Bayes Pharma Conference
Bethesda – October 14th, 2022

Vaccine immunogenicity demonstration on cellular responses using Bayesian zero-inflated beta regression to improve preclinical decision-making on small sample sizes

Hakem Ben Addi & **Cédric Taverne**
Vaccine Research Statistics, GSK

Contact: cedric.x.taverne@gsk.com

gsk.com

Disclaimer

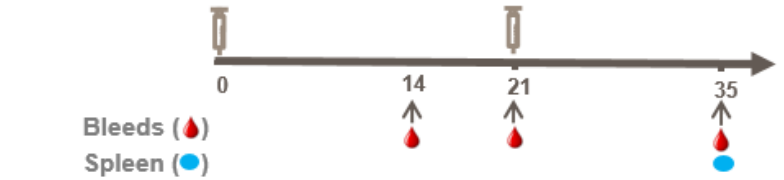
- This work was sponsored by GlaxoSmithKline Biologicals SA
- Hakem Ben Addi and Cédric Taverne are employees of the GSK group of companies
- But any opinions expressed in this presentation are those of the presenter

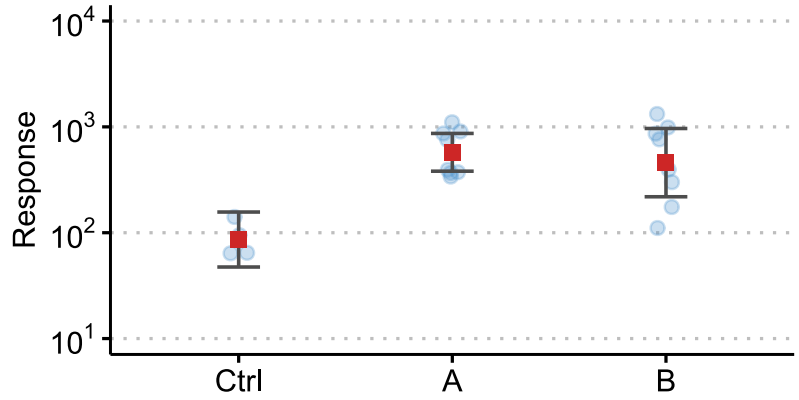
GSK

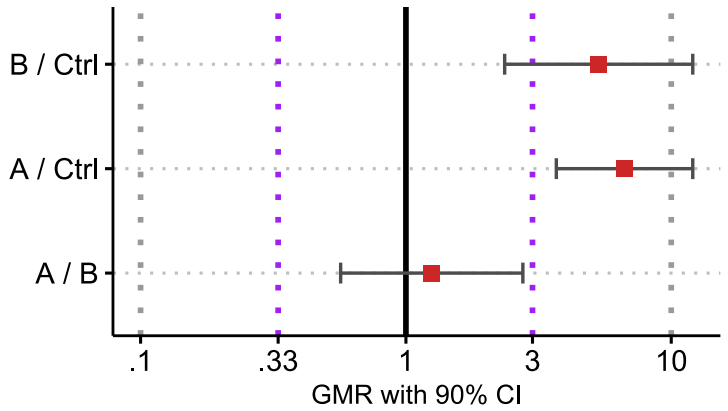
Vaccine Immunogenicity

- Vaccine immunogenicity
  = Ability to induce an immune response
- In preclinical research, mainly assessed in animal models on small sample sizes (≤ 10 / group)
- Post-vaccination levels are compared to pre-vaccination or control group levels
- Responses: Antigen-specific antibodies in sera, functional responses in sera, cellular responses (e.g., T-cells, B-cells) in lymphoid organs…
- Try to demonstrate non-inferiority with respect to a benchmark, superiority above a biologically relevant threshold, equivalence between lots, or to find the optimal dose-response…
- Default statistical approach: Log_{10}-normal hierarchical models

