Statistical and mathematical modelling for sero-epidemiological data of tropical infectious diseases

Nuno Sepúlveda, nuno.sepulveda@pw.edu.pl

Faculty of Mathematics & Information Science, Warsaw University of Technology

CEAUL - Centro de Estatística e Aplicações da Universidade de Lisboa

POLITECHNIKA WARSZAWSKA

WARSAW UNIVERSITY OF TECHNOLOGY

CEAUL

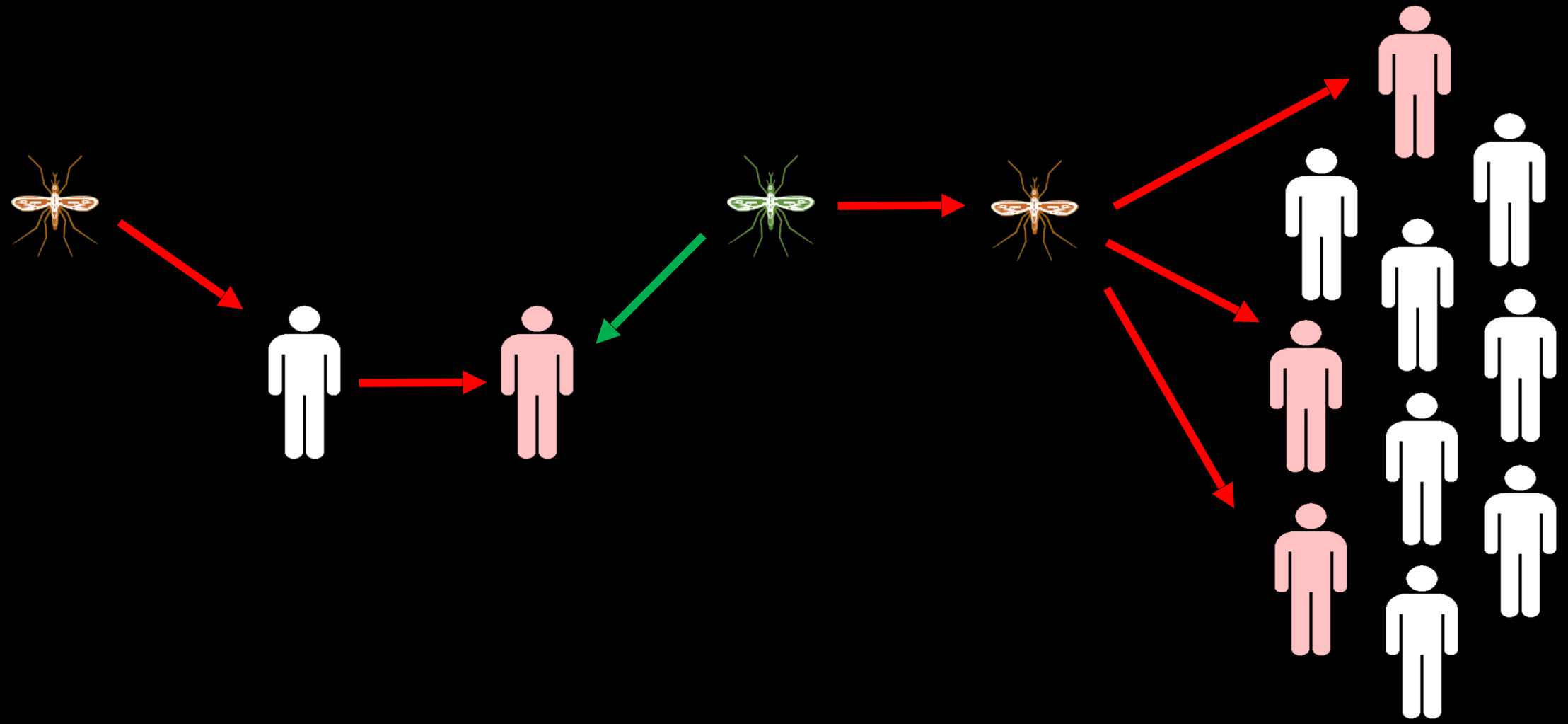
Centro de Estatística e Aplicações
Universidade de Lisboa

Workshop content

1. Defining seropositivity (using two-Gaussian mixture models) and estimating seroprevalence
2. Estimating seroconversion rate (using reversible catalytic models)
3. Calculating sample size for controlling precision of seroconversion rate

GitHub:
https://github.com/immune-stats/Workshop_Malaria_PALOP_2025

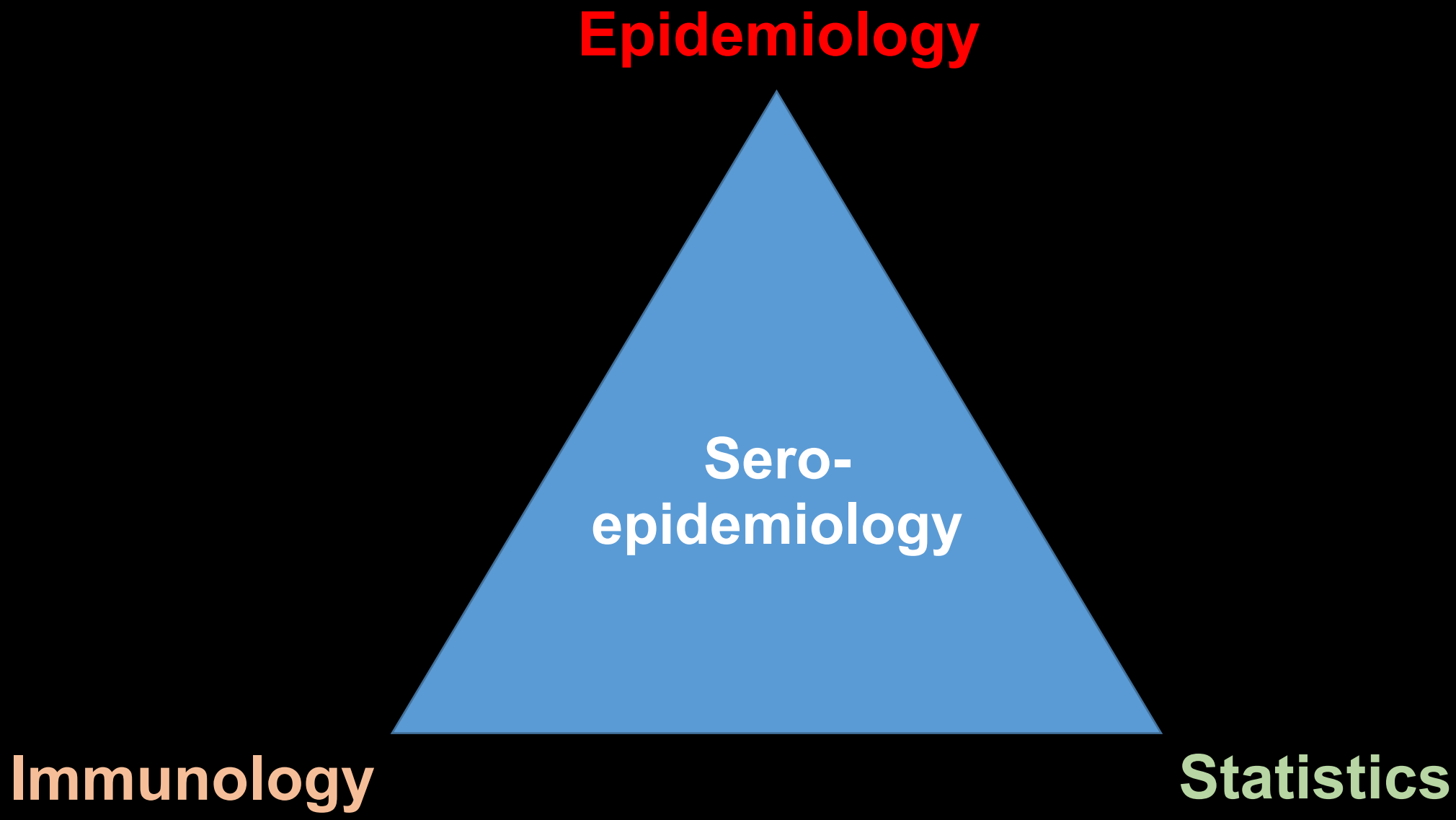
How to measure malaria transmission?



How to measure malaria transmission?

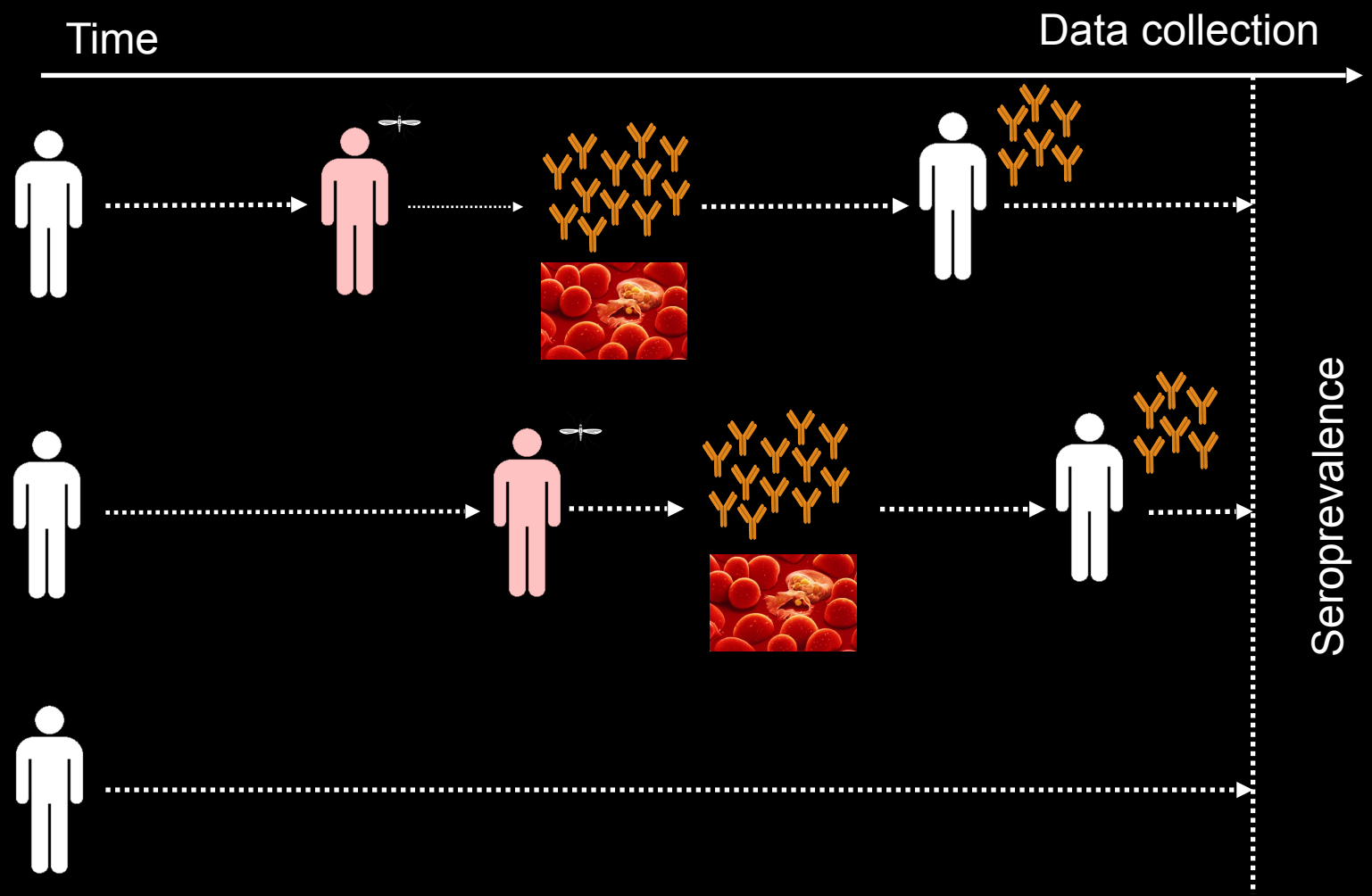
1. Prevalence of infection or parasite rate (non-informative when disease transmission intensity is low)
2. Entomological inoculation rate (trick to estimate)
3. Seroprevalence (prevalence of exposure)
4. Seroconversion rate (proxy of transmission intensity)

