independent.

Model 4: combine information from all 3 studies

Model 4 combined information from all studies by treating the RCT (model 1) and the case control study (model 2) as “sub-models” and linking them into a joint model where both “sub-models” contribute to calculating the OR based on the common log(OR) parameter. Priors for this model included a vague beta prior on the risk of death, a vague normal prior on the probability of receiving technique B and an informative normal prior for log(OR) using the results from the cross-sectional study (model 3). Assumptions for this model:

- patients from all studies are similar; and
- times taken to resuscitate patients in each study design are similar; and
- each trial gave equal weighting to the results.

Results

Median ORs are likely to be more reliable than mean ORs since posterior distributions on the OR scale tend to be asymmetric from the lower bound at zero, see figure 1. The OR measures the odds of death with technique B compared to the odds of death with technique A. In table 2, the median OR is greater than 1 for all models, indicating death is more likely with technique B. In Bayesian statistics, the 95% credible interval (CrI) is the range on the posterior distribution where the OR of dying with technique B compared to dying with technique A lies with 95% probability. A 95% CrI that contains the null value of 1 (no difference between the odds of death in treatment B compared to the odds of death in treatment A) may suggest that the resuscitation technique (A or B) may not be particularly strongly associated with the risk of death, see figure 2 for a visual display. In this analysis, the first three models include the null value of 1 in the 95% CrI, the level of uncertainty this creates can be quantified by the percent of the posterior distribution that is greater than the null value. In other words, amongst all models, the lowest estimated probability that the OR favours technique A is 89% or up to an 11% chance that the OR does not favour technique A.

Figure 1. Posterior density plots for log(OR) (lor) and OR parameters for all models

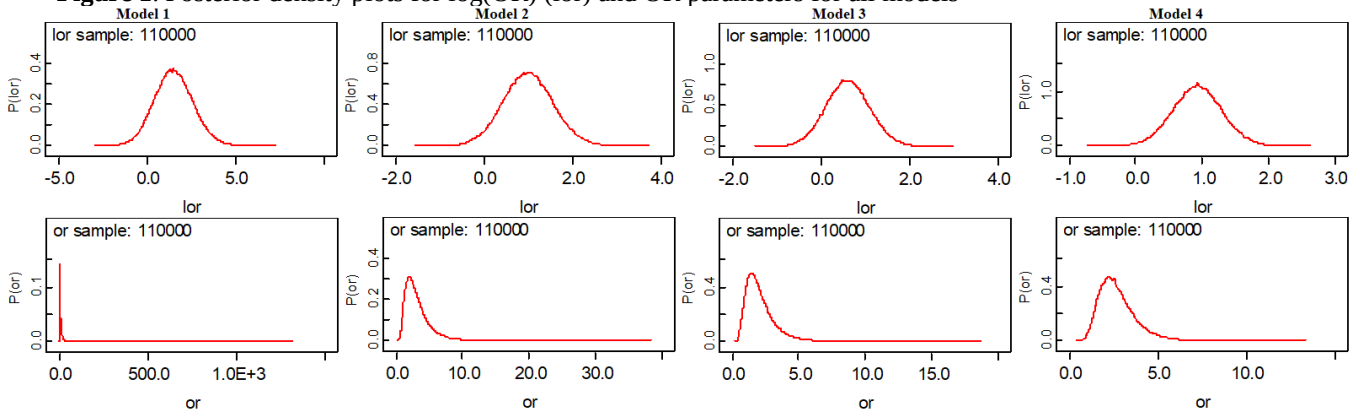


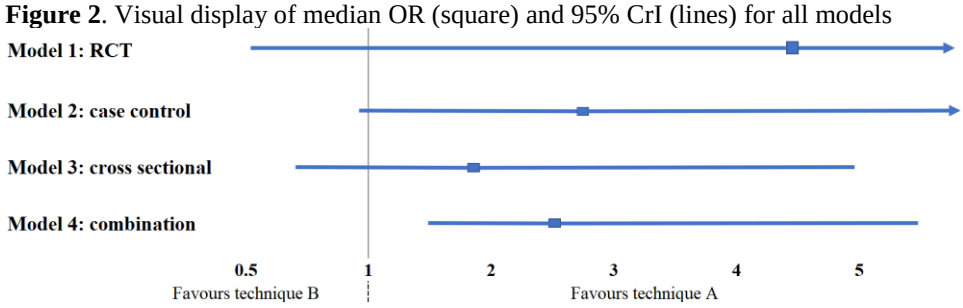
Table 2. Estimated odds ratios for each model using Bayesian MCMC methods

	Odds ratio (OR)				Posterior		Effective
Model	Median	Mean (SD)	95% CrI[1]	MCSE[2]	Probability[3]	sample size	sample size
RCT only	4.432	8.432 (14.6)	(0.55, 41.06)	0.062	91.89%	110,000	54,938
Case control only	2.705	3.189 (1.993)	(0.91, 8.39)	0.010	96.34%	110,000	36,583
Cross-sectional only	1.826	2.080 (1.124)	(0.70, 4.94)	0.003	88.97%	110,000	109,740
All	2.529	2.705 (1.030)	(1.23, 5.22)	0.004	99.43%	110,000	73,780

[1] 95% Credible Interval [2] Monte Carlo Standard Error [3] Posterior probability that the odds ratio favours technique A

The RCT has zero patients that died with technique A. Even though Bayesian methods handle small numbers better than frequentist methods, the zero values contribute to the very wide 95% CrI and makes model 1 sensitive to the "vague" variance on the prior for log(OR). The influence of zero values is attenuated in the joint model 4 and the effect would be further alleviated by including more data on the risk of death with both resuscitation techniques.