Group	N	Formulation	Dose
Ctrl	4	Saline	/
A	8	A	10 µg
B	8	B	10 µg

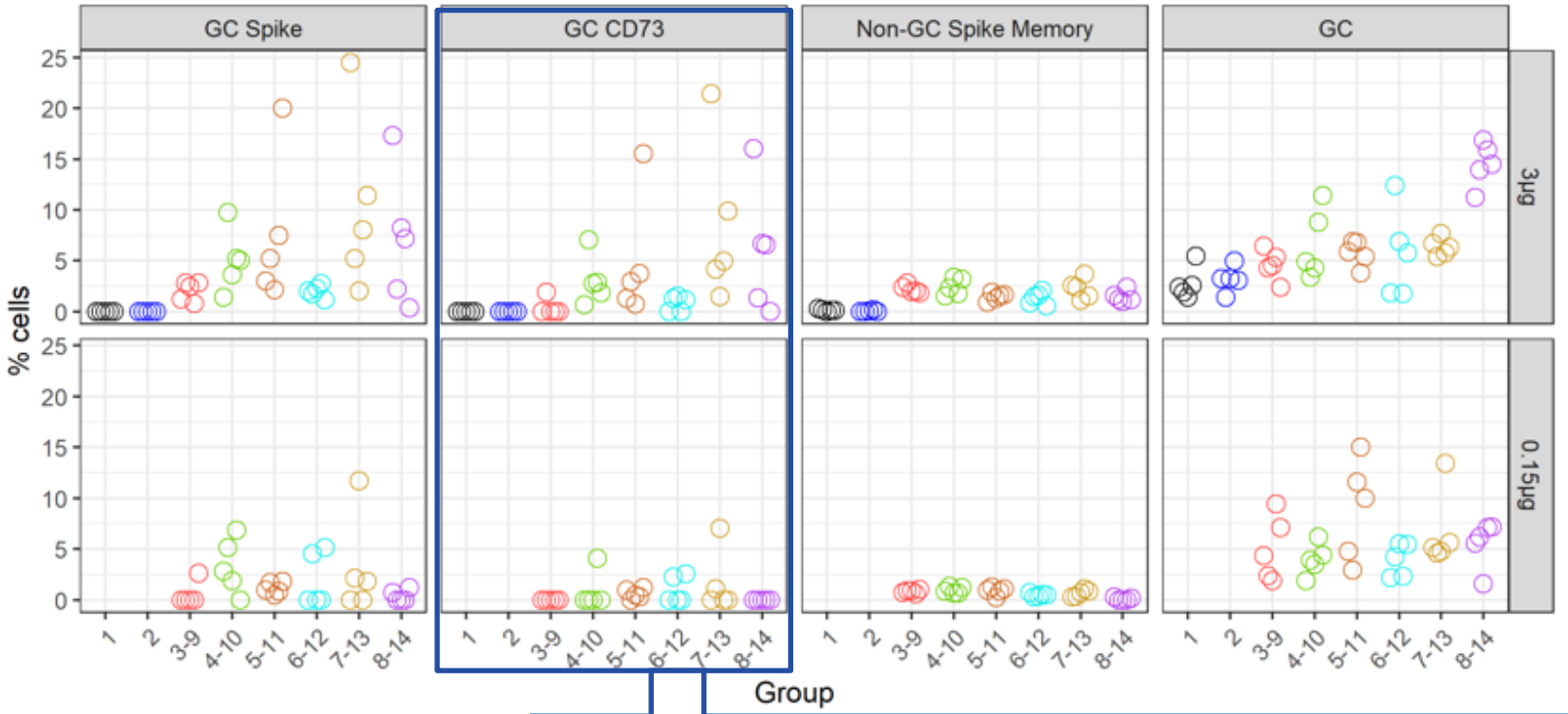


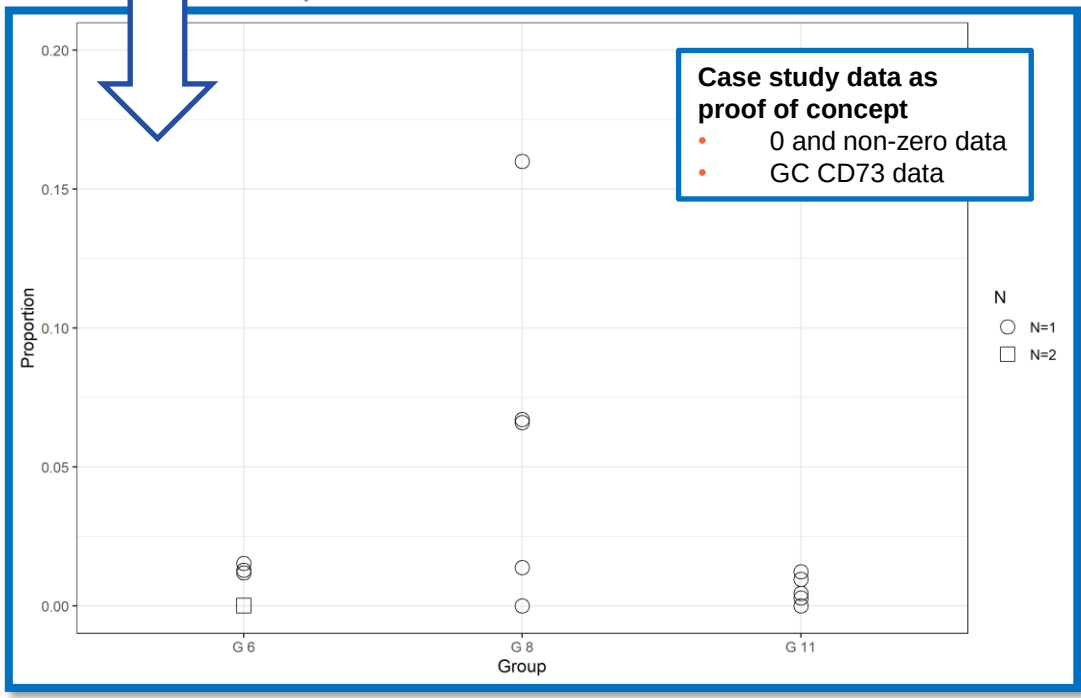




Cellular responses

- Immunogenicity through cellular responses are measured by flow cytometry
- Multidimensional data:
  - Several cell types (e.g., T-cells, B-cells)
  - Several sub-populations (e.g., Germinal centres)
- Data are provided as:
  - Cell counts (not recommended as total counts are specific to each sample)
