Modelling the viral dynamics of SARS-CoV-2 in the general community in a context of emerging variants

Maxime Beaulieu[1]

Florence Débarre[2], François Blanquart[3], Jérémie Guedj[1]

[1]*Université Paris Cité and Université Paris Sorbonne Paris Nord, Inserm, IAME, F-75018 Paris, France*
[2]*Institute of Ecology and Environmental Sciences Paris (IEES Paris), Sorbonne Université, CNRS, IRD, INRAE, 75005 Paris, France*
[3] *Centre interdisciplinaire de recherche en biologie, Collège de France, 75231 Paris, France*

4 July 2025

Université Paris Cité

iame RESEARCH CENTER ON INFECTIOUS DISEASES

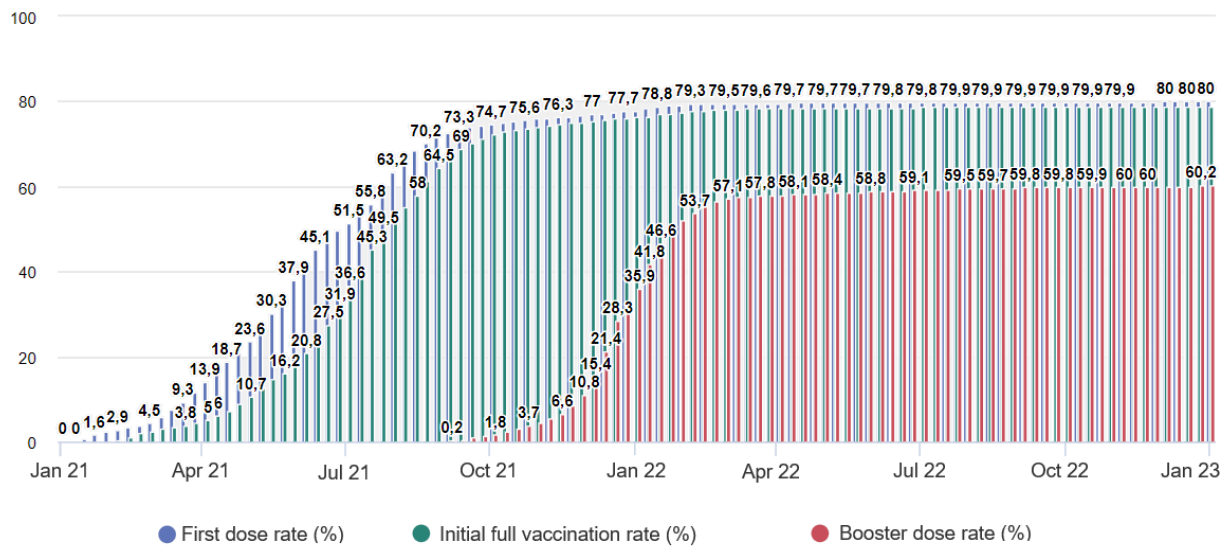
Inserm La science pour la santé From science to health

BIOGROUP biologie médicale

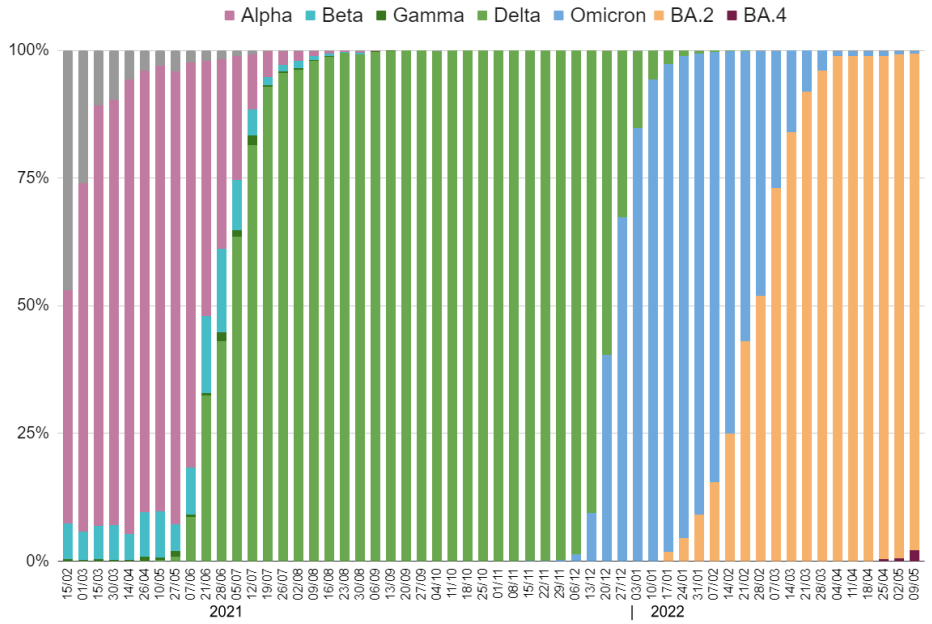
The COVID-19 pandemic – a changing landscape

From early 2021, the epidemic as been affected by:

- a strong **vaccination** campaign,
- and the **emergence of variants of concerns** (VoCs)



Vaccination rate in France since 2021 [1]



Successive waves of VoCs in France since 2021 [2]

[1] datavaccin-covid.ameli.fr (2023)

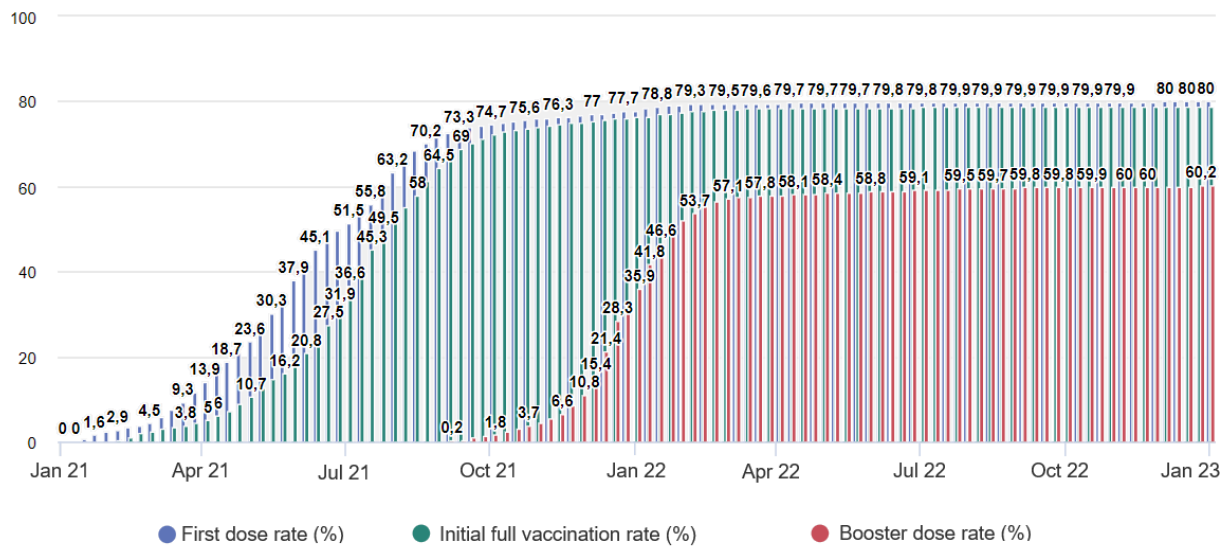
[2] N. Berrod, data from Santé publique France (2022)

The COVID-19 pandemic – a changing landscape

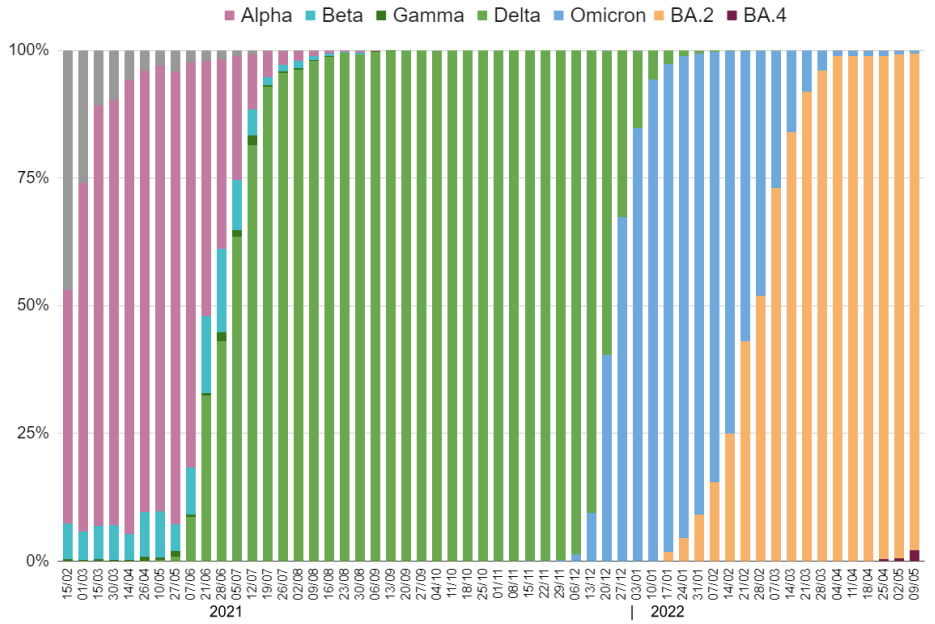
From early 2021, the epidemic as been affected by:

- a strong **vaccination** campaign,
- and the **emergence of variants of concerns** (VoCs)

→ Modified the **severity** and **transmission** of the virus



Vaccination rate in France since 2021 [1]

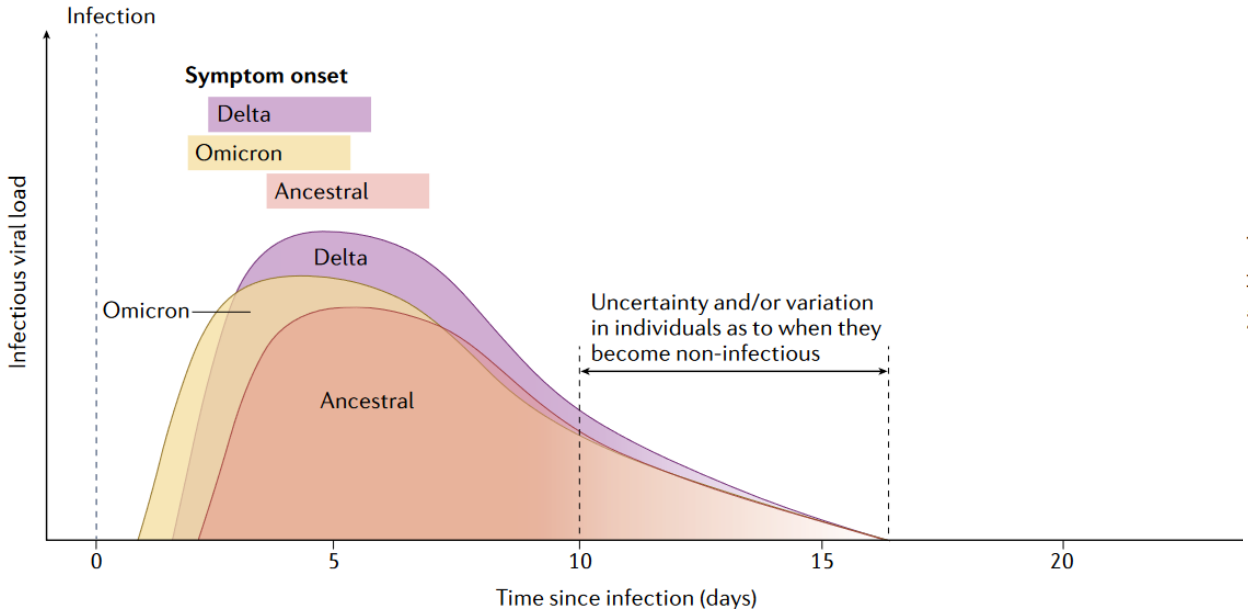


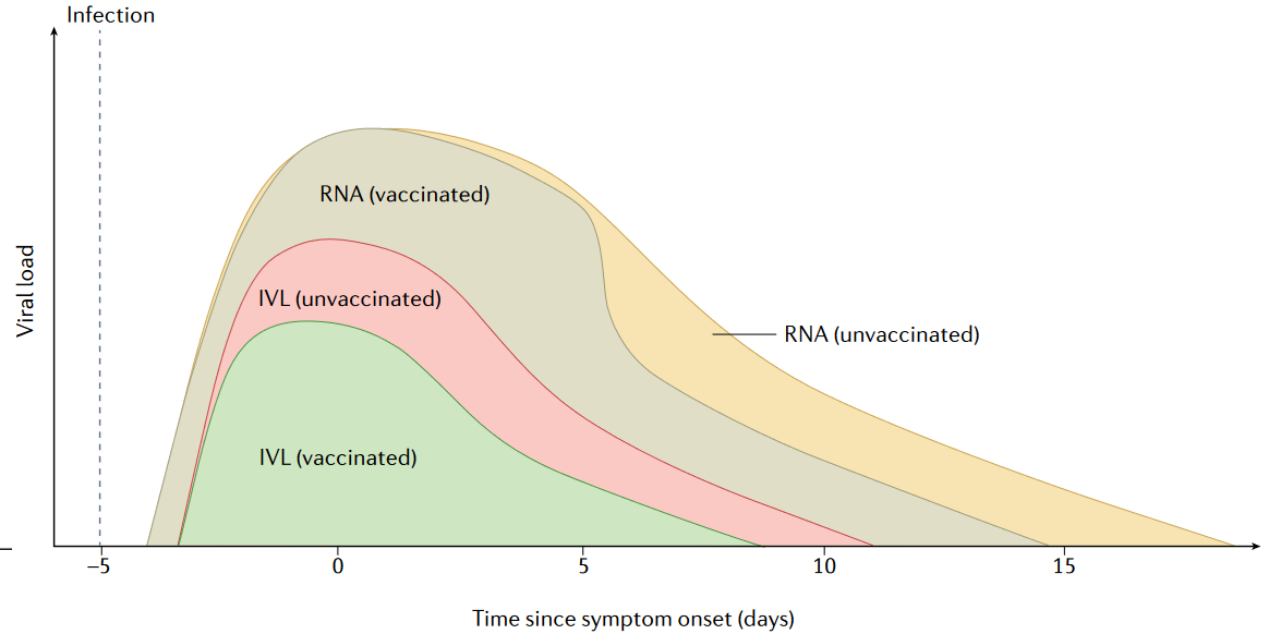
Successive waves of VoCs in France since 2021 [2]

