KU Leuven

Project

HIV Vaccine Test in Thailand

Bayesian Statistics

[G0B74a]

Gabriel Bénédict
gabriel.benedict@student.kuleuven.be - r0692805
Carl Elgin
carljames.elgin@student.kuleuven.be - r0644084
Daniel Gil Sanchez
daniel.gilsanchez@student.kuleuven.be - r0689432
Laura Juliana Guerrero
laurajuliana.guerrerovelasquez@student.kuleuven.be - r0689435
Benjamin Olsen
benjamin.olsen@student.kuleuven.be - r0653272

Supervised by Prof. Dr. Christel Faes

January 10, 2018

List of Figures

List of Tables

Problem Statement

In 2009, the RV144 randomized, double-blinded, efficacy trial in Thailand reported that a prime-boost human immunodeficiency virus (HIV) vaccine regimen conferred about 30% protection against HIV acquisition, but different analyses seemed to give conflicting results. Here, focus is on individuals in the general population in Thailand, mostly at heterosexual risk, 61% of which were men, randomized in the intention to treat population, excluding subjects found to be HIV positive at the time of randomization. The press release reported r1 = 51 infected among n1 = 8197 who have taken the vaccine, to be compared with r2 = 74 infected among n2 = 8198 who have taken a placebo.
We use the following model in OpenBugs for Bayesian inference:

```
model{
    for (i in 1:2){
        r[i] ~ dbin(theta[i],n[i])
        theta[i] ~ dbeta(0.5,0.5)
    }
}
```

1. Explain the model that is given in the code. What is the likelihood (write it out)? What prior is assumed for the parameters?

The model aims to estimate the posterior distribution of the proportion of infected patients for both control and treatment groups. Given the data, the proportions can be estimated for each group as the number of HIV positive subjects over the total number of participants, which follows a binomial distribution

$$L(\theta_k|r_k) \sim \binom{n_k}{r_k}\theta_k^{r_k}(1-\theta_k)^{n_k-r_k},\ for\ k\ \in 1,2$$

Assuming independent distributions for θ_1 and θ_2, in other terms, independence of the two groups of study, the likelihood can be written as the product of two binomials:

$$L(\theta_1,\theta_2|y) \sim \binom{n_1}{r_1}\theta_1^{r_1}(1-\theta_1)^{n_1-r_1}\binom{n_2}{r_2}\theta_2^{r_2}(1-\theta_2)^{n_2-r_2}$$

$$L(\theta_1,\theta_2|y) \propto \theta_1^{r_1}(1-\theta_1)^{n_1-r_1}\theta_2^{r_2}(1-\theta_2)^{n_2-r_2}$$

And the θ_k parameters are assumed to have a non-informative beta prior distribution $Beta(0.5, 0.5)$ (Jeffreys prior), so:

$$p(\theta_1,\theta_2) \propto \theta_1^{-1/2}(1-\theta_1^{-1/2})\theta_2^{-1/2}(1-\theta_2^{-1/2})$$

2. Use 10,000 MCMC iterations with a burn-in of 1,000 MCMC iterations for estimation of the parameters. What is the MCMC error? Check convergence of the MCMC chain for all parameters. Discuss.

As can be seen on table 1, MCMC error for both estimations is relatively low (for θ_1 is 9.22E-06 and for θ_2 is 1.05E-5). In fact, given the rule of thumb that this error should be smaller than 5% of the posterior standard deviation, it can be concluded that both posterior estimations are accurate.

Regarding the convergence of the MCMC chains for θ_1 and θ_2 (Figure 1 and 2 show trace plots), both converge to the posterior distribution. In addition, autocorrelation plots (Figure 3 and 4), indicate that the correlations between samples decreases. Indeed, after 2 iterations the chain has already forgotten the history of the MCMC sampler.

