Bayesian analysis of efficacy of the ChAdOx1 nCoV-19 (AZD1222) vaccine

Roudranil Das

Roll no: 20, Sem: VI
CIN: 19-300-4-04-0020
Paper Code: HMTDS6043D
Supervisor: Prof. Sucharita Roy
St. Xavier's College, Kolkata (Autonomous)
Department of Mathematics.

April 10, 2022

Abstract

2020 saw the onset of the Covid-19 pandemic cause by the SARS-CoV-2 virus, widely dubbed by many to be the first great pandemic of the information era. The pandemic forced the world into isolation and shut down all form of institutions. With the number of daily affected people and number of deaths due to the pandemic climbing sharply, a vaccine was the need of the hour.

In this paper we consider one such vaccine against Covid-19, the ChAdOx1 nCoV-19 (AZD1222) vaccine (known as Covishield in India), and we investigate its efficacy based on studies of four randomised controlled trials held in Brazil, South Africa, and the United Kingdom [1]. We use the Bayesian paradigm to model the posterior distribution of the vaccine efficacy and calculate its credible interval.

In our analysis, we use the open-source tools and packages NumPy, SciPy, Matplotlib and PyMC3. After performing Bayesian inferencing, we compare our findings with the results of the referenced studies. From there we conclude that the results obtained by the Bayesian paradigm is in agreement with the results obtained by frequentist methods.

Keywords: Bayesian, Covid-19, vaccine efficacy, beta-binomial, incidence rate ratio, credible interval, AstraZeneca.

Declaration: *I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.*

Roudranil Das

Contents

1 Introduction

The efficacy of Covid-19 vaccines is one of the most important parameters which governs whether a certain vaccine should be administered to the general populace, and it is even more important for a vaccine which is needed as urgently as this. However, in almost all studies which analyse the efficacy of these vaccines, the most common statistical approaches chosen are frequentist methods. In this paper we attempt to provide an alternate analysis mainly with Bayesian methods and MCMC sampling techniques to show that this too is an effective method of analysis.

Here, we analyse the efficacy of the AZD1222 vaccine (Covishield) based on the results of the clinical trials carried out independently in the UK, Brazil, and South Africa. The methods, findings, and all related data are outlined in "Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK"[1], by Voysey et al.

As prerequisites we shall be introducing the concepts of conditional probability and Bayes' Theorem in brief, before moving on to priors, likelihoods, and posterior distributions. Then we shall define what do we mean by vaccine efficacy. Also, for purely convenience's sake, we shall be referring to the vaccine as simply 'Covishield' hereafter.

1.1 Conditional probability and Bayes' Theorem

Conditional probability forms the heart of the statistical methods we shall be using. Essentially, we would like to find out the probability of an event B occurring, when we have the knowledge that an event A occurs.

From the Classical definition of probability, if we have N number of mutually exclusive, exhaustive, and equally likely outcomes of a random experiment, and if $n(A)$ outcomes are favourable to the event A, then the probability that A occurs is:-

$$p(A) = \frac{\text{number of outcomes favourable to } A}{\text{total number of outcomes}} = \frac{n(A)}{N} \tag{1.1}$$

In our case, we may consider that the total number of mutually exclusive, exhaustive, and equally likely outcomes of the random experiment reduces to $n(A)$ after A occurs and out of them $n(B \cap A)$ outcomes are favourable to the event B (as all outcomes favourable to B are favourable to A as well). Then we can say

$$\begin{aligned} p(B \mid A) &= \frac{n(B \cap A)}{n(A)} = \frac{{}^{n(B \cap A)}/_{N}}{{}^{n(A)}/_{N}} \\ &= \frac{p(B \cap A)}{p(A)} \end{aligned} \tag{1.2}$$

Now we turn our attention to Bayes' Theorem. Bayes' Theorem provides us with the tools to calculate the posterior probability of an event: Essentially, with conditional probability we calculate the probability of an *"effect"* happening due to an observed *"cause"*. With Bayes' Theorem, we calculate the probability that a certain *"cause"* has resulted in the observed *"effect"*.

In the above mathematical formulation, we calculated the conditional probability of B occurring when A has already occurred. Bayes' Theorem states that the probability of A occurring when B has occurred is given by:

$$p(A \mid B) = \frac{p(A) \cdot p(B \mid A)}{p(B)} \tag{1.3}$$

1.2 The philosophy of Bayesian inference

The core principle in Bayesian inference is *preserving uncertainty* about the parameters we want to infer about. This is in a direct contrast to the frequentist mode of approach where we want to deliver

our results in very certain values. The frequentist approach will perform a random experiment and will conclude that the observed frequency of an event occurring, in the long run is its probability. Bayesians, on the other hand, shall have a *prior belief* regarding the probability of the event, and based on the results of the random experiment, shall update it with those evidence, and generate the posterior distribution for the probability of the event.

In very modest terms, Bayesian data analysis is no more than finding out in how many ways a certain data could have occurred based on our assumptions (recalling our cause-and-effect example previously).

The difference between Bayesian and frequentist methods can be illustrated with the following short example of a chess player:

Say a chess player A is preparing for a match and plays N warmup games before the match begins. Suppose we know that this player is strong and seldom loses. He plays the warmup games and wins all of them. We pass the following argument to our frequentist function: "Player A has won all N of his warmup games: will he win the match?". To this our frequentist function would return a "YES". However, if we pass the following arguments to our Bayesian function: "Player A seldom loses and has won all N of his warmup games: will he win the match?". To this we would get the answer: "YES with probability 0.9 and NO with probability 0.1". We should notice how we passed our prior belief that the player seldom loses as an argument to the Bayesian function. There lies the difference between the answers, and the crux of the philosophy behind Bayesian inference. A numerical example is provided in section 1.3.

