# 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*

# Why?


- The Covid-19 pandemic caused great distress since it began in 2020. Educational, government and business institutions were forced to shift all activity to the online mode.
- By 2021, several vaccines were developed worldwide for immediate use, one of them being the Oxford/AstraZeneca vaccine (AZD1222). This was approved for use in India under the name Covishield.
- And hence the need to correctly gauge the efficacy of this vaccine.

# How?

- Based on data of four randomised controlled clinical trials of the vaccine, we use the Bayesian paradigm to model the posterior distribution of the vaccine efficacy.
- We impose certain prior distributions on the incidence rates of infection and choose a suitable likelihood distribution in order to achieve the above.
- We calculate the mean vaccine efficacy, and it's 95% Bayesian Credible Interval.

# Philosophy of Bayesian Inference

The core principle in Bayesian inference is ***preserving uncertainty*** about the parameters we want to infer about. Say we are interested in some population parameter $\boldsymbol{\theta}$ and $\boldsymbol{X}$ (observed samples drawn from the population). Then:

**Prior distribution of $\boldsymbol{\theta}$:** Probability distribution denoting the prior beliefs of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$

**Likelihood of $\boldsymbol{X}|\boldsymbol{\theta}$:** Likelihood of observing $\boldsymbol{X}$ given $\boldsymbol{\theta}$, $p(\boldsymbol{X}|\boldsymbol{\theta})$

**Posterior distribution of $\boldsymbol{\theta}$:** Probability distribution of $\boldsymbol{\theta}$ given $\boldsymbol{X}$, $p(\boldsymbol{\theta}|\boldsymbol{X}) \propto p(\boldsymbol{\theta}) \cdot p(\boldsymbol{X}|\boldsymbol{\theta})$

# Results of the clinical trials

- Trials carried out on participants divided into 2 cohorts: **Vaccine** and **Control**
- Across trials, two different dosages of vaccine were administered:
  - Low Dose/Standard Dose (LD/SD)
  - Standard Dose/Standard Dose (SD/SD)

| | Total number of cases | Vaccine cohorts n/N (%) | Control cohorts n/N (%) | Vaccine Efficacy % (95% CI) |
|---|---|---|---|---|
| Overall | 131 | 30/5807 (0.5%) | 101/5829 (1.7%) | 70 (54.8 to 80.6) |
| 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) |

Data taken from Table 2 of "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," by Voysey et al. For overall vaccine efficacy the CI is of 95.8%.

# Terms and definitions

**Incidence rate:**

$$\frac{Number\ of\ confirmed\ Covid{-}19\ cases\ in\ a\ cohort}{Total\ number\ of\ participants\ in\ the\ cohort}$$

For the two cohorts we have:
**Vaccine Incidence Rate** and **Control Incidence Rate**

**Incidence rate ratio (IRR):**

$$\frac{Vaccine\ Incidence\ Rate}{Control\ Incidence\ Rate} = \frac{\theta_V}{\theta_C}$$

**Vaccine efficacy (VE):**

$$100 \cdot (1 - IRR)\% = 100 \cdot \left(1 - \frac{\theta_V}{\theta_C}\right)\%$$

# Beta-Binomial model

- To model the posterior distribution of the $VE$, we need the posterior distributions of $\theta_V$ and $\theta_C$.
- Say $N$ people are participating in the trials.
- Let,
  $\theta \rightarrow$ the probability of a person being infected by Covid-19.
  $X \rightarrow$ the number of people infected by Covid-19 during the surveillance period.
- Considering that appropriate self-distancing protocols are maintained, each participants probability of being infected is independent of the others. Thus given $X \mid \theta \sim Bin(N, \theta)$.
- We want our prior beliefs regarding $\theta$ to be fairly weak. Moreover as the likelihood of observed data is following a Binomial distribution, we choose that $\theta \sim Beta(\alpha, \beta)$, which is a conjugate prior to the binomial distribution.
- Hence the posterior distribution of $\theta$ is: $\theta \sim Beta(x + \alpha, N - x + \beta)$.

Hence:-

- Our belief is that on an average $\theta$ is very low initially, $\approx 2\%$. We assume $\beta = 1$ as that allows for higher probability density for low values of $\theta$. Thus, equating $\frac{\alpha}{\alpha+1} = 0.02$, we get $\alpha = 0.020408$.
- Our prior belief is also that $\theta_V = \theta_C$.