[1] datavaccin-covid.ameli.fr (2023) [2] N. Berrod, data from Santé publique France (2022)

Study the impact of this changing landscape on viral dynamics in the population

- **Variant of infection** and **patient characteristics** may also **shape the viral dynamics** [3,4]
  - Studies often conducted on small specific cohorts (symptomatic, commorbidities…) → Potential selection bias



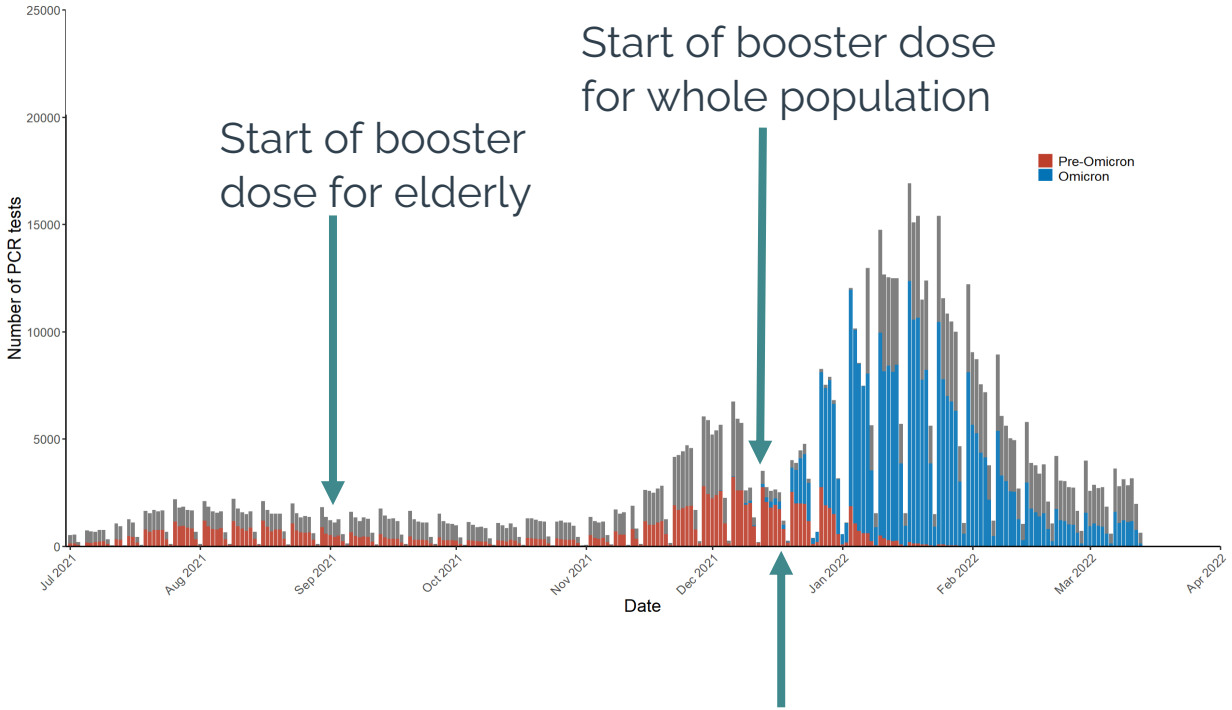


[3] Puhach et al, *Nat. Rev. Microbiol* (2023)
[4] Yang et al, *The Lancet Microbe* (2023)

Study the impact of this changing landscape on viral dynamics in the population

- **Variant of infection** and **patient characteristics** may also **shape the viral dynamics** [3,4]
  - Studies often conducted on small specific cohorts (symptomatic, commorbidities…) → Potential selection bias

➔ Analyzing millions of PCR tests performed **in community labs**



324,428 individuals (**407,375** obs) with:

- Date of **symptom onset**,
- **Vaccination status,**
- **Variant** of infection

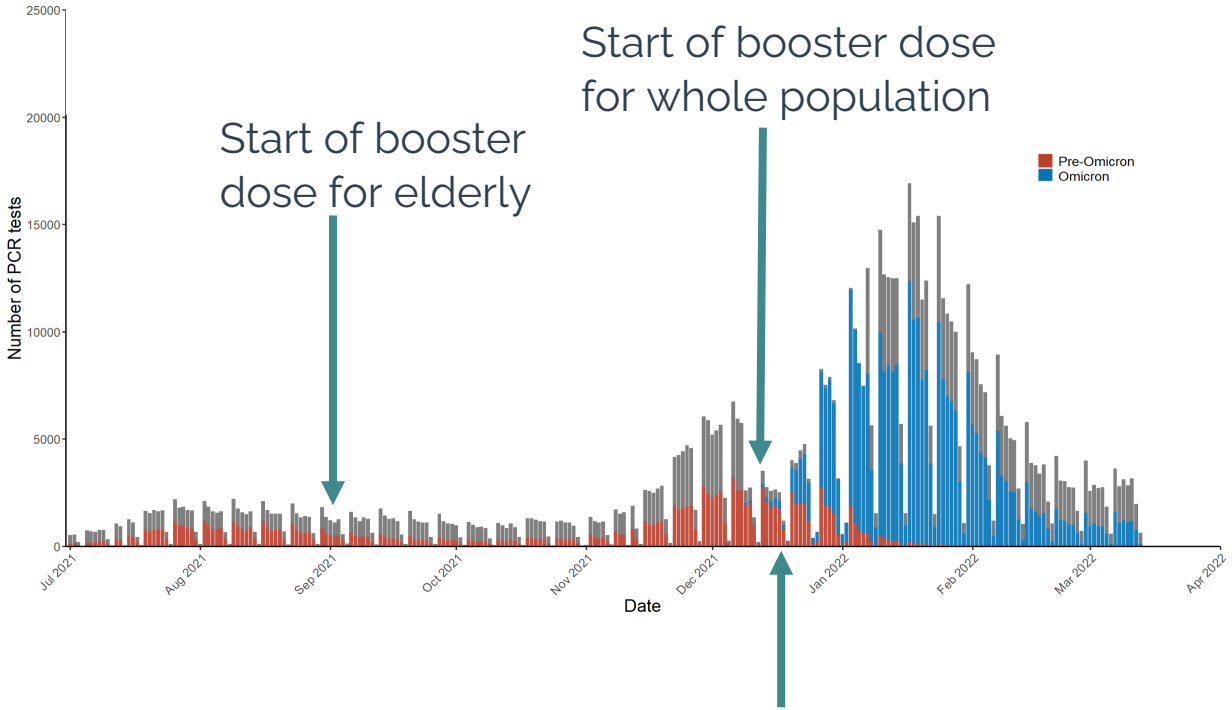
BIOGROUP
biologie médicale

[3] Puhach et al, *Nat. Rev. Microbiol* (2023)
[4] Yang et al, *The Lancet Microbe* (2023)

Study the impact of this changing landscape on viral dynamics in the population

- **Variant of infection** and **patient characteristics** may also **shape the viral dynamics** [3,4]
  - Studies often conducted on small specific cohorts (symptomatic, commorbidities…) → Potential selection bias

➔ Analyzing millions of PCR tests performed **in community labs**



324,428 individuals (**407,375** obs) with:

- Date of **symptom onset**,
- **Vaccination status,**
- **Variant** of infection

BIOGROUP
biologie médicale

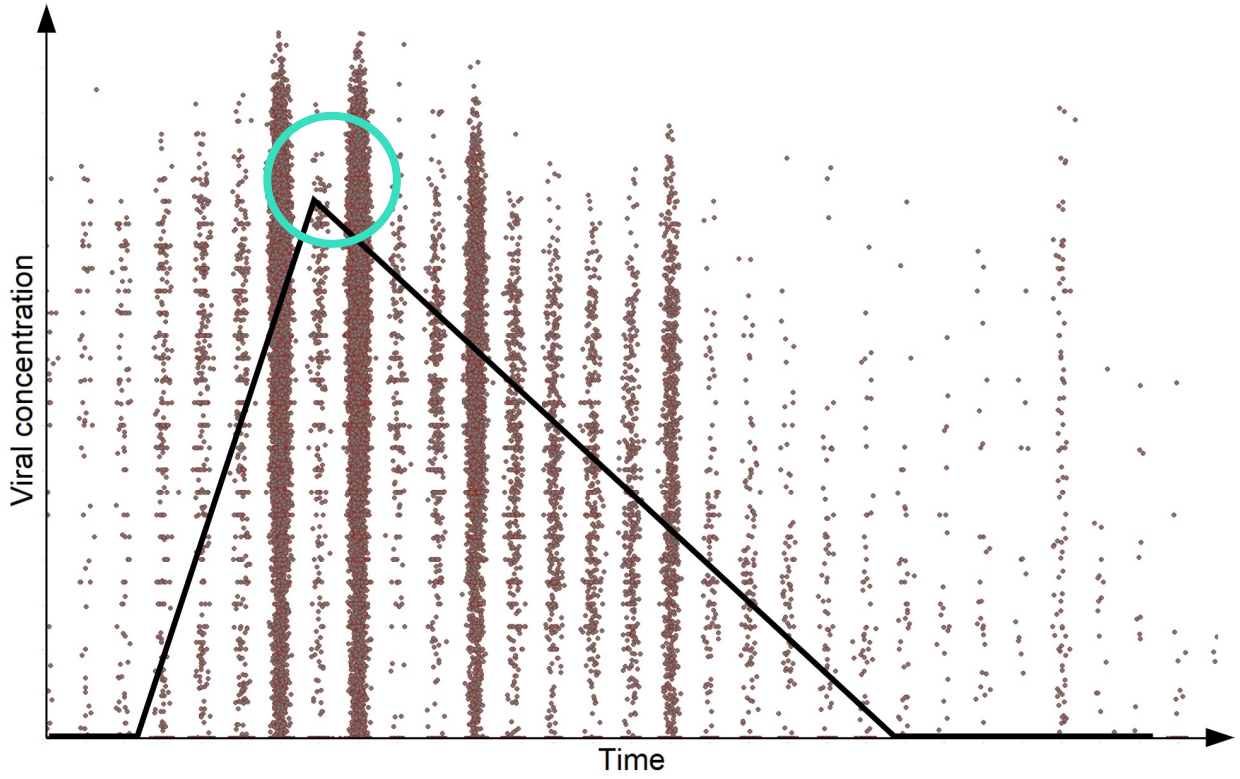
Can we model the community labs tests to identify patterns in viral load ?

