Optimizing Age-Structured Sampling for Estimating the Seroconversion Rate in Malaria Seroepidemiology: A Simulation Study

Authors

Yura K Ko[1,2], Shilei Li[1], Tom Britton[3], Wataru Kagaya[4], Akira Kaneko[1,5]

[1]Department of Microbiology, Tumor and Cell Biology (MTC), Karolinska Institutet, Sweden

[2]Department of Vector Ecology and Environment, Institute of Tropical Medicine (NEKKEN), Nagasaki University, Japan

[3]Department of Mathmatics, Stockholm University, Sweden

[4]Department of Ecoepidemiology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, Japan

[5]Department of Virology and Parasitology, Graduate School of Medicine, Osaka Metropolitan University, Japan.

Table of Contents

Instructions for the shiny application

To run the shiny application locally on your computer, follow these steps:

0. Install R, Julia, and RStudio:

If you have not already installed R, Julia, and RStudio, download and install them.

- **R**: https://cran.r-project.org/
- **Julia**: https://julialang.org/
- **RStudio**: https://posit.co/download/rstudio-desktop/

1. Download the shiny app files:

You can download all the necessary files from here (https://github.com/KoKYura/malaria_seroepi_sampling)

2. Install required R packages:
Open Rstudio and install the necessary packages by running the following commands in the R console.

```
Install.packages("shiny")

Install.packages("shinydashboard")

Install.packages("DT")

Install.packages("JuliaCall")
```

3. Run the application:
Open the "app.R" in Rstudio, and run the application either clicking the "Run App" button in Rstudio or running the following command in the R console:

```
shiny::runApp()
```

4. Select a scenario from the left panel:

Choose the simulation scenario based on your interest: 1) Stable SCR or 2) Change in SCR.

5. Set parameters and run simulations:

Modify the following parameters:

For the Stable SCR scenario,

- Number of samples
- Seroconversion rate (SCR)
- Seroreversion rate (SRR)
- Estimating parameter (SCR only, SCR and SRR)

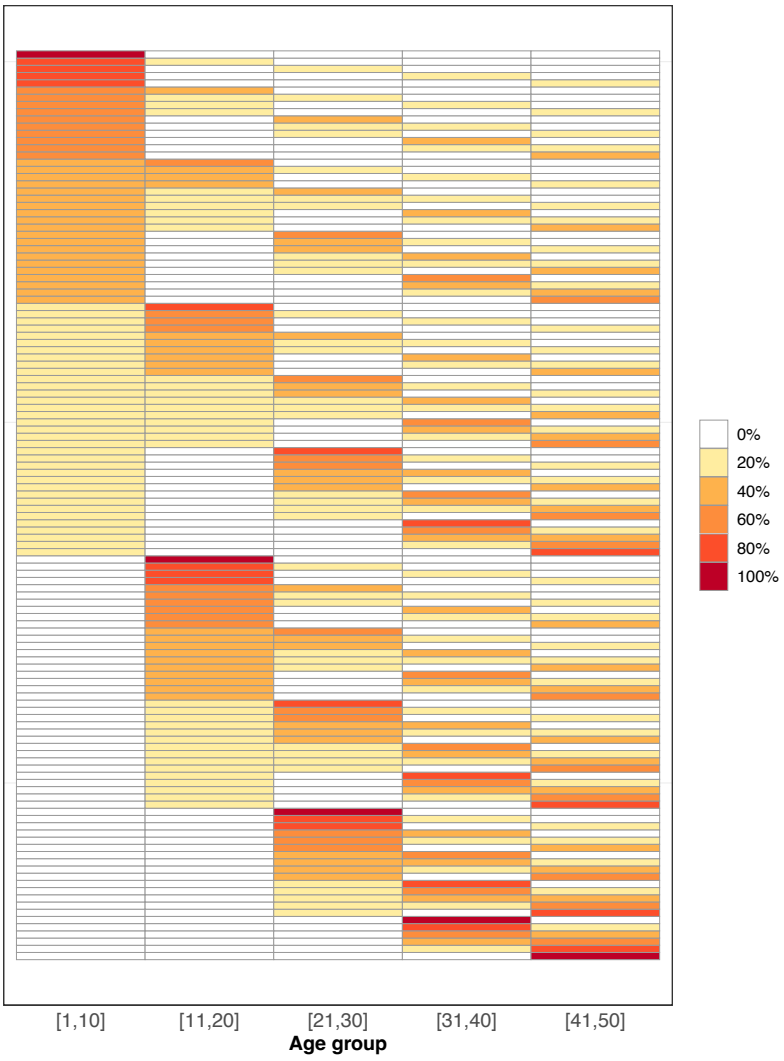
For the Change in SCR scenario,

- Number of samples

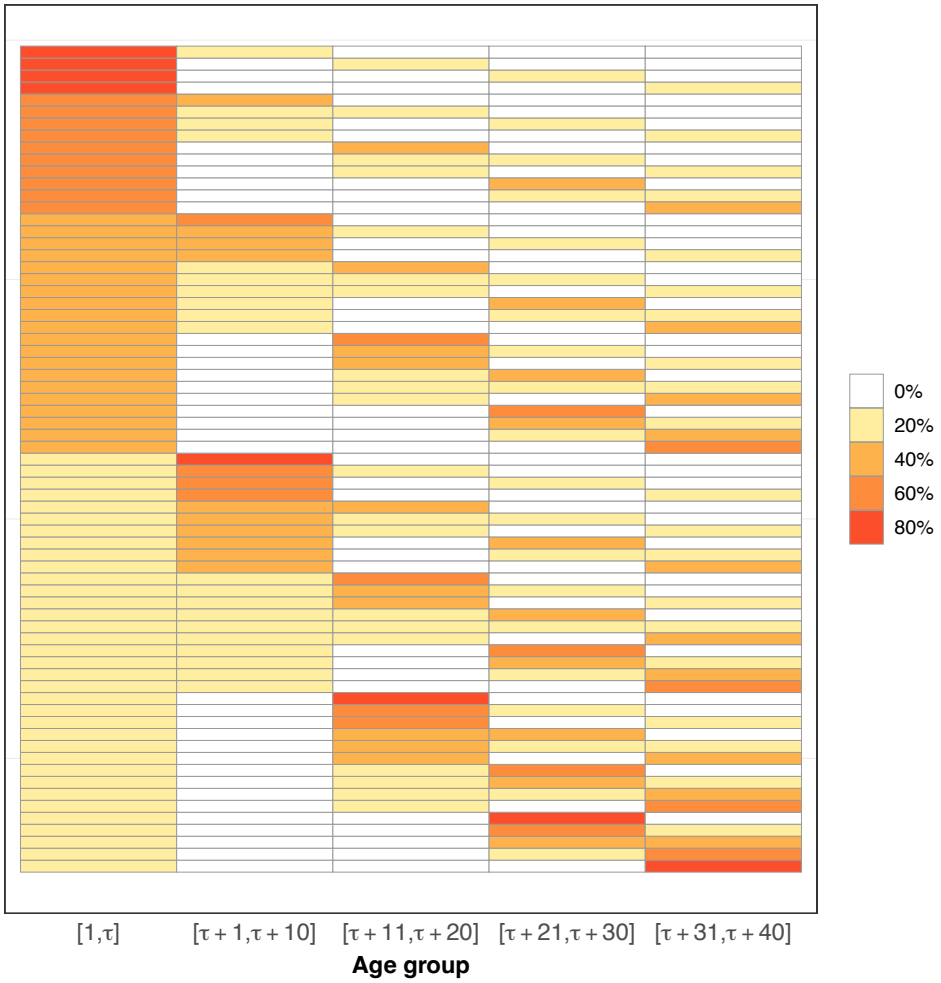
- SCR before the change point
- SCR after the change point
- SRR
- Chnge point (Years)
- Estimating parameter (SCRs only, SCRs and SRR)