  - Cell proportions or percentages per sample (preferred)
- Data often contains many zeros, especially when looking at combinations of cellular responses





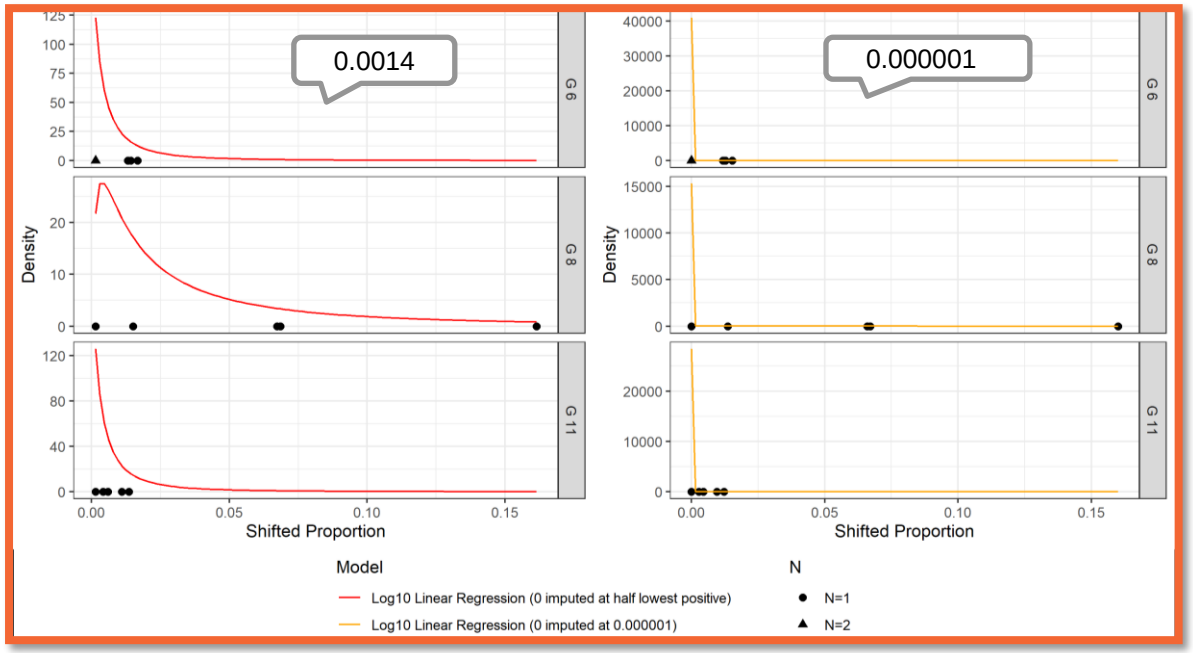
Log-linear model

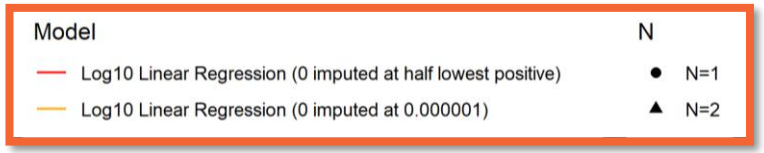
$$\log_{10} Y_{ij} \sim N\left(\mu_i, \sigma_i^2\right)$$

with index $i = 1,2,\ldots,K$ for groups and $j = 1,2,\ldots,n_i$ for samples.
The hierarchical structure depends on the survey design