[3] Puhach et al, *Nat. Rev. Microbiol* (2023)
[4] Yang et al, *The Lancet Microbe* (2023)

Simple viral load dynamics model to identify patterns

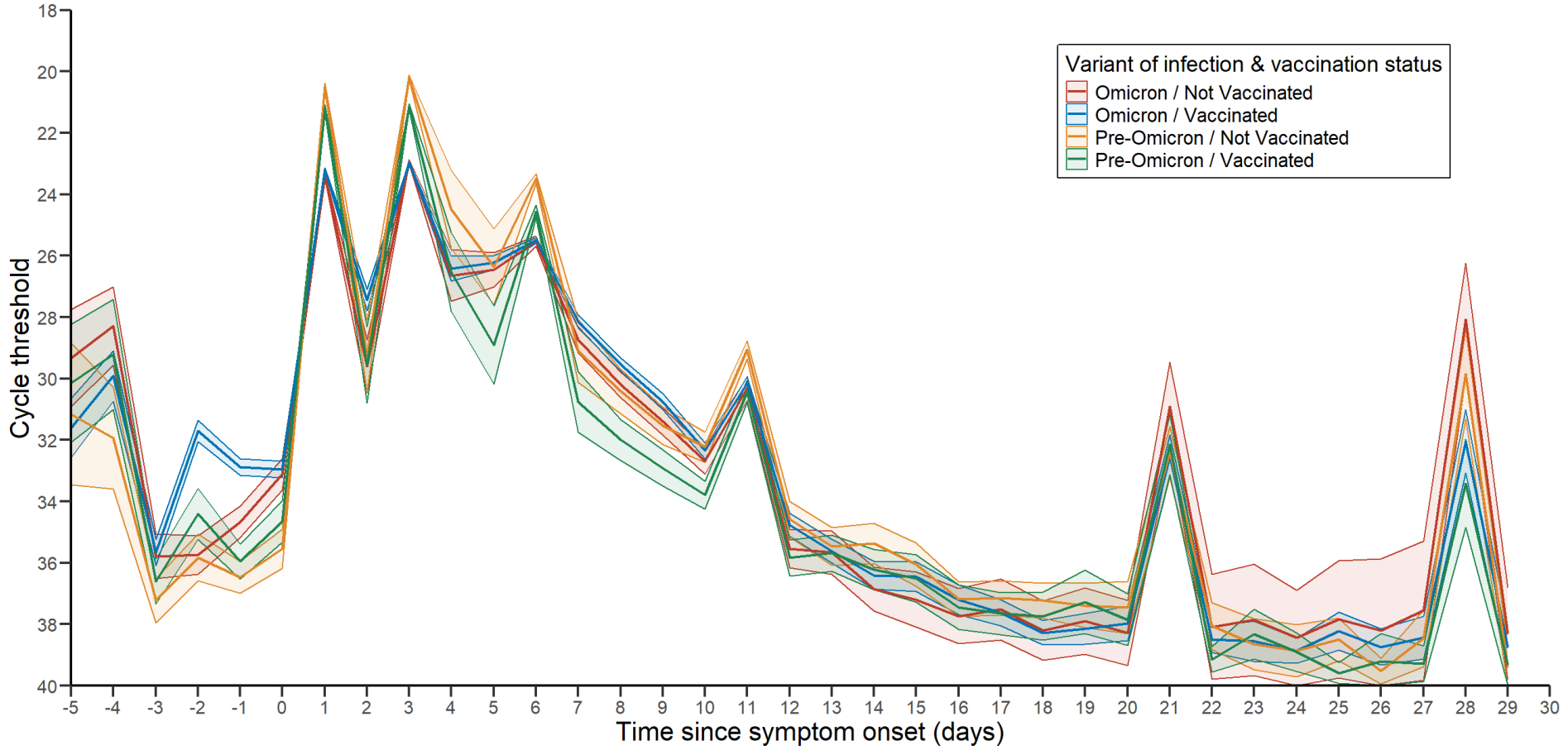
Modelling the viral load dynamics : reconstruct the individual viral load trajectories with mathematical models

Empirical models



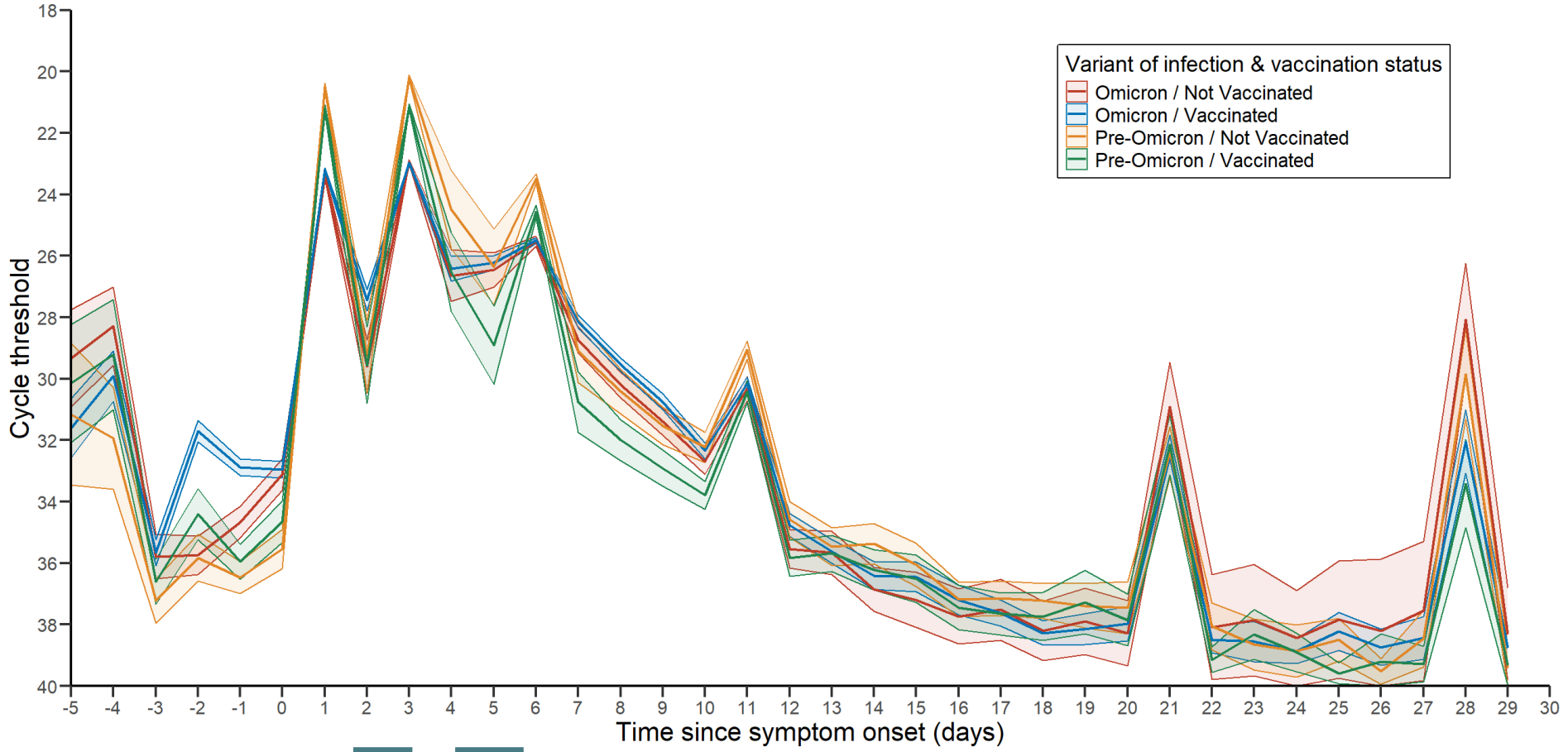
- ➔ Identify how viral dynamics patterns are impacted by variants and vaccination
- ➔ Using simple mathematical models
- ➔ Estimate subset of key parameters related to viral dynamics
- ➔ Adapted to massive datasets

Description of the data from community labs



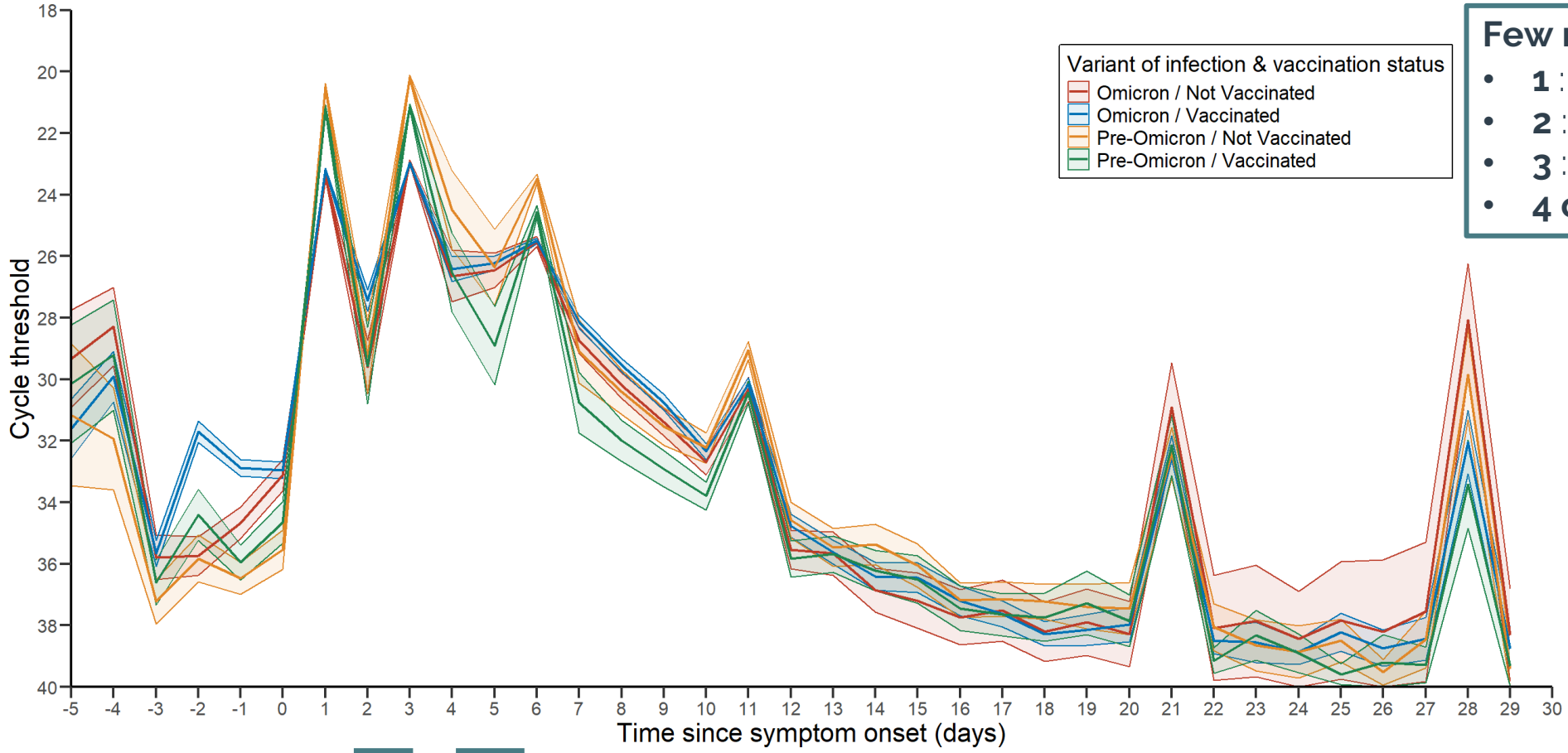
	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Omicron / Not Vaccinated	108	152	413	540	997	1125	24016	373	26776	269	553	5574	942	1010	808	890	1217	301	238	161	119	88	89	72	55	52	131	30	43	35	31	26	18	78	34
Omicron / Vaccinated	306	425	1192	2453	4153	3972	62326	2205	110467	1166	2987	19890	3772	3460	2546	2525	4127	919	804	579	485	417	445	338	260	246	538	219	161	217	174	151	182	292	128
Pre-Omicron / Not Vaccinated	63	128	306	410	775	650	13439	295	26501	188	196	6072	247	342	450	693	2093	344	294	256	218	213	230	192	145	104	367	90	69	74	63	56	52	168	53
Pre-Omicron / Vaccinated	81	107	348	434	725	604	11100	249	24604	170	180	4841	253	389	475	708	1705	322	312	239	188	175	178	132	102	113	286	69	86	98	70	56	59	142	48

Description of the data from community labs



	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Omicron / Not Vaccinated	108	152	413	540	997	1125	24016	373	26776	269	553	5574	942	1010	808	890	1217	301	238	161	119	88	89	72	55	52	131	30	43	35	31	26	18	78	34
Omicron / Vaccinated	306	425	1192	2453	4153	3972	62326	2205	110467	1166	2987	19890	3772	3460	2546	2525	4127	919	804	579	485	417	445	338	260	246	538	219	161	217	174	151	182	292	128
Pre-Omicron / Not Vaccinated	63	128	306	410	775	650	13439	295	26501	188	196	6072	247	342	450	693	2093	344	294	256	218	213	230	192	145	104	367	90	69	74	63	56	52	168	53
Pre-Omicron / Vaccinated	81	107	348	434	725	604	11100	249	24604	170	180	4841	253	389	475	708	1705	322	312	239	188	175	178	132	102	113	286	69	86	98	70	56	59	142	48