1.3 Prior, Posterior, Likelihood and Uncertainty Intervals

Before we begin, we shall state some conventions: θ will denote the *unknown population parameters* of interest (i.e., in case of multiple parameters, it will be the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ where $k \in \mathbb{N}$), $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ where $n \in \mathbb{N}$ will denote the observed samples drawn from the population.

The prior beliefs of $\boldsymbol{\theta}$ is a probability distribution, called the *prior probability distribution*, which is represented using the pdf (or pmf) $p(\boldsymbol{\theta})$. The observed data $\boldsymbol{X}$ follows a probability distribution, which depends on these k unknown population parameters. Thus, the likelihood of observing $\boldsymbol{X}$, given the parameters $\boldsymbol{\theta}$ is denoted by the pdf (or pmf) $p(\boldsymbol{X} \mid \boldsymbol{\theta})$.

Finally applying Bayes' Theorem, we get the posterior density of $\boldsymbol{\theta}$ to be:

$$p(\boldsymbol{\theta} \mid \boldsymbol{X}) = \frac{p(\boldsymbol{\theta}) \cdot p(\boldsymbol{X} \mid \boldsymbol{\theta})}{p(\boldsymbol{X})} \implies p(\boldsymbol{\theta} \mid \boldsymbol{X}) \propto p(\boldsymbol{\theta}) \cdot p(\boldsymbol{X} \mid \boldsymbol{\theta}) \tag{1.4}$$

$$\text{where } p(\boldsymbol{X}) = \int p(\boldsymbol{\theta}) p(\boldsymbol{X} \mid \boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{1.5}$$

Thus, we can say that the posterior density of $\boldsymbol{\theta}$ is proportional to the prior times likelihood. This is because, once we observed the data $\boldsymbol{X}$, the probability, $p(\boldsymbol{X})$ in (1.5), becomes a constant. It is also known as the *marginal distribution* of $\boldsymbol{X}$, or the *prior predictive distribution*.

We should also take note of conjugate priors. A class Π of prior distributions is said to form a *conjugate family* if the posterior density is in class Π whenever the prior distribution is in Π.

The uncertainty intervals we will be using are the *95% Credible Interval* and the *95% Highest Density Interval*. A 95% Credible Interval is the interval in which the concerned unobserved parameter falls with probability 0.95. The 95% Highest Posterior Density Interval is the minimum width 95% Credible Interval.

A numerical example of Bayesian Inference: Suppose we receive an urn with 4 balls, each ball either black (denoted by B) or white (denoted by W) in colour. It is unknown to us the number of balls of each colour. Hence we have the 5 possible scenarios: $\theta_0 : 4W$, $\theta_1 : 3W + 1B$, $\theta_2 : 2W + 2B$, $\theta_3 : 1W + 3B$, $\theta_4 : 4B$. Suppose, *a priori*, we assume that each scenario is equally possible, i.e., $\mathrm{P}(\theta_i) = \frac{1}{5}$. This is our prior distribution. We initially draw 3 balls from the urn, with replacement, to get the following sequence: $X : B, W, B$. The likelihood of observing X given scenario θ_i is given as:

Scenario θ_i	$\mathrm{P}(X \mid \theta_i)$
θ_0	0
θ_1	$3/64$
θ_2	$8/64$
θ_3	$9/64$
θ_4	0

Our goal is to compute the posterior probability of scenario θ_1 given that X is observed. Using Bayes' theorem,

$$\begin{aligned}\mathrm{P}(\theta_1 \mid X) &= \frac{\mathrm{P}(X \mid \theta_1) \cdot \mathrm{P}(\theta_1)}{\sum_{i=0}^{5} \mathrm{P}(X \mid \theta_i) \cdot \mathrm{P}(\theta_i)} \\ &= \frac{3/64 \cdot 1/5}{3/64 \cdot 1/5 + 8/64 \cdot 1/5 + 9/64 \cdot 1/5} \\ &= \frac{3}{20}\end{aligned}$$

We get the posterior probability of θ_1 given X to be $\frac{3}{20}$.

1.4 Some commonly used probability distributions

Probability distributions are mathematical functions which gives the probability of the occurrence of the different outcomes when a random experiment is performed. If we have a random variable X, which stores the outcomes of a random experiment, then the probability distribution, represented by the function $f(x)$, gives the probability or probability density at $X = x$.

Based on whether the random variable is discrete or continuous, the probability distribution function is called a probability mass function (pmf) or a probability density function (pdf). Here we are interested primarily in two probability distributions: the *Binomial distribution* and the *Beta distribution*.

1.4.1 Binomial Distribution

A discrete random variable X which stores the number of successes of n independent Bernoulli trials, is said to follow the Binomial distribution with the parameters n, p (where p is the probability of success in one trial). Notation wise, we say that $X \sim Bin(n, p)$, and the probability mass function (pmf) of X is given by:

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & , \ x = 0(1)n, \ 0 < p < 1, \ p + q = 1 \\ 0 & , \ otherwise \end{cases} \tag{1.6}$$

We can compute the mean and variance of this distribution as follows:

$$Mean = \mathrm{Exp}[X] = \sum_{x=0}^{n} x f(x) = np \tag{1.7}$$

$$Variance = \mathrm{Var}(X) = \mathrm{Exp}[X^2] - \mathrm{Exp}[X]^2 = \sum_{x=0}^{n} x^2 f(x) - np = npq \tag{1.8}$$

1.4.2 Beta Distribution

When a random variable X follows a Beta distribution, we say that $X \sim Beta(\alpha, \beta),\ \alpha, \beta > 0$. The pdf, mean and variance of the beta distribution are:

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} & , \ 0 < x < 1,\ \alpha, \beta\ > 0 \\ 0 & , \ otherwise \end{cases} \quad \text{where } \mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{1.9}$$

$$Mean = \mathrm{Exp}[X] = \int_0^1 x f(x) \mathrm{d}x = \frac{\alpha}{\alpha+\beta} \tag{1.10}$$

$$Variance = \mathrm{Var}(X) = \mathrm{Exp}[X^2] - \mathrm{Exp}[X]^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{1.11}$$

We should also note that, if we choose $\alpha = 1$ and $\beta = 1$, we have the *continuous Uniform distribution* in (0,1).