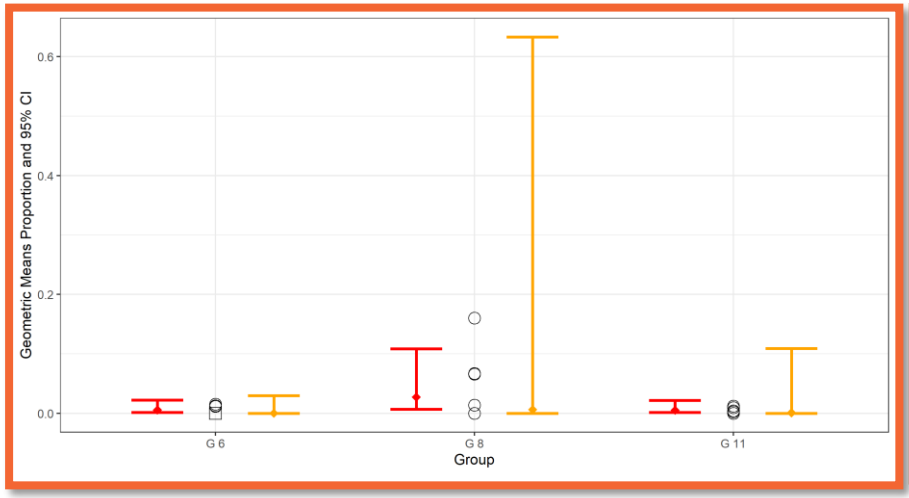
- Aligned with other response types (e.g., antibody responses) and clinical statistical models
- Zeros cannot be included without shifting (all) data by a constant
  - ➢ Impact model outcomes
  - ➢ Sometimes poorly fitting proportions and/or small sample sizes data

Frequentist predictive distributions with different imputations of zeros

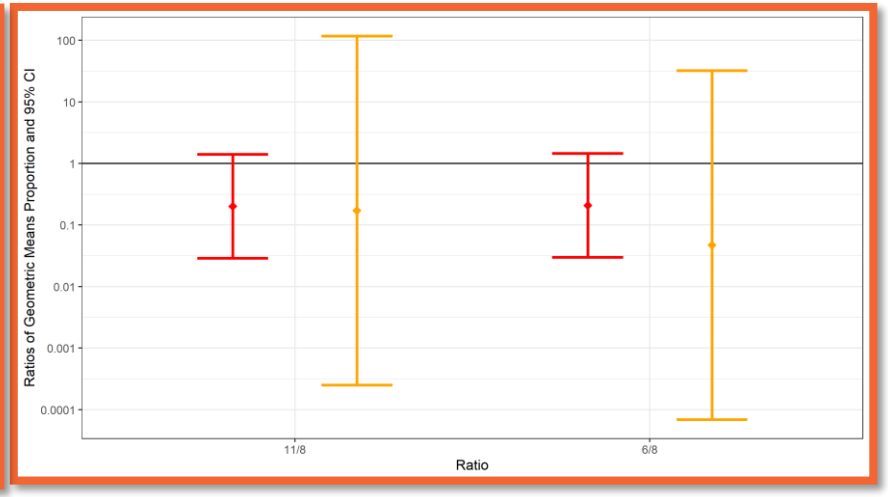




Log-linear model geometric mean proportions



Log-linear model geometric mean ratios (GMR)



Beta distribution and beta regression model

- **Beta distribution** $Y \sim Beta(p, q)$
  - Continuous probability distribution with various (skewed) shapes and support on $(0,1)$
  - $E(Y) = \frac{p}{p+q}$ and $Var(Y) = \frac{pq}{(p+q)^2(p+q+1)}$
- **Beta regression model**
  - Location-scale model on a reparameterization[1] of the distribution
  - Location $E(Y) = \boldsymbol{\mu}$, with $0 < \mu < 1$
  - Scale (precision) $Var(Y) = \mu(1-\mu)\,/(\boldsymbol{\phi} + 1)$ with $\phi > 0$
  - Link to covariates: $h_1(\mu) = f_1(X, \beta)$ $\quad h_2(\phi) = f_1(Z, \gamma)$

    with h_1 generally defined as logit and h_2 as ln
  - Limitation: The support of the beta excludes 0 and 1 or you need to constrain $\mathrm{p}, \mathrm{q} > 1$ → Either omit, or shift zeros

Various shapes of the beta distribution
(https://en.wikipedia.org/wiki/File:Beta_distribution_pdf.svg)