After setting all parameters, click “Run” to start the simulation. Once the simulation is complete, you will obtain the the 95%CI for the SCR(s) along with the age-based sampling structures that yield the highest and lowest precision.

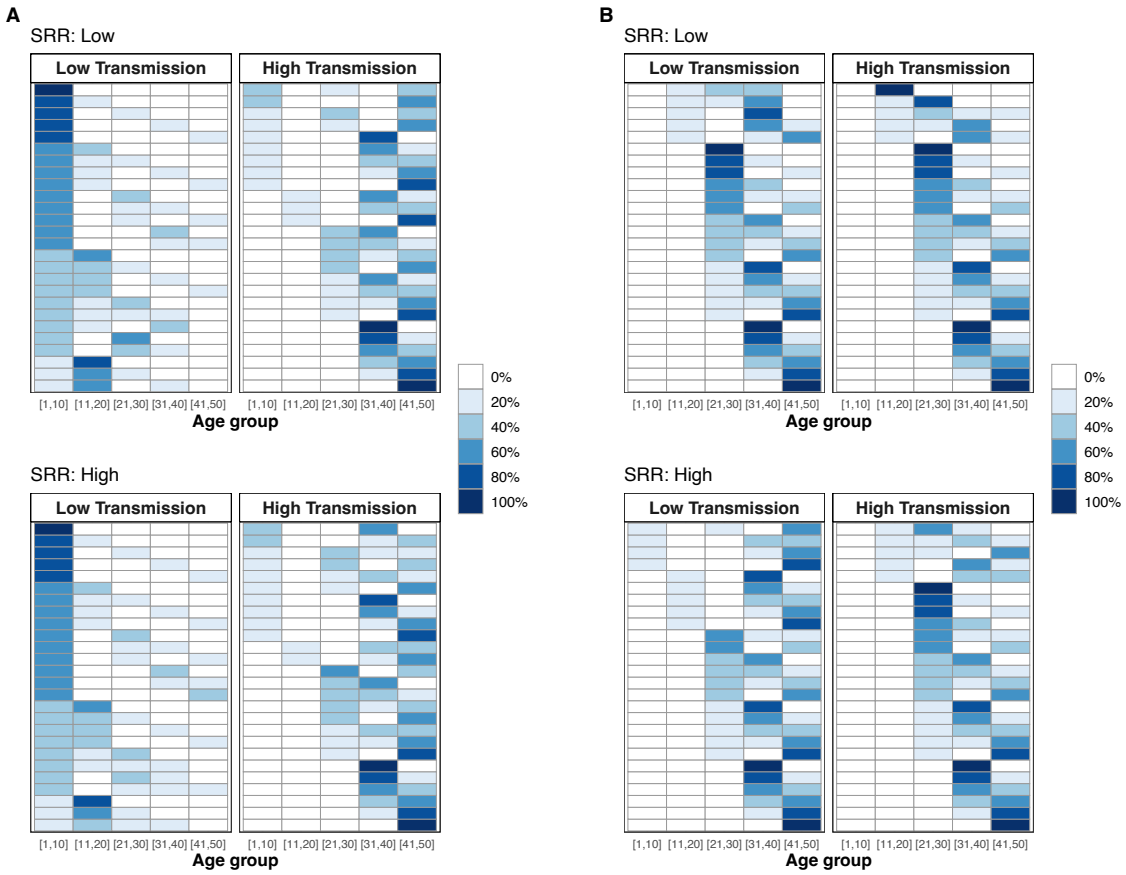
Supplementary Figures



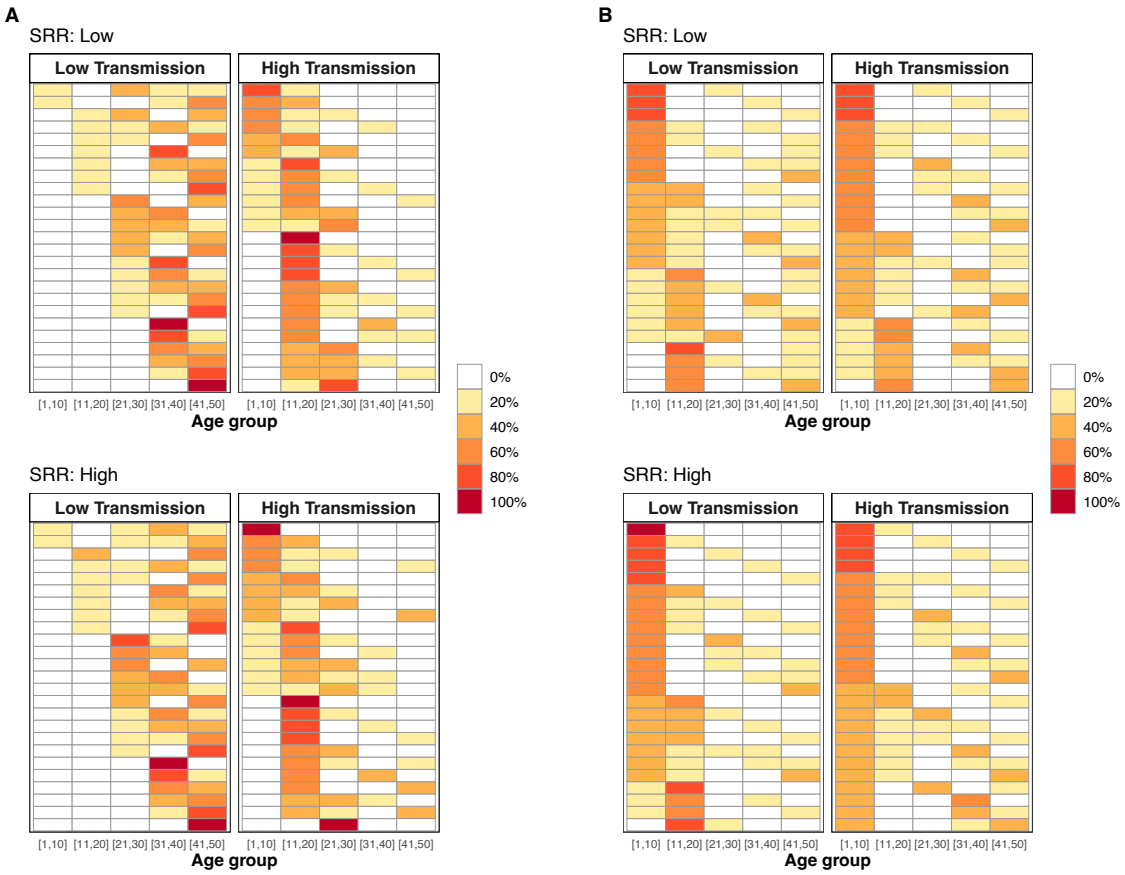
Supplementary Figure 1**.** The 126 possible combinations of five age groups ([1,10], [11, 20], [21, 30], [31, 40], and [41, 50]) with proportions ranging from 0% to 100% in increments of 20%.



