Analyzing the Economic Impact of Gender Equality Across Countries

— Project Report —
Advanced Bayesian Data Analysis

236442 Rachel John Christopher
257929 Sixuan Wei

Group 16

March 14, 2025

TU Dortmund University

Contents

Abstract

This study investigates the relationship between gender equality and economic growth across countries using Bayesian data analytics. The primary motivation is to identify the variables that significantly influence a country's GDP, with a specific focus on gender equality. Based on the data structure and research objectives, multilevel Bayesian models were employed, including Random Intercept and Random Slope models to explore different grouping factors and to capture variations across income groups and world regions. This study aims to provide an understanding of the complexity of gender equality factors' economic implications and the importance of considering heterogeneous effects in cross-country analyses.

The results suggest that gender equality indicators, such as female labor force participation and secondary education enrollment, are positively associated with GDP per capita, though the strength of the association varies by income level. Meanwhile, the WBL Index does not show a significant effect, and the relationship between female labor participation and economic growth differs across income groups. These findings highlight the complexity of gender equality factors' economic implications and the importance of considering heterogeneous effects in cross-country analyses.

1 Introduction

1.1 Motivation

Economic growth is influenced by multiple factors, and gender equality has emerged as an important factor correlated with economic outcomes. Identifying the significant predictors of GDP growth is crucial for shaping effective policies. This study aims to assess the role of gender equality using Bayesian methods, leveraging insights from Kabeer and Natali (2013), who emphasize that gender parity in education, labor force participation, and wages contributes to economic development (1).

However, the effect of gender equality on GDP varies across regions and income levels. Diachkova and Kontoboitseva (2022) found a positive impact in both the EU and BRICS countries (2), while Agyina and Osei-Fosu (2020) showed a negative effect in West Africa (3). These differences highlight the need for region or income level-specific policies. To explore these differences, we use multilevel Bayesian models, which allow for a deeper examination of how these effects vary across income groups and global regions, while accounting for country-level heterogeneity.

1.2 Objective

- To identify the variables that are associated with a country's GDP, with a focus on gender equality.

- To use multilevel Bayesian models and compare their suitability for the data using model evaluation techniques.
- To examine how the association of gender equality variables and the GDP per capita of a country varies across different regions and income groups.

2 Data Description

The dataset used in this study is sourced from the World Bank Dataset (4) covering 32 years (1991–2022) with a total of 5,888 records from 184 countries. We adopt the World Bank's income classification (low, lower-middle, upper-middle, and high income) and also group countries by 7 regions (e.g., East Asia & Pacific, Sub-Saharan Africa) to facilitate our multilevel analysis. The variables were selected and collected from within World Bank Dataset, and were merged to create a comprehensive dataset for an in-depth analysis.

2.1 Variable explanation

Dependent Variable

- **GDP per capita (PPP, constant 2021 $)**: Measures economic output per person, adjusted for purchasing power parity.

Independent Variables

- **WBL Index(WBL)**: Captures legal and policy support for women (0-100 scale).
- **Female Labor Force Participation (Lf)**: Represents women's workforce involvement, which is correlated with higher economic productivity.
- **Female Secondary School Enrollment (Es, f)**: Indicates access to education, which often enhances job prospects.
- **Female Waged and Salaried Workers (Wf)**: Reflects formal employment, associated with economic stability and growth.

These variables provide insights into gender equality's economic impact across different regions and income levels.

2.2 Data cleaning

2.2.1 Handling Missing Values

Most variables, except the WBL index, had missing values. Education data had the highest percentage of missing values (up to 64%). Countries with more than 50% missing GDP, Labor Force Participation, or Waged female data or more than 70% missing education data were excluded, reducing the dataset from 169 to 127 countries.

To handle the remaining missing values, we applied interpolation methods:

- NOCB (Next Observation Carried Backward) for missing values at the end of the time series.
- LOCF (Last Observation Carried Forward) for missing values at the beginning.
- Spline Interpolation for missing values in between.

We avoided mean imputation because it can distort covariance structures and lead to inaccurate model estimations (5). Although multiple imputation or full Bayesian data augmentation may offer more robust approaches, we opted for interpolation methods to preserve temporal trends under the time constraints of this applied project.