Description of the data from community labs

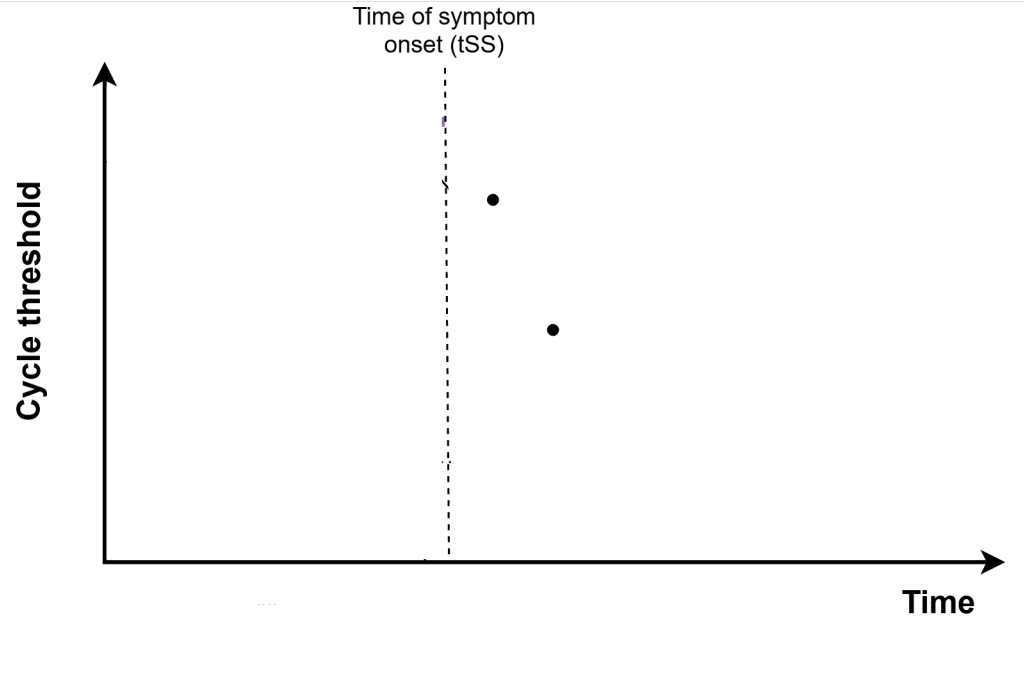


Few repeated tests :

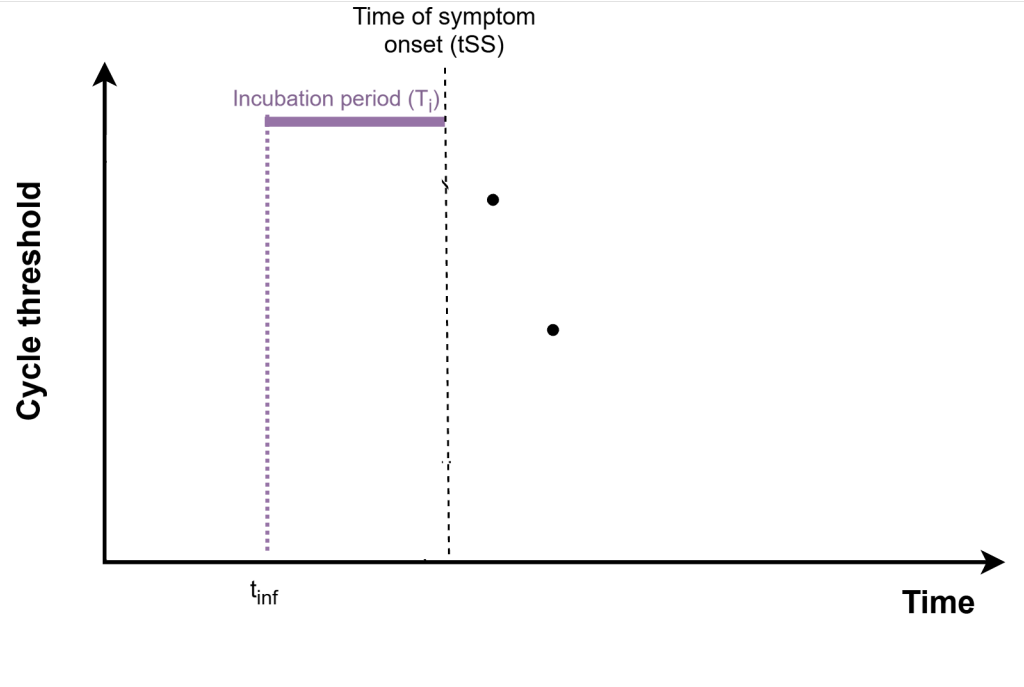
- **1** : 79% (N = 256,650)
- **2** : 17% (N = 55,560)
- **3** : 3 % (N = 9,957)
- **4 or more** : 1% (N = 2,261)

	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Omicron / Not Vaccinated	108	152	413	540	997	1125	24016	373	26776	269	553	5574	942	1010	808	890	1217	301	238	161	119	88	89	72	55	52	131	30	43	35	31	26	18	78	34
Omicron / Vaccinated	306	425	1192	2453	4153	3972	62326	2205	110467	1166	2987	19890	3772	3460	2546	2525	4127	919	804	579	485	417	445	338	260	246	538	219	161	217	174	151	182	292	128
Pre-Omicron / Not Vaccinated	63	128	306	410	775	650	13439	295	26501	188	196	6072	247	342	450	693	2093	344	294	256	218	213	230	192	145	104	367	90	69	74	63	56	52	168	53
Pre-Omicron / Vaccinated	81	107	348	434	725	604	11100	249	24604	170	180	4841	253	389	475	708	1705	322	312	239	188	175	178	132	102	113	286	69	86	98	70	56	59	142	48

Piecewise linear mixed-effects model to capture the viral load dynamics



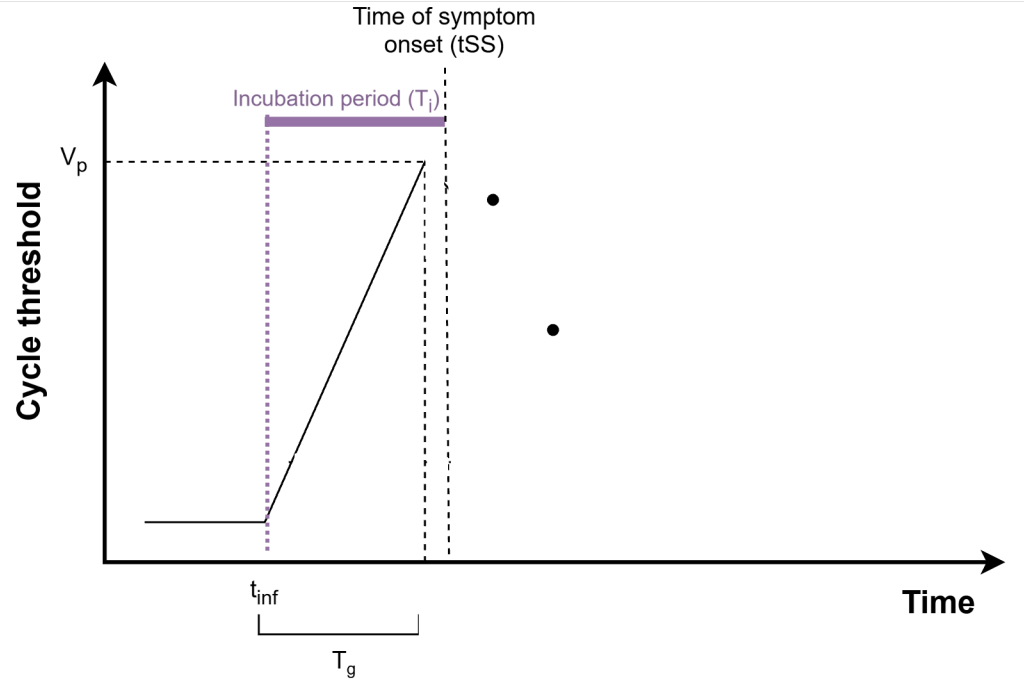
Piecewise linear mixed-effects model to capture the viral load dynamics



We estimate 4 parameters:

- Incubation period (days), $\boldsymbol{T_i}$

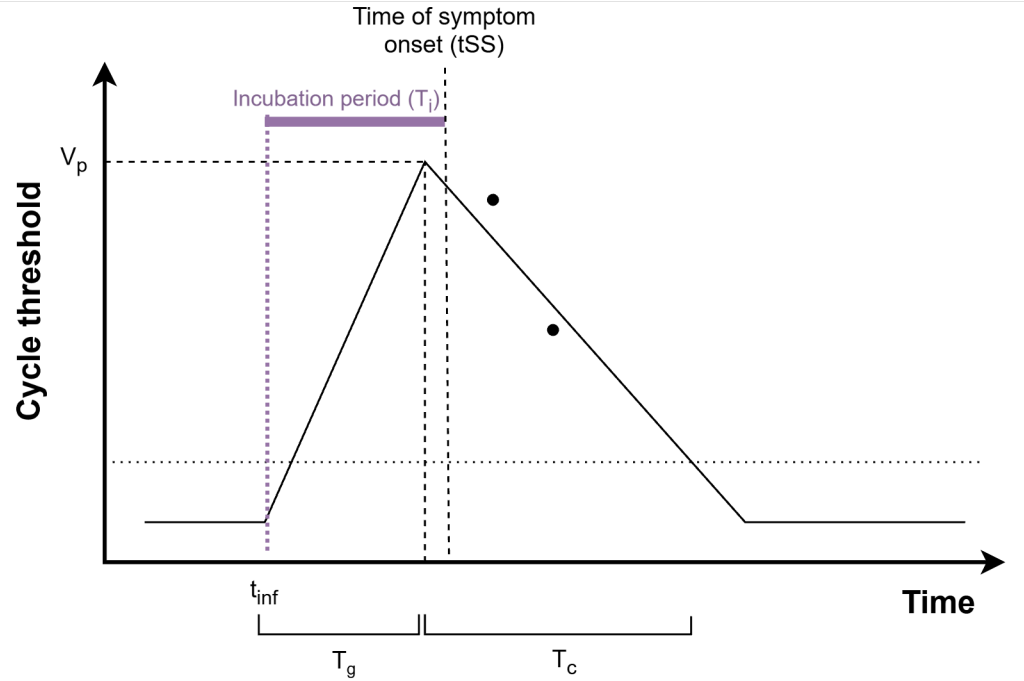
Piecewise linear mixed effects model to capture the viral load dynamics



We estimate 4 parameters:

- Incubation period (days), $\boldsymbol{T_i}$
- Proliferation phase (days), $\boldsymbol{T_g}$
- Viral load at peak (Ct), $\boldsymbol{V_p}$

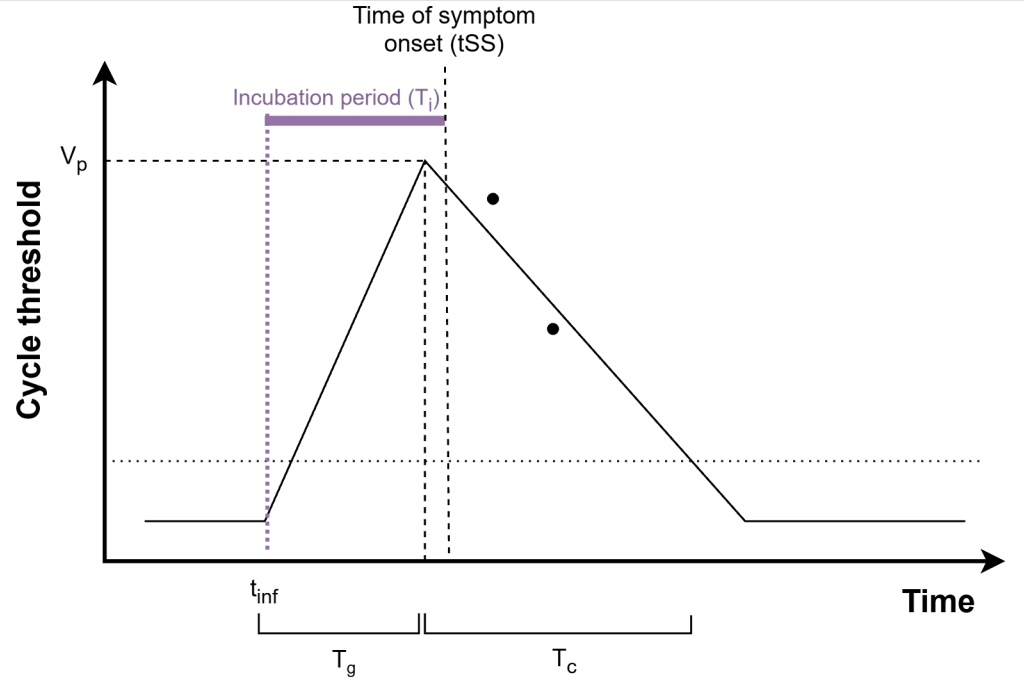
Piecewise linear mixed-effects model to capture the viral load dynamics



We estimate 4 parameters:

- Incubation period (days), $\boldsymbol{T_i}$
- Proliferation phase (days), $\boldsymbol{T_g}$
- Viral load at peak (Ct), $\boldsymbol{V_p}$
- Clearance phase (days), $\boldsymbol{T_c}$

Piecewise linear mixed-effects model to capture the viral load dynamics



We estimate 4 parameters:

- Incubation period (days), $\boldsymbol{T_i}$
- Proliferation phase (days), $\boldsymbol{T_g}$
- Viral load at peak (Ct), $\boldsymbol{V_p}$
- Clearance phase (days), $\boldsymbol{T_c}$

