# Mathematical Statistics II

## STA2212H S LEC9101

Week 7

March 3 2021

Start recording!

Behind the numbers: what does it mean if a Covid vaccine has '90% efficacy'?

*David Spiegelhalter and Anthony Masters*

Confusion surrounds the vaccines' effectiveness. The leading Cambridge professor clarifies the data behind the trials

▲ People rest in Salisbury Cathedral, England, after receiving the Pfizer/BioNTech vaccine. Photograph: Neil Hall/EPA

# Vaccine efficacy

Guardian, Jan 24

Link to Guardian

Pfizer-BioNTech vaccine trial:

vaccine: 22000 subjects, 8 cases

placebo: 22000 subjects, 162 cases

$8/162 = 5\% \Longrightarrow 95\%$ efficacy

data released November 18 2020 link

published December 31 2020 in NEJM link

Behind the numbers: what does it mean if a Covid vaccine has '90% efficacy'?

*David Spiegelhalter and Anthony Masters*

Confusion surrounds the vaccines' effectiveness. The leading Cambridge professor clarifies the data behind the trials

▲ People rest in Salisbury Cathedral, England, after receiving the Pfizer/BioNTech vaccine. Photograph: Neil Hall/EPA

# Pfizer publication

NEJM

Editor's Note: This article was published on December 10, 2020, at NEJM.org.

ORIGINAL ARTICLE

## Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Absalon, M.D., Alejandra Gurtman, M.D., Stephen Lockhart, D.M., John L. Perez, M.D., Gonzalo Pérez Marc, M.D., Edson D. Moreira, M.D., Cristiano Zerbini, M.D., Ruth Bailey, B.Sc., Kena A. Swanson, Ph.D., et al., for the C4591001 Clinical Trial Group*

Article Figures/Media Metrics

13 References 263 Citing Articles Letters

December 31, 2020
N Engl J Med 2020; 383:2603-2615
DOI: 10.1056/NEJMoa2034577
Chinese Translation 中文翻译

Results: A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6).

# Pfizer publication

NEJM

**Table 2.** Vaccine Efficacy against Covid-19 at Least 7 days after the Second Dose.*

| Efficacy End Point | BNT162b2 | | Placebo | | Vaccine Efficacy, % (95% Credible Interval)‡ | Posterior Probability (Vaccine Efficacy >30%)§ |
|---|---|---|---|---|---|---|
| | No. of Cases | Surveillance Time (n)† | No. of Cases | Surveillance Time (n)† | | |
| | (N=18,198) | | (N=18,325) | | | |
| Covid-19 occurrence at least 7 days after the second dose in participants without evidence of infection | 8 | 2.214 (17,411) | 162 | 2.222 (17,511) | 95.0 (90.3–97.6) | >0.9999 |
| | (N=19,965) | | (N=20,172) | | | |
| Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection | 9 | 2.332 (18,559) | 169 | 2.345 (18,708) | 94.6 (89.9–97.3) | >0.9999 |

* The total population without baseline infection was 36,523; total population including those with and those without prior evidence of infection was 40,137.

† The surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.

‡ The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

§ Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

**STA 2212S: Mathematical Statistics II**
**Syllabus** Spring 2021

Updated Mar 3

| Week | Date | Methods | References |
|---|---|---|---|
| 1 | Jan 13/15 | Review of parametric inference | AoS Ch 9 |
| 2 | Jan 20/22 | ~~Significance testing~~Hypothesis testing | AoS Ch 10.1,2,~~6~~,7; SM Ch 7.3.2, 4 |
| 3 | Jan 27/29 | Significance testing | AoS Ch 10.2, 6; SM Ch 7.3.1, Ch 4 |
| 4 | Feb 3/5 | Goodness of fit testing, Intro to multiple testing | AoS Ch 10.3,4,5,8; SM p.327-8 (hard) |
| 5 | Feb 10/12 | Multiple testing and FDR | AoS Ch 10.7, EH Ch 15.1,2 |
| | Feb 17/19 | Break | |
| 6 | Feb 24/26 | Bayesian Inference | AoS Ch 11.1-4; SM Ch 11.1,2; EH Ch 3, 13 |
| 7 | Mar 3/5 | Bayesian Inference | AoS Ch 11.5-9; SM Ch 11.4 |
| 8 | Mar 10/12 | Empirical Bayes | EH Ch 6, SM Ch 11.5 |
| 9 | Mar 17/19 | Statistical decision theory | AoS Ch 12, SM Ch 11.5.2 |
| 10 | Mar 24/26 | Multivariate Models | AoS Ch 14; SM Ch 6.3 |
| 11 | Mar 31 | Causal Inference and Graphical Models | AoS Ch 16, 17 |
| 12 | Apr 7 | Recap | |

# Recap 2

- Bayes theorem – conditional probability
- twin boys, diagnostic tests
- ingredients for inference: prior, posterior
- Bernoulli, normal correlation coefficient
- point estimates, equi-tailed posterior intervals, HPD intervals
- transformation of parameters
- approximate normality of posterior

discrete

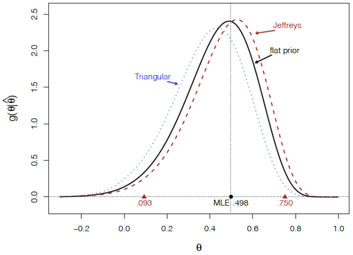


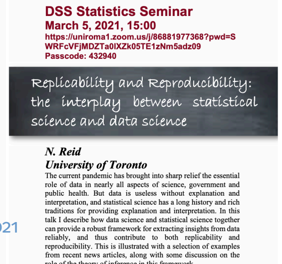

**Figure 3.2** Student scores data; posterior density of correlation $\theta$ for three possible priors.

**Start Recording**

1. Bayesian inference Part II
   Multi-par Bayes; noninformative priors; subjective priors; philosophy; Laplace approx; ; AoS 11.6-9; SM 11.1.3;11.2;11.311.4
2. Friday – I owe you: proof of BH; careful $\chi^2$; careful LRT; notes from Feb 26