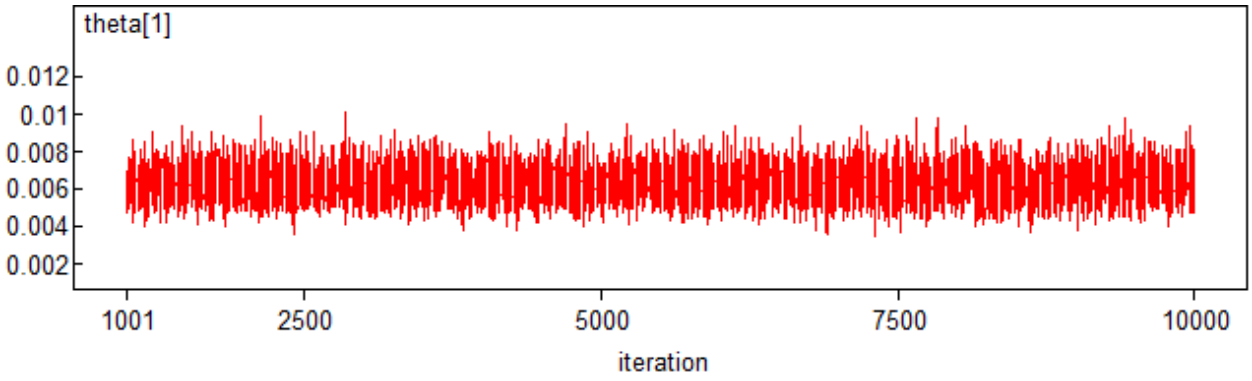

Figure 1: History Plot - Vaccine group (θ_1)

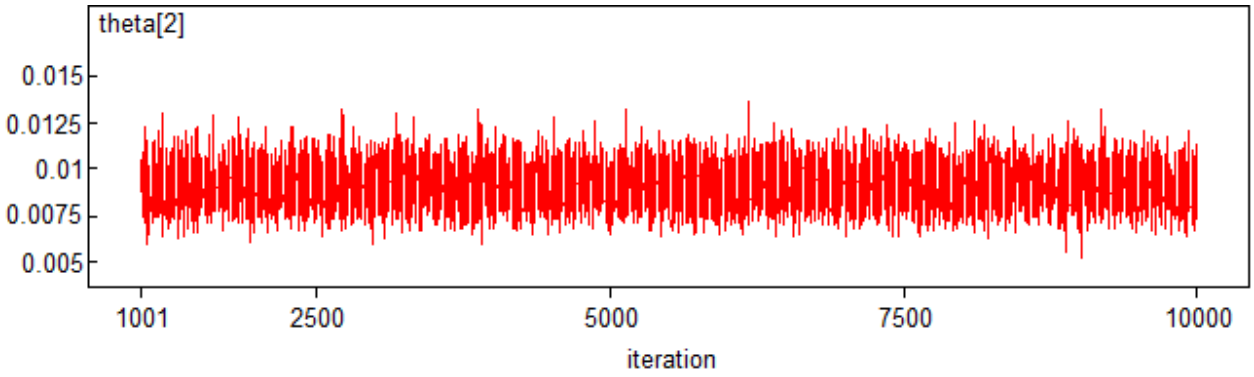

Figure 2: History Plot - Placebo group (θ_2)

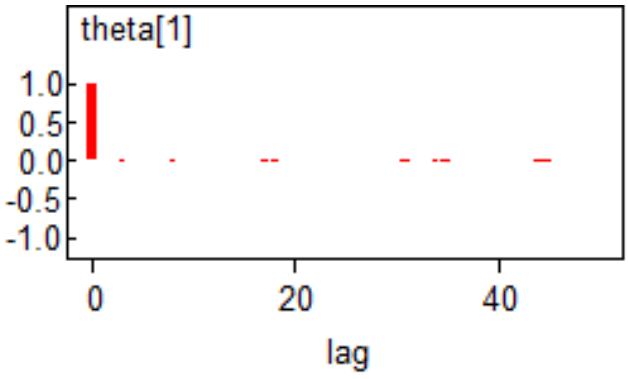

Figure 3: Auto correlation plot θ_1

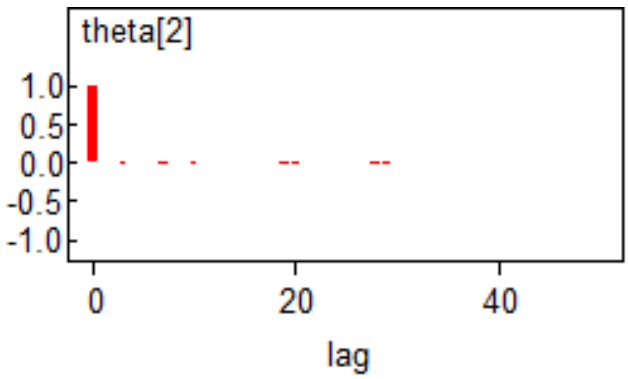

Figure 4: Auto correlation plot θ_2

3. Give the posterior mean and median for the proportions θ_1 and θ_2. Give also 95% credible intervals.

Node	Median	Mean	Sd	MC Error	95% lower (eq)	95% upper (eq)
θ_1	0.006245	0.006288	8.81E-04	9.22E-06	0.004674	0.008169
θ_2	0.009042	0.009091	0.001059	1.05E-05	0.007124	0.011320

Table 1: Posterior summary statistics for θ_1 and θ_2

The equal tail 95% credible interval for both estimations contains the 95% most plausible parameter values a posteriori.