2.2.2 Addressing Skewed Data

Since GDP values are typically large and its distribution is highly skewed, directly using raw GDP may lead to difficulties in sampling convergence during Bayesian inference. Hence, a log transformation was applied to reduce skewness while preserving relative differences. As shown in Figure 1, this transformation effectively reduced the right-skewness, resulting in a distribution closer to normal. Consequently, we assume that Log(GDP) follows a normal distribution given the predictors and specify the Gaussian family for the likelihood in our models.

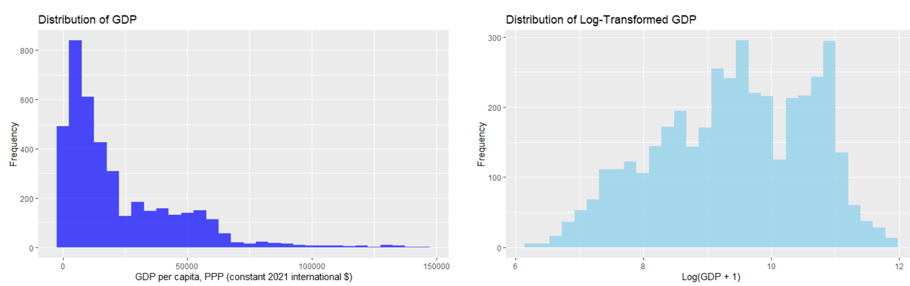

Figure 1: Histogram of GDP (Original vs. Log Transformed)

3 Modeling Phase

To appropriately capture the hierarchical nature of our data and research question, we structure our models using a Bayesian multilevel framework. The levels are defined as follows:

- **Level 1:** Panel data (country-year observations).
- **Level 2:** Country-level effects, accounting for national differences.

- **Level 3:** Regional or income group variations.

We begin with the simplest linear regression model and progressively introduce hierarchical structures to better account for country-level heterogeneity and group-specific variations.

We implement our models using the `brms` package in `R`, with CmdStanR as the backend. Bayesian inference is performed using Markov Chain Monte Carlo (MCMC) sampling.

3.1 Simple Linear Regression Model

We first estimate a Bayesian linear regression model to establish a baseline relationship between gender equality variables and GDP:

$$log(GDP_{it}) = \beta_0 + \beta_1 WBL_{it} + \beta_2 Lf_{it} + \beta_3 Es, f_{it} + \beta_4 Wf_{it} + \beta_5 Year_t + \epsilon_{it}$$

where i denotes country, t denotes year, and the predictors are defined as in Section 2.1. For simplicity, in the following sections, we denote $\beta_1 WBL_{it} + \beta_2 Lf_{it} + \beta_3 Es, f_{it} + \beta_4 Wf_{it} + \beta_5 Year_t$ as $\beta_1 X_{it}$.

To ensure a weakly informative prior structure, we specify the following priors:

- **Intercept:** $\beta_0 \sim \mathcal{N}(9.4,\ 0.5)$, centered on the sample's mean log(GDP), allowing moderate deviation.
- **Slopes:** $\beta \sim \mathcal{N}(0, 1)$, ensuring that the regression coefficients are centered around zero with flexibility.
- **Residual Std. Dev. (σ):** $\sigma \sim \mathcal{N}(0,\ 2.5)$ $[\sigma > 0]$, which in practice is interpreted as a broad half-normal shape. We set a broader prior for the model to capture noise or unexplained variance without forcing overly tight predictions, thus reducing the risk of overfitting to random fluctuations in the data.

We included $Year_t$ as a predictor to capture the general upward trend in GDP over time. While Year is not directly related to gender equality, incorporating it prevents us from mistakenly attributing overall economic growth to our core independent variables. However, the model ignores country-level heterogeneity, potentially overstating confidence in estimates. This limitation motivates the multilevel models discussed next.

3.2 Random Intercept Model

To address unobserved country-level heterogeneity, we extend the simple linear model by introducing country-specific intercepts:

$$log(\text{GDP}_{it}) = \beta_0 + \alpha_i + \beta_1 X_{it} + \epsilon_{it}, \quad \alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$$

where α_i represents country-specific intercepts.