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} y^{p-1}(1-y)^{q-1}$$

with $0 < \mathrm{y} < 1$ and $\mathrm{p}, \mathrm{q} > 0$

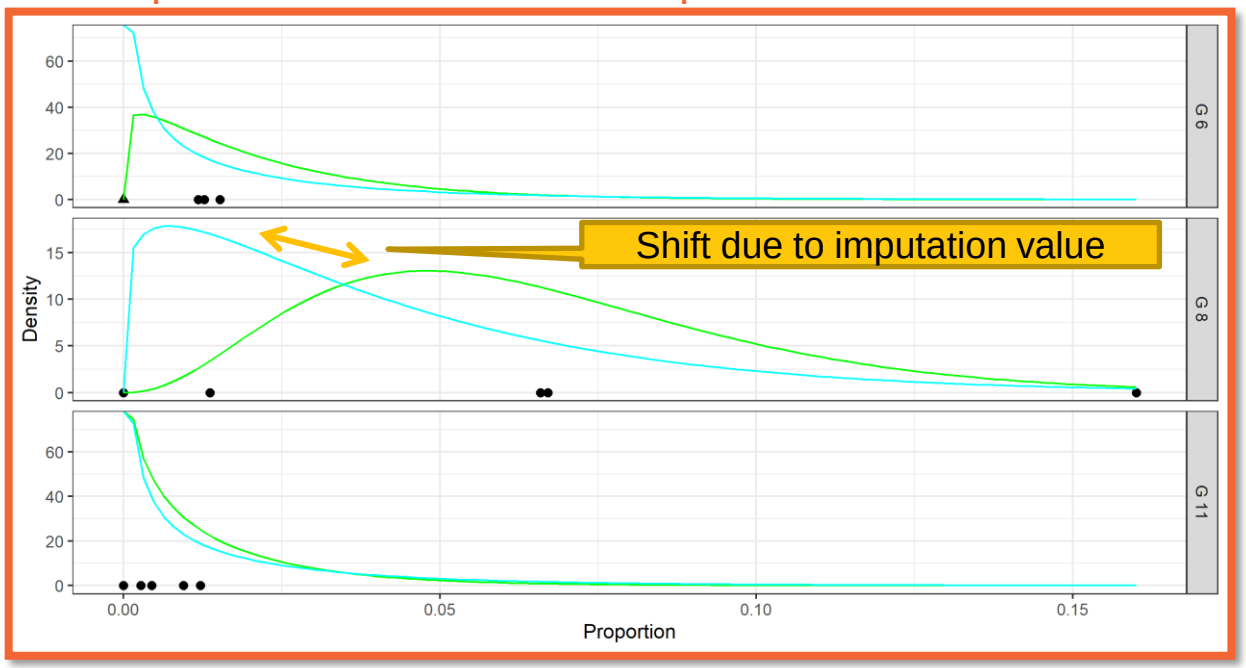
and $\Gamma(\cdot)$ is the gamma function

1. Simas AB, Barreto-Souza W, Rocha AV (2010). "Improved Estimators for a General Class of Beta Regression Models." Computational Statistics & Data Analysis, 54(2), 348–366.

Beta regression model

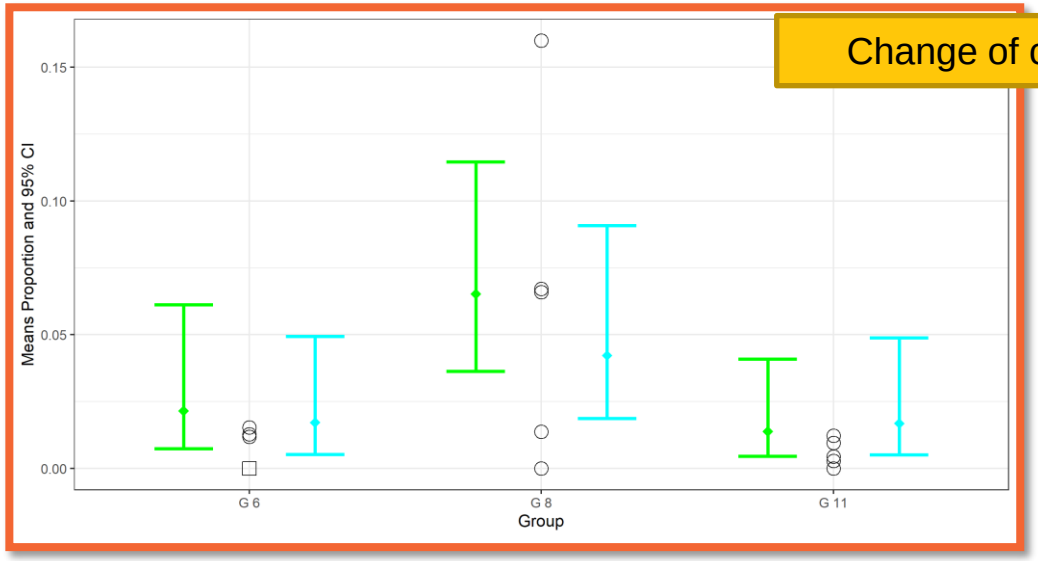
- More realistic fit of the data
- Limitation: Support excludes 0 and 1
  - Either omit, or shift zeros
  - Imputation impact the outcomes

Posterior predictive distribution with/without imputation of zeros

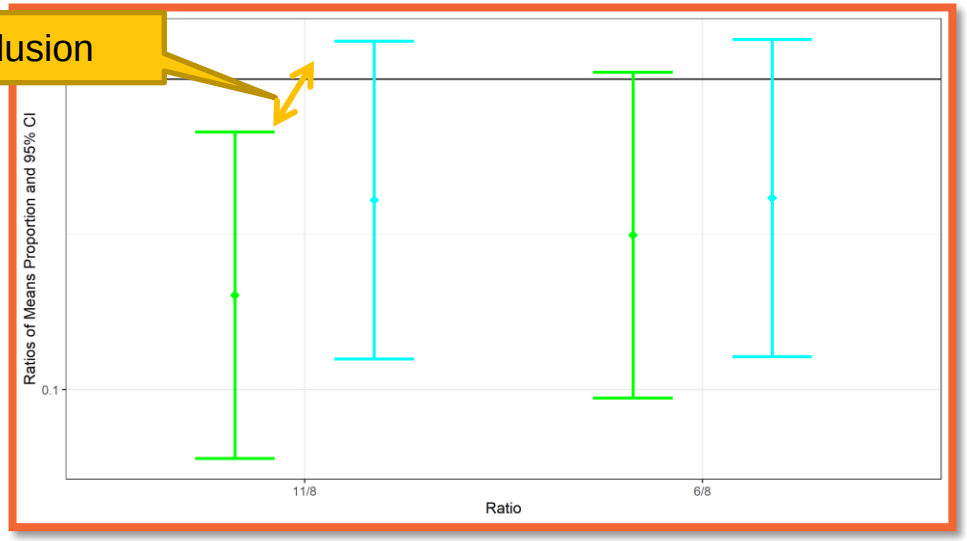


Model	N
Beta Regression (0 not included)	N=1
Beta Regression (Shifted by half lowest positive)	N=2