Seroepidemiology

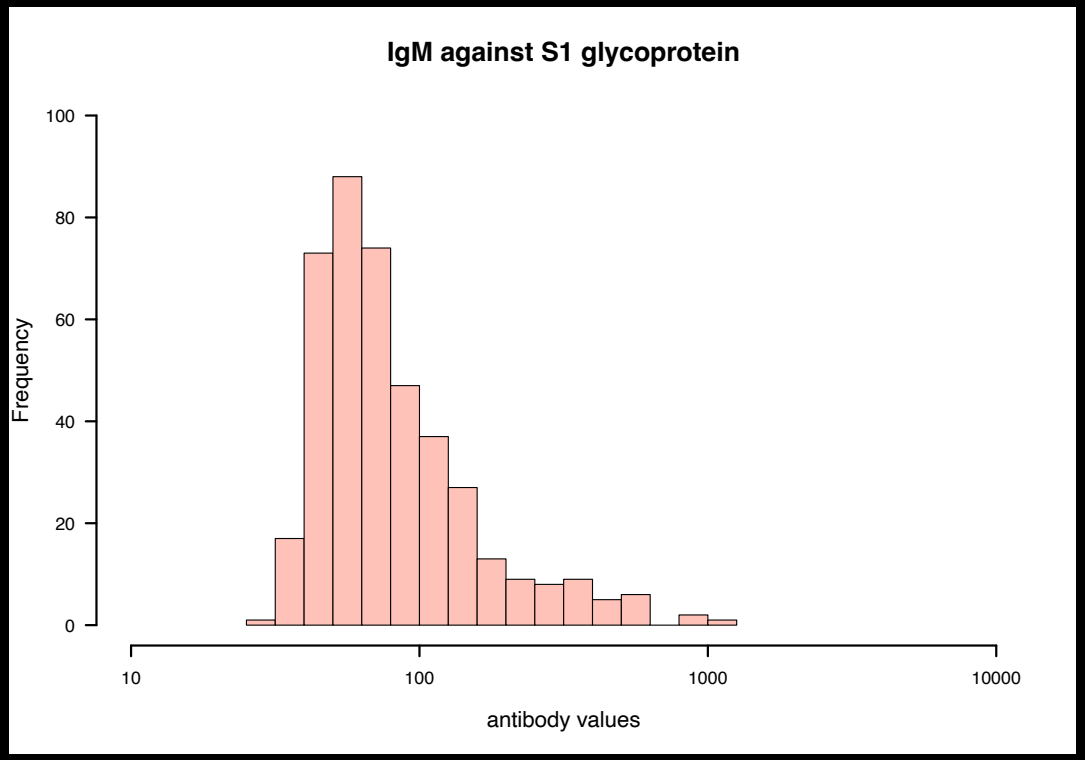

Epidemiology
Sero-
epidemiology
Immunology
Statistics

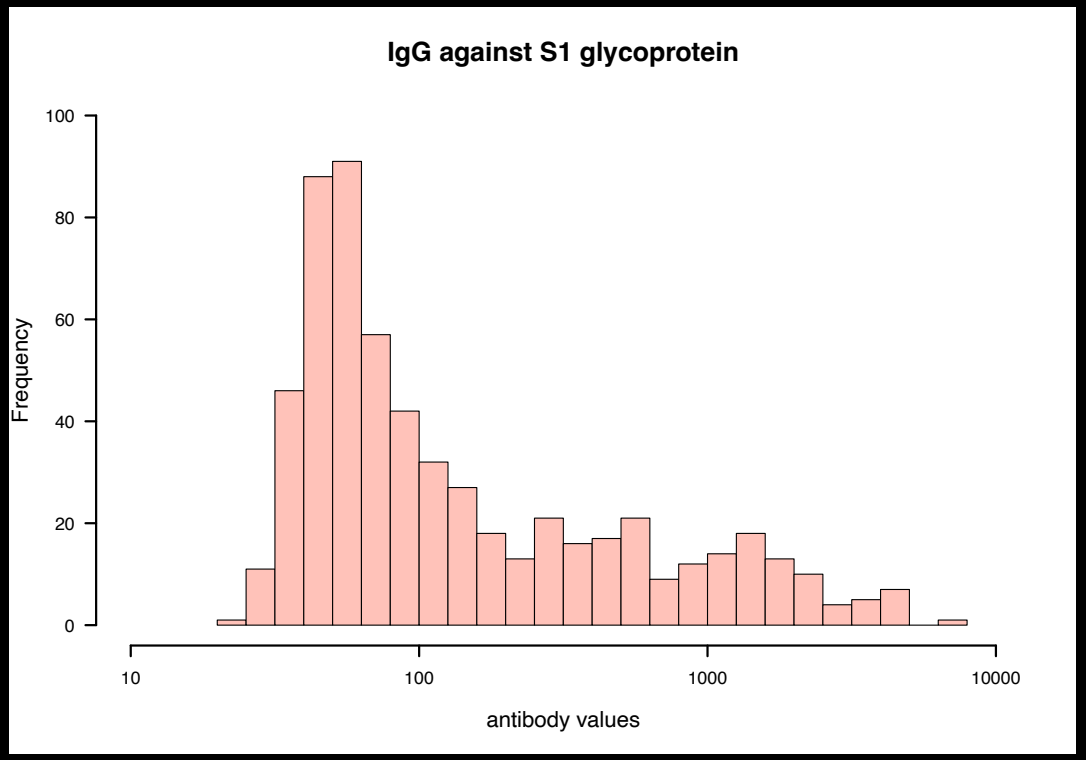
1.
Defining seropositivity (using two-Gaussian mixture models) and estimating seroprevalence

Basic principle



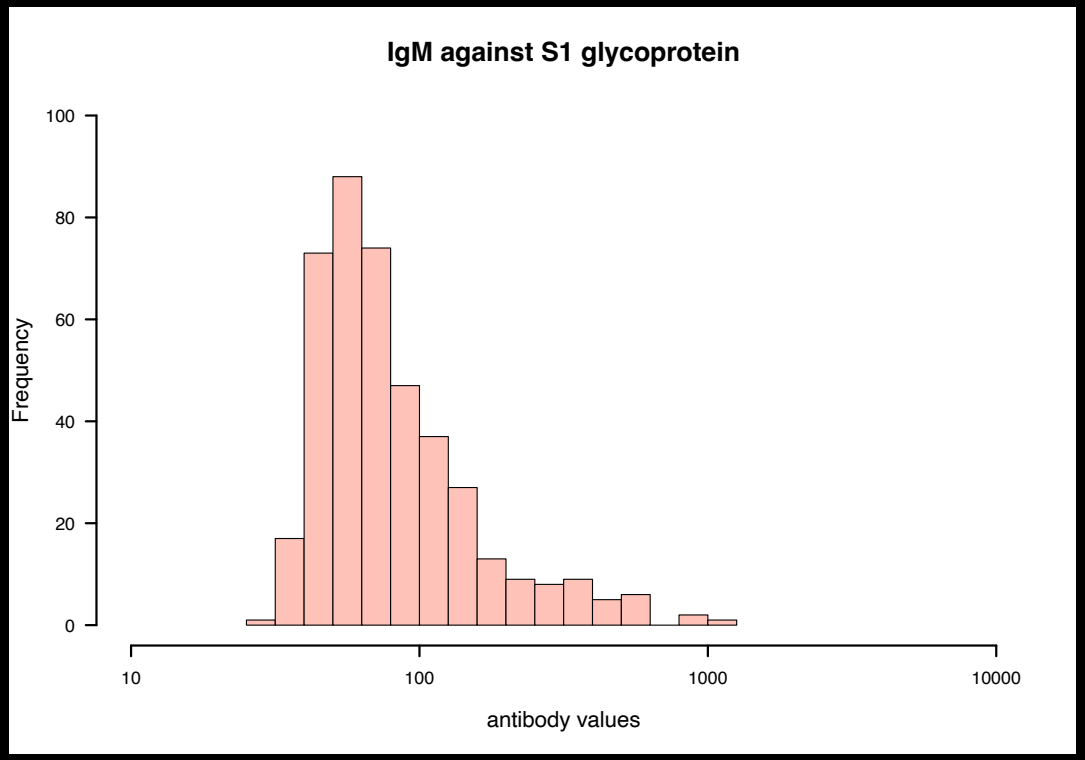
Antibody data are intrinsically quantitative

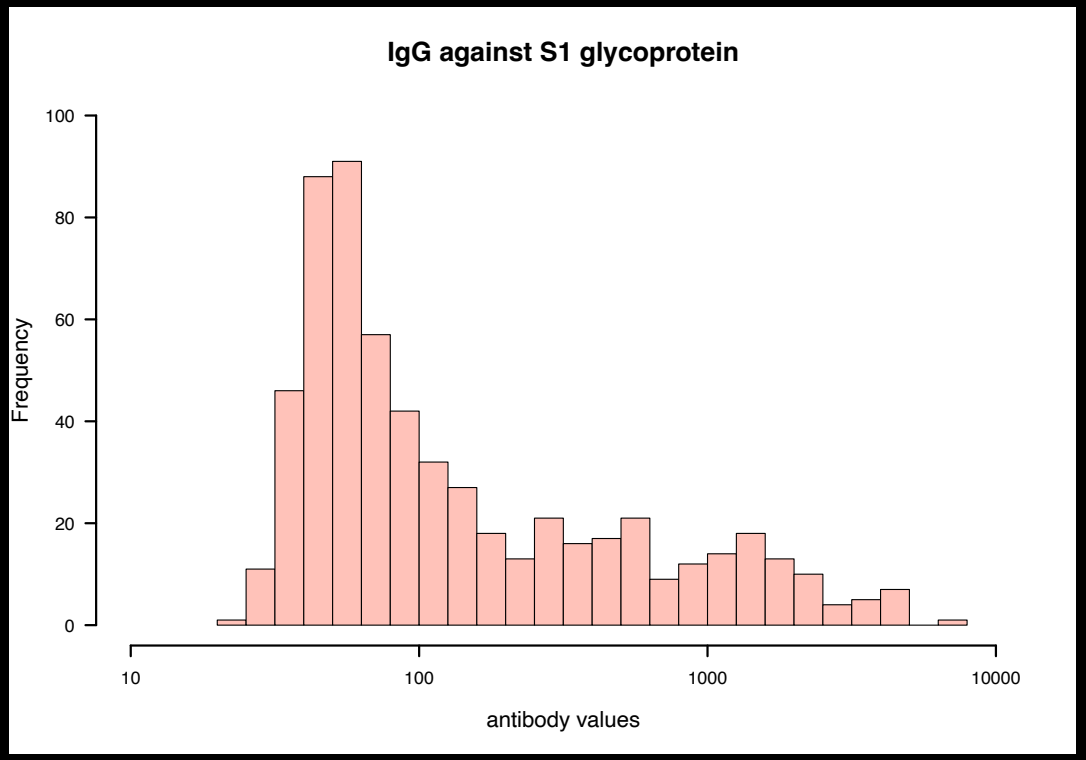




Rosado et al (2020). Serological signatures of SARS-CoV-2 infection: Implications for antibody-based diagnostics. medRxiv 2020.05.07.20093963.