We maintain the priors described in Section 3.1 for the fixed effects, intercept, and residual standard deviation. Additionally, we introduce a prior for the **Random Intercept Std. Dev.(α_i) for Country:**

$$\mathrm{sd}(\alpha_i) \sim \mathcal{N}(0,\, 0.5)\ [\mathrm{sd}(\alpha_i) > 0]$$

The model assumes uniform predictor effects across countries, which may oversimplify reality. For instance, the effect of female labor participation could differ between high- and low-income groups. This limitation motivates our next step: allowing predictor effects to vary by region or income group (Section 3.3 and 3.4).

3.3 Fixed Interaction Model

To examine how the effects of gender equality variables differ across income groups, we extended the random intercept model by introducing fixed interactions between predictors and income groups. The model specification is:

$$log(\mathrm{GDP}_{it}) = \beta_0 + \alpha_i + \sum_{g=1}^{4} \gamma_g(\text{Income Group}_g \times X_{it}) + \epsilon_{it}, \quad \alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$$

where γ_g represents the slope of predictors for income group g (High Income, Upper Middle Income, Lower Middle Income, Low Income). With other priors remaining the same, the prior of **Fixed Slopes for Income Groups** is also set to be the same with slopes in section 3.1 and 3.2:

$$\gamma_g \sim \mathcal{N}(0,\, 1)$$

The fixed interaction approach is theoretically appropriate since income groups are discrete and mutually exclusive categories rather than samples from a larger population(6).

3.4 Random Slopes Model

Another way to allow predictor effects to vary by grouping factors is by random slopes model:

$$log(\mathrm{GDP}_{it}) = \beta_0 + \alpha_i + (\beta_1 + \eta_g)X_{it} + \epsilon_{it}, \quad \alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)\ ,\ \eta_g \sim \mathcal{N}(0, \sigma_\eta^2)$$

where η_g represents group-specific deviations from the global slope β_1, allowing for heterogeneous predictor effects. Here we set the prior of **Random Slopes Std. Dev.:**

$$\mathrm{sd}(\eta_g) \sim \text{Student-}t(3,\, 0,\, 0.5)\ [\mathrm{sd}(\eta_g) > 0],$$

allowing heavier tails for group-level slope variations (e.g., Region or Income Group), in case certain groups deviate strongly from the global slope.

Here we tested two grouping strategies:

- Separate Grouping: introducing random slopes separately for Regions (7 categories) and Income Groups (4 categories).
- Combined Grouping: combining Region and Income Group into a single factor, resulting in 17 distinct categories.

With only 4 income groups or 7 regions, random slope estimation can be unstable if group sample size is small(7). Combining them yields 17 categories, which may provide more variation but can still pose challenges if certain group intersections have very few countries. We compare these two grouping approaches in following chapter.

4 Model Comparison

4.1 Convergence Analysis

For all models, we ran 4 parallel chains with 2000 iterations each, including 1000 warm-up iterations, resulting in a total of 4000 post-warmup samples. Convergence was assessed via $\hat{R} \approx 1)$, effective sample size and trace plots, indicating stable mixing, but the **separate grouping** method in Model 3.4 showed much slower convergence. This may be due to the small number of groups (only 4 or 7) and high complexity of model, which made variance estimation unstable.

In the **separate grouping approach**, all coefficient credible intervals (CIs) were very wide and often included zero. This suggests that the model could not reliably estimate group effects. Studies (7) & (8) have shown that small group numbers lead to high between-group variance estimates, which increases uncertainty.

In contrast, the **combined grouping method** (17 groups) improved convergence and reduced variance, though some small groups (with only 1 or 2 countries) still had wide CIs.

	Separate Grouping	**Combined Grouping**
Posterior Mean (β_3)	0.18	0.17
95% Credible Interval	[-0.09, 0.43]	[0.09, 0.23]
Between-Group Variance (τ^2)	0.856	0.588

Table 1: Posterior mean and 95% credible intervals for selected coefficients in random slopes models (separate vs. combined grouping).

4.2 Posterior Predictive Checks (PPC)

Posterior predictive checks compare observed outcomes with draws from the posterior predictive distribution, indicating whether a model reproduces key features of the data. To evaluate model fit, we checked:

- **Residual Plots**: Model 3.1 showed a clear linear pattern in residuals vs observed value, meaning it failed to capture country-level heterogeneity. Models 3.2-3.4 had similar residual distributions, without strong patterns.
- **Posterior Predictive Distributions**: All models produced similar predictive distributions.