4. Interest is in the vaccine efficacy, defined as one minus the relative hazard rate of HIV in vaccine versus placebo group, or

$$VE = 1 - \frac{\theta_1}{\theta_2}$$

4.1. Obtain a 95% credible interval for VE. How can we interpret this interval?

Node	Median	Mean	Sd	MC Error	95% lower (eq)	95% upper (eq)
VE	0.3102	0.2987	0.13	0.001388	0.001933	0.5224

Table 2: Posterior summary statistics for VE

The vaccine efficacy shows the percentage reduction of infection in the vaccinated group compared to the placebo group. Table 2 shows the vaccine reduces 31.02% of the risk of acquiring the virus. Also, the 95% CI for the most plausible values show the vaccine could actually decrease the risk from 0.2% up to a 52%. It is important to mention that this estimation has an MC error lower than the posterior standard deviation (5%) and that the trace plot shows convergence of the chain. Thus the estimation is reliable.

4.2. Give a plot for the posterior density of VE.

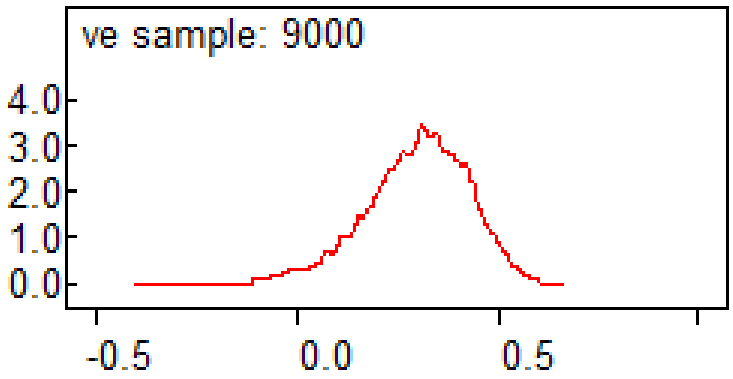

Figure 5: Posterior Density of Vaccine Efficacy

5. What is the posterior probability that the proportion θ_1 of infected amongst those vaccinated is actually smaller than the proportion θ_2 of infected among those that take the placebo?

It is of interest to test if $\theta_1 \leq \theta_2$, so it is necessary to evaluate the posterior of $p(\theta_1 - \theta_2|y)$. In order to do so, samples are taken from $p(\theta_1|y)$ and $p(\theta_2|y)$, and then the difference of the parameters for each sample is computed. i.e. the posterior probability of θ_1 being smaller than θ_2 is calculated based on an indicator variable that counts the number of samples where this condition holds, which is about 98% (see Table 3).

Node	Median	Mean	Sd	MC Error	95% lower (eq)	95% upper (eq)
$\theta_1 - \theta_2$	-0.002799	-0.002803	1.39E-03	1.48E-05	-0.005542	-1.40E-05
p		0.976				

Table 3: Posterior summary statistics for the difference

6. Give a plot of the posterior predictive distribution for a future 100 subjects that are vaccinated, as well as the posterior predictive distribution for a future 100 subjects that take the placebo. Give also a 95% posterior predictive set.

The posterior predictive distribution of a model that have prior beta distribution and binomial likelihood, is a beta-binomial distribution. Although, it is not possible to take samples from this distribution in BUGS, it is important to notice that this distribution corresponds to a binomial distribution in which the probability parameter is random (following a *Beta* distribution). As a result, the samples of the posterior predictive distribution are computed using a Binomial distribution. In Figure 6 and 7 the results are shown.

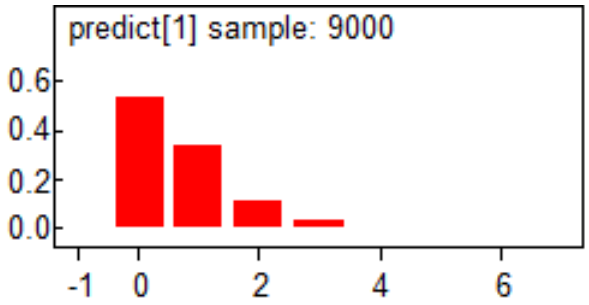

Figure 6: Posterior predictive distribution for vaccine population

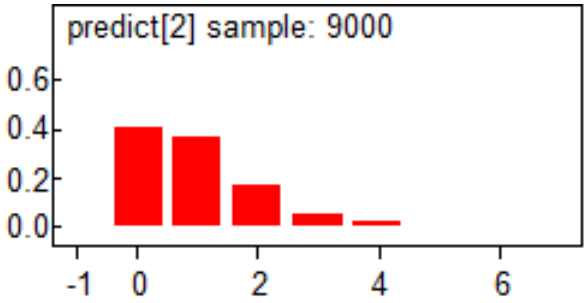

Figure 7: Posterior predictive distribution for placebo population

The posterior predictive distribution is concentrated around zero, meaning that in a new population of 100 individuals that are vaccinated, it is more probable to find non-infected individuals. A similar situation will occur with a group that take the placebo.
It is important to mention that the 95% posterior predictive set for both groups is between 0 and 3, meaning that in a new group of 100 individuals (regardless whether placebo or vaccine was taken), it is expected to find between zero and 3 people infected. However, for the placebo group the median is 1 instead of zero for the vaccinated group (see Table 4). This would confirm that vaccination decreases risk of HIV.