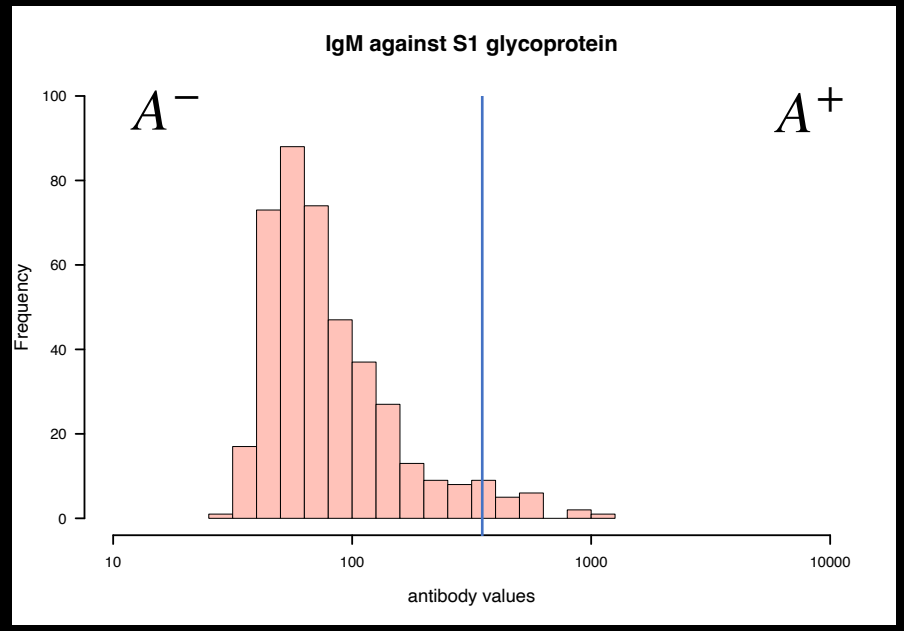
Who are the seropositive individuals?

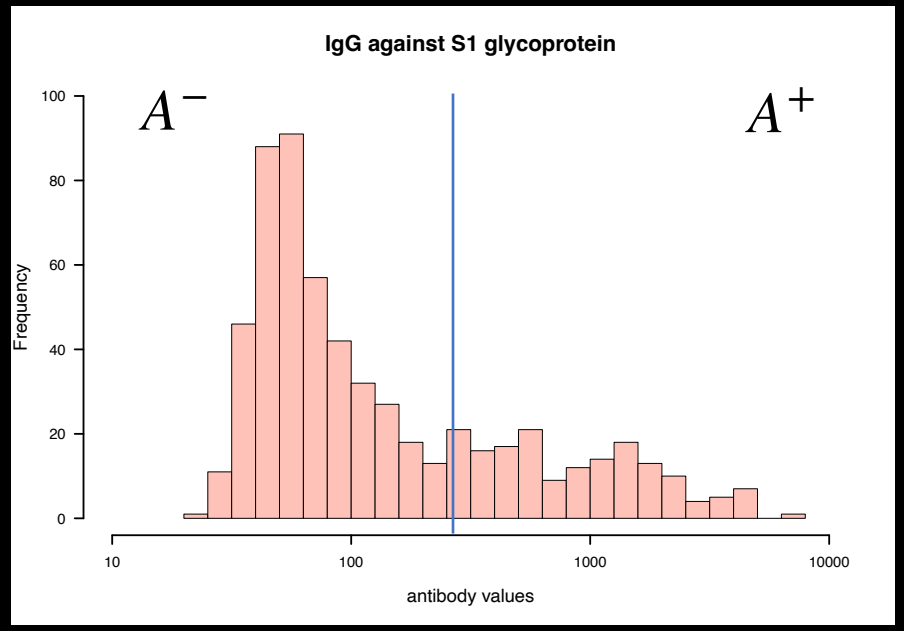


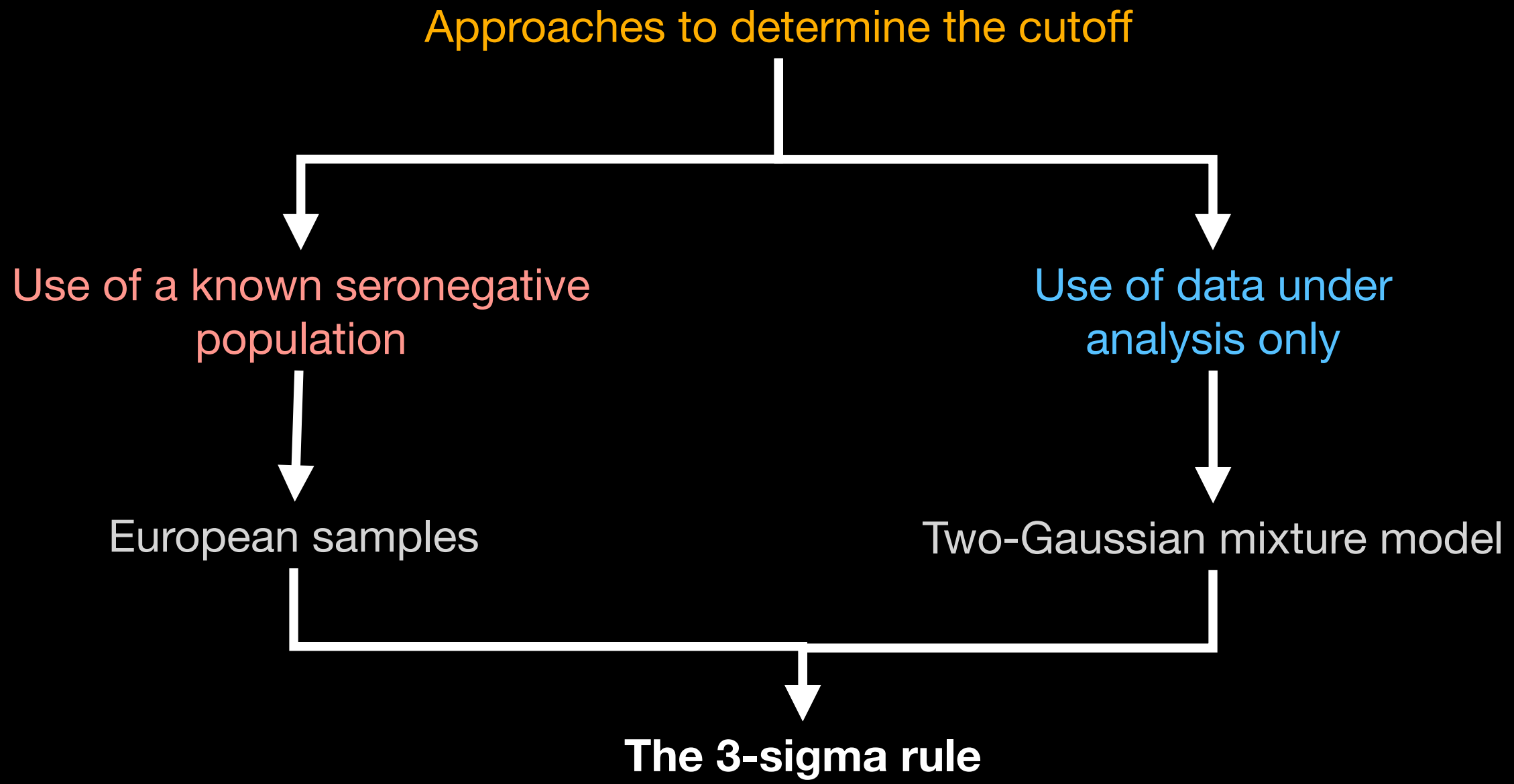


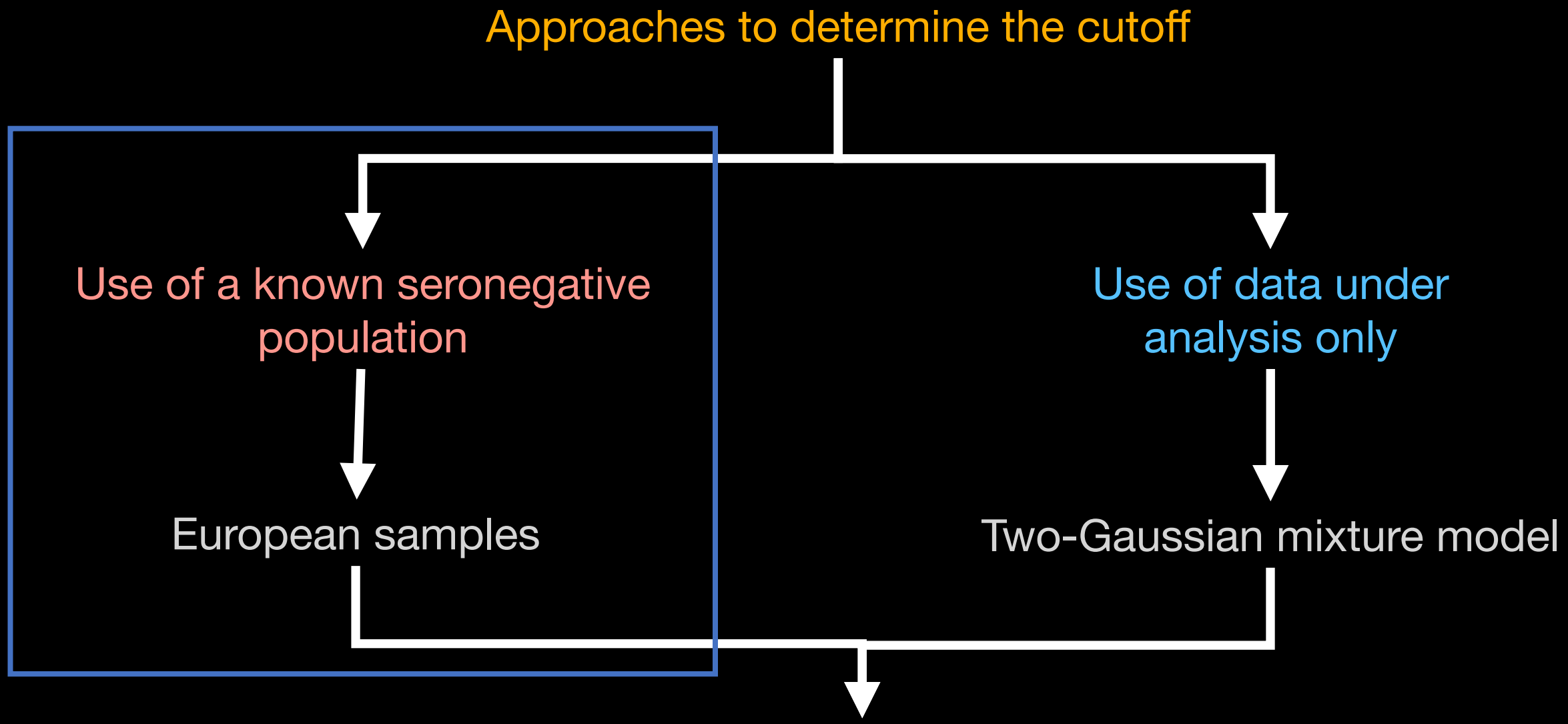
Rosado et al (2020). Serological signatures of SARS-CoV-2 infection: Implications for antibody-based diagnostics. medRxiv 2020.05.07.20093963.

How to determine the cut-off?