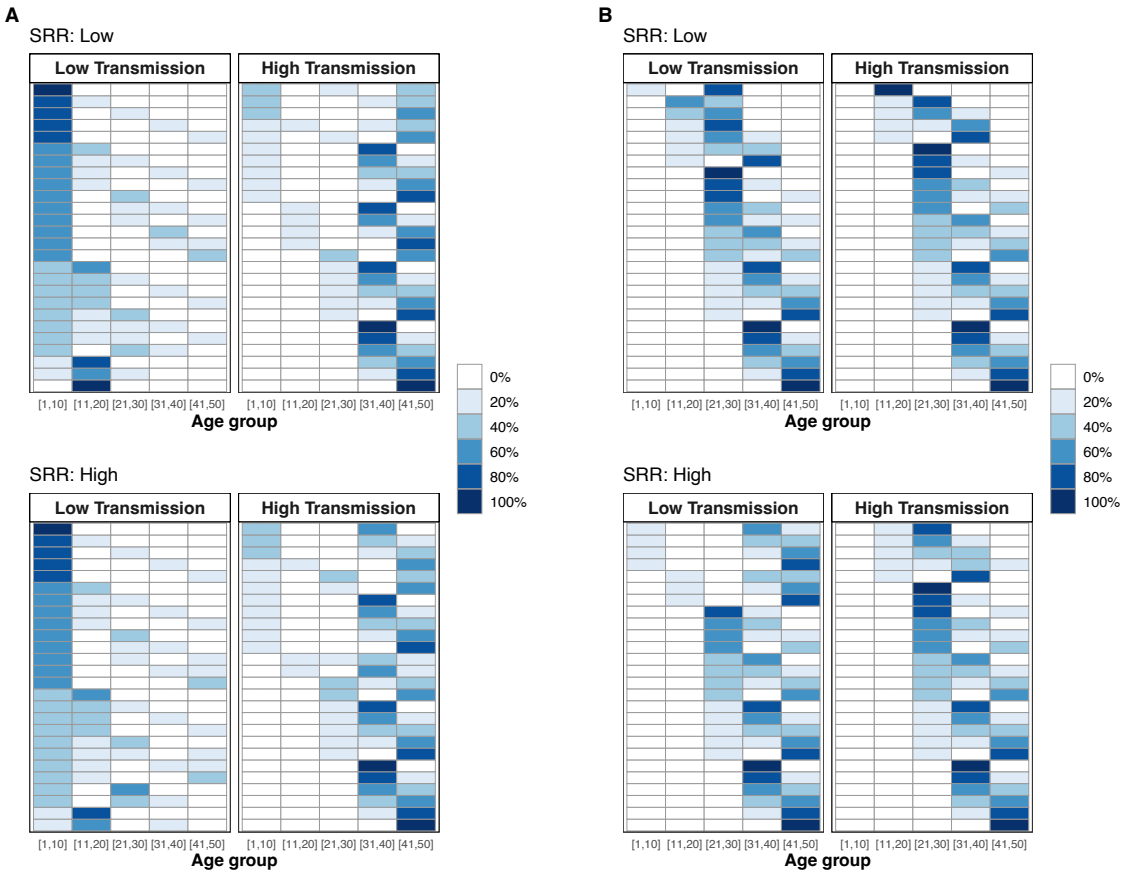
Supplementary Figure 2**.** The 69 possible combinations of five age groups ($[1, \tau]$, $[\tau + 1, \tau + 10]$, $[\tau + 11, \tau + 20]$, $[\tau + 21, \tau + 30]$, and $[\tau + 31, \tau + 40]$) under a single change point in malaria transmission intensity scenario. τ corresponds the time (in years) since the change point. The proportion of the first age group was varied from 20 to 80% in increments of 20%, while the other age groups were varied from 0 to 80% to capture information from both before and after the change point.



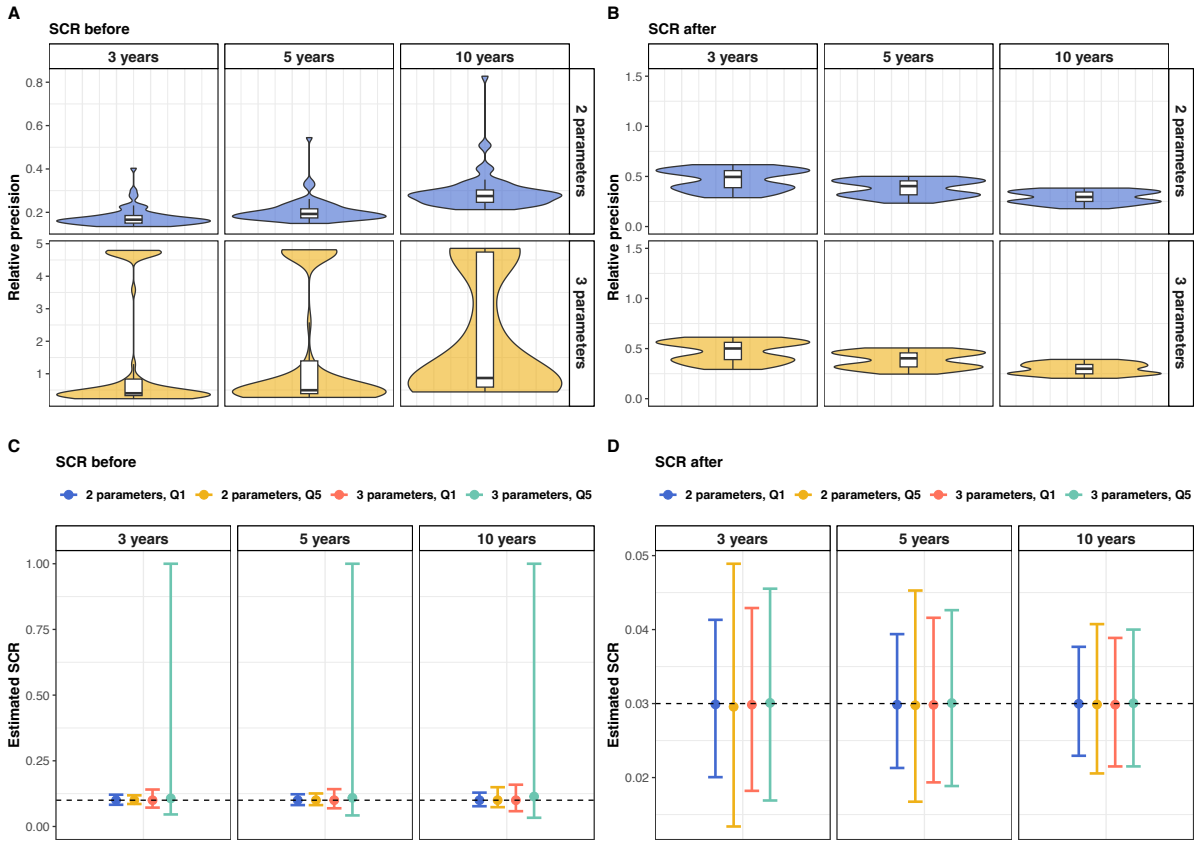
Supplementary Figure 3**.** Age-based sampling structures that yielded the bottom 20% relative precision under the stable malaria transmission intensity scenario with a total sample size of 500. A) Only the seroconversion rate (SCR) was estimated. B) The SCR and the seroreversion rate (SRR) were jointly estimated.