2 The problem statement

This section lists the data to be used in the following sections, as well as some specialised definitions stemming from epidemiology.

2.1 Data used

The data we will be using comes from blinded, randomised, controlled trials done in different groups across 3 countries. Here we are concerned with results of 3 groups from them, which are also the ones primarily studied in [1]. The following table outlines the baseline characteristics of the participants of the study, which mainly shows the the number of participants per age group in the study.

	Group 1 (COV002 - UK; LD/SD)		Group 2 (COV002 - UK; SD/SD)		Group 3 (COV003 - Brazil; SD/SD)	
Age(yr)	Vaccine (1367)	Control (1374)	Vaccine (2377)	Control (2430)	Vaccine (2063)	Control (2025)
18 - 55	1367	1374	1879	1922	1843	1833
55 - 69	0	0	285	293	209	187
$\geq$ 70	0	0	213	215	11	5

Table 1: Baseline population characteristics[1]

[1] Table 1 of [1]

The following table details out the number of fresh Covid-19 cases observed and the calculated vaccine efficacy in the primary efficacy population more than 14 days after they received the second dose of their vaccine.

	Total number of cases	Vaccine cohorts n/N (%)	Control cohorts n/N (%)	Vaccine efficacy% (95% CI)[2]
Overall	131	30/5807 (0.5%)	101/5829 (1.7%)	70 (54.8 to 80.6)[3]
LD/SD	33	3/1367 (0.2%)	30/1374 (2.2%)	90.0 (67.4 to 97.0)
SD/SD	98	27/4440 (0.6%)	71/4455 (1.6%)	62.1 (41.0 to 75.7)

Table 2: Efficacy calculation of primary efficacy population[4]

[2] 95% CI: Frequentist 95% Confidence Interval
[3] 95.8% CI
[4] Table 2 of [1]

As we can see from this table, the overall vaccine efficacy percentage is calculated to be about 70%. This vaccine efficacy was calculated as 1 – adjusted relative risk (vaccine vs control incidence rates).

2.2 Terminology used and definitions

The terms incident rate ratio, relative risk ratio and vaccine efficacy will be repeatedly used here. In epidemiology, *vaccine incidence rate* is the ratio of the number of confirmed Covid-19 cases per number of people in the vaccine cohort. The *control incidence rate* is similarly defined as the ratio of number of confirmed Covid-19 cases per number of people in the control cohort.

The *incidence rate ratio* is then defined as follows:

$$\text{Incidence Rate Ratio (IRR)} = \frac{\text{Vaccine Incidence Rate}}{\text{Control Incidence Rate}} \tag{2.1}$$

Following this the *vaccine efficacy* is defined as follows:

$$\text{Vaccine efficacy (VE)} = 100 \cdot (1 - \text{IRR}) \tag{2.2}$$

We shall be reporting the mean vaccine efficacy and then a 95% Credible Interval (95% CI) and then the 95% Highest Posterior Density Interval (95% HDI).

3 Bayesian Inference

Keeping in mind our primary goal of estimating the vaccine efficacy we need some reasonable assumptions for the prior distribution of the efficacy and the likelihood of it occurring given the observed data. In the following sections we outline three plausible models, each of which provides results which are more or less in general agreement with the results discussed in [1].

3.1 Beta-Binomial Model

The first model that we propose is the *Beta-Binomial model* [2], where the likelihood of observed data will be binomial and we shall have a beta prior. The justification is stated below.

We assume that we are performing our trials on N individuals. After s_N surveillance time has elapsed, let X be the random variable denoting the number of fresh cases of infection among the participants, and the probability that any individual gets infected is θ. We claim that the events that the participants get infected in independent. This is reasonable, as after they have received their doses, they would not come into contact with each other and hence they would not affect the probability of another participant getting infected in general. Thus $X \mid \theta \sim \textit{Binomial}(N, \theta)$.

We would want our prior beliefs regrading θ to be fairly weak, since we would want it to flexibly change give the observed data. Thus, we assume that $\theta \sim \textit{Beta}(\alpha, \beta), \alpha, \beta > 0$, which is a non-informative prior, has a support of $(0, 1)$, and also a conjugate prior for the Binomial distribution. Summing up,

$$\begin{aligned}
&\text{Prior distribution for } \theta && : p(\theta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)};\\
&\text{Likelihood of observed data} && : p\left(X = x \mid \theta\right) = \binom{N}{x}\theta^x(1-\theta)^{N-x};\\
&\text{Posterior density of } \theta && : p\left(\theta \mid X = x\right) \propto p(\theta)p\left(X = x \mid \theta\right)\\
&\Rightarrow && p\left(\theta \mid X = x\right) \propto \binom{N}{x}\frac{x^{\theta^x(1-\theta)^{N-x}\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\\
&\Rightarrow && p\left(\theta \mid X = x\right) \propto \theta^{x+\alpha-1}(1-\theta)^{N-x+\beta-1}
\end{aligned}$$

[As $\binom{N}{x}$ and $\mathrm{B}(\alpha, \beta)$ are constants, removing them doesnot affect proportionality.]

$$\therefore \text{The posterior density of } \theta \ : \ p\left(\theta \mid X = x\right) \propto \theta^{x+\alpha-1}(1-\theta)^{N-x+\beta-1} \tag{3.1}$$