Node	Median	Mean	Sd	MC Error	95% lower (eq)	95% upper (eq)
predict[1]	0	0.6387	0.8017	0.00852	0	3
predict[2]	1	0.909	0.9584	0.01048	0	3

Table 4: Summary statistics for the posterior predictive distribution

-

7. Do a sensitivity analysis of the prior. How do results change in case you use the following priors:

7.1. Use a (non-informative) uniform prior for both θ_1 and θ_2.

For the non-informative uniform prior, the prior distribution is changed to a $Beta(1, 1)$.

Node	Median	Mean	Sd	MC Error	95% lower (eq)	95% upper (eq)
θ_1	0.006292	0.006337	8.80E-04	9.48E-06	0.004734	0.008152
θ_2	0.009114	0.009155	0.00106	1.01E-05	0.007176	0.01138
VE	0.3107	0.2983	0.1294	0.001261	0.007801	0.5172
$\theta_1 - \theta_2$	-0.002835	-0.002819	0.001388	1.33E-05	-0.005515	-5.29E-05
p		0.9767				
predict[1]	0	0.6458	0.8053	0.008706	0	3
predict[2]	1	0.9136	0.962	0.01052	0	3

Table 5: Posterior summary statistics with a non-informative prior

In Table 5 the results of the analysis with a non-informative prior distribution are shown. It can be seen that these results do not change significantly compared with the previous one (Jeffreys prior which is also non-informative). This can also be explained by the large sample size of the likelihood, and how it dominates the prior distribution in the estimation of the posterior distribution.

7.2. Use a prior that reflects a previous study. In the previous study, there are r1 = 56 infected among n1 = 8202 who have taken the vaccine, and r2 = 76 infected among n2 = 8200 who have taken a placebo.

To reflect the results of a previous study in the prior distribution, the parameters α and β are changed using the fact that a $Beta(\alpha, \beta)$ prior distribution is equivalent to a binomial experiment with $\alpha - 1$ successes in $\alpha + \beta - 2$ experiments. Thus, the priors for vaccine and placebo groups are modified to Beta(57, 8147) and Beta(77,8125), respectively.

Node	Median	Mean	Sd	MC Error	95% lower (eq)	95% upper (eq)
θ_1	0.006573	0.006589	6.40E-04	7.34E-06	0.005381	0.007876
θ_2	0.009181	0.009203	7.44E-04	8.55E-06	0.007785	0.01072
VE	0.2845	0.2793	0.0922	9.95E-04	0.08345	0.4463
$\theta_1 - \theta_2$	-0.002613	-0.002614	9.89E-04	1.09E-05	-0.004551	-6.88E-04
p		0.9959				
predict[1]	0	0.6701	0.8255	9.63E-03	0	3
predict[2]	1	0.9187	9.49E-01	9.60E-03	0	3

Table 6: Posterior summary statistics with a prior that reflects a previous study

In Table 6 the results of the analysis with a prior distribution from a previous study are shown. Once again, it can be seen that these results do not change significantly compared with the previous one (Jeffreys prior). In this case the likelihood and the prior are close, so it is natural to have a posterior distribution that does not present considerable changes compared with previous results.

Appendix - BUGS Code

```
model{
  # Specification data model
  for (i in 1:2){
    # Jeffrey's prior
    theta[i] ~ dbeta(0.5,0.5)
    # NI Uniform prior
  # theta[i] ~ dbeta(1,1)
    # Previous study prio
  # theta[i] ~ dbeta(alpha[i],beta[i])
    # Likelihood
    r[i] ~ dbin(theta[i],n[i])
    # Prediction distribution
    predict[i] ~ dbin(theta[i],m[i])
  }
  # Vaccine efficacy
  ve <- 1-(theta[1]/theta[2])
  # testing if theta_1 is smaller than theta_2
  diff <- theta[1]-theta[2]
  P <- 1 - step(diff)
}

#        Data
list(n=c(8197,8198),
         r=c(57,74),
         m=c(100,100),
         alpha=c(57,77),
         beta=c(8147,8125))

#        Initial values
list(theta=c(0.3,0.3),predict=c(0,0))
```