Approaches to determine the cutoff
Use of a known seronegative population
Use of data under analysis only
European samples
Two-Gaussian mixture model
The 3-sigma rule


Approaches to determine the cutoff
Use of a known seronegative population
Use of data under analysis only
European samples
Two-Gaussian mixture model
The 3-sigma rule

The 3-sigma rule

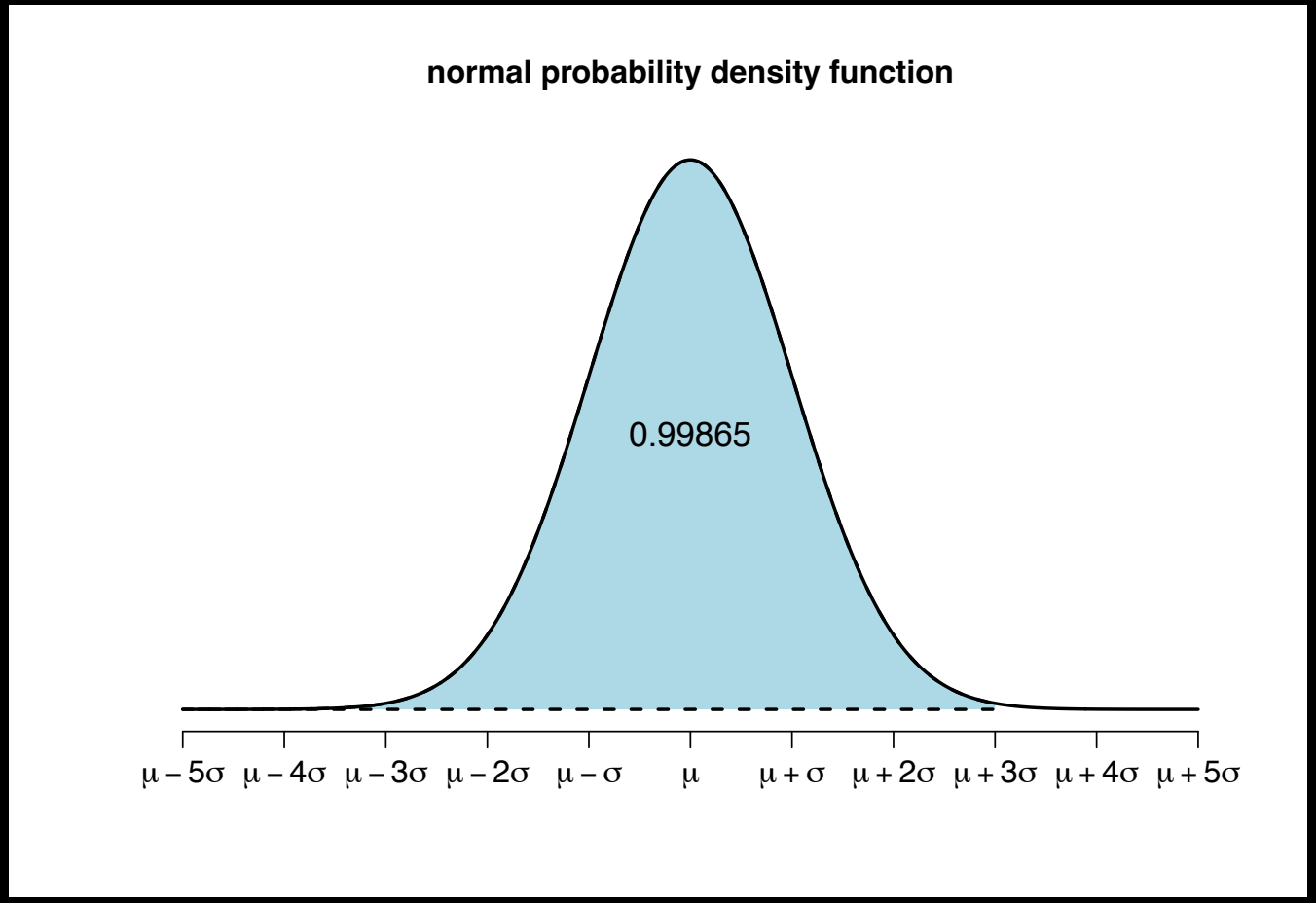
$$\mu_{A^-} = E\left[X|A^-\right]$$

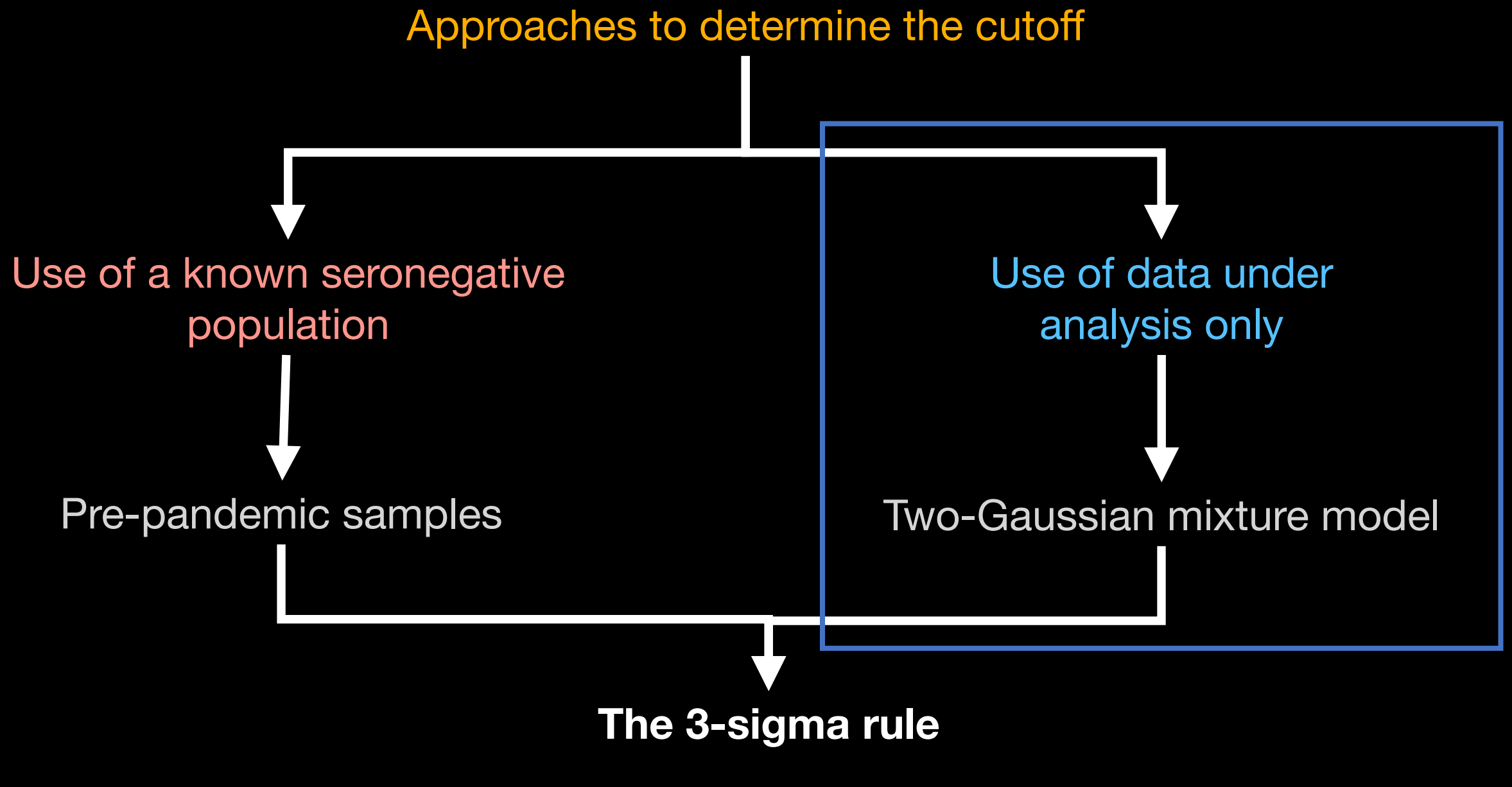
$$\sigma_{A^-} = \sqrt{Var\left[X|A^-\right]}$$

Seronegative, if $X_i \leq \mu_{A^-} + 3\sigma_{A^-}$

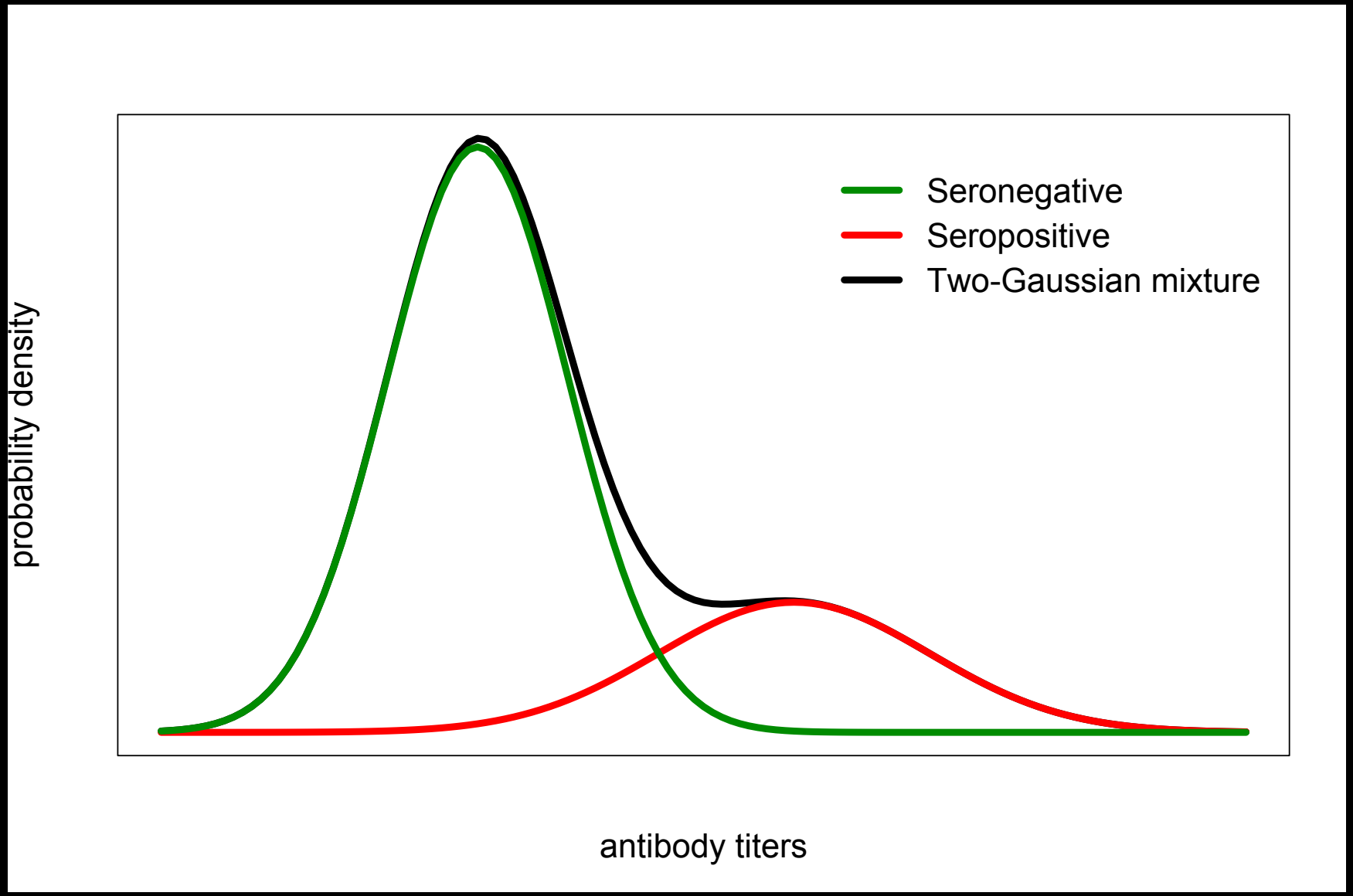
Seropositive, otherwise

The link to the Normal distribution




Approaches to determine the cutoff
Use of a known seronegative population
Use of data under analysis only
Pre-pandemic samples
Two-Gaussian mixture model
The 3-sigma rule

Two-Gaussian mixture model



Two-Gaussian mixture models

$$f_X(x) = (1-\pi) f_{N(\mu_{S^-},\sigma_{S^-})}(x) + \pi f_{N(\mu_{S^+},\sigma_{S^+})}(x)$$

Definition of $S^- \Rightarrow \mu_{S^-} < \mu_{S^-}$

In general:

$$f_X(x) = \sum_{i=1}^{k} \pi_i f_{N(\mu_i,\sigma_i)}(x) \qquad \text{where} \sum_{i=1}^{k} \pi_i = 1$$

Estimation of the model by maximum likelihood method

EM (Expectation-Maximization) Algorithm

1. Start with initial estimates for the parameters
2. E-Step - calculate the probability of each individual belonging to a given subpopulation according to estimates at 1.
3. M-Step - re-estimate the parameters using these probabilities and repeat the E-step with these new estimates
4. Stop with the increment in the log-likelihood is below a given tolerance error.