Figure 2. Visual display of median OR (square) and 95% CrI (lines) for all models



Discussion

The purpose of this DMC report is to understand if there is enough information to conclusively state one technique is superior to the other. Although all the estimated median odds ratios favour technique A, it would be interesting to investigate potential confounders which may contribute to these results such as age and sex of the patient. Furthermore, an investigation as to the safety or adverse events that are experienced with each technique would be beneficial. Given these additional investigations and the wide posterior distributions, there is not overwhelming evidence for one technique over the other and since there does not appear to be harm done to patients nor are future patients denied an effective treatment, I would recommend the trial to recruit more patients and hold another interim analysis when more data, including safety, can be analysed.

Technical Appendix

A1: Mathematical equations

The posterior probability that the odds ratio favours technique A was calculated as $\mathrm{P}(\mathrm{OR} > 1)$.

Model 1: information from the randomised clinical trial only

Log(OR) of death for patients that had resuscitation technique B compared to patients that had technique A:

$$\log(\mathrm{OR}) = \mathrm{logit}(r_2) - logit(r_1)$$

Priors:

Risk of death with technique A: $r_1 \sim Beta(1,1)$

- A uniform beta distribution gave equal weight to all probabilities

Log(OR) of death for patients with technique B compared to technique A: $\log(\mathrm{OR}) \sim \mathrm{Normal}(0, 3.03)$

- A neutral normal prior centred on no effect; and
- constructed in a way where 95% of the prior OR values would be between 1/30 and 30

Likelihood:

$Y_i \sim Binomial(r_i, n)$ where Y_i represents deaths from resuscitation technique A or B, n is the fixed number of patients randomised to each arm (n=5) and r_i indicates the risk of death ($0 \leq r_i \leq 1$) for i,

$$\mathrm{i} = \begin{cases} 1 = technique\ A \\ 2 = technique\ B \end{cases}$$

Model 2: information from the case control study only

Log(OR) for probability of a case exposed to technique B compared to a control exposed to technique B:

$$\log(\mathrm{OR}) = \mathrm{logit}(p_1) - logit(p_0)$$

Priors:

Probability of receiving technique B amongst controls: $logit(p_0) \sim \mathrm{Normal}(0, 3.03)$

- Log odds calculation for the probability of receiving technique B:
  Among controls: $\mathrm{logit}(p_0) = log\left(\frac{p_0}{1-p_0}\right)$ Among cases: $\mathrm{logit}(p_1) = log\left(\frac{p_1}{1-p_1}\right)$; and
- A neutral normal prior centred on no effect; and
- constructed in a way where 95% of the prior OR values would be between 1/30 and 30

Log(OR) of exposure to technique B among cases compared to controls: $\log(\mathrm{OR}) \sim \mathrm{Normal}(0, 3.03)$

- A neutral normal prior centred on no effect; and
- constructed in a way where 95% of the prior OR values would be between 1/30 and 30

Likelihood:

$X_i \sim Binomial(p_i, Y_i)$ where p_i represents the probability of being exposed to technique B for i,

$$\mathrm{i} = \begin{cases} 0 = in\ controls \\ 1 = in\ cases \end{cases}$$

and X_0 is the number of controls who had technique B, X_1 is the number of controls who had technique B, Y_0 is the number of controls and Y_1 is the number of cases in the case control study.

Model 3: information from the cross-sectional study

The odds ratio comparing the odds of death with technique B to the odds of death with technique A:

$$\mathrm{OR} = \frac{p_2 p_3}{p_1 p_4}$$

Where p_1=number of patients that died with technique A, p_2=number of patients that died with technique B, p_3=number of patients that survived with technique A, p_4=number of patients that survived with technique B.

Prior:

Vector of probabilities: $\mathbf{p} \sim \mathrm{Dirichlet}(1,1,1,1)$

- A vague prior gave equally small strength probability to each value in the vector of probabilities

Likelihood:

$\mathrm{N} \sim \mathrm{Multinomial}([p_1, p_2, p_3, p_4], 70)$ where $0 \leq p_i \leq 1$ and N is the total number of individuals studied.

Model 4: combined RCT and case control with information from the cross-sectional study

Prior: The vague beta prior for the risk of death with technique A (model 1) and the vague normal prior for the Probability of receiving technique B amongst controls (model 2) remained the same in the joint model.