Thus, the posterior density of θ follows a $Beta(x+\alpha, N-x+\beta)$ distribution (hence conjugate prior). In our case, θ signifies the *incidence rate* among vaccinated individuals. Since we have two cohorts: the vaccine cohort and the control cohort, we shall have two likelihoods and the respective priors for the two cohorts. Let:

- θ_V : probability of getting infected in the vaccine cohort
- θ_C : probability of getting infected in the control cohort
- X_V : number of individuals infected in the vaccine cohort
- X_C : number of individuals infected in the control cohort

Our prior belief is that θ_V and θ_C are the same, i.e., they follow a Beta distribution with the same parameters α and β. We can reasonably assume that the average probability of getting infected initially is low, say $\approx 2\%$ (with data on positivity rates in UK and Brazil). We can also choose $\beta = 1$ as that allows higher density for values of θ close to 0 and much lower for values close to 1. Thus, we have,

$$\frac{\alpha}{\alpha+1} = \frac{2}{100} \Rightarrow \alpha = \frac{2}{98} \approx 0.020408$$

Thus our model finally stands at:

$$\begin{aligned} &\bullet \text{ Likelihood :} && X_V \sim Binomial(N_V, \theta_V) \\ &&& X_C \sim Binomial(N_C, \theta_C) \\ &\bullet \text{ Priors :} && \theta_V, \theta_C \overset{\text{ind}}{\sim} Beta(0.020408, 1) \\ &VE = 100 \cdot \left(1 - \frac{\theta_V}{\theta_C}\right) \end{aligned} \tag{3.2}$$

And the posterior densities finally look like:

$$\begin{aligned} \theta_V &\sim Beta(x_V + 0.020408, N_V - x_V + 1) \\ \theta_C &\sim Beta(x_C + 0.020408, N_C - x_C + 1) \end{aligned} \tag{3.3}$$

Note that in our model, VE is not a *stochastic*, but a *deterministic* variable as function of θ_V and θ_C, i.e., its value is completely determined once values of θ_V and θ_C are known.

3.2 Hierarchical Beta-Binomial Model

In the previous model we had taken a single stage prior, that is $\theta \sim Beta(\alpha, \beta)$. Here we consider hyperpriors for α and β. We assume that α and β are drawn from their own probability distributions. This additional level in defining the priors for this Bayesian model allows for further flexibility. The directed acyclic graph in Figure 1 displays the relationship between the random variables, the priors and the hyperpriors.

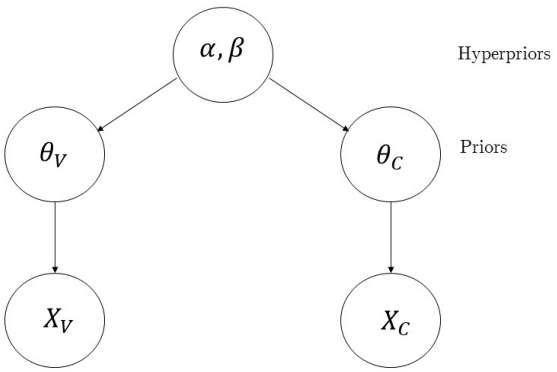


Figure 1: Simplified DAG for the hierarchical model

However we can't define the probability distributions on α and β as is. We will need to reparameterise them, which we do as follows:

Note that for a $Beta(\alpha, \beta)$ distribution, α denotes the number of success and β denotes the number of failures. Thus we define :=

$$\mu = \frac{\alpha}{\alpha + \beta}, \text{ i.e., the prior sample mean}$$

$$\eta = \alpha + \beta, \text{ i.e., the prior sample size}$$

From this, we can derive that: $\alpha = \eta \cdot \mu$ and $\beta = \eta \cdot (1 - \mu)$. Now, given that the values of α and β are known when they have been sampled from the hyper-population, the posterior of θ takes the form $Beta(\alpha + x, \beta + N - x)$ (3.3). Hence the posterior mean of θ, given x is:

$$\begin{aligned} \text{Exp}[\theta \mid X = x] =& \frac{\alpha + x}{\alpha + \beta + N} \\ =& \frac{\eta \cdot \mu + x}{\eta + N} \end{aligned} \tag{3.4}$$

We define the *shrinkage fraction* λ to be the degree by which the posterior mean of θ "shrinks" away from the sample proportion $\frac{x}{N}$ towards the prior sample mean μ. Hence we claim the following.

Claim. *Shrinkage fraction* $\lambda = \dfrac{\eta}{\eta + N}$

Proof. From (3.4) we have,

$$\begin{aligned} \text{Exp}[\theta \mid X = x] &= \frac{\eta \cdot \mu + x}{\eta + N} \\ &= \frac{\eta \cdot \mu}{\eta + N} + \frac{x}{\eta + N} \\ &= \frac{\eta}{\eta + N} \cdot \mu + \frac{N}{\eta + N} \cdot \frac{x}{N} \\ &= \lambda \cdot \mu + (1 - \lambda) \cdot \frac{x}{N} \end{aligned}$$

Here $\lambda = \dfrac{\eta}{\eta + N}$ is the shrinkage fraction. □

We should note that this claim also shows that the posterior mean is the weighted average of the prior sample mean, and the sample proportion.

We assume that λ can take any value between 0 and 1 with equal probability, and hence we can assume $\lambda \sim Uniform(0, 1)$. By performing a transformation of random variables to get the probability distribution function for η we get $p(\eta) = \frac{n^*}{(n^* + \eta)^2}, \eta > 0$ for a representative sample size n^* [3]. Further, by taking the transformation $\ln(\eta)$ and again performing transformation of random variables, we get that $\ln(\eta) \sim Logistic(\ln(n^*), 1)$ [3].