Calculate the cutoff for seropositivity according to $\hat{\mu}_{S^-}$ and $\hat{\sigma}_{S^-}$

Package mixtools

time!

OPEN ACCESS Freely available online

PLoS one

Serological Markers Suggest Heterogeneity of Effectiveness of Malaria Control Interventions on Bioko Island, Equatorial Guinea

Jackie Cook[1], Immo Kleinschmidt[2], Christopher Schwabe[3], Gloria Nseng[4], Teun Bousema[1], Patrick H. Corran[1], Eleanor M. Riley[1], Chris J. Drakeley[1]*

1 Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom, **3** Medicinal Care Development International, Silver Spring, Maryland, United States of America, **4** Ministry of Health and Social Welfare, Malabo, Equatorial Guinea

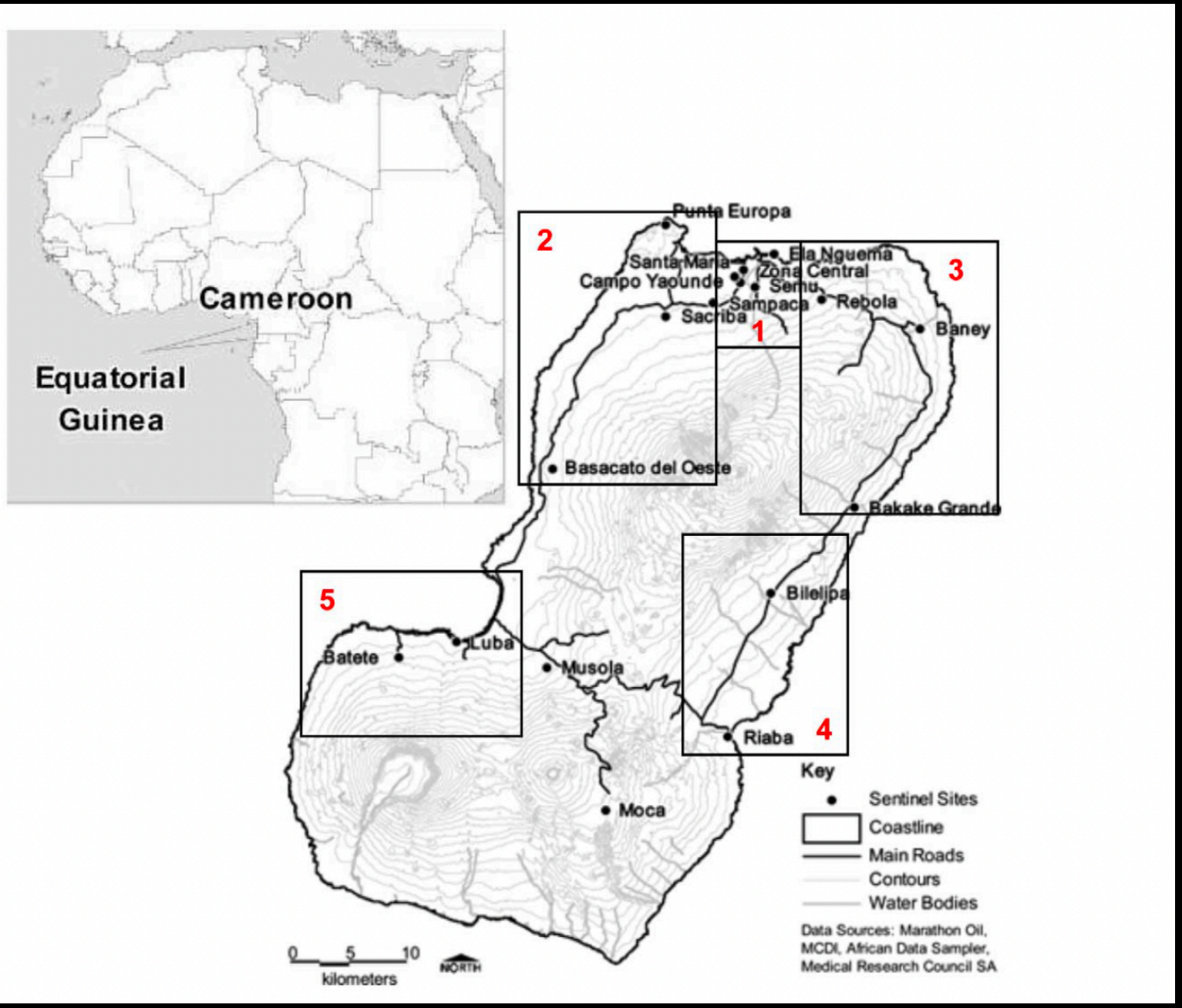

Cameroon
Equatorial Guinea
Punta Europa
2
Santa Maria
Ela Nguema
Zona Central
Campo Yaounde
Sendu
Sampaca
Rebola
3
Sacriba
1
Baney
Basacato del Oeste
Bakake Grande
5
Bilelipa
Batete
Luba
Musola
Riaba
4
Moca
Key
Sentinel Sites
Coastline
Main Roads
Contours
Water Bodies
Data Sources: Marathon Oil, MCDI, African Data Sampler, Medical Research Council SA
0 5 10
kilometers
NORTH

Table 1. Demographic characteristics of the study population.

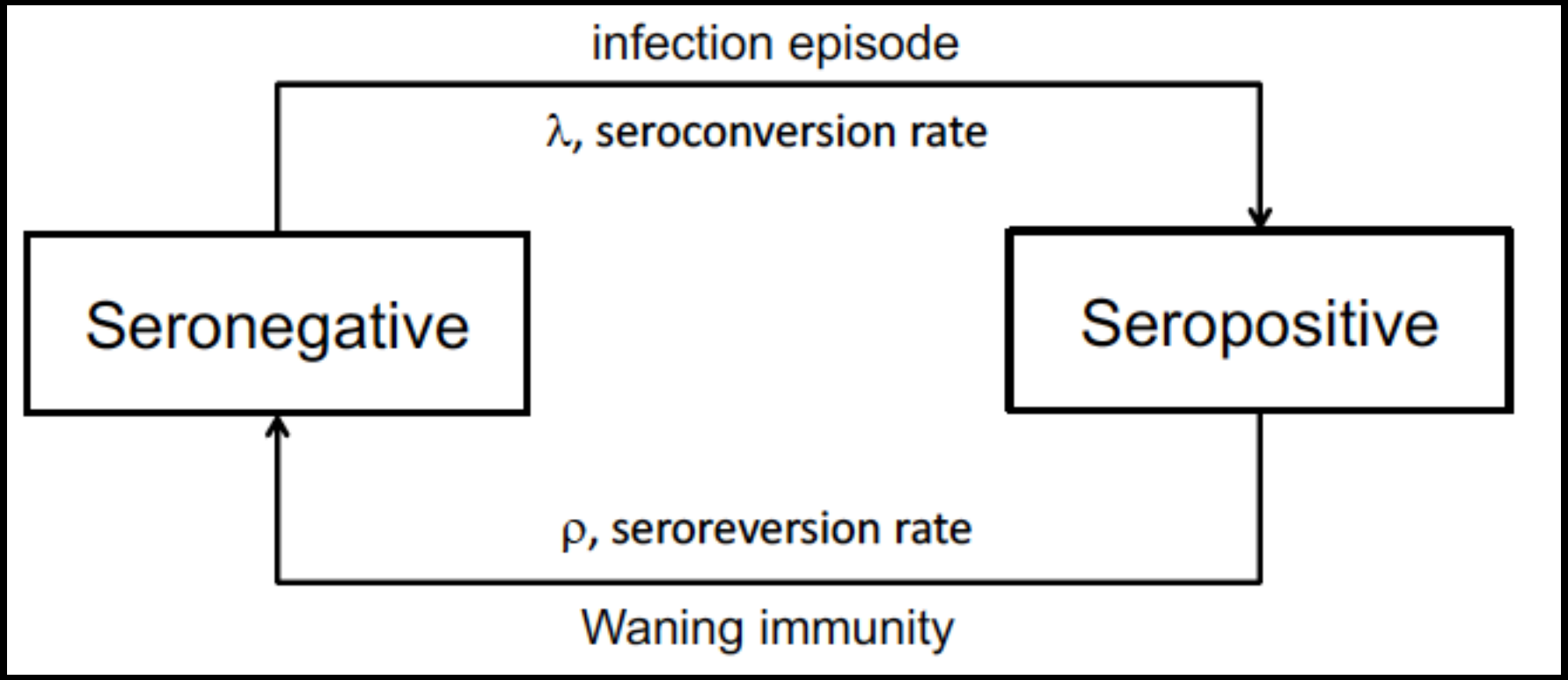
		% [n]						
		Malabo N = 2328	**North West N = 1749**	**North East N = 1323**	**South East N = 700**	**South West N = 588**	**Other** N = 699**	**Total N = 7387**
Age (years)	0–1	14.1 [324]	10.5 [182]	10.4 [137]	12.2 [85]	10.0 [58]	7.5 [52]	11.4 [838]
	1–5	21.1 [458]	18.0 [312]	19.8 [261]	14.6 [102]	16.8 [97]	15.2 [106]	18.6 [1363]
	5–15	26.3 [605]	30.6 [531]	30.1 [396]	21.0 [146]	24.0 [139]	28.7 [200]	27.5 [2017]
	15–90	38.6 [890]	41.0 [712]	39.8 [524]	52.2 [364]	49.2 [285]	48.6 [338]	42.5 [3113]
Sex	Female	61.2 [1410]	54.2 [932]	61.1 [805]	55.8 [389]	58.4 [338]	54.8 [382]	58.2 [4256]
House recently sprayed[1]	Yes	74.2 [1580]	81.2 [1306]	85.6 [1076]	81.7 [519]	89.5 [477]	87.9 [574]	81.2 [5532]
Slept under ITN[2]	Yes	82.6 [1629]	68.0 [988]	65.8 [797]	63.3 [404]	73.1 [385]	71.4 [449]	72.4 [4652]
Parasite positive	Yes	14.8 [300]	27.0 [374]	7.9 [94]	21.7 [135]	18.6 [97]	12.1 [75]	16.9 [1075]

[1]- within the previous 6 months.
[2]- on the night before the survey.
**Moca and Musola kept separate due to their high altitude.
doi:10.1371/journal.pone.0025137.t001