Beta regression model mean proportions



Beta regression model mean ratios



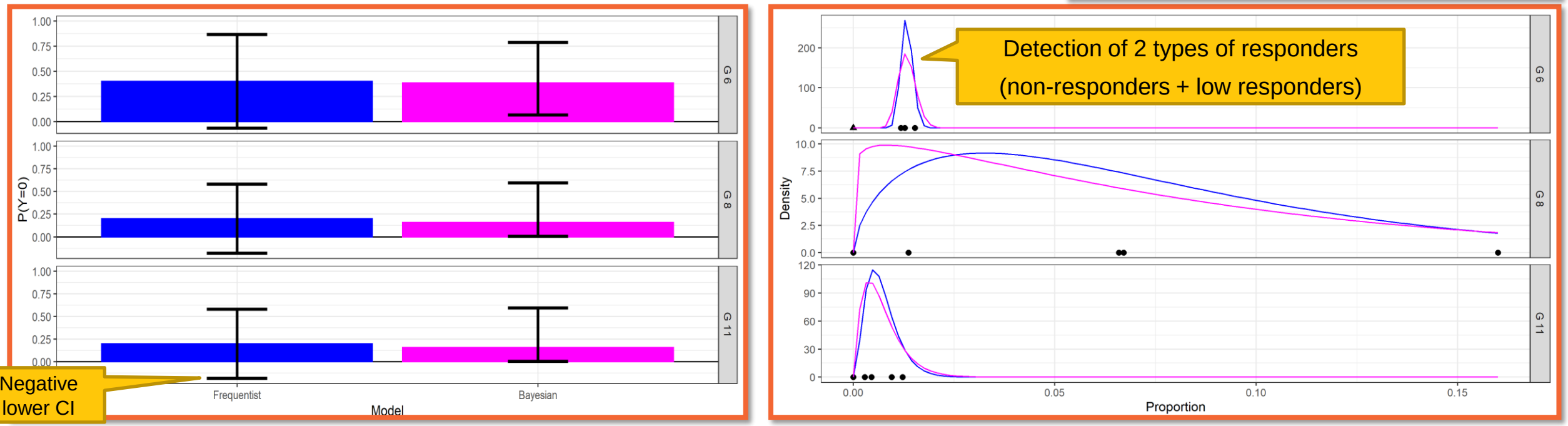
GSK

Zero-inflated Beta regression model

- A probability mass in zero, $0 \leq \alpha \leq 1$, is added to the location-scale model (μ, ϕ) to get a mixed continuous discrete distribution model[1] with $f(y;\alpha,\mu,\phi) = \begin{cases} \alpha & \text{if } y = 0 \\ (1-\alpha) f(y;\mu,\phi) & \text{if } y \in (0,1) \end{cases}$ with $f(y;\mu,\phi)$ = beta density
- All parameters are linked to (different) covariates: $h_0(\alpha) = f_0(W,\rho)$; $h_1(\mu) = f_1(X,\beta)$; $h_2(\phi) = f_1(Z,\gamma)$
- Existing Bayesian R package *zoib*

Zero-Inflated Beta regression model distribution fitting (on median of posterior)

Probability mass in zero **Continuous part**



1. Ospina R. and Ferrari SLP. *A general class of zero-or-one inflated beta regression models*. Computational Statistics & Data Analysis. 2012 June; 56(6): 1609-1623

Zero-inflated Beta regression model

- Address log-normal and non-inflated beta regression models limitations
- Bayesian framework allows fit for purpose outcomes
- Discontinuous Highest Posterior Density (HPD) intervals might be relevant

Model

Frequentist Inflated Beta Regression (NLMixed)

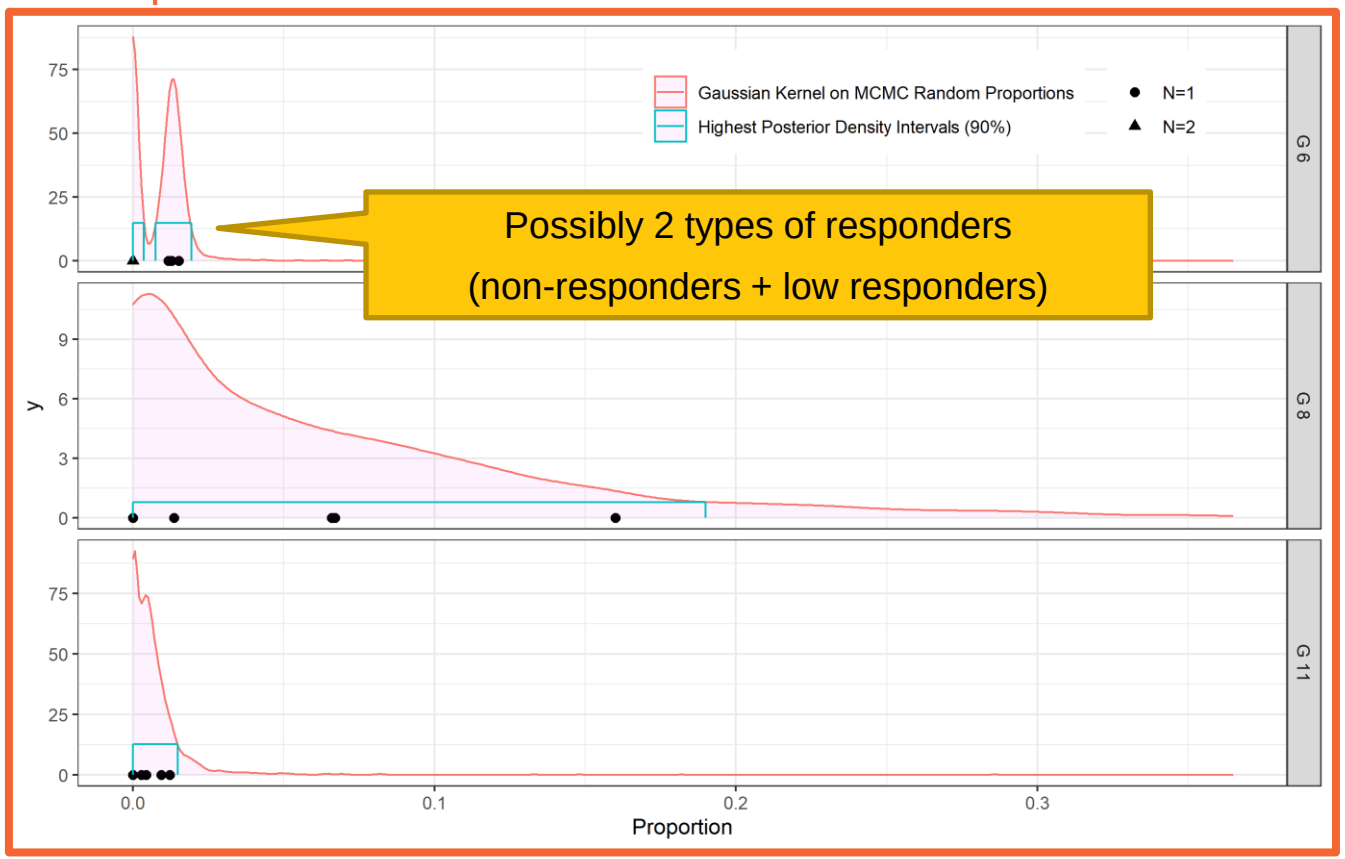
Bayesian Inflated Beta Regression (Pct CI)