Final model:

Likelihood:

$$X_V \sim Bin(N_V, \theta_V)$$
$$X_C \sim Bin(N_C, \theta_C)$$

Priors:

$$\theta_V, \theta_C \sim Beta(0.020408, 1)$$

Posteriors:

$$\theta_V \sim Beta(x_V + 0.020408, N_V - x_V + 1)$$
$$\theta_C \sim Beta(x_C + 0.020408, N_C - x_C + 1)$$

$$VE = 100 \cdot \left(1 - \frac{\theta_V}{\theta_C}\right)$$

# Results of the Beta-Binomial model

| Category | | Original values | Model results |
|---|---|---|---|
| Overall | Mean ± sd<br>95.8% CI | 70<br>54.8 to 80.6 | 69.85 ± 6.30<br>55.4 to 80.9 |
| LD/SD | Mean ± sd<br>95% CI | 90.0<br>67.4 to 97.0 | 89.56 ± 6.45<br>73.5 to 97.9 |
| SD/SD | Mean ± sd<br>95% CI | 62.1<br>41.0 to 75.7 | 61.32 ± 8.79<br>41.8 to 76.2 |

# Hierarchical Beta-Binomial model

- We improve upon the Beta-Binomial model by considering hyperpriors for $\theta_V, \theta_C$.
- Reasoning: An assumption of a fixed average probability of infection does not always hold. Based on positivity rates during the peaks of infection waves, we can see the positivity rate can fluctuate between as much as 0.2% and 40% based on surveillance times. We need more flexibility in the model to account for this fluctuation.
- We have $\theta \sim Beta(\alpha, \beta)$. Let,

$$\eta = \alpha + \beta$$
$$\mu = \frac{\alpha}{\alpha + \beta}$$

- Thus $\eta \rightarrow$ prior sample size, $\mu \rightarrow$ prior sample mean. Hence the posterior mean on $\theta$ is given by

$$E[\,\theta \mid X = x\,] = \frac{\mu \cdot \eta + x}{\eta + N}$$

- We define:

  $\lambda \rightarrow$ The shrinkage fraction, i.e., the degree by which the posterior mean shrinks away from the prior sample mean $\mu$ towards the sample proportion $\frac{x}{N}$. We can show that

$$\lambda = \frac{\eta}{\eta + N}$$

- Assuming $\lambda \sim Uniform(0,1)$, by applying a transformation of random variables, we get the probability density function for $\eta$ to be $p(\eta) = \frac{n^*}{(n^*+\eta)^2}$, for some representative sample size $n^*$. By taking a log transformation, we get $\ln(\eta) \sim Logistic(\ln(n^*), 1)$.

- Based on fluctuations of positivity rates, we can take $\mu \sim Uniform(0, 0.4)$.

Final model:

Likelihood:

$$X_V \sim Bin(N_V, \theta_V)$$
$$X_C \sim Bin(N_C, \theta_C)$$

Priors:

$$\theta_V, \theta_C \sim Beta(\alpha, \beta)$$

Hyperpriors:

$$\eta = \alpha + \beta$$
$$\mu = \frac{\alpha}{\alpha + \beta}$$
$$\ln(\eta) \sim Logistic(\ln(n^*), 1)$$
$$\mu \sim Uniform(0, 0.4)$$