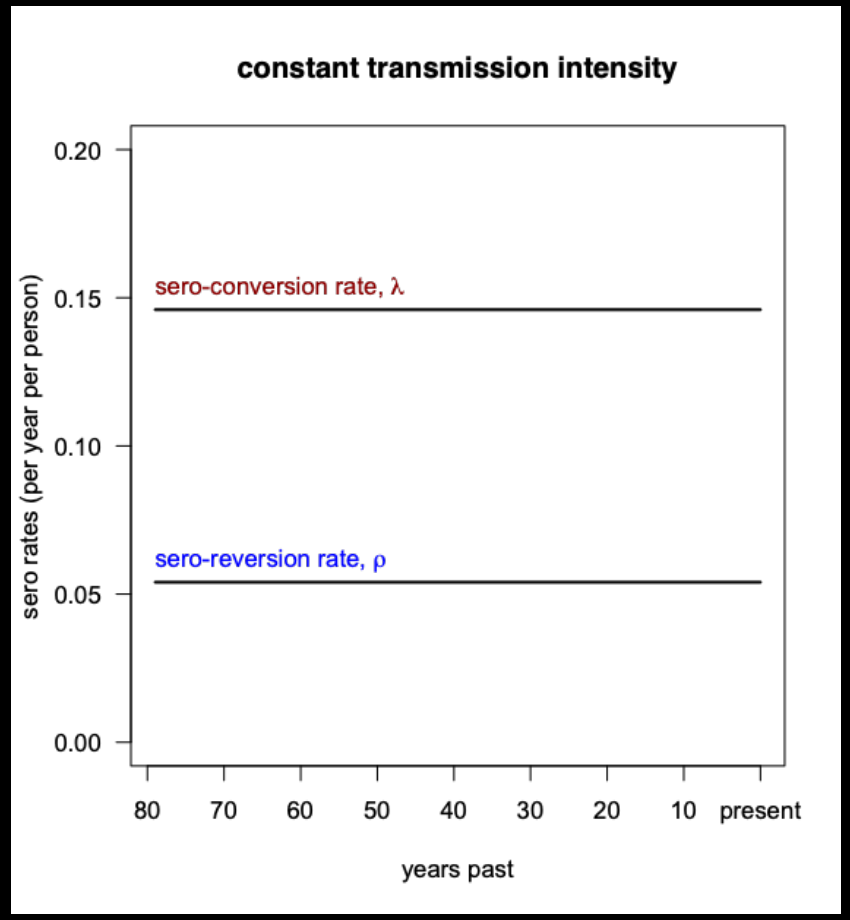
Variables

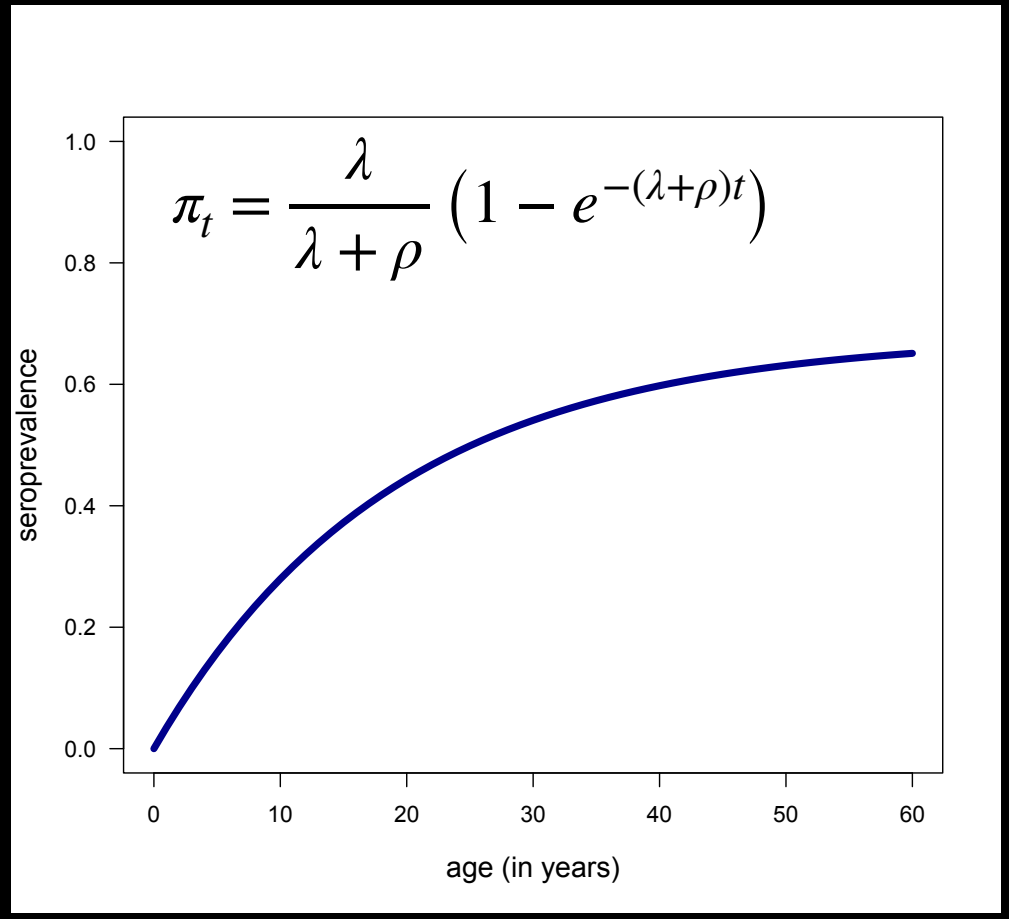
2.
Estimating seroconversion rate (using reversible catalytic models)

Reversible catalytic models

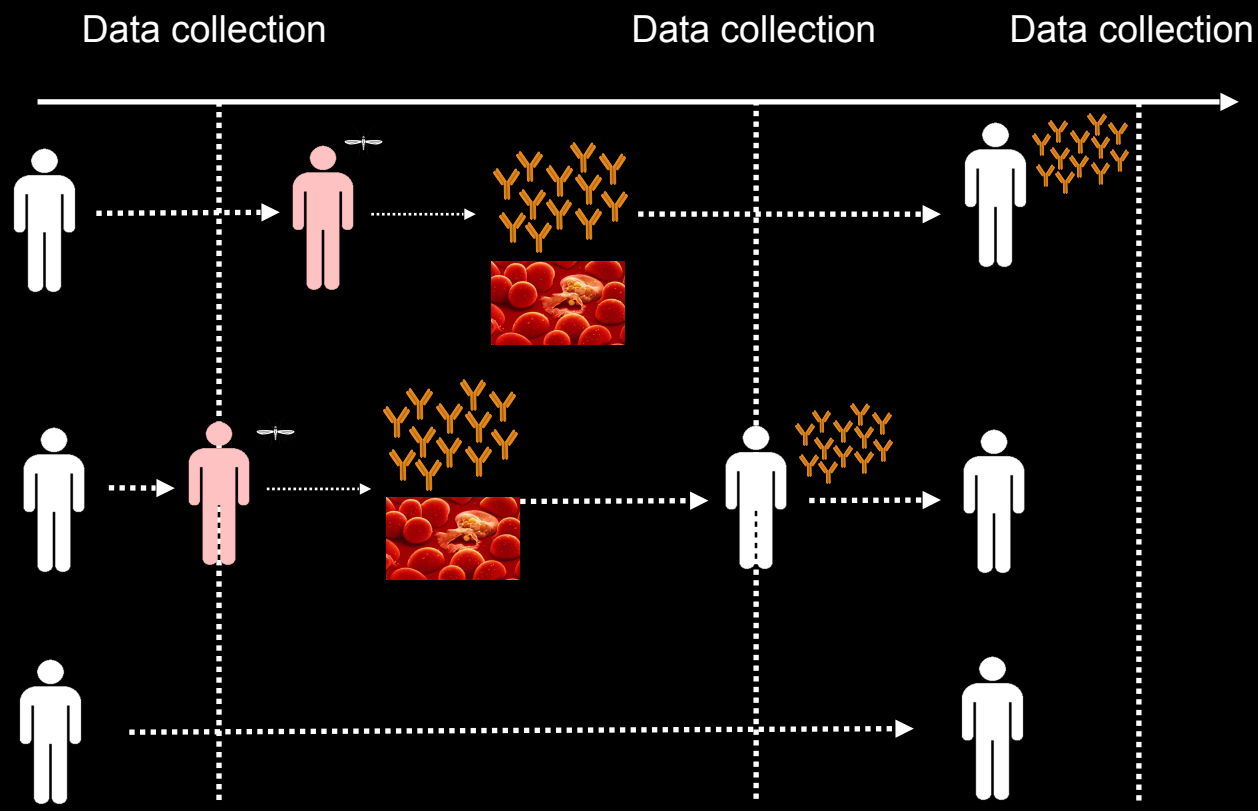


Constant transmission intensity





Longitudinal surveys



Statistical information: ++++

Direct observation of serological transitions

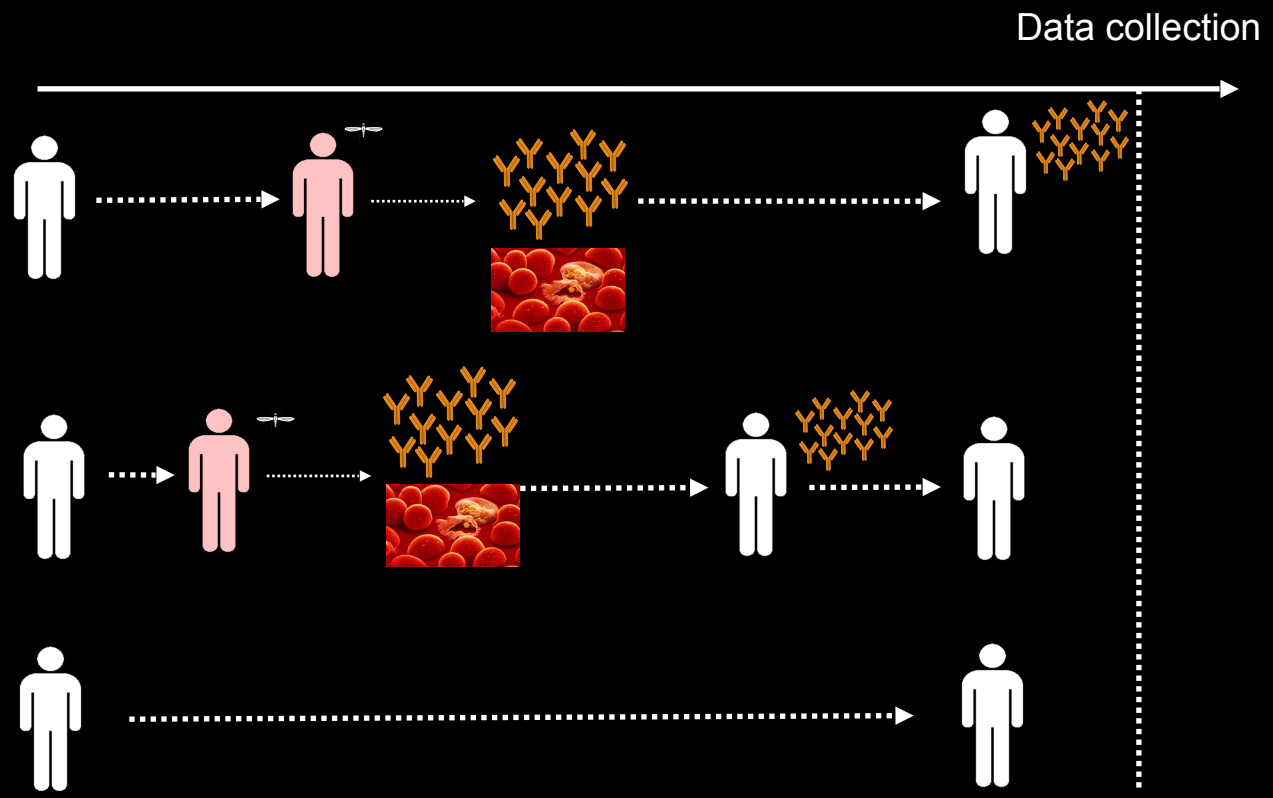
Execution difficulty: ++++

Time consuming

Sampling intensive

Participation adherence/drop-outs

Cross-sectional surveys



Statistical information: ++

No direct observation of serological transitions
Age as proxy of time

Execution difficulty: ++

Easy to engage participation
Quick sampling

Longitudinal versus cross-sectional surveys

Type of Study	Seroconversion rate	Seroreversion rate
Longitudinal	**0.021** (0.001-0096)	**0.163** (0.001,0.729)
Cross-sectional	**0.023** (0.001,0.052)	**0.0001** (0.001,0.255)

Seroreversion rate is difficult to be estimated!
So it is often fixed at 0 or an good estimate, say 0.017.