- We use the symptom onset to estimate the start of the infection ($\boldsymbol{t_{inf}}$):

$$t_{inf} = tSS - T_i$$

- We deduct the time to peak ($\boldsymbol{t_p}$) as followed:

$$t_p = t_{inf} + T_g$$

Proliferation phase

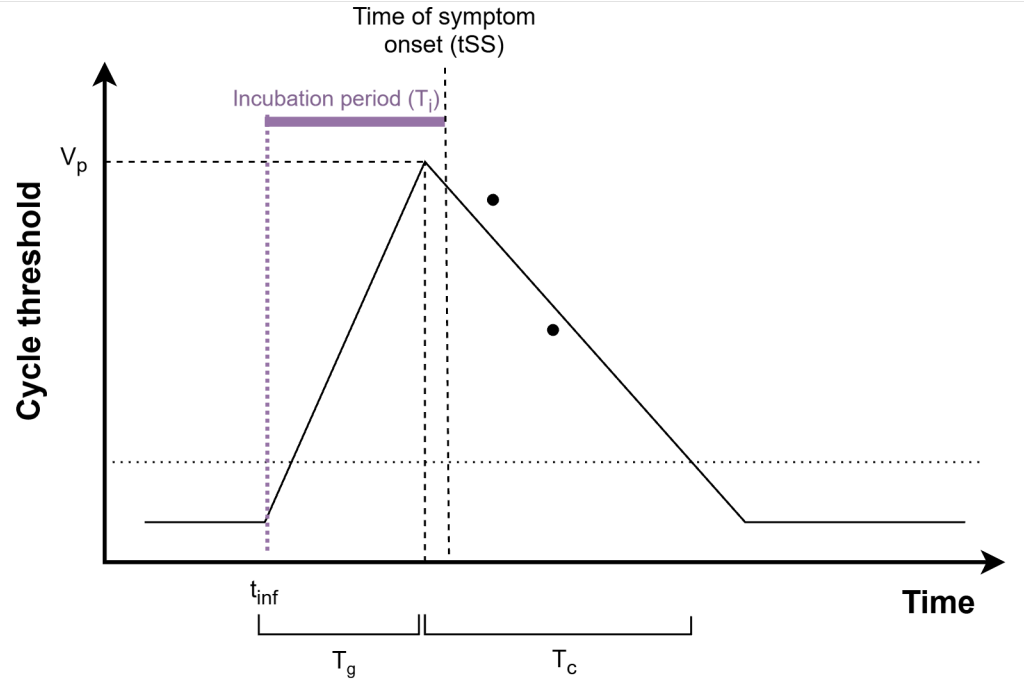
$$\text{E}[\Delta\text{Ct(t)}] = \begin{cases} 0 & t = t_{inf} \\ Vp \times \left(\frac{t - t_{inf}}{t_p - t_{inf}}\right) & t_{inf} < t < t_p \\ Vp & t = t_p \end{cases}$$

Clearance phase

$$\text{E}[\Delta\text{Ct(t)}] = \begin{cases} Vp & t = t_p \\ Vp + (V_{inf} - LOD - Vp) \times \left(\frac{t - t_p}{t_c - t_p}\right) & t_p < t < t_c \\ V_{inf} - LOD & t = t_c \end{cases}$$

with $\Delta\text{Ct(t)} = V_{inf} - Ct(t)$, $V_{inf} = 50$, and $LOD = 40$

Piecewise linear mixed-effects model to capture the viral load dynamics



We estimate 4 parameters:

- Incubation period (days), $\boldsymbol{T_i}$
- Proliferation phase (days), $\boldsymbol{T_g}$
- Viral load at peak (Ct), $\boldsymbol{V_p}$
- Clearance phase (days), $\boldsymbol{T_c}$

- We use the symptom onset to estimate the start of the infection ($\boldsymbol{t_{inf}}$):

$$t_{inf} = tSS - T_i$$

- We deduct the time to peak ($\boldsymbol{t_p}$) as followed:

$$t_p = t_{inf} + T_g$$

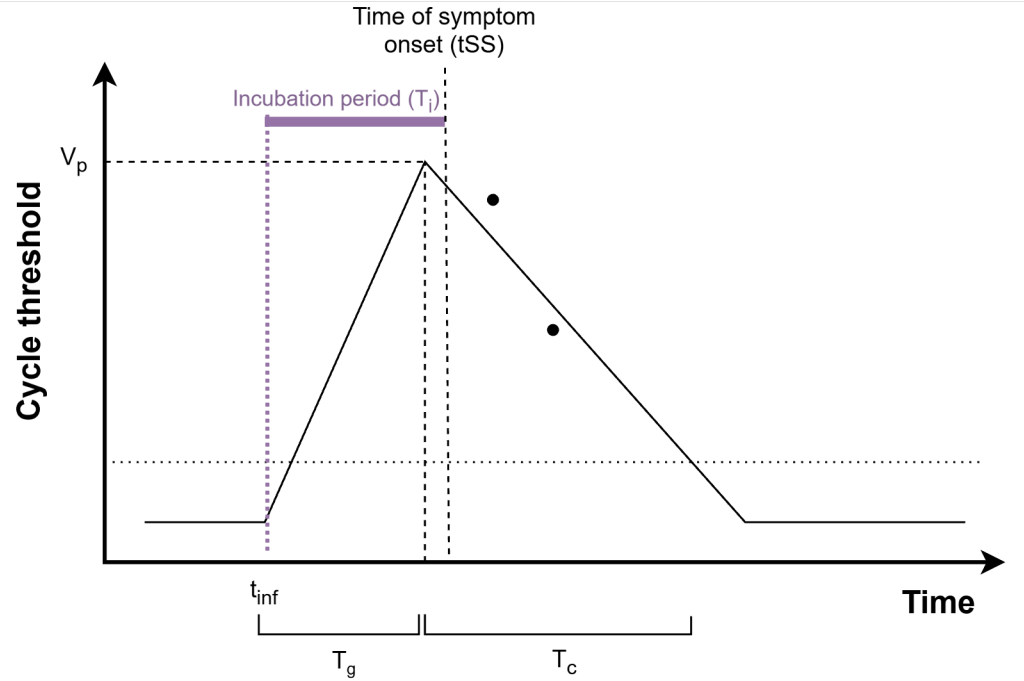
- Vector of individual parameters $\boldsymbol{\theta}[i] = \{T_i[\mathrm{i}], T_g[\mathrm{i}], T_c[\mathrm{i}], V_p[\mathrm{i}]\}$ is defined as followed:

$$\log(\boldsymbol{\theta}[i]) = \boldsymbol{exp}(\log(\mu_{\boldsymbol{\theta}}) + \boldsymbol{\eta}_{\boldsymbol{\theta}[i]})$$

Fixed effects: $\mu_{\boldsymbol{\theta}} \sim \boldsymbol{N}^{+}(\bar{m}_{\boldsymbol{\theta}}, \boldsymbol{\sigma}_{\boldsymbol{\theta}})$

Random effects: $\boldsymbol{\eta}_{\boldsymbol{\theta}[i]} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\omega}^{\mathbf{2}})$

Piecewise linear mixed-effects model to capture the viral load dynamics



Bayesian inference framework, HMC NUTS algorithm in Stan

Add information with prior distributions

We estimate 4 parameters:

- Incubation period (days), $\boldsymbol{T_i \sim N^+(5,1)}$
- Proliferation phase (days), $\boldsymbol{T_g \sim N^+(6,1)}$
- Viral load at peak (Ct), $\boldsymbol{V_p \sim N^+(25,2)}$
- Clearance phase (days), $\boldsymbol{T_c \sim N^+(15,2)}$

Weakly informative prior distributions (centered on simulation value)

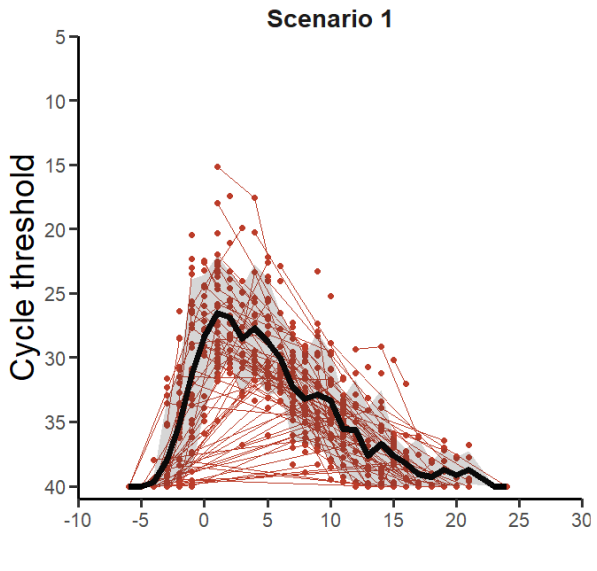
Simulation study

- 50 simulated datasets of 1000 individuals
- 50% of the population is infected
- Few repetead tests

Simulation study

- 50 simulated datasets of 1000 individuals
- 50% of the population is infected
- Few repetead tests

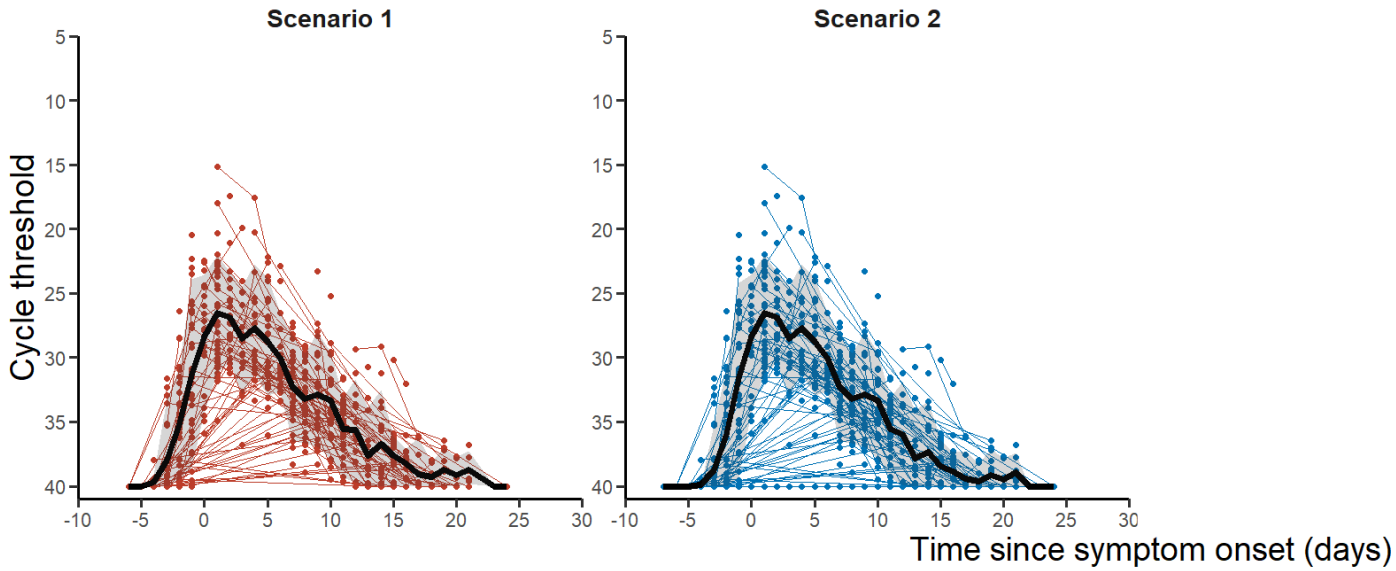
	Inclusion criteria	Percentage of infected individuals (P_{inf})	Timing of testing
Scenario 1	≥ 1 positive PCR	100%	Uniform from infection to clearance



Simulation study

- 50 simulated datasets of 1000 individuals
- 50% of the population is infected
- Few repetead tests

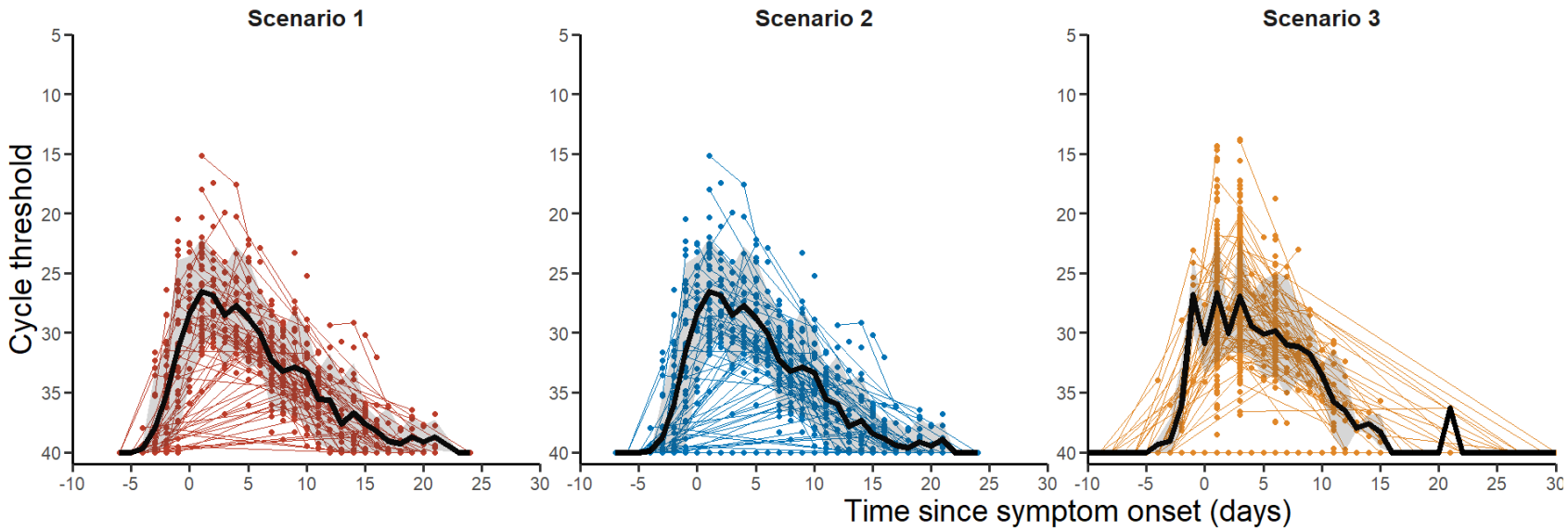
	Inclusion criteria	Percentage of infected individuals (P_{inf})	Timing of testing
Scenario 1	≥ 1 positive PCR	100%	Uniform from infection to clearance
Scenario 2	Entire population	50%	Uniform from infection to clearance