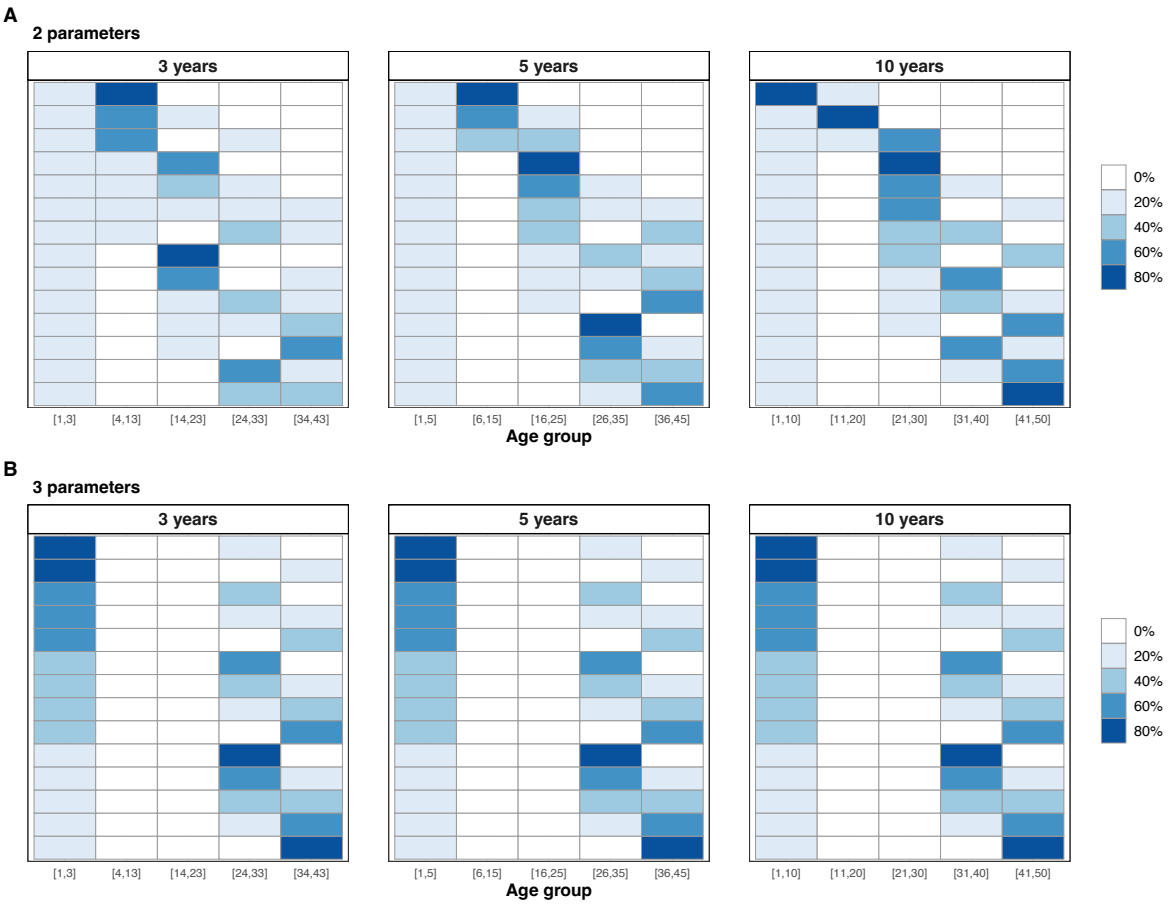
Supplementary Figure 4**.** Age-based sampling structures that yielded the top 20% relative precision under the stable malaria transmission intensity scenario with a total sample size of 1,000. A) Only the seroconversion rate (SCR) was estimated. B) The SCR and the seroreversion rate (SRR) were jointly estimated.