Since Models 3.2-3.4 had similar PPC results, the main difference comes from LOO analysis.

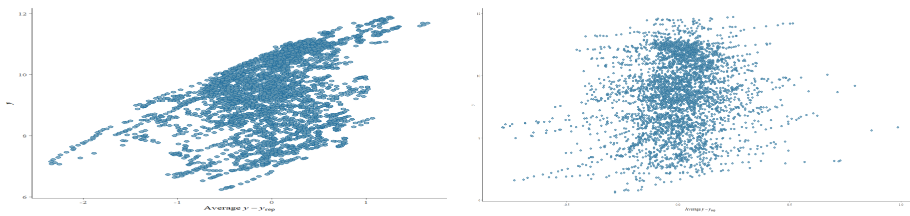

Figure 2: Residual plots for the simple linear model (3.1) and random intercept model (3.2), highlighting the linear structure in 3.1

4.3 Model Fit via LOO-CV

We compare models using **Leave-One-Out Cross-Validation (LOO-CV)** to evaluate predictive performance. Table 2 reports the estimated expected log pointwise predictive density (**elpd_diff**) and its standard error (**se_diff**) for each model. The best model (highest elpd) is used as the reference, with a difference of 0.0.

Model	**elpd_diff**	**se_diff**
3.4 Random Slopes Model (Combined Grouping)	0.0	0.0
3.4 Random Slopes Model (Separate Grouping)	-228.1	31.9
3.4 Random Slopes Model (Income Group)	-600.3	47.1
3.3 Fixed Interaction Model (Income Group)	-677.4	51.3
3.2 Random Intercept Model	-838.7	61.0
3.1 Simple Linear Model	-5,314.3	97.2

Table 2: Leave-One-Out Information Criterion (LOO-IC) for each model. Higher values indicate better predictive performance.

Key findings:

- The **3.4 Random Slopes Model (Combined Grouping)** achieved the best predictive performance.
- The **Separate Grouping** approach in Model 3.4 performed worse (elpd_diff $= -231.0$), likely due to the small number of groups and the high complexity leading to unstable variance estimates.

- The **Income Group-only model** (3.4) and **Fixed Interaction Model** (3.3) performed similarly, suggest that random effect doesn't perform better when the number of group is small(7).
- The **Random Intercept Model** (3.2) was significantly worse than models incorporating predictor interactions or random slopes.
- The **Simple Linear Model (3.1)** had the poorest performance, with an elpd difference of $-5,314.3$, confirming that ignoring country-level heterogeneity leads to poor predictions.

These results suggest that allowing predictor effects to vary by **combined grouping** (Model 3.4) results in the best trade-off between flexibility and predictive performance. However, models with fewer groups such as **separate grouping** (Model 3.4) suffer from high variance estimates due to limited group counts. The detailed implications of these results will be discussed in later sections.

5 Prior Sensitivity Analysis

To evaluate prior sensitivity, besides the prior specification in Chapter 3, we set two alternative prior strengths:

Parameter	**Default Priors**	**Weak Priors**	**Strong Priors**
Slopes (β)	$\mathcal{N}(0,1)$	$\mathcal{N}(0,10)$	$\mathcal{N}(0,0.5)$
Intercept (β_0)	$\mathcal{N}(9.4,0.5)$	$\mathcal{N}(9.4,2)$	$\mathcal{N}(9.4,0.2)$
Residual Std. Dev. (σ)	$\mathcal{N}(0,2.5)$	$\mathcal{N}(0,2.5)$	$\mathcal{N}(0,1)$
Random Intercept Std. Dev. (α_i)	$\mathcal{N}(0,0.5)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,0.3)$
Random Slope Std. Dev. (τ)	Student-t$(3,0,0.5)$	Student-t$(3,0,1)$	Student-t$(3,0,0.3)$

Table 3: Comparison of Default, Weak, and Strong Prior Distributions

We label them 'weak' and 'strong' based on the variance of their distributions, with a larger variance (e.g. $\beta \sim \mathcal{N}(0,10)$, imposing minimal prior constraint on slopes, and a smaller variance (e.g. $\beta \sim \mathcal{N}(0,0.5)$) providing stronger regularization.