Arnold et al (2014). PLoS One

Fixed seroreversion rate at 0

$$\rho = 0 \Rightarrow \pi_t = 1 - e^{-\lambda t}$$

$$\Rightarrow \log 1 - \pi_t = -\lambda t$$

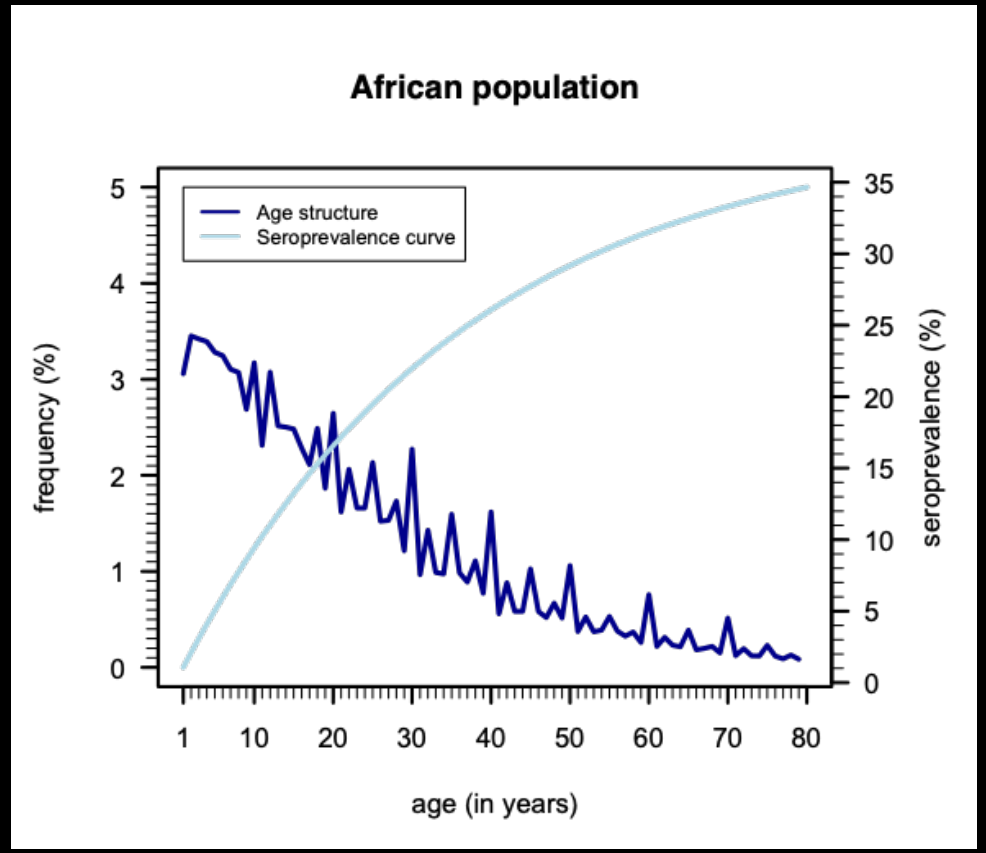
$$\Rightarrow \log(-\log(1 - \pi_t)) = \log \lambda + \log t$$

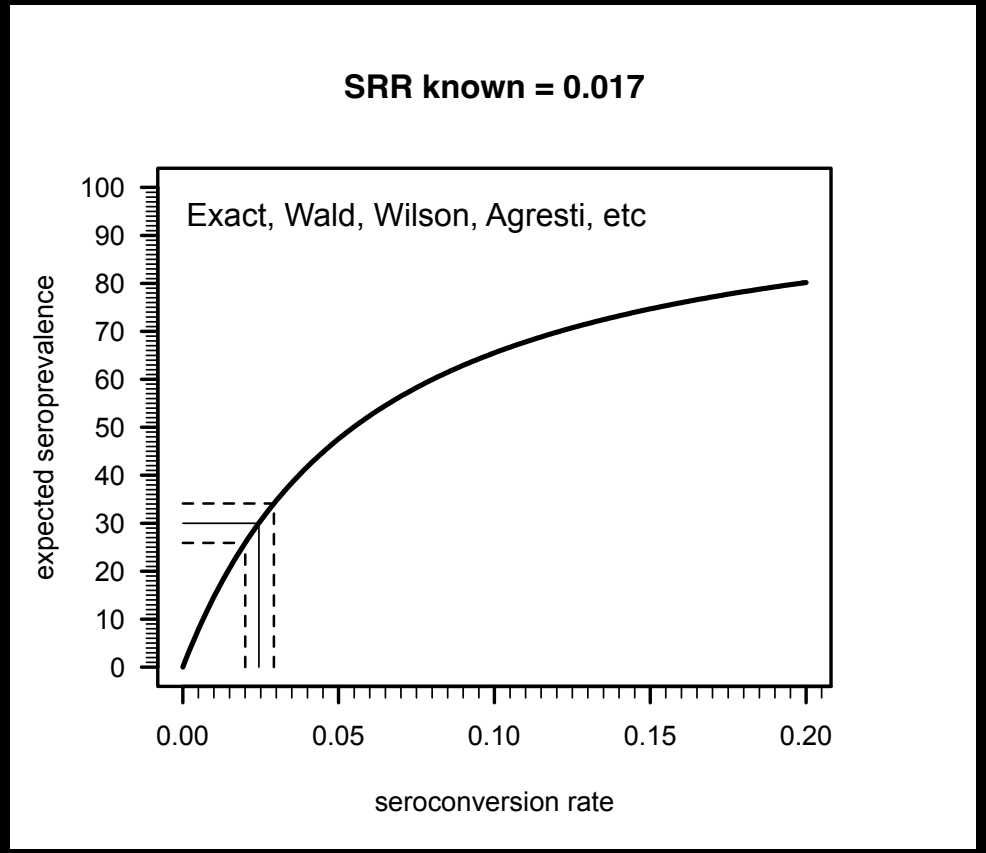
Complementary log-log model

time!

3.
Calculating sample size for controlling precision of seroconversion rate estimate

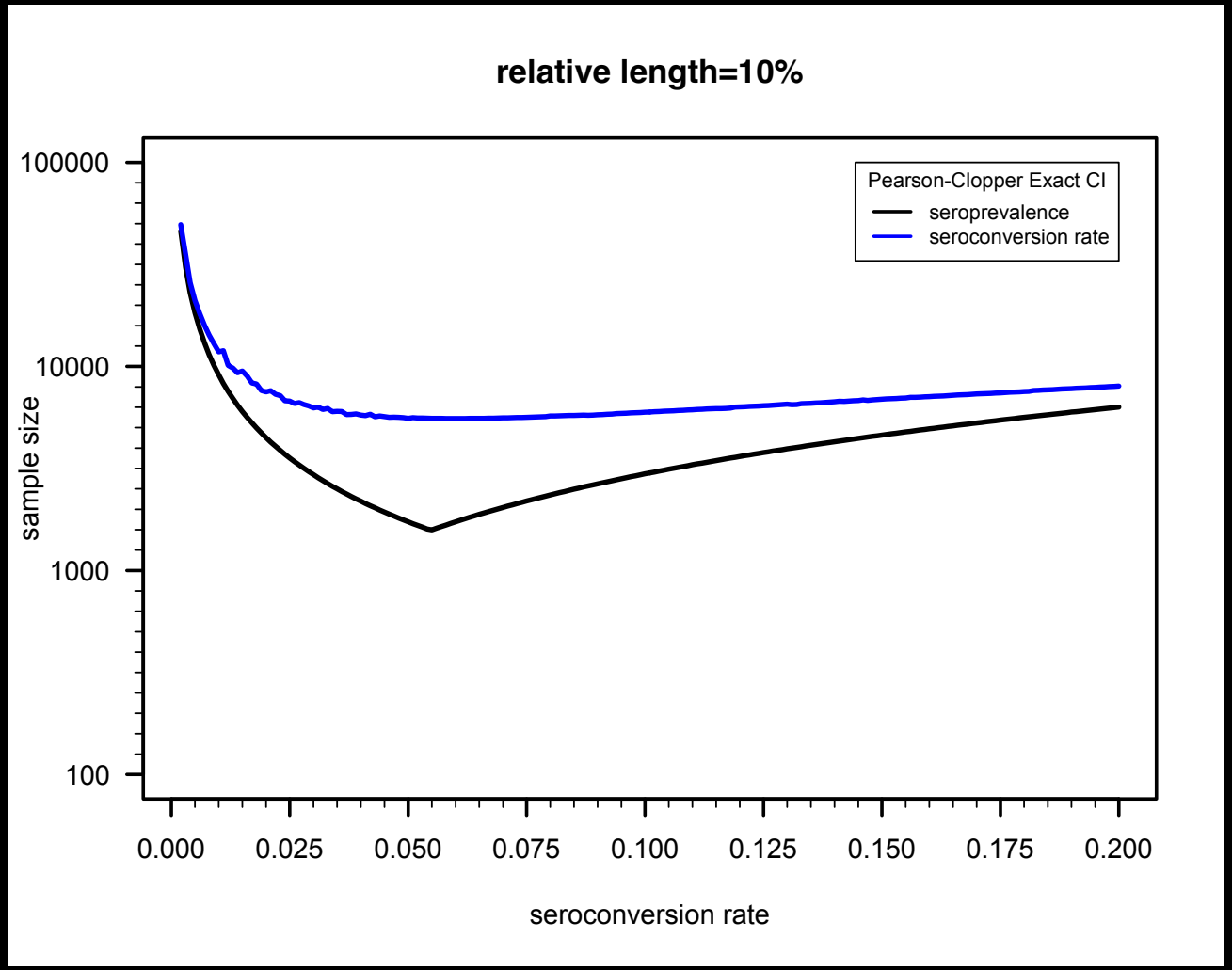
Sample size calculation for seroconversion rate





Sepúlveda et al (2015). Malaria Journal

Sample size calculation for seroconversion rate



Sample size calculation in practice

1. Desired precision
2. Antibody with known seroreversion rate
3. Transmission intensity of the population
4. Age structure associated with sampling scheme
5. Type of confidence interval to be used