N

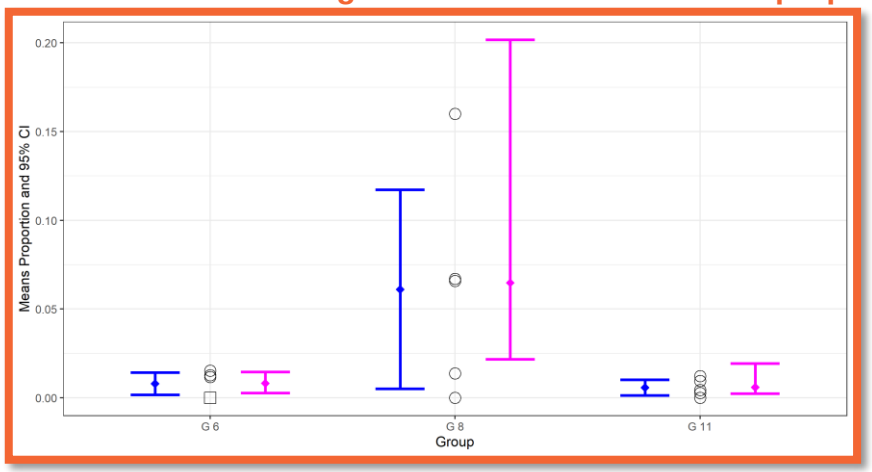
N=1

N=2

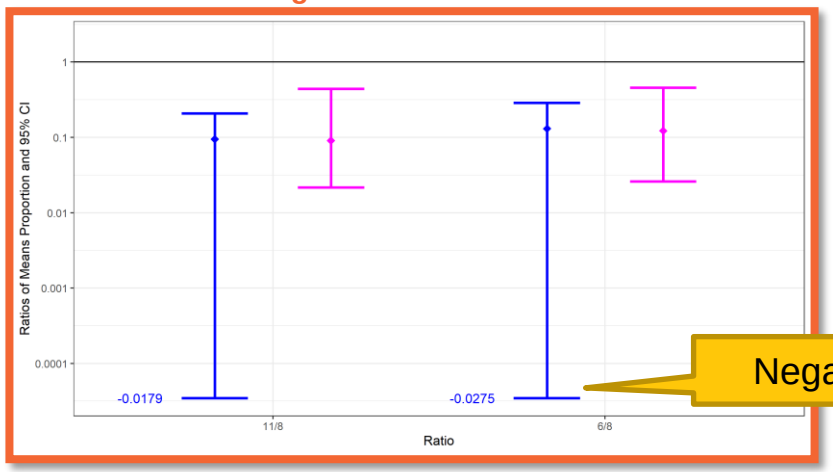
Posterior predictive distribution



Zero-inflated Beta regression model based mean proportions

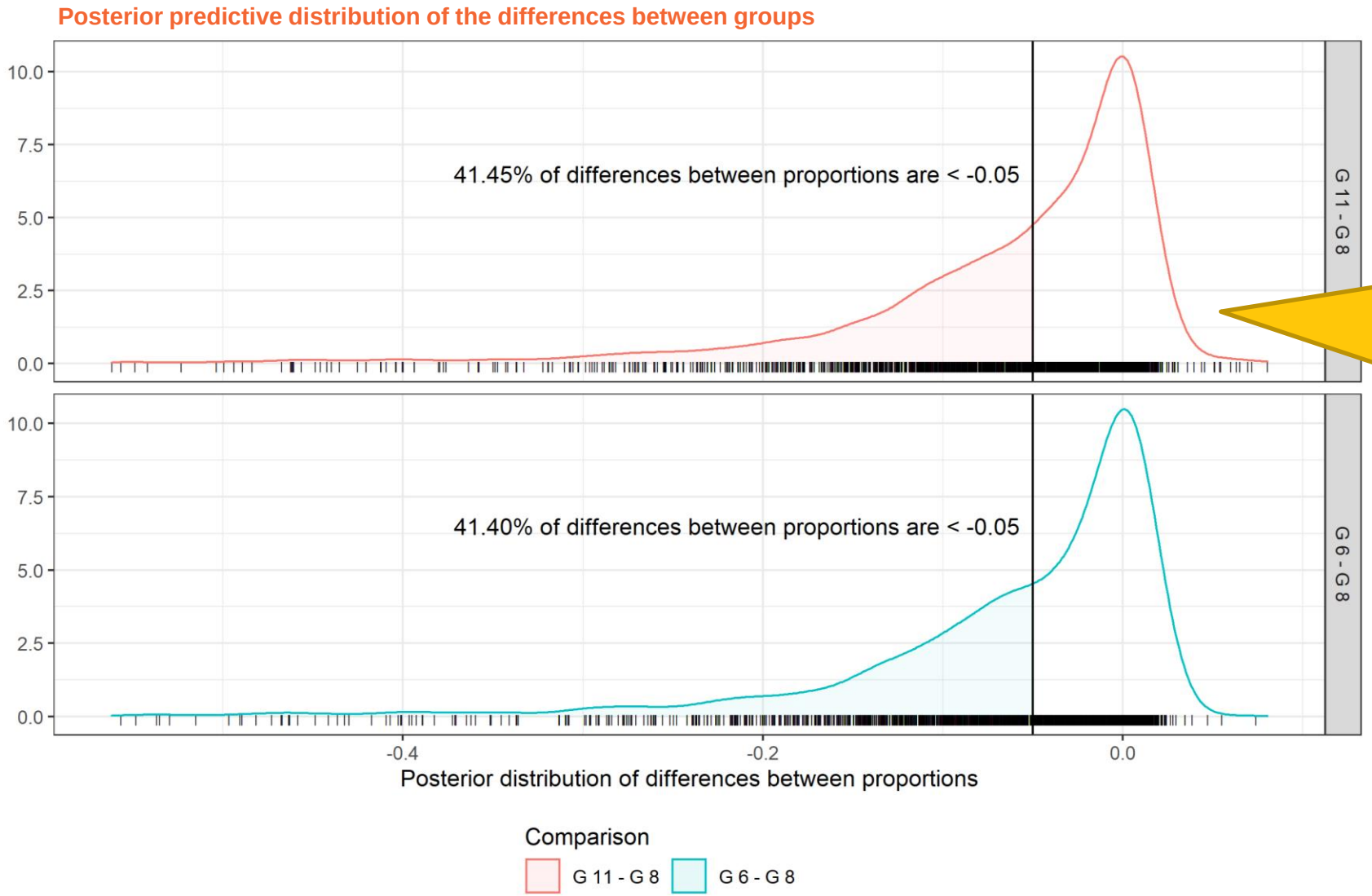


Zero-inflated Beta regression model based means ratio



Zero-inflated Beta regression model

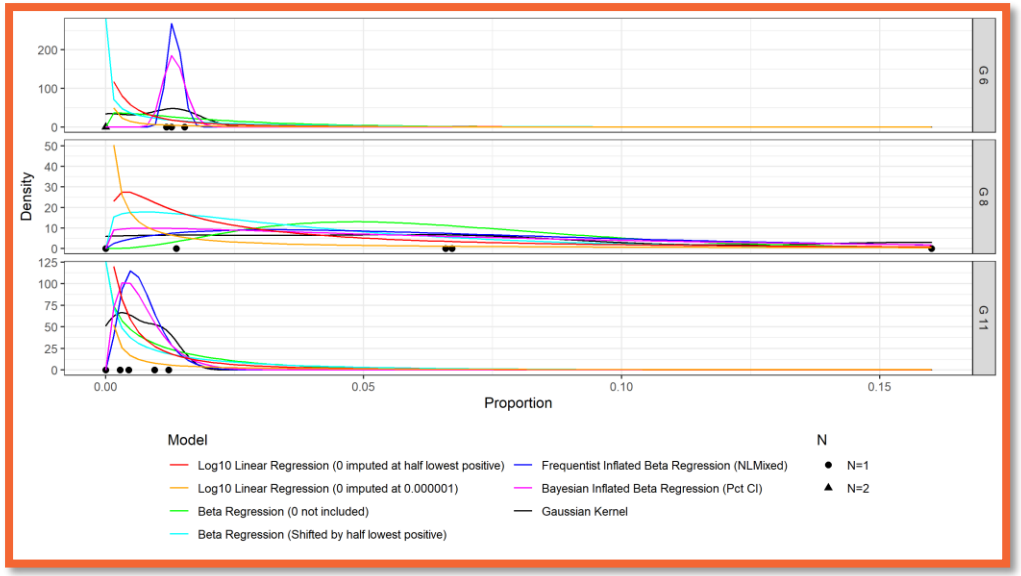
- Posterior predictive distribution allows to produce more fit-for-purpose outcomes than just hypothesis testing