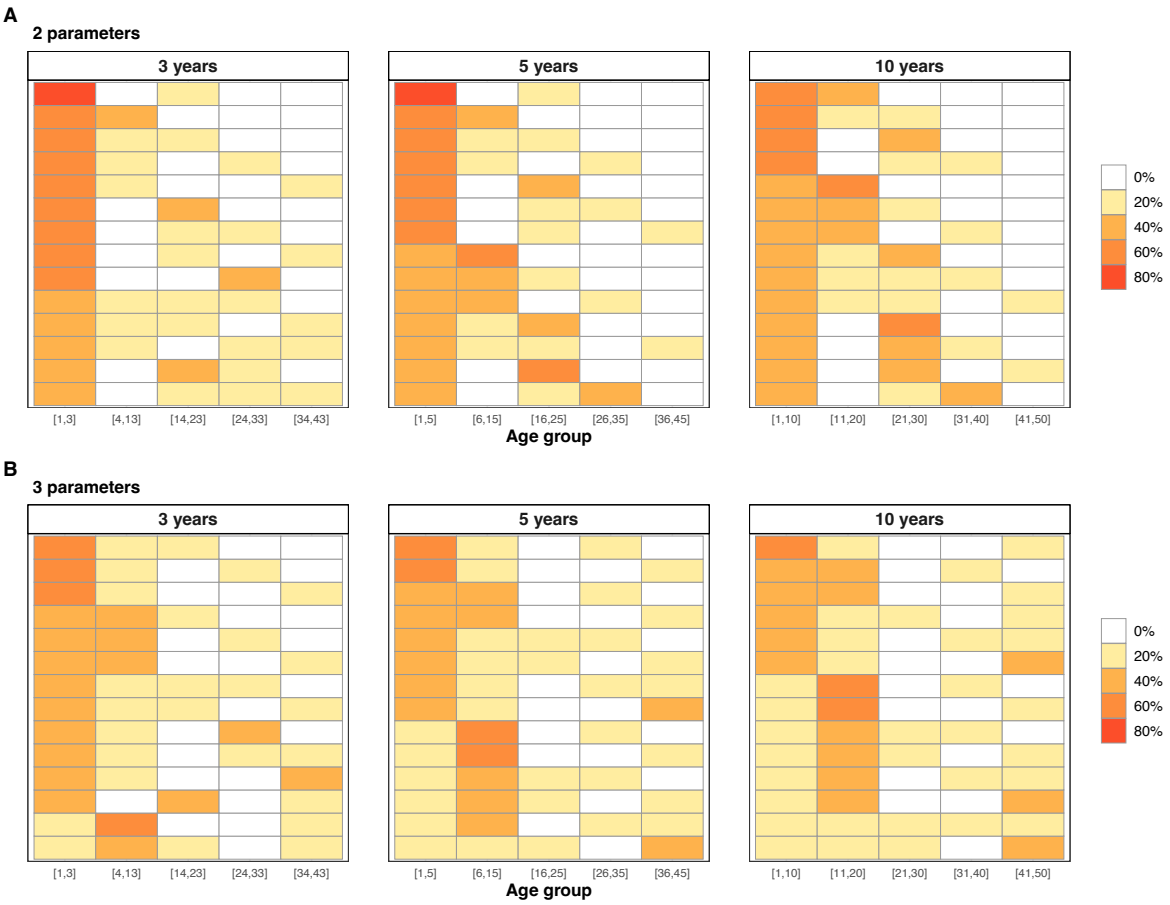
Supplementary Figure 5**.** Age-based sampling structures that yielded the bottom 20% relative precision under the stable malaria transmission intensity scenario with a total sample size of 1,000. A) Only the seroconversion rate (SCR) was estimated. B) The SCR and the seroreversion rate (SRR) were jointly estimated.



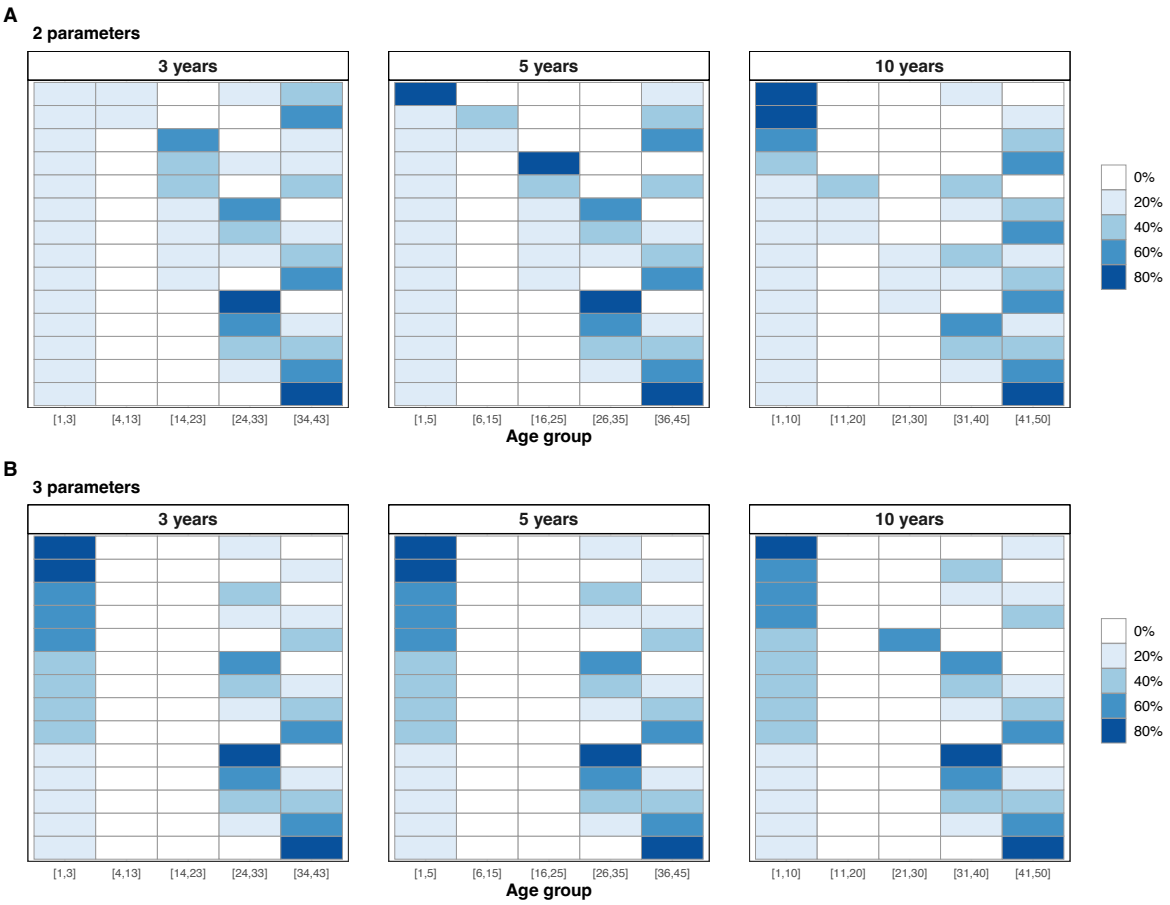
Supplementary Figure 6**.** Relative precisions (A and B) and 95% confidence intervals (C and D) of the estimated seroconversion rates (SCRs) before (A and C) and after (B and D) a single change point in malaria transmission intensity. The analysis assumed a total sample size of 1,000 with high underlying SCRs (SCR = 0.1 before and 0.03 after the change point) and a low seroreversion rate (SRR = 0.017). Results are stratified by 1) the time since the change point (3, 5, and 10 years) and 2) the number of parameters estimated (2 parameter: SCRs before and after the change point; 3 parameters: SCRs and SRR). Panels C and D are further stratified by the age structures that yielded the top 20% joint precision (Q1) and the bottom 20% joint precision (Q5). Each dot in Panels C and D represents the median estimated SCR within each group.