On the other hand, μ denotes the prior sample mean, i.e., in our case, assumed prior average probability of being infected by Covid-19. However, this prior probability may and does vary by place and time, for example, some countries have witnessed much higher number of daily cases, and most countries experienced "waves", where the number of fresh Covid-19 cases per day surpassed all previously recorded values, implying extremely high rates of infection. Average positivity rate of Covid-19, across countries and different surveillance periods, fall between 0.2% and 40% (observed during peaks of waves). Positivity rate has been shown to be a better measure of the spread of the pandemic and hence the probability of getting infected [4]. For our purpose, we can hence consider $\mu \sim Uniform(0, 0.4)$, in order to consider all possible probabilities of infection.

Thus summarising our current hierarchical model,

$$\begin{aligned}
&\bullet \text{ Likelihood :} && X_V \sim \textit{Binomial}(N_V, \theta_V) \\
& && X_C \sim \textit{Binomial}(N_C, \theta_C) \\
&\bullet \text{ Priors :} && \theta_V, \theta_C \overset{\text{ind}}{\sim} \textit{Beta}(\alpha, \beta) \\
&\bullet \text{ Hyperpriors :} && \eta = \alpha + \beta \\
& && \mu = \frac{\alpha}{\alpha + \beta} \\
& && \ln(\eta) \sim \textit{Logistic}(\ln(n^*), 1) \\
& && \mu \sim \textit{Uniform}(0, 0.4) \\
& VE = 100 \cdot \left(1 - \frac{\theta_V}{\theta_C}\right)
\end{aligned} \tag{3.5}$$

Further, we can define other different choices for distributions on η. For example, we can take the distribution to be $\textit{Exponential}(0.05)$ [5].

4 Model simulation with Monte Carlo methods

In Section 3, we stated the Bayesian models, which would be used to model the posterior distribution of θ_V, θ_C and subsequently VE. In order to do this, we shall be drawing samples from the posterior distributions of the aforementioned random variables using MC techniques, and then compute their mean, and the corressponding Bayesian credible intervals.

4.1 Brief note on MC methods used

Monte Carlo methods are a broad class of computational methods, which involves random sampling from probability distributions in order to generate numerical results. By the *Law of large numbers* the expected value of a random variable can be approximated by taking the sample mean of a large number of independent samples of the variable. Thus, we generally use Monte Carlo methods to randomly sample from the target probability distribution of the random variable and draw inferences from thereon.

One such class of methods to sample are the *Markov Chain Monte Carlo* methods. We shall be focussing only on random walk MCMC algorithms here. This essentially involves arbitrarily choosing a set of points in the probability space of the concerned random variables, and *randomly walking* towards points according to the decided algorithm and assigning higher probability to those points. These algorithms generate *Markov chains* which are then used to draw inferences.

MCMC methods were primarily developed in order to solve multi-dimensional problems which generic algorithms could not. However in cases of higher dimensions, MCMC algorithms tend to suffer from the curse of dimensionality, and the random walk algorthms fail to converge upon the optimal points. As a result, algorithms such as *Hamiltonian MC* were developed, which uses various methods of reducing auto-correlation between points in the Markov chain. Hamiltonian Monte Carlo is an algorithm which avoids the random walk behaviour and sensitivity to correlated parameters. This allows it to converge to target distributions in higer-dimensional spaces more quickly. *No-U-Turn-Sampling* [6] is an extension to Hamiltonian MC which reduces it's required number of steps. This improves performance and hence is also computationally faster.

4.2 Results of the Beta-Binomial Model

We apply the Beta-Binomial model from (3.2) to the data discussed in section 2.1. We compare the results obtained after the application of the following 3 MC algorithms: Monte Carlo Sampling, Hamiltonian Monte Carlo and NUTS. Markov chains are generated from the target posterior distribution of VE. In case of direct Monte Carlo sampling, the analytically calculated posterior distribution of θ_V and θ_C is used from (3.3). Following that, we can obtain the posterior distribution of the vaccine efficacy

using the relation from (3.2). From there we compute the relevant statistics. The results obtained corressponding to the vaccine efficacy are outlined in the table below, alongside the original observations:

Category		**Original**	**Monte Carlo sampling**	**Hamiltonian MC**	**NUTS**
Overall	Mean ± sd	70.4	69.88 ± 6.28	69.88 ± 6.24	69.85 ± 6.30
	95.8% CI	54.8 to 80.6	55.5 to 80.9	55.4 to 80.9	55.4 to 80.9
LD/SD	Mean ± sd	90	89.54 ± 6.39	89.56 ± 6.45	89.56 ± 6.45
	95% CI	67.4 to 97.0	73.6 to 97.9	73.5 to 97.9	73.5 to 97.9
SD/SD	Mean ± sd	62.1	61.49 ± 8.69	61.32 ± 8.72	61.32 ± 8.79
	95% CI	41.7 to 75.7	42.1 to 75.8	42.1 to 75.9	41.8 to 76.2

Table 3: Vaccine efficacy as calculated by the Beta-Binomial Model

As we can see from this table, the mean vaccine efficacy calculated using the Beta-Binomial model is very much similar to the mean vaccine efficacy as reported by [1]. This also extends to the concerned uncertainty intervals as well. Furthermore, the Bayesian paradigm results in a much more precise 95% CI of $\approx (73.5, 97.9)$ for the LD/SD category, compared to the originally calculated $(67.4, 97.0)$. Overall mean θ_V was calculated to be 0.5%, while mean θ_C was calculated to be 1.7% by the model. As for computational runtime, NUTS required on an average 44 seconds to complete its execution, compared to 75 seconds for Hamiltonian MC.