Simulation study

- 50 simulated datasets of 1000 individuals
- 50% of the population is infected
- Few repetead tests

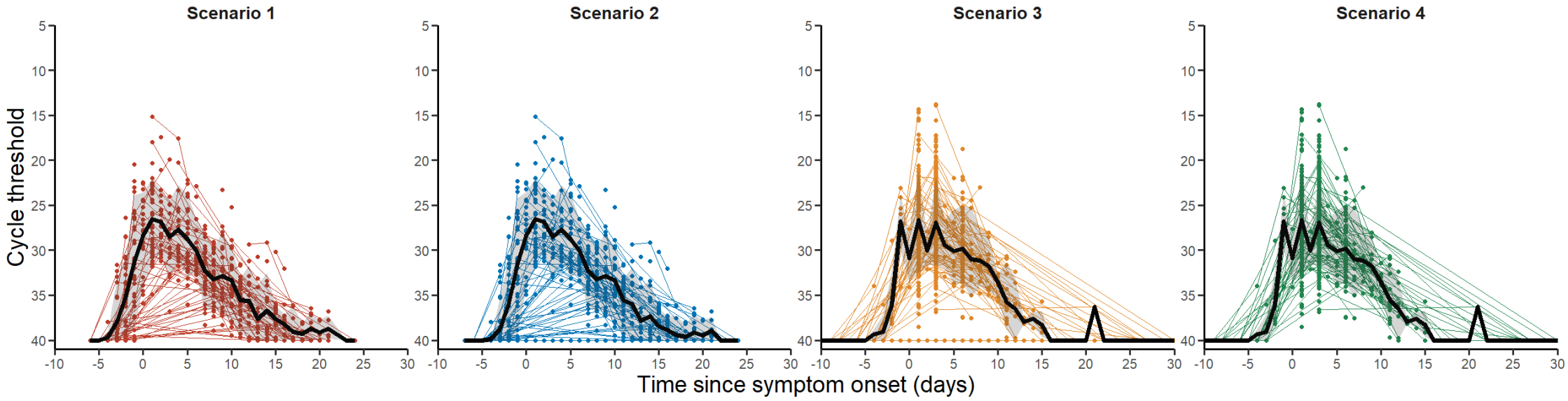
	Inclusion criteria	Percentage of infected individuals (P_{inf})	Timing of testing
Scenario 1	≥ 1 positive PCR	100%	Uniform from infection to clearance
Scenario 2	Entire population	50%	Uniform from infection to clearance
Scenario 3	Entire population	50%	Mostly at symptom onset



Simulation study

- 50 simulated datasets of 1000 individuals
- 50% of the population is infected
- Few repetead tests

	Inclusion criteria	Percentage of infected individuals (P_{inf})	Timing of testing
Scenario 1	≥ 1 positive PCR	100%	Uniform from infection to clearance
Scenario 2	Entire population	50%	Uniform from infection to clearance
Scenario 3	Entire population	50%	Mostly at symptom onset
Scenario 4	≥ 1 positive PCR	100%	Mostly at symptom onset



Likelihood definition in each scenario

- Modelling the **viral dynamics of the infected individuals only** (**Scenario 1**, **4**)

$$L(y_{i,t}|\psi_i) = \mathbb{1}_{\{uncensored\}} f_N\left(x \middle| E[y_{i,t}|\psi_i], \sigma\right) + \mathbb{1}_{\{censored\}} F_N\left(LOD \middle| E[y_{i,t}|\psi_i], \sigma\right)$$

- f_N the Normal PDF, F_N the Normal CDF,
- *LOD* = Limit of detection

Contribution if infected

Likelihood definition in each scenario

- Modelling the **viral dynamics of the infected individuals only** (**Scenario 1**, **4**)

$$L(y_{i,t}|\psi_i) = \mathbb{1}_{\{uncensored\}} f_N\left(x|E[y_{i,t}|\psi_i], \sigma\right) + \mathbb{1}_{\{censored\}} F_N\left(LOD|E[y_{i,t}|\psi_i], \sigma\right)$$

- **Determining infectious status** and **modeling viral dynamics for all individuals** (**Scenario 2**, **3**)

$$L(y_{i,t}|\psi_i) = P_{inf}\left[\mathbb{1}_{\{uncensored\}} f_N\left(x|E[y_{i,t}|\psi_i], \sigma\right) + \mathbb{1}_{\{censored\}} F_N\left(LOD|E[y_{i,t}|\psi_i], \sigma\right)\right] + (1 - P_{inf})\left[\mathbb{1}_{\{uncensored\}} P(false\ positive) + \mathbb{1}_{\{censored\}} P(true\ negative)\right]$$

- f_N the Normal PDF, F_N the Normal CDF,
- LOD = Limit of detection
- $P(false\ positive)$ =0,0002, $P(true\ negative) = 0{,}9998$,
- P_{inf} the percentage of infected individuals in the dataset

Contribution if infected

Contribution if not infected

Evaluation of the model

- **Chains convergence**: R-hat [5]
  - Ratio of intra-chain to inter-chain variance
  - Must be less than 1.1 for each parameter

- **Error estimation**: (Relative Estimates Error, REE in %) [6]

$$REE\left(\hat{\theta}^k\right) = \frac{\hat{\theta}^k - \theta^*}{\theta^*} \times 100$$

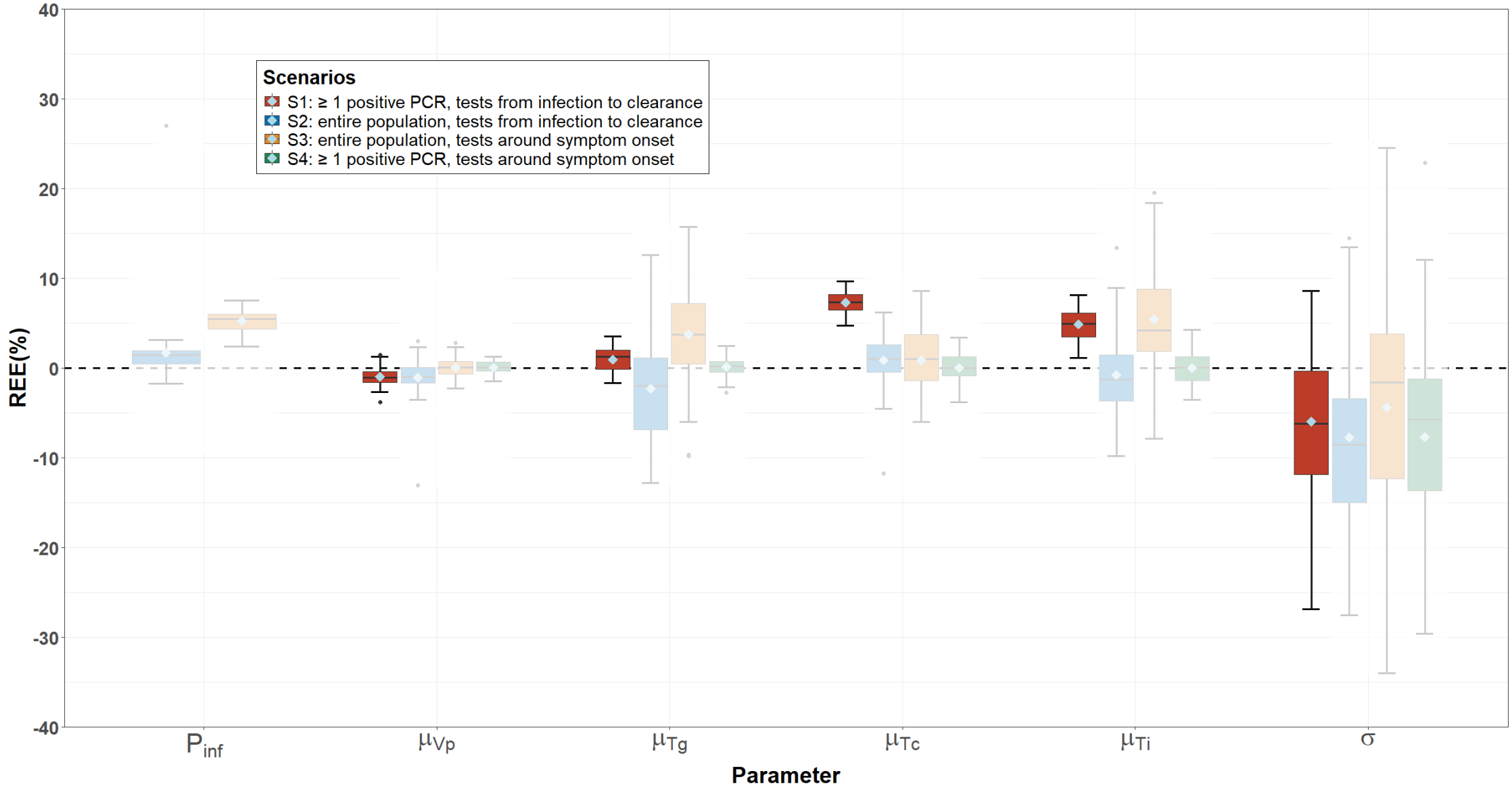
- **Estimation accuracy**: (Coverage rate, CR) [6]

$$CR_{(1-\alpha)}(\theta) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}_{\{\theta^* \in \widehat{CI}^k_{(1-\alpha)}\}}$$

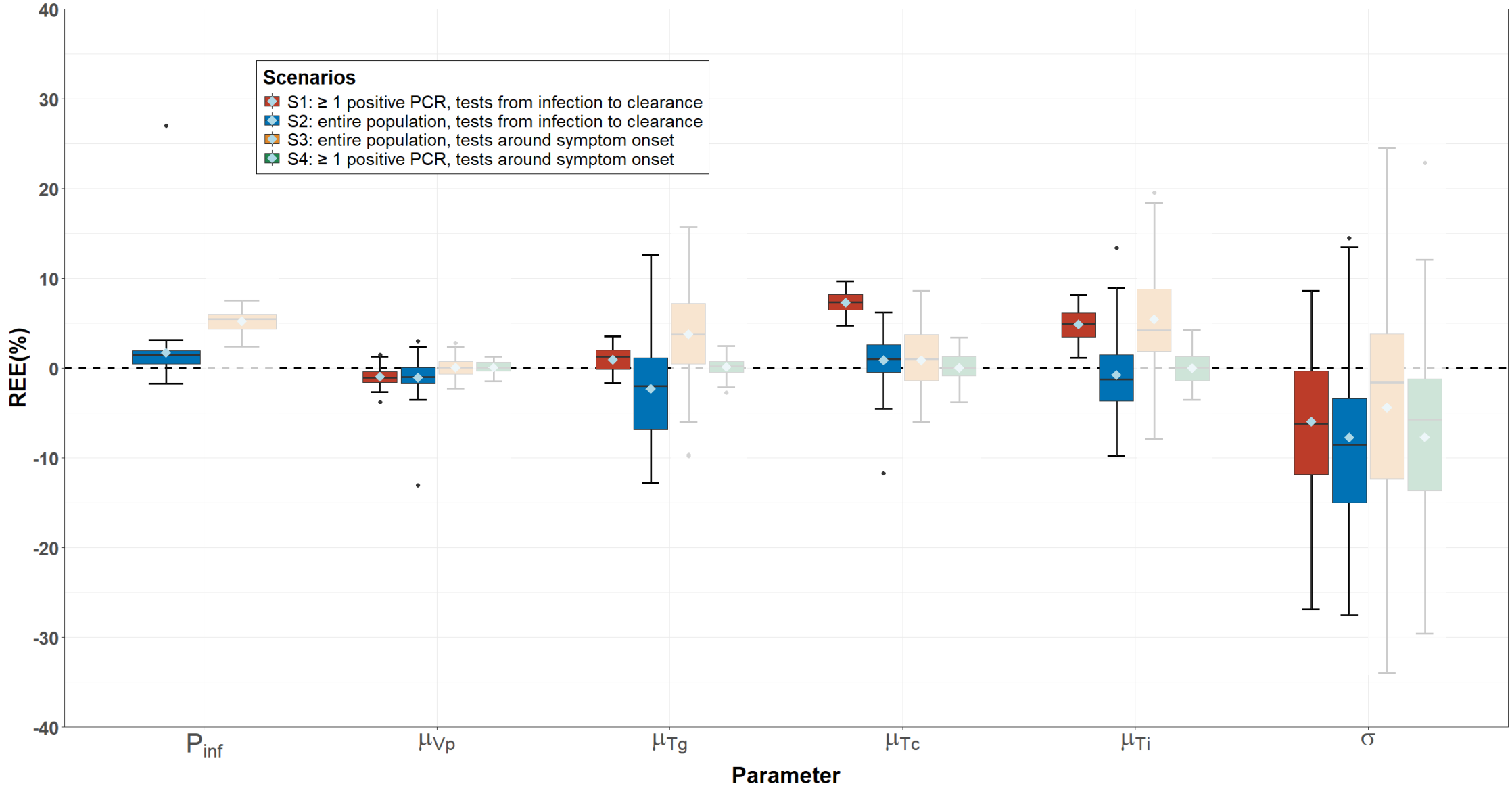
- **Goodness of fit**: (Posterior Predictive Check)