To assess prior sensitivity, we compared model performance using **LOO-CV**. Table 4 summarizes the results.

Model	**elpd_diff**	**se_diff**
Weak Priors	0.0	0.0
Default Priors	-1.4	0.5
Strong Priors	-1.6	0.6

Table 4: LOO Information Criterion for different prior strengths.

The results show minimal variation in LOO scores across different prior choices. The largest difference observed is only -1.6 in expected log pointwise predictive density (elpd), suggesting that posterior inference remains stable regardless of prior choice.

Moreover, to visualize how the priors influence posterior distributions, we generated an **MCMC density overlay** for a selected coefficient (`b_WBL.Index`). Figure 3 shows the density of the prior samples (chain 1) compared to the posterior samples (chain 2).

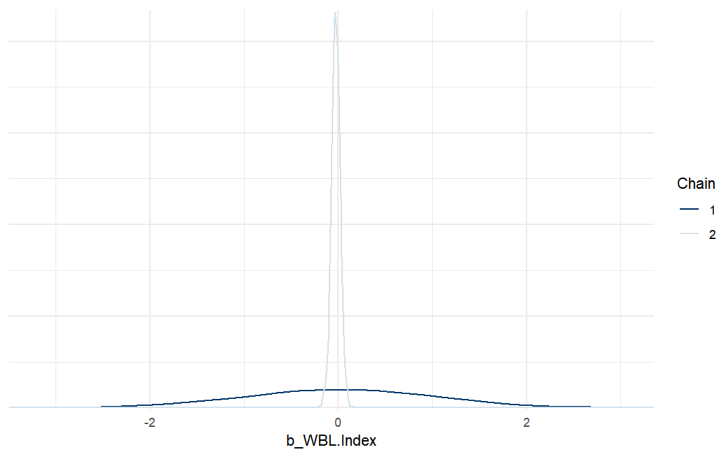

Figure 3: MCMC density overlay for the WBL Index coefficient

The prior distribution (chain 1) is wide and flat, indicating weak informativeness, while the posterior distribution (chain 2) is narrower and more peaked, showing that the data dominates inference.

6 Results & Discussion

In this chapter, we discuss the result of the Random Slope Model chosen in Chapter 4. While our Bayesian multilevel approach highlights significant associations, we emphasize that these findings do not establish causality. Here we discuss only apparent patterns but not specific values to avoid over explaining.

We present the results using **point-range plots** of the posterior distributions. These plots show the estimated slopes for each of the 17 groups, where dots indicate the posterior mean estimates and horizontal lines represent 95% credible intervals.

To ensure interpretability, we extracted the coefficient `coef()` instead of `ranef()` while generating the plot, so the estimates are centered around the **fixed effects** rather than zero.

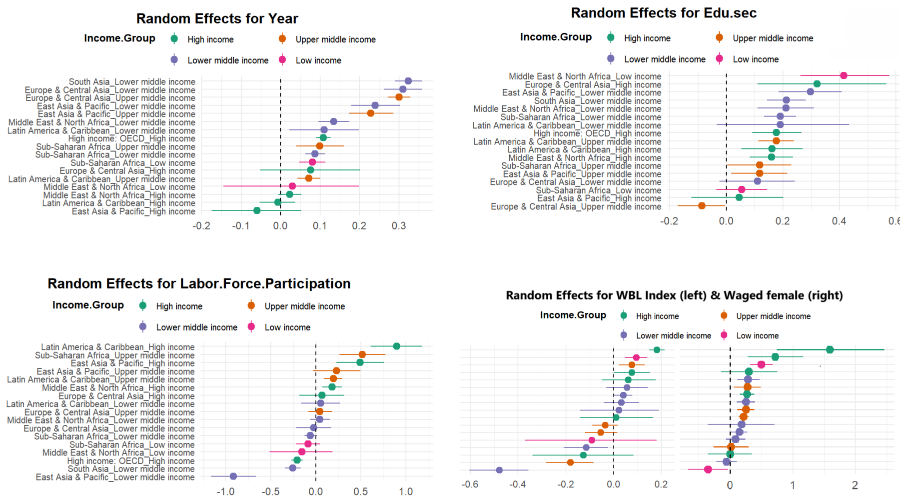

Figure 4: Point-range plots of different predictors' associations with GDP.