4.3 Results of the Hierarchical Beta-Binomial Model

We apply the Hierarchical Beta-Binomial Model discussed in (3.5) to the data discussed in section 2.1. As before, we compare the results obtained by using 2 different MC algorithms: Hamiltonian MC and NUTS. We compute the relevant statistics of θ_V, θ_C and VE as before. The following table outlines the results and the original observations corressponding to the vaccine efficacy.

Category		**Original**	**Hamiltonian MC**	**NUTS**
Overall	Mean ± sd	70.4	69.65 ± 6.33	69.67 ± 6.31
	95.8% CI	54.8 to 80.6	55.2 to 80.8	55.3 to 80.8
LD/SD	Mean ± sd	90	88.65 ± 6.75	88.62 ± 6.72
	95% CI	67.4 to 97.0	72.0 to 97.6	72.0 to 97.6
SD/SD	Mean ± sd	62.1	60.88 ± 8.87	60.96 ± 8.85
	95% CI	41.7 to 75.7	41.2 to 73.8	41.2 to 75.8

Table 4: Vaccine efficacy as calculated by the Hierarchical Beta-Binomial Model

As we can see from this table, this model too produces results not unlike the originaly calculated, with a deviation of 2 percentage-points atmost between the mean values.

4.4 Comparison of results

Table 3 and Table 4 outline the results obtained by the Beta-Biomial model and the Hierarchical Beta-Binomial model respectively. A quick comparison shows that both models produce almost similar results. Graph 2 compares the posterior densities of overall vaccine efficacy as calculated by both models. We can observe that both of them report almost identical densities.Graph 3 compares the results of the 3 different MC algorithms when used with the Beta-Binomial model. As expected, the algorithms all produce the same posterior densities, as shown by the curves almost overlapping with each other. Graph 4 shows the same trend for the Hierarchical Beta-Binomial model.

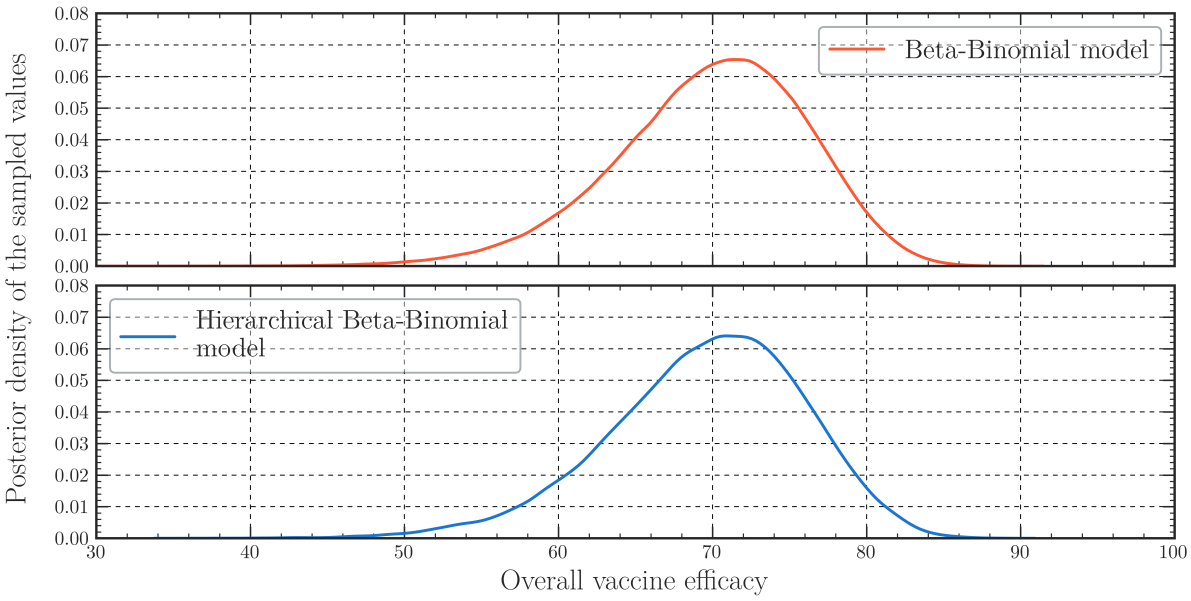


Figure 2: Posterior densities for overall vaccine efficacy as calculated by the two models

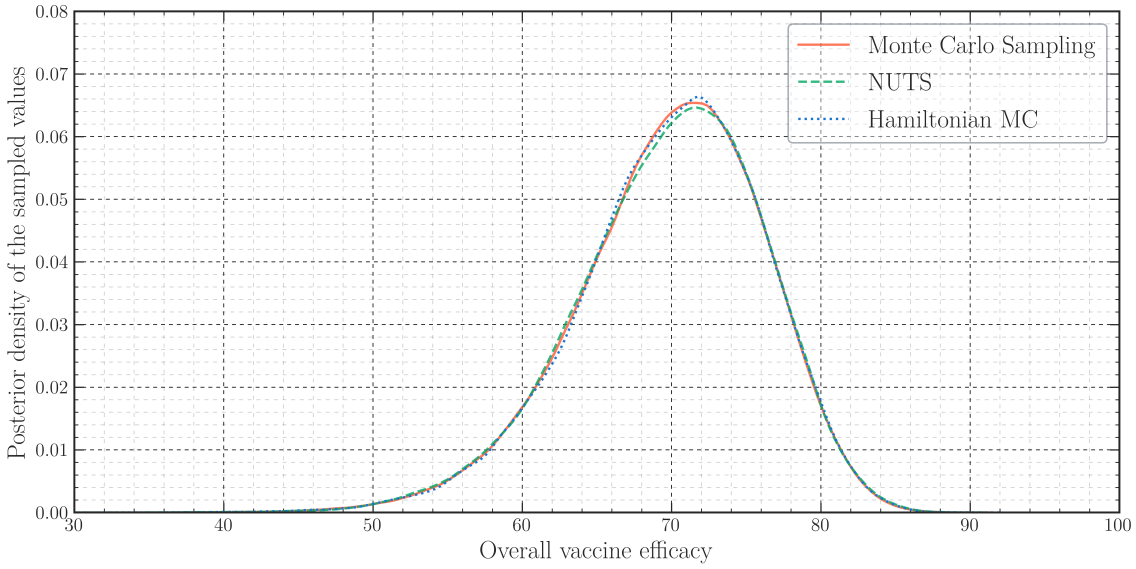


Figure 3: Comparison between the posterior densities for overall vaccine efficacy as calculated by the Beta-Binomial model using the 3 different algorithms