[5] Gelman et al, *Bayesian Data Analysis* (1995)
[6] Morris et al, *Statistics in Medicine* (2018)

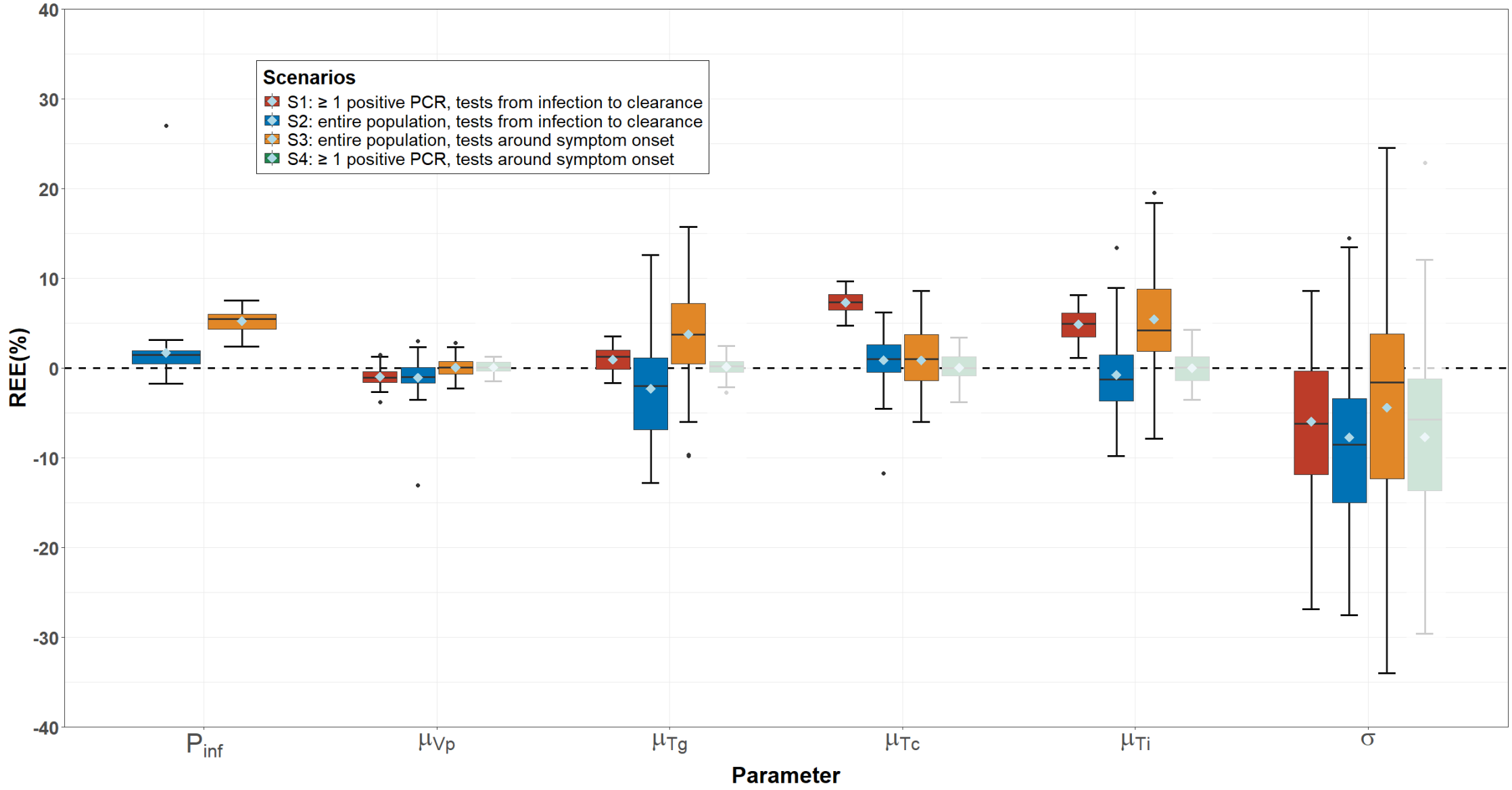
Relative Estimates Errors (REEs)



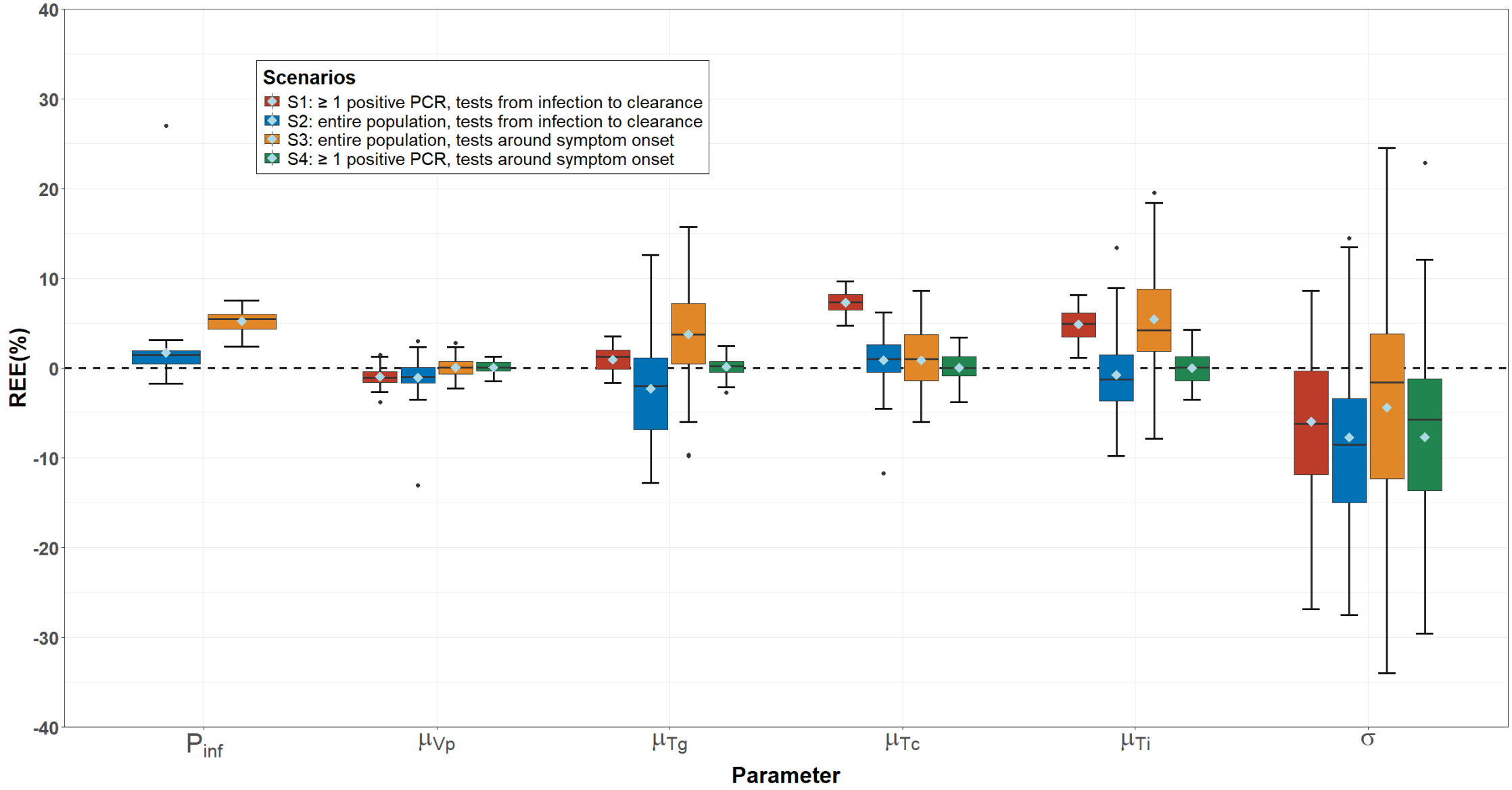
Relative Estimates Errors (REEs)



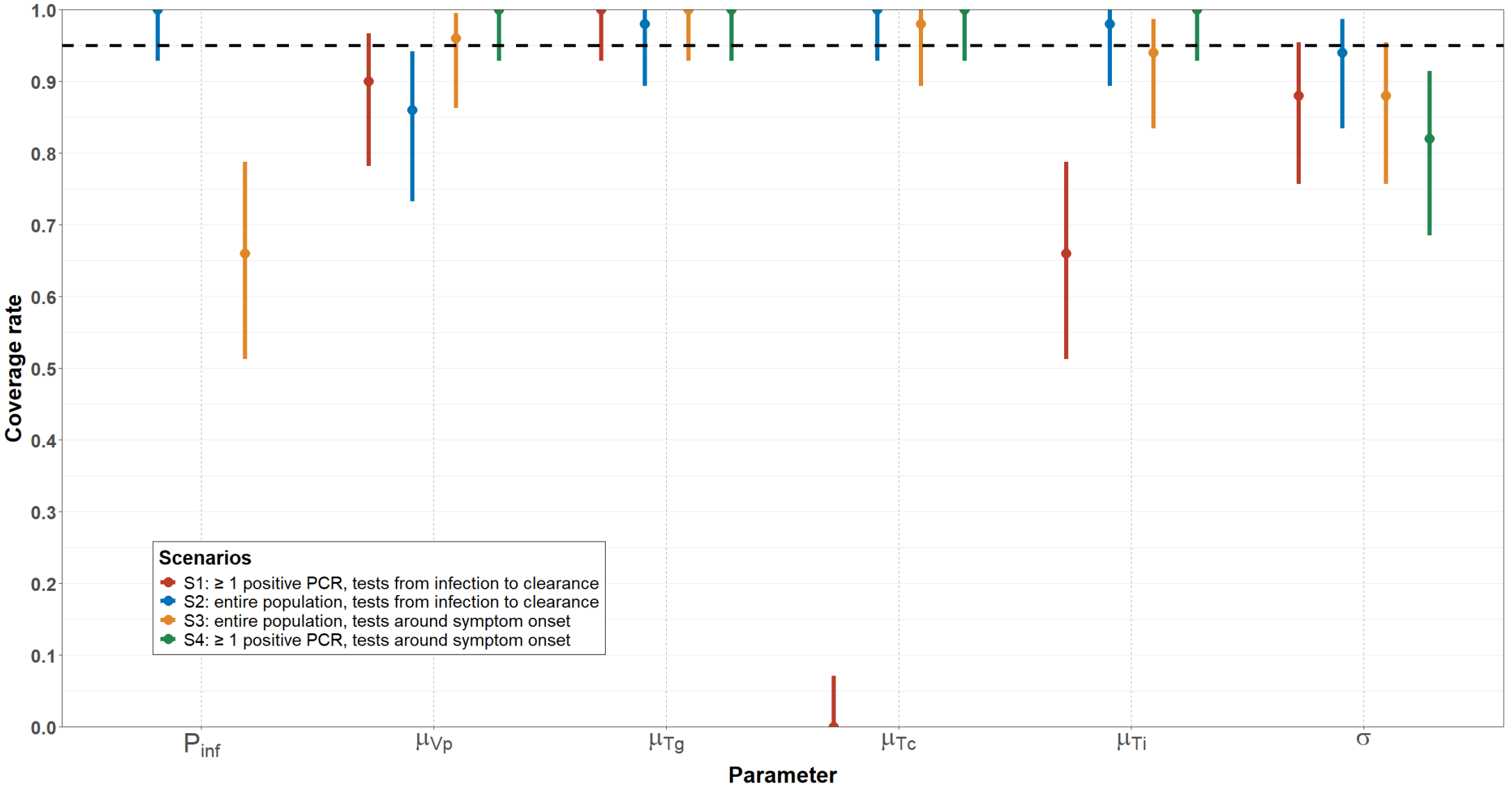
Relative Estimates Errors (REEs)