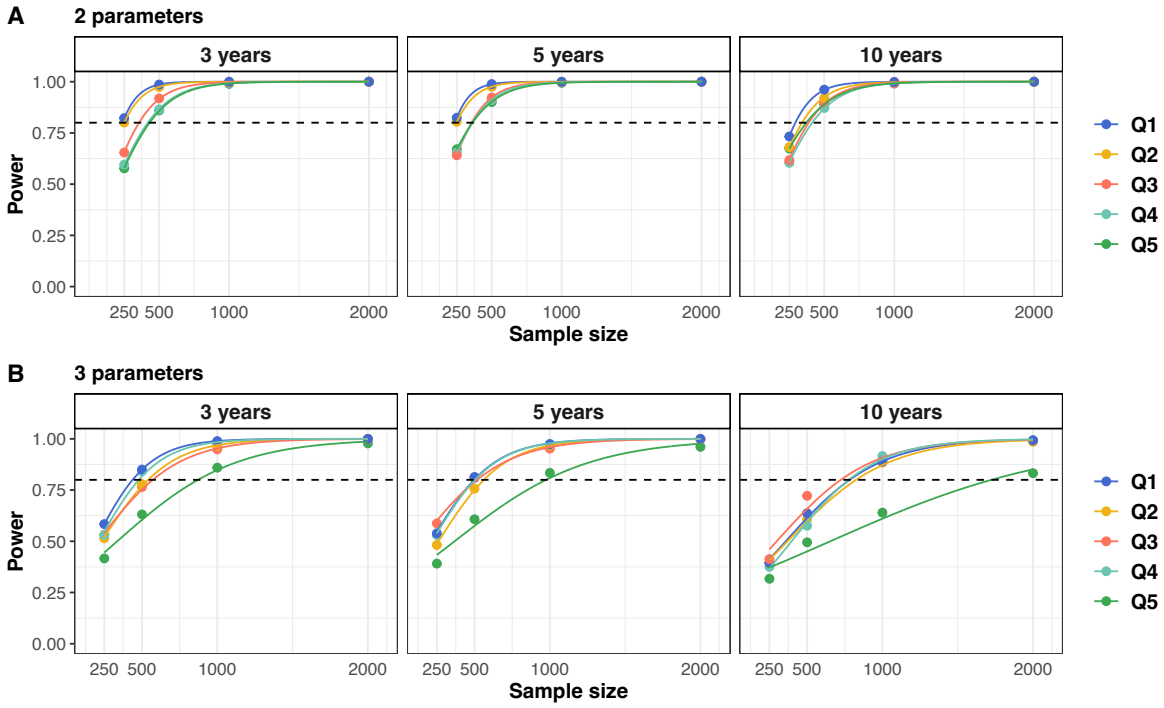
Supplementary Figure 7**.** Age-based sampling structures that yielded the bottom 20% joint precision under a single change point in malaria transmission intensity scenario. The analysis assumed a total sample size of 1,000 with low underlying SCRs (SCR = 0.03 before and 0.01 after the change point) and a low seroreversion rate (SRR = 0.017). Results were stratified by the time since the change point (3, 5, and 10 years). A) Only the SCR before and after the change point were estimated. B) The SCRs and the seroreversion rate (SRR) were jointly estimated.



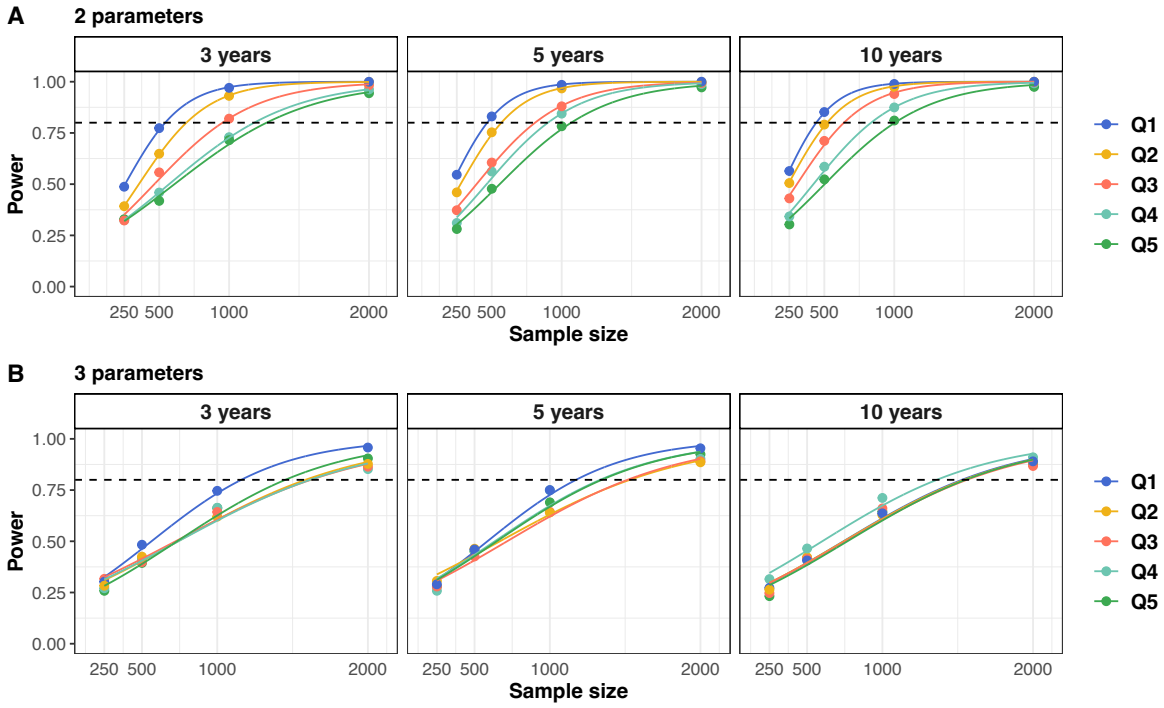
Supplementary Figure 8**.** Age-based sampling structures that yielded the top 20% joint precision under a single change point in malaria transmission intensity scenario. The analysis assumed a total sample size of 1,000 with high underlying SCRs (SCR = 0.1 before and 0.03 after the change point) and a low seroreversion rate (SRR = 0.017). Results were stratified by the time since the change point (3, 5, and 10 years). A) Only the SCR before and after the change point were estimated. B) The SCRs and the seroreversion rate (SRR) were jointly estimated.