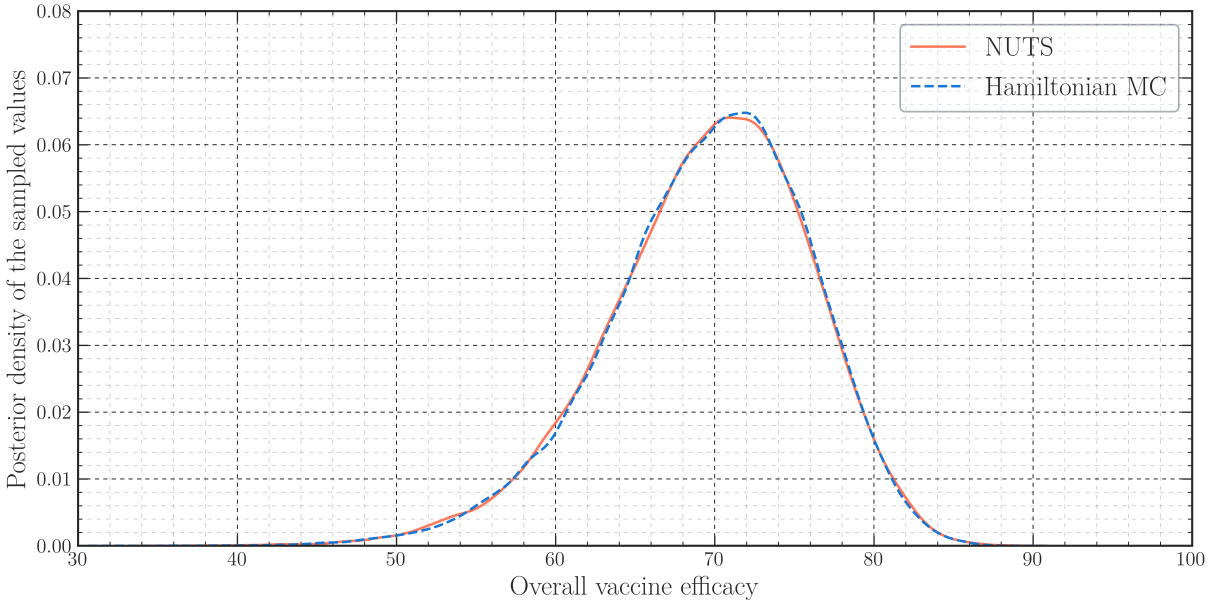
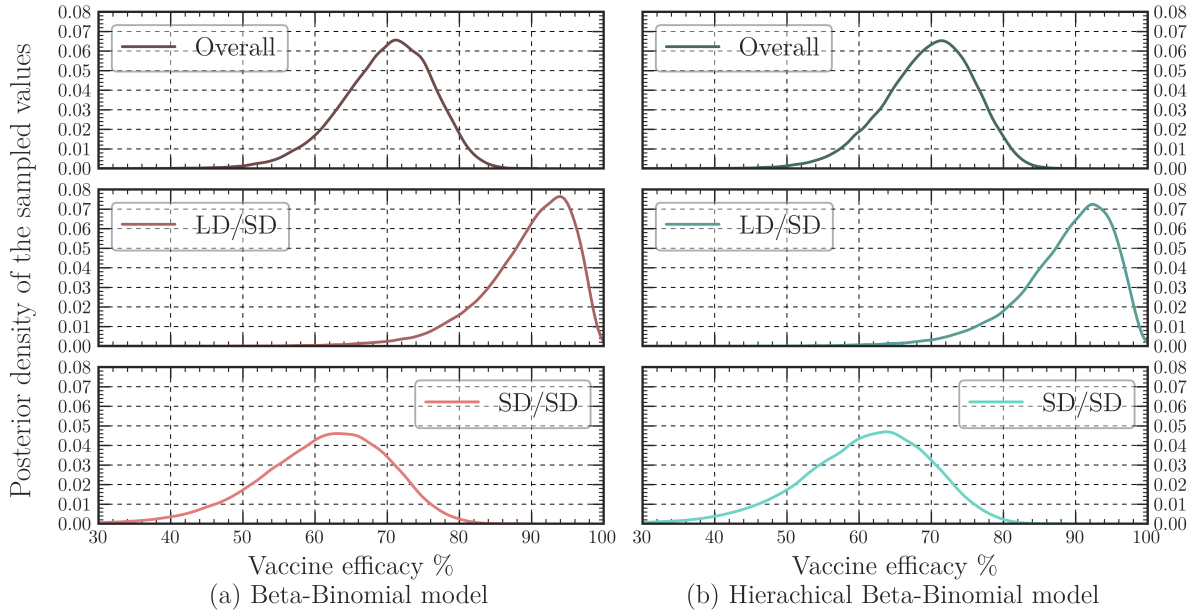