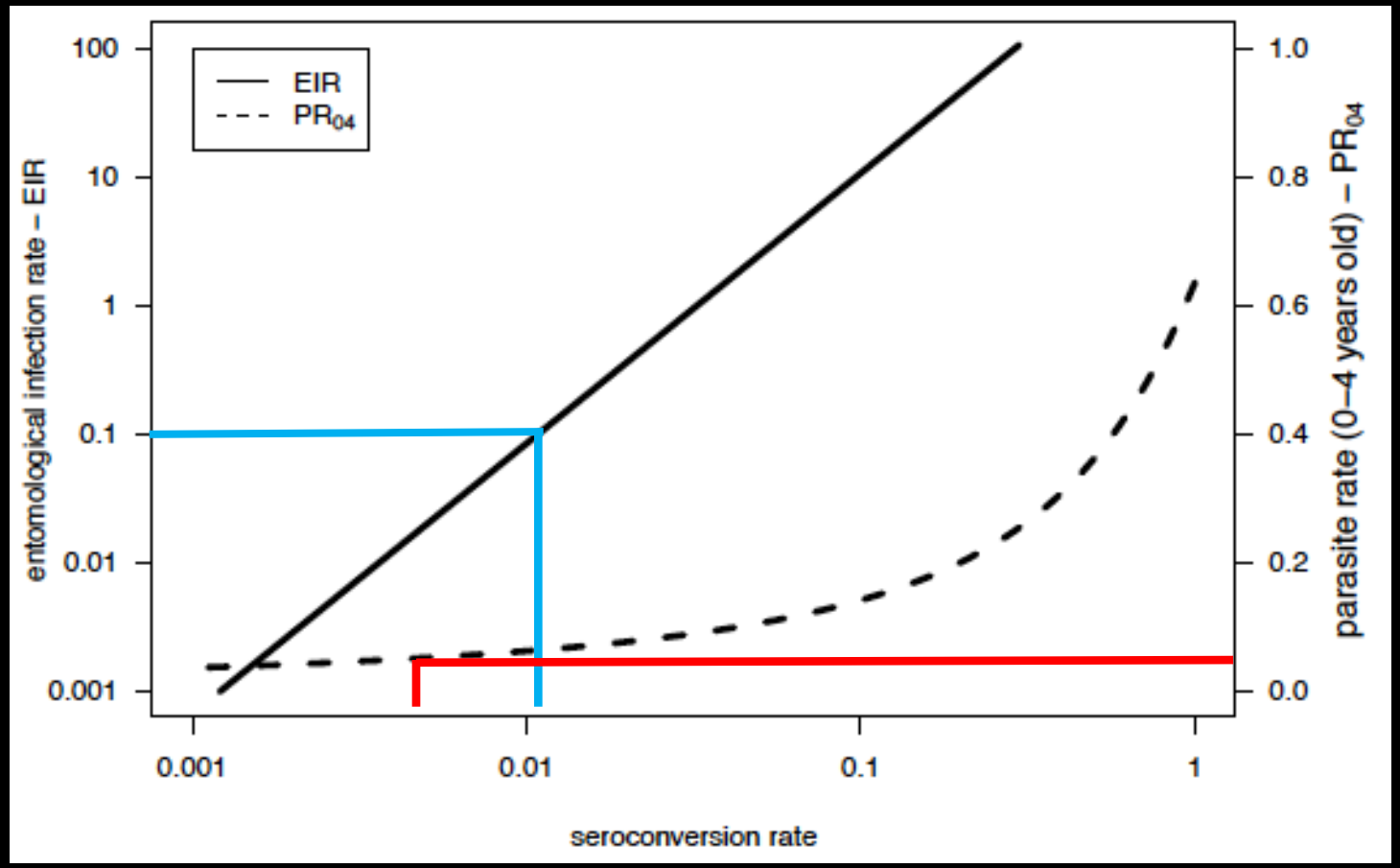
Identification of transmission intensity

Table 1 Expected relationship between EIR, PR_{04}, SCR and SP in African (AFR), Southeast Asian and South American (SEA + SA) populations where seroreversion rate was fixed at 0.017

			Seroprevalence	
EIR	PR_{04}	SCR	AFR	SEA + SA
0.01	0.050	0.0036	0.057	0.073
0.10	0.073	0.0108	0.156	0.195
1.00	0.119	0.0324	0.365	0.437
10.0	0.231	0.0969	0.647	0.720
100.0	0.625	0.2900	0.860	0.896

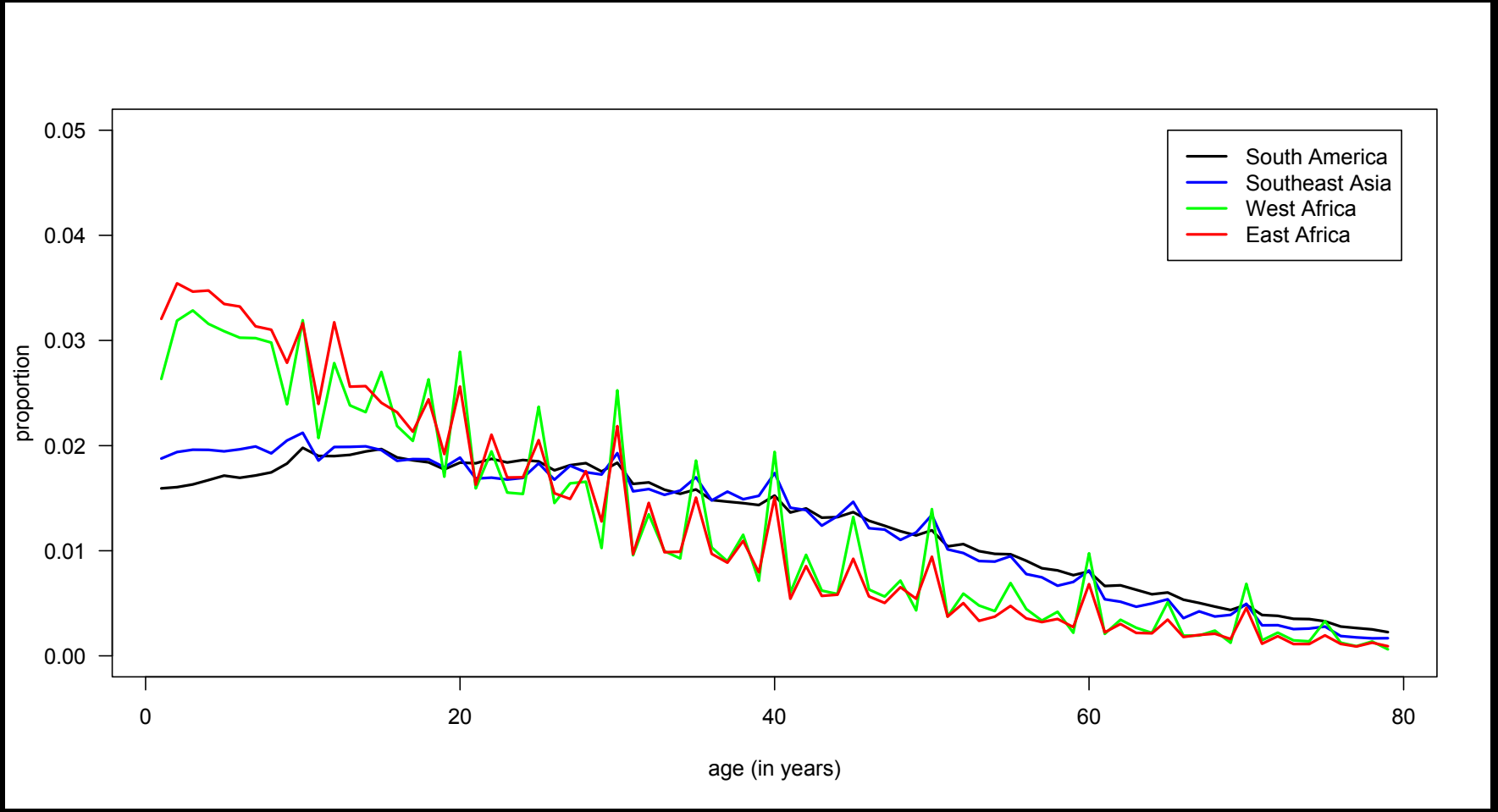
Sepúlveda & Drakeley (2015). Malaria J.

Identification of transmission intensity



Sepúlveda & Drakeley (2015). Malaria J.

Identification of age structure



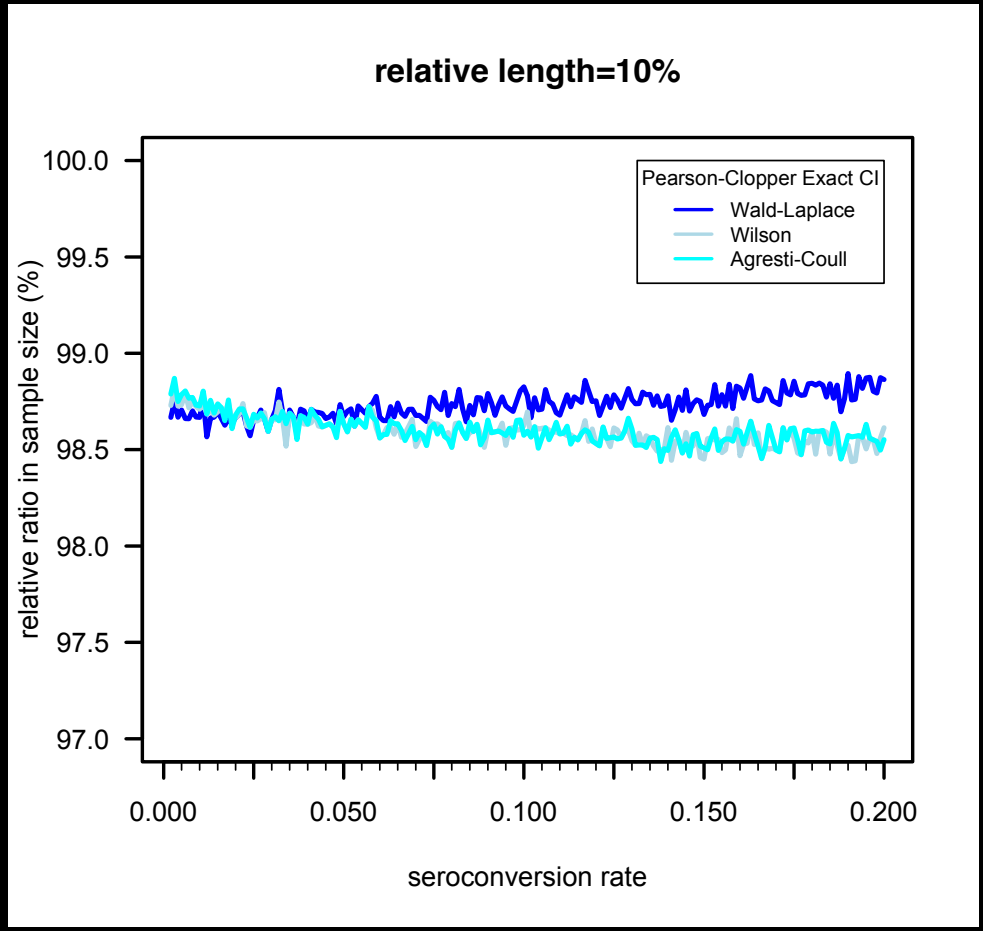
Type of confidence interval

Pearson-Clopper exact
Coverage higher than nominal confidence level

Wald
Degenerate when x=0 or x=n
Overshooting

Wilson
Good coverage in extreme probabilities

Agresti-Coull
Overshooting
Better coverage than the exact CI



Sepúlveda & Drakeley (2016). Proceedings of Portuguese Society for Statistics

time!

References

Cook J, Kleinschmidt I, Schwabe C, et al. Serological markers suggest heterogeneity of effectiveness of malaria control interventions on Bioko Island, equatorial Guinea. PLoS One. 2011;6(9):e25137. doi:10.1371/journal.pone.0025137

Sepúlveda N, Paulino CD, Drakeley C. Sample size and power calculations for detecting changes in malaria transmission using antibody seroconversion rate. Malar J. 2015;14:529. Published 2015 Dec 30. doi:10.1186/s12936-015-1050-3

Sepúlveda N, Stresman G, White MT, Drakeley CJ. Current Mathematical Models for Analyzing Anti-Malarial Antibody Data with an Eye to Malaria Elimination and Eradication. J Immunol Res. 2015;2015:738030. doi:10.1155/2015/738030