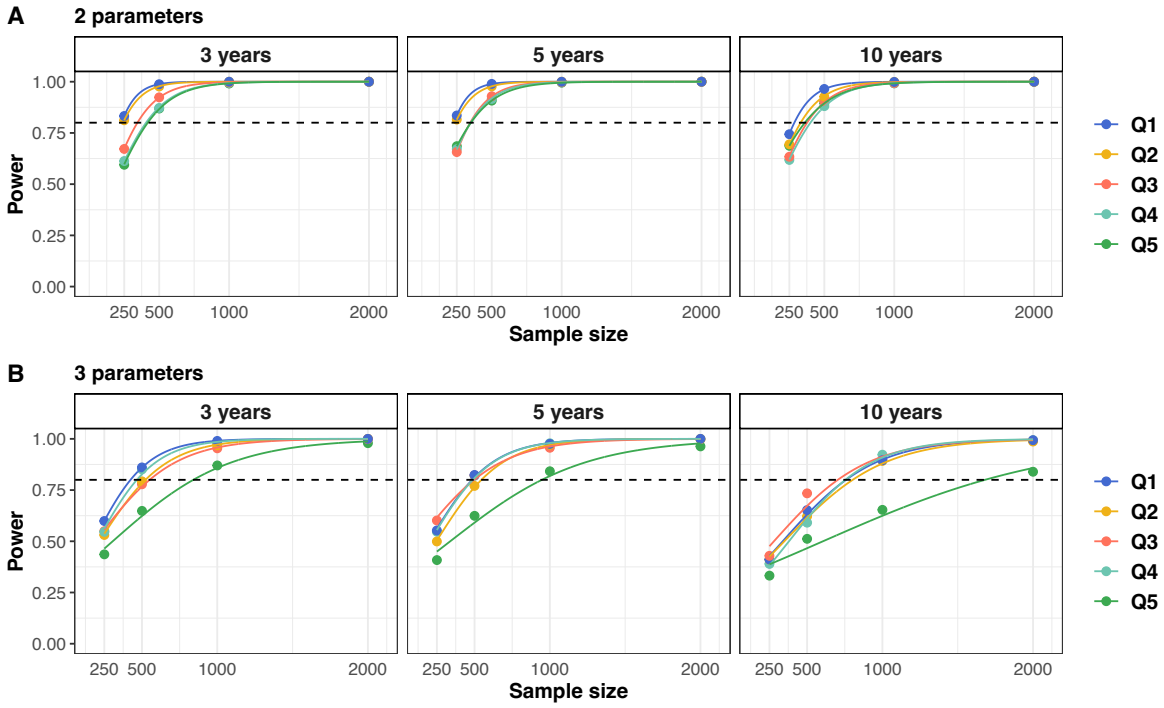
Supplementary Figure 9**.** Age-based sampling structures that yielded the bottom 20% joint precision under a single change point in malaria transmission intensity scenario. The analysis assumed a total sample size of 1,000 with high underlying SCRs (SCR = 0.1 before and 0.03 after the change point) and a low seroreversion rate (SRR = 0.017). Results were stratified by the time since the change point (3, 5, and 10 years). A) Only the SCR before and after the change point were estimated. B) The SCRs and the seroreversion rate (SRR) were jointly estimated.



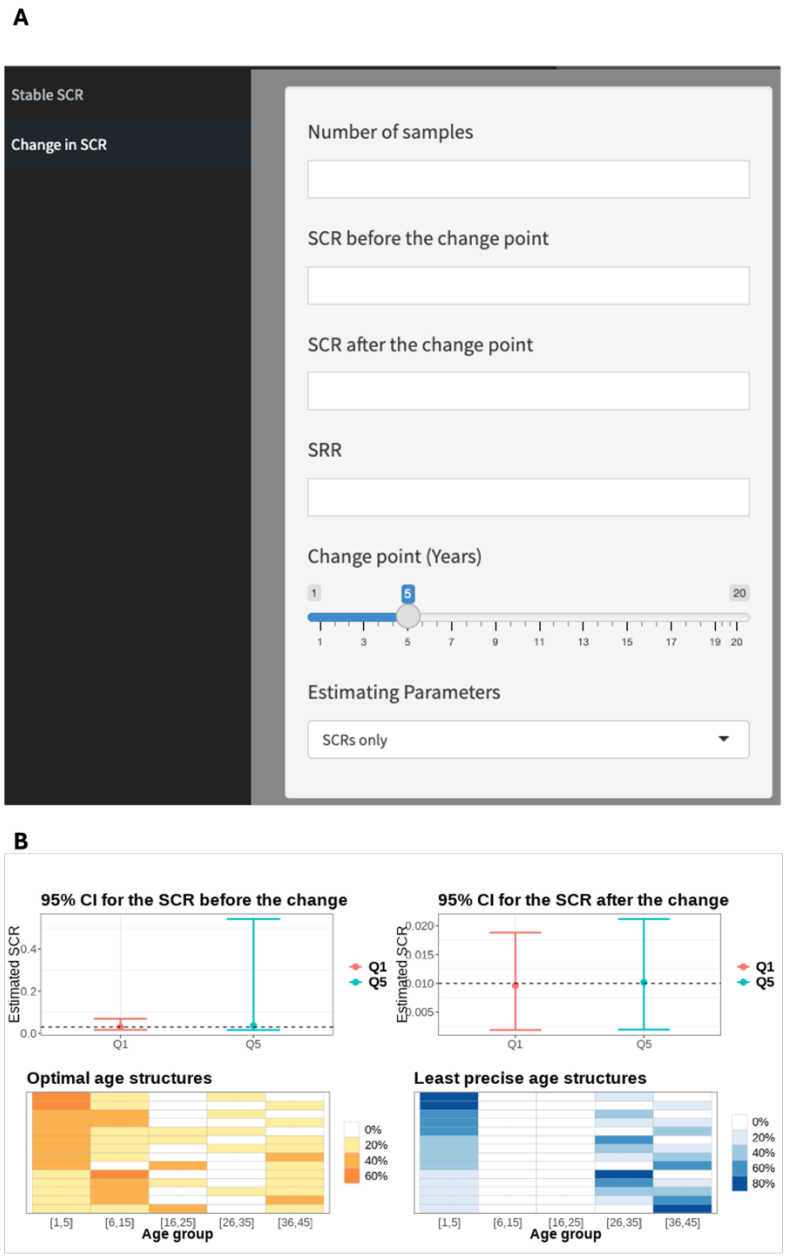
Supplementary Figure 10**.** Statistical power to detect a change in seroconversion rate in high transmission settings (SCR, 0.1 before and 0.03 after the change point) using Akaike Information Criterion (AIC)-based significance criteria. Solid lines represent the fitted logistic functions applied to the corresponding simulated results (filled circles), stratified by the quintiles of joint precsion (Q1–Q5). A) Only the SCR before and after the change point were estimated. B) The SCRs and the seroreversion rate (SRR) were jointly estimated.