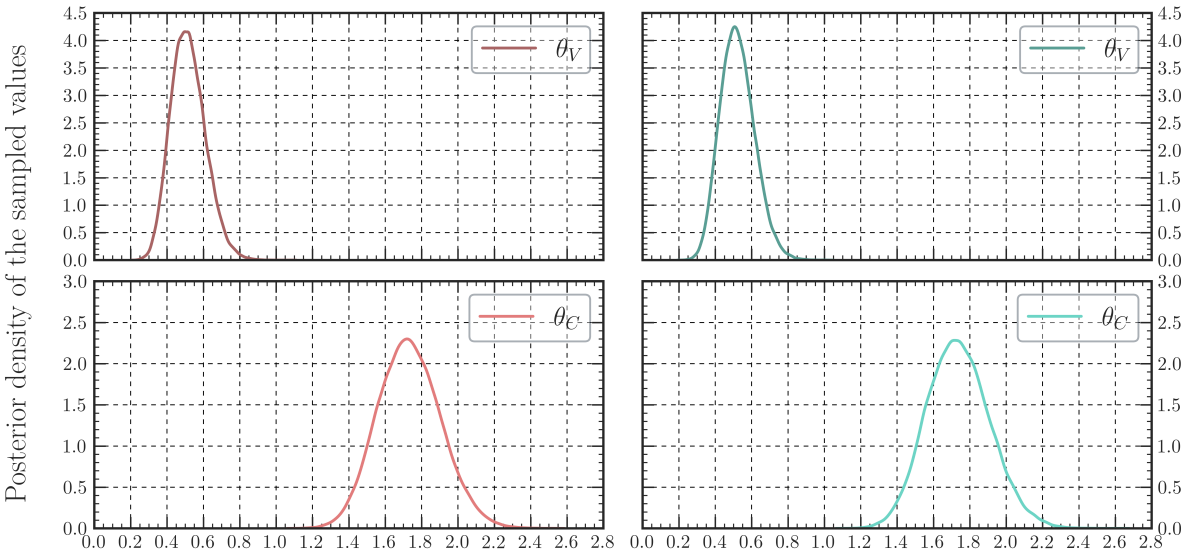
Figure 4: Comparison between the posterior densities for overall vaccine efficacy as calculated by the Hierarchical Beta-Binomial model using the 2 different algorithms

Graph 5 does a side-by-side comparison of the posterior densities of vaccine efficacy of the different categories between the two models. Graph 6 compares the incidence rates, θ_V and θ_C as calculated by both models. This shows that despite the difference in the prior distributions in both models, the posterior densities are identical. Graph 6 also shows that overall mean $\theta_V \approx 0.5\%$, while mean $\theta_C \approx 1.7\%$.



(a) Beta-Binomial model (b) Hierachical Beta-Binomial model

Figure 5: Comparison between overall, LD/SD and SD/SD vaccine efficacy posterior denisties of both models



(a) Beta-Binomial model (b) Hierachical Beta-Binomial model

Figure 6: Comparison between posterior densities of overall θ_V and θ_C calculated by both models

5 Concluding remarks

The graphical and tabular comparisons in the previous sections prove that the Bayesian paradigm results in almost identical values as initially observed in [1]. Both models produce almost same results and posterior densities for the concerned parameters. This proves that the Bayesian paradigm is also an effective alternate method of analysis.

Since the number of random variables involved is less, usage of different Monte Carlo methods will result in improvements in computation time atmost. As seen in the graphs in the previous section, different algorithms will produce almost similar results in this case.

However, hierarchical Bayesian models will be better suitable for analysis of such parameters as they allow for increased flexibility and can incorporate more details than frequentist methods. Future research in this area will allow us to obtain much more inclusive and precise results using hierarchical Bayesian models.

Acknowledgements

I would sincerely like to thank **Prof. Sucharita Roy, HOD, Department of Mathematics, St. Xavier's College (Autonomous), Kolkata** for her valuable inputs during the completion of this paper. Without her valuable inputs and insights, the paper would not have made it to completion.

I would also like to thank my parents **Mr. Susanta Das** and **Mrs. Minakshi Das** for their unfettered support throughout the duration of the project. It was their efforts which enabled me to complete my project without having to think of anything else.

References

[1] M. Voysey, S. A. C. Clemens, S. A. Madhi, L. Y. Weckx, P. M. Folegatti, and P. K. Aley, "Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: An interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK," *The Lancet*, vol. 397, no. 10269, pp. 99–111, 2021. DOI: 10.1016/S0140-6736(20)32661-1. [Online]. Available: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)32661-1/fulltext.

[2] B. Skorulski, *Bayesian Statistics of Efficacy of Pfizer-BioNTech COVID-19 Vaccine-part I*, 2021. [Online]. Available: https://towardsdatascience.com/bayesian-statistics-of-efficacy-of-pfizer-biontech-covid-19-vaccine-part-i-efac8d4e0539.

[3] J. Albert and J. Hu, *Probability and Bayesian modeling*, 1st ed. 2019, ch. 10.3 Hierarchical Beta-Binomial Modeling.

[4] A. A. Dallal, U. AlDallal, and J. A. Dallal, "Positivity rate: An indicator for the spread of covid-19," *Current Medical Research and Opinion*, vol. 37, no. 12, pp. 2067–2076, 2021, PMID: 34517740. DOI: 10.1080/03007995.2021.1980868. eprint: https://doi.org/10.1080/03007995.2021.1980868. [Online]. Available: https://doi.org/10.1080/03007995.2021.1980868.

[5] J. Niemi, "Hierarchical models," p. 12, 2019. [Online]. Available: https://www.jarad.me/courses/stat544/slides/Ch05/Ch05a.pdf.

[6] M. D. Hoffman and A. Gelman, "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," 2011. arXiv: 1111.4246 [stat.CO].