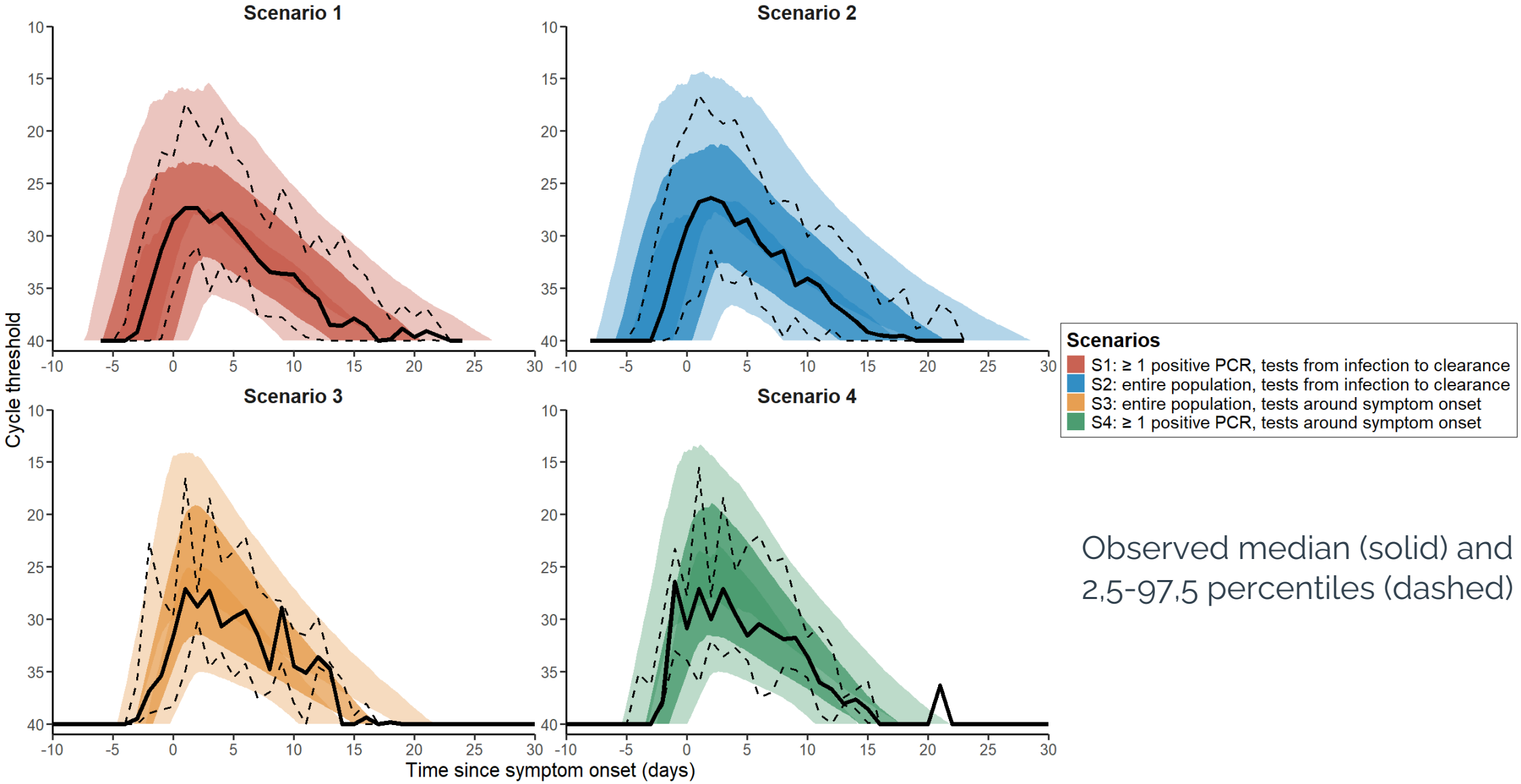
Relative Estimates Errors (REEs)



Coverage rate

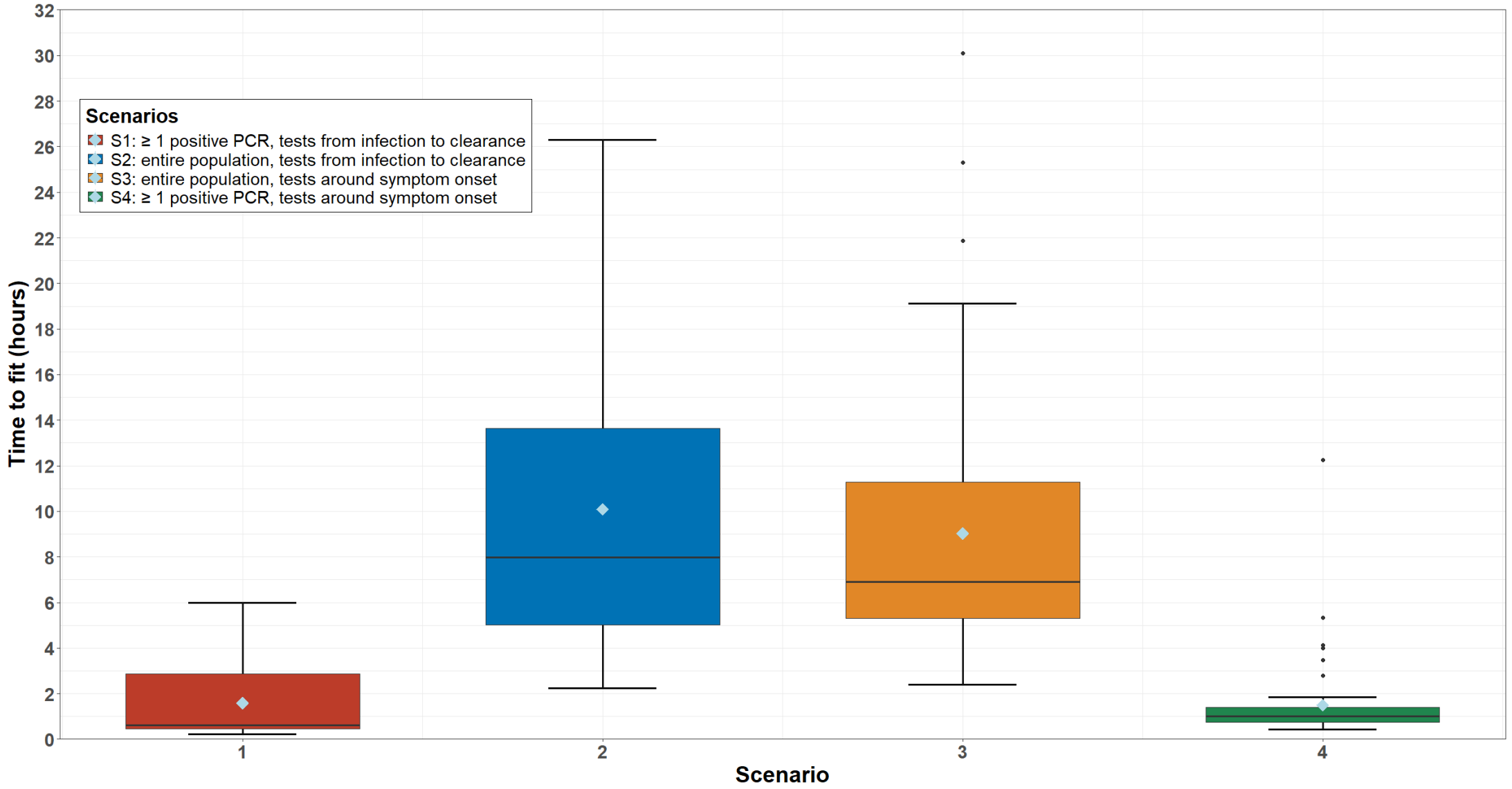


Posterior predictive checks

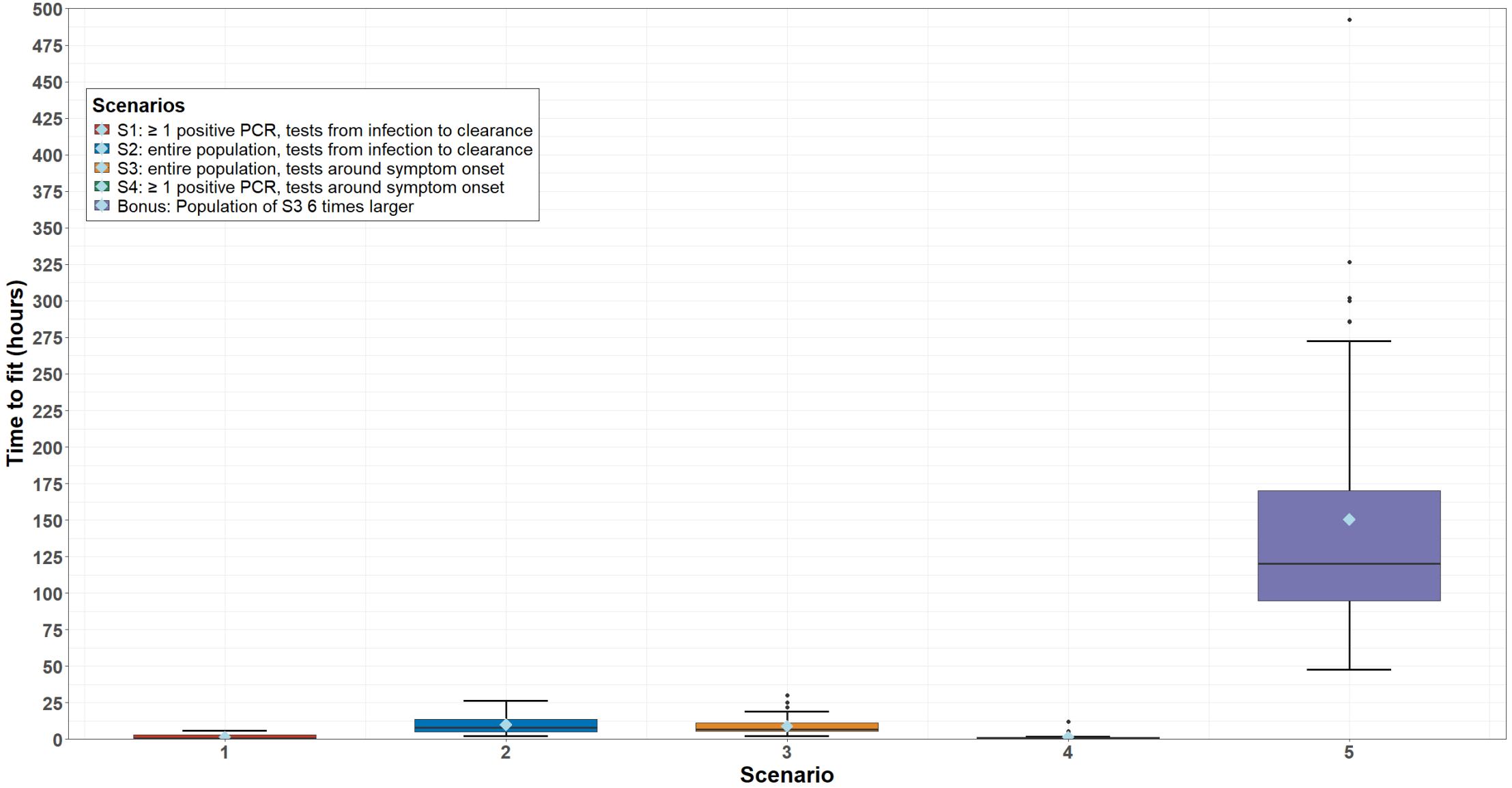


Observed median (solid) and 2,5-97,5 percentiles (dashed)

High computation time



High computation time



How to run faster ?

- Vectorization of the code
- Within-chain parallelization

How to run faster ?

- Vectorization of the code
- Within-chain parallelization
- Data reduction

  The idea : 2 individuals with the same Ct value at the same time from symptom onset contributes exactly the same way in the LL

  ➔ weight the LL contribution by the number of individuals having exactly the sames observations at the same time

  With 80% of individuals with only one positive PCR test ➔ **could be highly valuable**

How to run faster ?

- Vectorization of the code
- Within-chain parallelization
- Data reduction

  The idea : 2 individuals with the same Ct value at the same time from symptom onset contributes exactly the same way in the LL

  ➔ weight the LL contribution by the number of individuals having exactly the sames observations at the same time

  With 80% of individuals with only one positive PCR test ➔ **could be highly valuable**

Original dataset		Reduced dataset	
Number of individuals	Number of observations	Number of individuals	Number of observations
586,635	738,403	38,368	104,542

Divided by 15

We were unable to run ➔ The chains did not mix

Discussion

Findings

- We can identify the main patterns of viral load with a piecewise linear model

Limitations

- High computation time due to Bayesian framework
- Prior distributions are weakly informative but centred on the true value of the parameter

Perspectives

- Impact of variant of infection and vaccination in patterns of viral load ?
- Model other acute respiratory diseases (Influenza and RSV)

  M2 internship Laura MULAS (April-October 2025)

Acknowledgements

Jérémie Guedj (IAME)

Florence Débarre (IEES Paris, CNRS)

François Blanquart (CIRB, Collège de France)

Assil Merlaud (IAME)

Bob Carpenter (Flatarion Institute) for the responses in the Stan forum

Biogroup team

Université Paris Cité to fund this PhD