The findings suggest distinct patterns across different economic contexts:

- **Time Trend**: Most groups exhibit a positive time trend, particularly in middle-income groups (purple and orange), suggesting worldwide economic growth in 1991-2022.

- **Secondary Education Enrollment**: A generally positive and significant association with GDP, especially in lower-income groups (purple), suggesting that improving education access may be particularly relevant for economic growth in these economies.

- **Female Labor Force Participation**: Higher-income groups (green and orange) tend to show a positive association, while lower-income groups exhibit a mixed or weak relationship, indicating that structural labor market differences may play a role.

- **WBL Index and Waged Female Percentage**: The WBL Index does not show a clear pattern across different groups, and its relationship with GDP is often weak or insignificant. However, the percentage of waged female workers is positively associated with GDP, suggesting that wage employment opportunities for women may be more strongly linked to economic outcomes than legal equality measures alone.

It was noticed that the groups with only one or two countries have much wider credible intervals. For instance, the Middle East & North Africa – Low Income group, which includes Syria, shows high uncertainty due to the country's economic fluctuations.

These results suggest that gender equality-related factors are associated with economic growth in different ways depending on income levels and structural conditions. Further policy analysis should consider both legal and labor market aspects to better understand these relationships and ensure inclusive economic development.

7 Summary and Limitations

This study explored how four gender equality-related factors are associated with GDP using Bayesian multilevel models that account for regional and income-level differences. The results show that education and the percentage of waged women workers are generally positively associated with GDP. However, the WBL Index showed mostly weak or insignificant associations, while female labor force participation had mixed relationships—being positively associated with GDP in high-income groups but sometimes negatively associated in low-income groups.

Despite these insights, the analysis had some limitations. Data gaps restricted the scope due to the lack of complete datasets on key variables such as education, labor force participation, and waged women workers. The missing data may introduce bias or reduce variance unpredictably. Furthermore, the linear nature of the model might oversimplify the complex relationship between gender equality and economic growth, potentially missing bidirectional influences. Lastly, lagged effects and serial correlation were not considered, meaning that changes in indicators such as the WBL Index or education could take years to fully relate to GDP. These constraints highlight the need for cautious interpretation of our findings and suggest directions for future research.

Overall, these findings highlight the multifaceted ways gender equality and economic development align, while underscoring the importance of accounting for heterogeneity and data limitations in cross-country analyses.

References

[1] Naila Kabeer and Luisa Natali. *Gender Equality and Economic Growth: Is there a Win-Win?*. IDS Working Paper 417, 2013, Volume 2013, Pages 1-58.

[2] Diachkova, A., Kontoboitseva, A. (2022). Economic Benefits of gender equality: comparing EU and BRICS countries. Economic consultant, 37 (5), 4-15. doi: 10.46224/ecoc.2022.1.1

[3] Agyina, Joseph, and Anthony KofiOsei-Fosu. "The Effect of Gender Inequality on Economic Growth in West Africa." (2020).

[4] World Bank Dataset, `https://data.worldbank.org/`

[5] Heikki Junninen, Harri Niska, Kari Tuppurainen, Juhani Ruuskanen, Mikko Kolehmainen. *Methods for imputation of missing values in air quality data sets*. Elsevier, 2004, Volume 38, Issue 18, Pages 2895-2907.

[6] Gelman, A., Hill, J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, 2006, Page 259-263

[7] Gomes, Dylan G.E. Including random effects in statistical models in ecology: fewer than five levels? 2021

[8] Snijders, Tom AB and Bosker, Roel, Multilevel analysis: An introduction to basic and advanced multilevel modeling, 2011, Chapter 10, Page 140, 144, 150

[9] OECD: Jonas Fluchtmann, Willem Adema and Mark Keese. Gender Equality and Economic Growth, 2023.

[10] Eric A. Hanushek and Ludger Woessmann. *The Knowledge Capital of Nations: Education and the Economics of Growth*. MIT Press, 2015. ISBN: 9780262029179. Address:Cambridge, MA.

[11] Gerry Gatawa. The Effect of Social Factors to Economic Growth. Advance, 2022. `https://doi.org/10.1101/2021.04.11.439357`