- Mar 5 12.00 pm EST Olufunmilayo I. Olopade
  “What African Genomes Tell Us
  About the Origins of Breast Cancer” Stage ISSS

- Mar 5 9.00 am EST ME



**DSS Statistics Seminar**
**March 5, 2021, 15:00**
https://uniroma1.zoom.us/j/86881977368?pwd=SWRFcVFjMDZTa0lXZk05TE1zNm5adz09
Passcode: 432940

Replicability and Reproducibility: the interplay between statistical science and data science

*N. Reid*
*University of Toronto*

The current pandemic has brought into sharp relief the essential role of data in nearly all aspects of science, government and public health. But data is useless without explanation and interpretation, and statistical science has a long history and rich traditions for providing explanation and interpretation. In this talk I describe how data science and statistical science together can provide a robust framework for extracting insights from data reliably, and thus contribute to both replicability and reproducibility. This is illustrated with a selection of examples from recent news articles, along with some discussion on the role of the theory of inference in this framework.

STAGE ISSS: Olufunmilayo I. Olopade

# Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(x^n \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid x^n) = \dfrac{f(x^n \mid \theta)}{f(x^n)}$ $\qquad x^n = (x_1, \ldots, x_n)$
- $\pi(\theta \mid x^n) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\}; \qquad \pi(\theta \mid x^n) \approx N\{\tilde{\theta}, \tilde{\jmath}(\tilde{\theta})\}$
- careful statement — Berger, 1985; Ch.4
  - For any $a, b \in \mathbb{R}, a < b$
  - let $a_n = \hat{\theta}_n + a j^{-1/2}(\hat{\theta}_n)$, $b_n = \hat{\theta}_n + b j^{-1/2}(\hat{\theta}_n)$
  - $\hat{\theta}_n$ is the solution of $\ell'(\theta; x^n) = 0$, assumed unique, and $j(\theta) = -\ell''(\theta; x^n)$

  Then
$$\int_{a_n}^{b_n} \pi(\theta \mid x^n) \longrightarrow \Phi(b) - \Phi(a), \quad n \to \infty.$$

need $\pi(\theta) > 0, \pi'(\theta)$ continuous

# Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(x^n \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid x^n) = \dfrac{f(x^n \mid \theta)}{f(x^n)}$ $\qquad x^n = (x_1, \ldots, x_n)$
- $\pi(\theta \mid x^n) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\}; \qquad \pi(\theta \mid x^n) \approx N\{\tilde{\theta}, \tilde{j}(\tilde{\theta})\}$
- approximate posterior probability intervals
- exact posterior probability intervals $\qquad \tilde{\theta} \approx \hat{\theta}$

# Choosing a prior



- subjective
- conjugate
- “flat”
- “matching”
- convenience
- empirical

# Choosing a prior



- subjective
- conjugate
- “flat”
- “matching”
- convenience
- empirical

# Noninformative prior: Jeffreys'

AoS 11.6; SM 11.1.3

# Matching prior: scalar parameter

# Matching prior: scalar parameter

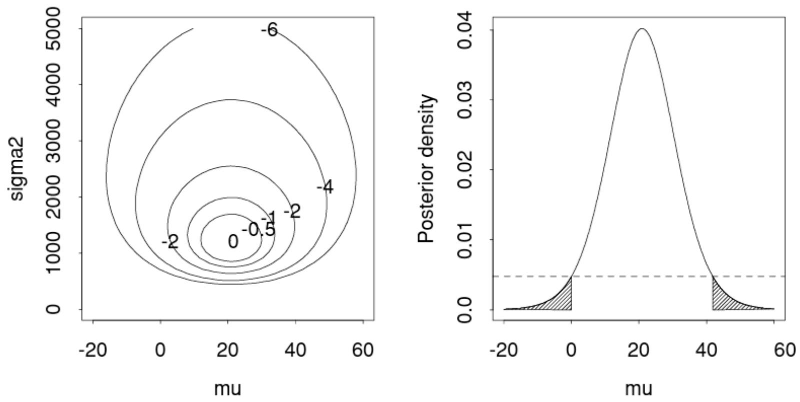



**Figure 11.2** Posterior densities of $(\mu, \sigma^2)$ of normal model for maize data. Left: contours of the normalized log joint posterior density. Right: marginal posterior density for $\mu$, showing 95% HPD credible set, which is the set of values of $\mu$ whose values of the posterior density $\pi(\mu \mid y)$ lie above the dashed line. The shaded region has area 0.05.

# Multi-parameter models

# Multi-parameter models

# Bayesian philosophy



- empirical probability

- epistemic probability

$\longrightarrow$ AoS.pdf 1

# Bayesian philosophy



- Example 11.8

- Example 11.10

$\longrightarrow$ AoS.pdf 2