Common prior for log(OR): $\log(\mathrm{OR}) \sim \mathrm{Normal}(0.6067, 0.249)$

- Informative normal prior using the results from the cross-sectional study (model 3)

Likelihood: The likelihood for deaths from resuscitation technique A or B (model 1) and the likelihood for the number of cases and controls who were resuscitated with technique B (model 2) remained the same when they were incorporated into the combined model.

A2: Inference validity

Methodology for how models were run.

The following method was applied to all methods after each model's mathematical equations were finalised:

- Assessment of convergence: visual inspection of the history and trace plots for a "fat hairy caterpillar" suggest when the Markov chain has reached stability with some random scatter around a mean value, this means the Markov chain contains sufficient information for reliable inferences.
- Identify length of burn-in: The Brooks-Gelman-Rubin (bgr) diagnostic was used to refine the approximate point of convergence, the point when the red line (ratio of the chains) is stable around 1 and the blue and green lines (within and between chain variability) start to reach stability. The auto-correlation tool further identified if convergence has been achieved or if a high degree of auto-correlation exists in the sample. Then, the number of iterations discarded as 'burn-in' was chosen.
- Model analysis: After 'burn-in' was discarded, models were updated according to the effective sample size of an autocorrelated chain (the amount of information the posterior distribution contains) using the formula: Effective sample size $= \left({}^{SD}/_{MCSE^{ac}}\right)^2$ where SD is the posterior standard deviation and MCSE is the Monte Carlo Standard Error in autocorrelated samples. Inferences were considered valid when the effective sample size was greater than 10,000, or the MSCE was observed to be approximately 2 orders of magnitude smaller than the posterior SD. Then posterior plots and sample statistics were produced.
- Sensitivity analysis: Priors that were intended to be vague will be changed to understand if unintended information is entering the model in addition to forward sampling on a range of plausible priors to evaluate the model's sensitivity to different assumptions

All models had a posterior sample size of 110,000 from two chains with a 'burn in' of 5000 iterations.

Model checking for model 1: RCT only

After the model, data and initial values were loaded, 10,000 iterations were run, and model convergence was investigated. Figure 2 shows the stabilising of the log(OR) chain through a trace and figure 3 shows its desired "hairy caterpillar" history plot, both plots suggest the model converged. Figure 4 shows the Brooks-Gelman-Rubin (bgr) diagnostic where the red line nears stabilisation at 1 and the blue and green lines are approaching stabilisation. The data for the bgr diagnostic fluctuates around 1, and for the iteration range 1901-3800 the ratio is 1.0. Figure 5 shows the auto-correlation plot for log(OR) with a rapid drop and small jumps in sampled values from one iteration to another. This represents low levels of autocorrelation and near-independence of the log(OR) values being sampled at consecutive iterations. This is an ideal plot as the accuracy of the inferences is dependent on the efficiency of the posterior sample, which decreases with an increasing level of autocorrelation. All plots confirm convergence, 5000 iterations were discarded as 'burn-in' and the model was conducted with 110,000 samples.

Figure 2. Trace plot for log(OR)

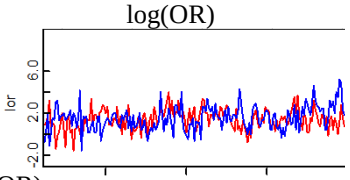


Figure 3. History plot for log(OR)

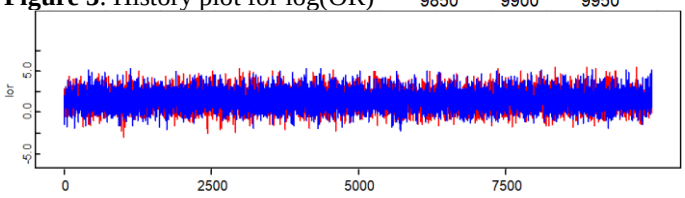


Figure 4. Brooks-Gelman-Rubin diagnostic for log(OR)

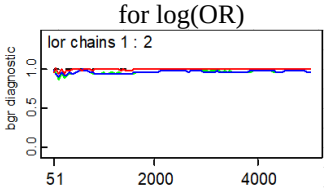
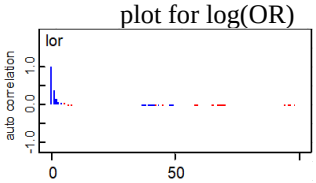


Figure 5. Auto-correlation plot for log(OR)



After the model was ran, a sensitivity analysis determined the amount of influence the priors had on the parameters and if any unintended information was being included with the selected priors. Since the RCT has zero deaths with technique A, the likelihood without any priors is infinity ([3*5/2*0]) making the results from model 1 highly sensitive to the width of the variance selected for the prior on the log(OR) parameter. Forwards sampling density plots in figure 6 indicated the selected vague beta prior allocated equal weights to the risk of death with technique A (r[1]) and some information was being incorporated for the risk of death with technique B (r[2]) using the vague normal prior for log(OR). This information weight centred around 0 and 1 increased as the variance selected for the vague normal log(OR) prior increased.

Figure 6. Forwards sampling for beta priors with log(OR) ~ N(0, 3.03) on the left and log(OR) ~ N(0, 100^2) on the right