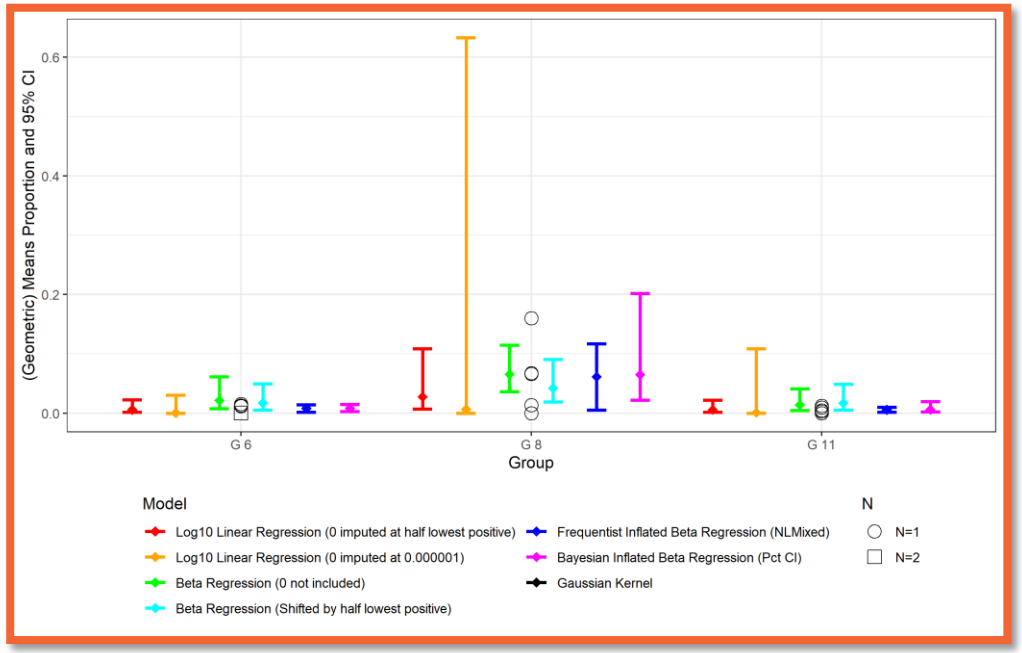


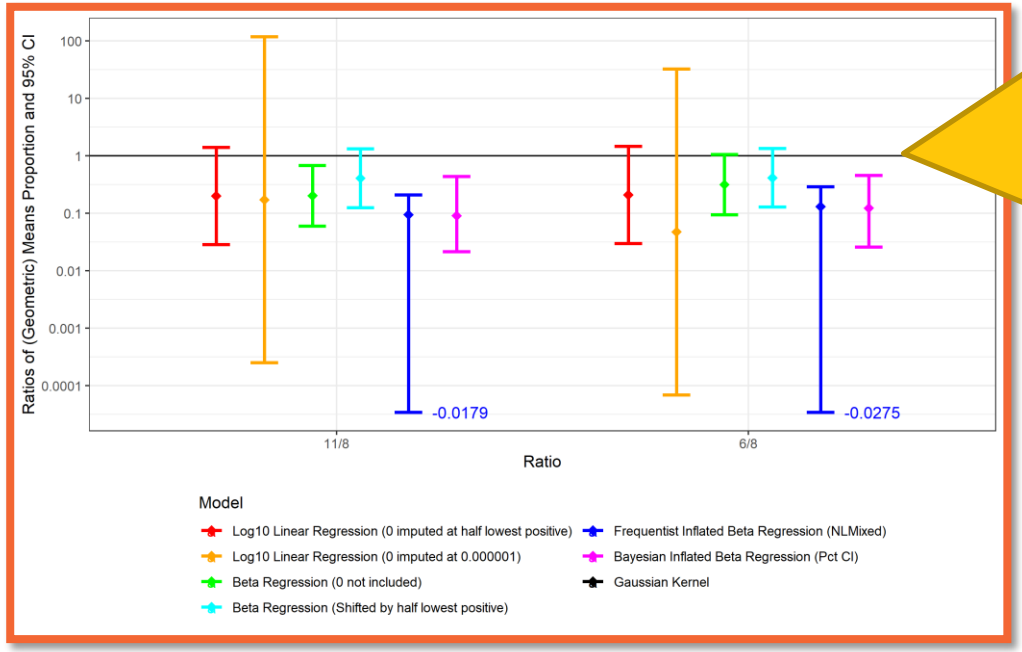
Despite the significant difference observed on means ratios, about 59% of the observations may NOT differ more than 0.05 (i.e., 5%) between group 8 and group 11

Model comparison recap

Posterior predictive distributions







This case study shows a change of conclusion between the classical log-linear model implementation (red) and the Bayesian zero-inflated beta regression model (pink)

Key questions and some answers

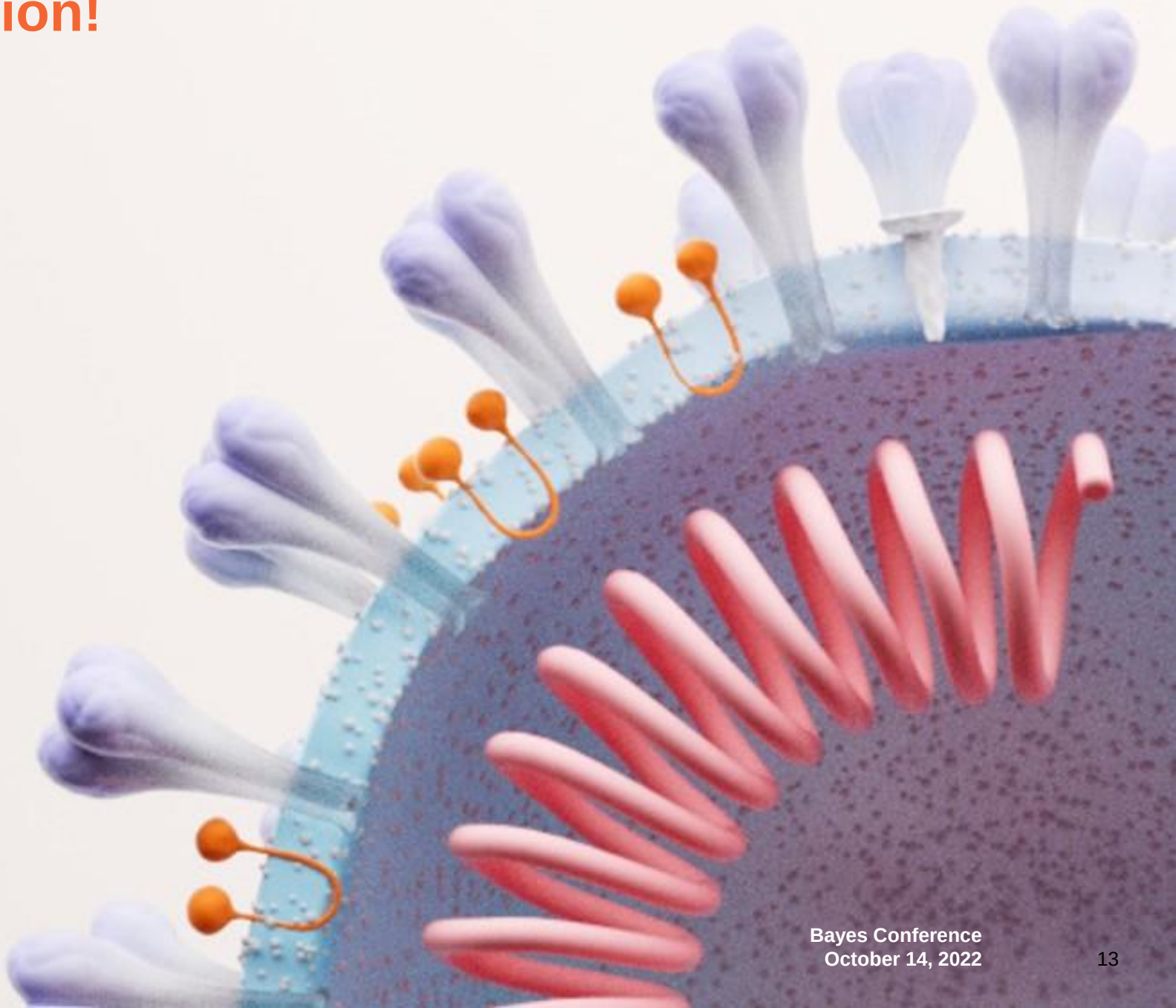
1. Is the classical log-linear method fit for purpose?
   - Log-linear models allow to address cellular data similarly to other endpoints
   - But might be heavily influenced by the imputation value (change of conclusion)
   - Might suffer of some lack of fit due to the unmet nature of the data
2. Does the Zero-inflated beta regression model address the key issues?
   - Zero-inflated beta regression models address the presence of zero while considering the correct support of proportion data
   - Bayesian approach improves data understanding despites limited sample sizes
3. Does the method significantly change the outcome?
   - Some conclusions are changed in the case study, but not systematic in applications
   - More comprehensive and fit-for-purpose outcomes with Bayesian statistics
4. Do we benefit from changing the approach?
   - Benefit/risk for both stats and lab scientists should be cautiously discussed
   - Bayesian framework brings lot of added values, far more than just including a prior
5. Next steps?
   - Study the impact of “non-informative” priors on results based on small sample sizes

GSK

Thank you for your attention!

Contact:

- Cédric Taverne
- Vaccine Research Statistics, GSK
- cedric.x.taverne@gsk.com



gsk.com