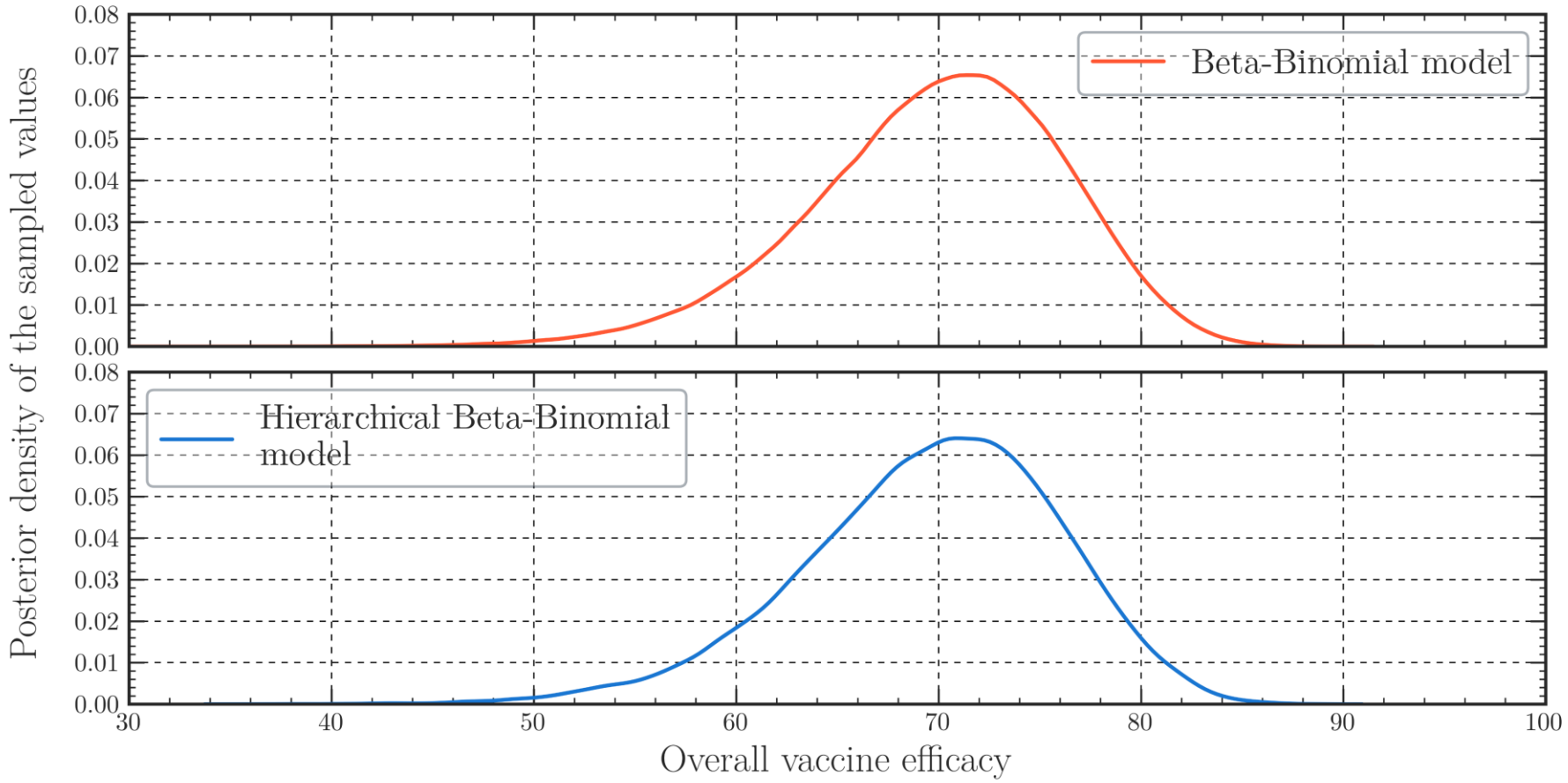
Posteriors:

$$VE = 100 \cdot \left(1 - \frac{\theta_V}{\theta_C}\right)$$

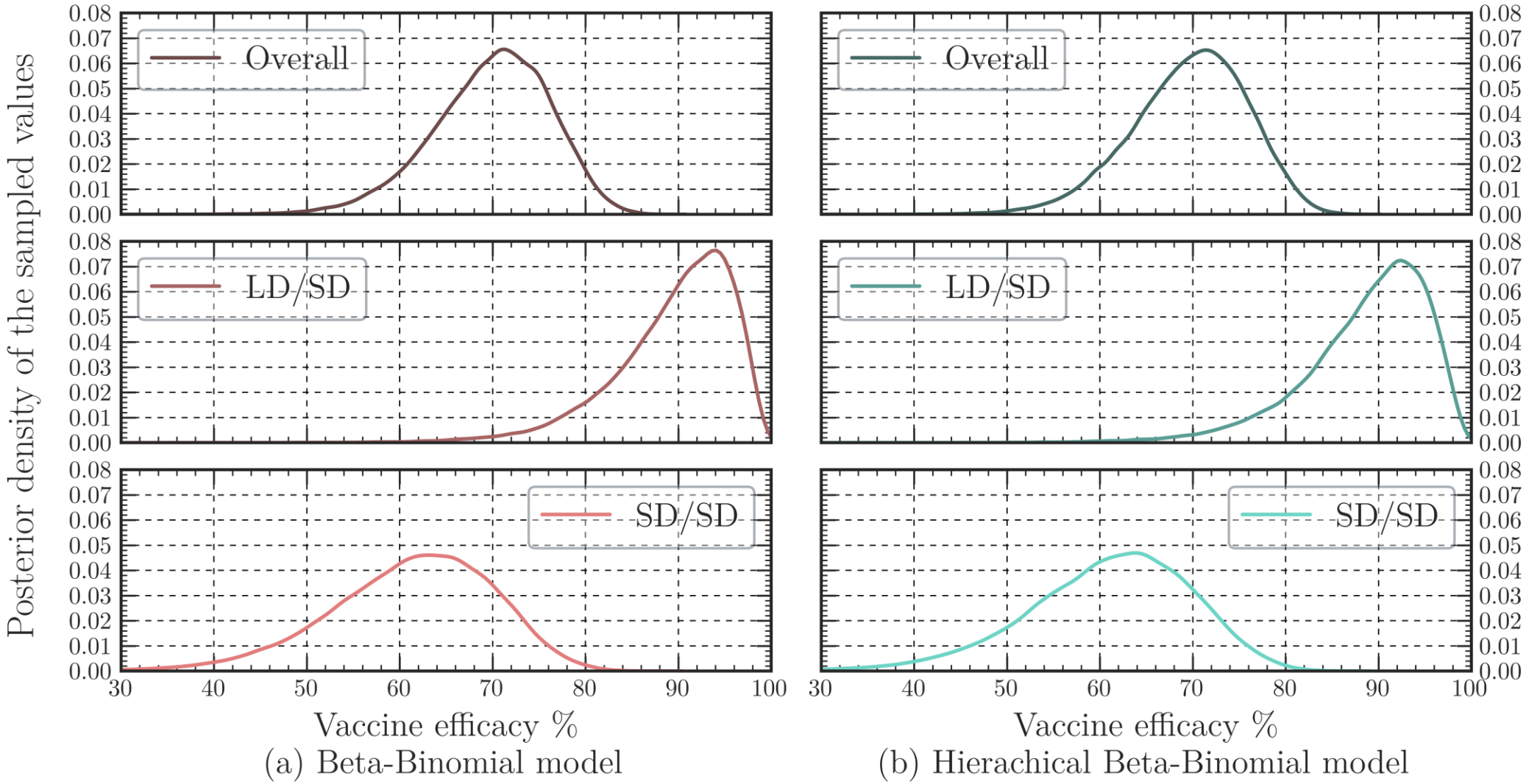
# Results of the hierarchical Beta-Binomial model

| Category | | Original values | Beta-Binomial model results | Hierarchical Beta-Binomial model results |
|---|---|---|---|---|
| Overall | Mean $\pm$ sd<br>95.8% CI | 70<br>54.8 to 80.6 | 69.85 $\pm$ 6.30<br>55.4 to 80.9 | 69.67 $\pm$ 6.31<br>55.3 to 80.8 |
| LD/SD | Mean $\pm$ sd<br>95% CI | 90.0<br>67.4 to 97.0 | 89.56 $\pm$ 6.45<br>73.5 to 97.9 | 88.62 $\pm$ 6.72<br>72.0 to 97.6 |
| SD/SD | Mean $\pm$ sd<br>95% CI | 62.1<br>41.0 to 75.7 | 61.32 $\pm$ 8.79<br>41.8 to 76.2 | 60.96 $\pm$ 8.85<br>41.2 to 75.8 |

# Graphical comparisons




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




Comparison between overall, LD/SD and SD/SD vaccine efficacy posterior densities of both models

# Concluding remarks

Thus we see that Bayesian analysis provides an equally feasible method of analysis of such problems, and in some cases returns much more precise bounds for confidence intervals at the same confidence level.

At the same time, we saw how we can the models much more flexible all the while retaining performance. 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.

# Thank you

The End