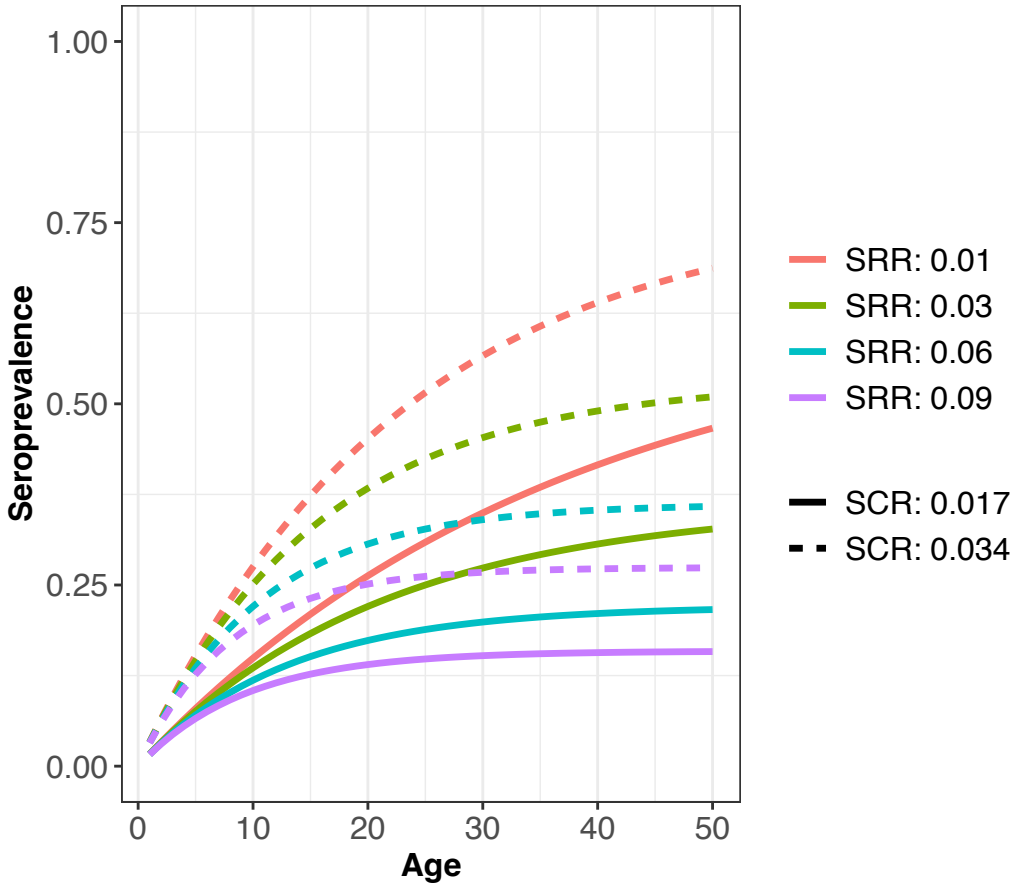
Supplementary Figure 11. Statistical power to detect a change in seroconversion rate in low transmission settings (SCR, 0.03 before and 0.01 after the change point) using likelihood ratio test (LRT)-based significance criteria. Solid lines represent the fitted logistic functions applied to the corresponding simulated results (filled circles), stratified by the quintiles of joint precsion (Q1–Q5). A) Only the SCR before and after the change point were estimated. B) The SCRs and the seroreversion rate (SRR) were jointly estimated.



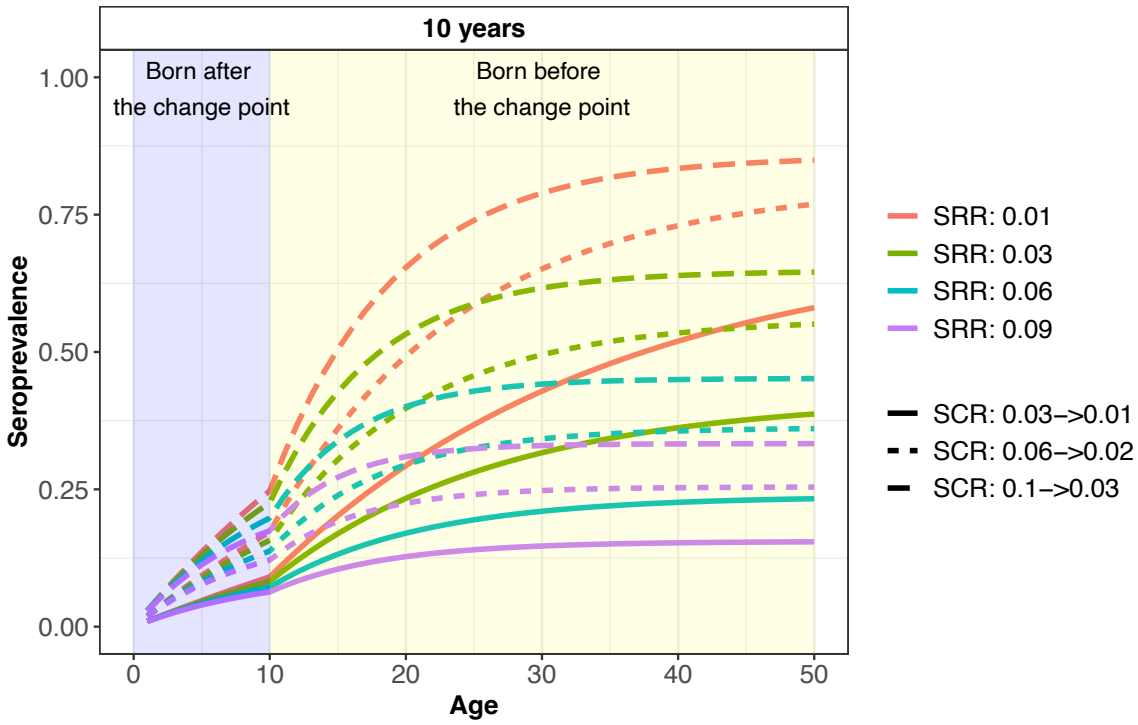
Supplementary Figure 12**.** Statistical power to detect a change in seroconversion rate in high transmission settings (SCR, 0.1 before and 0.03 after the change point) using likelihood ratio test (LRT)-based significance criteria. Solid lines represent the fitted logistic functions applied to the corresponding simulated results (filled circles), stratified by the quintiles of joint precsion (Q1–Q5). A) Only the SCR before and after the change point were estimated. B) The SCRs and the seroreversion rate (SRR) were jointly estimated.



Supplementary Figure 13**.** The user interface (A) and output (B) of the Shiny application. By specifying the scenario, values for seroconversion rate (SCR) and seroreversion rate (SRR), the change point (for the chage-in-SCR scenario), and the number of parameters to be estimated, the user can obtain the 95% CI for the SCR(s) along with the age-based sampling structures that yield the highest and lowest precision.



Supplementary Figure 14**.** Age-specific seroprevalence curves across different combinations of seroconversion rate (SCR) and seroreversion rate (SRR), as estimated by the reverse catalytic model (RCM).



Supplementary Figure 15**.** Age-specific seroprevalence curves across different combinations of seroconversion rate (SCR) before and after the change point (10 years before the sampling) and seroreversion rate (SRR), as estimated by the reverse catalytic model (RCM) with a single change point in the SCR.