Cross-Sectional HIV Incidence Estimation: Approaches and Challenges

Jacob Konikoff

UCLA Department of Biostatistics

May 11th, 2015

Motivation

- Why is estimating HIV incidence important?
- Why is it challenging to measure?
- Approaches to address the challenges

Incidence: The expected number of new infections per year at a given moment in time, among those uninfected up to that point in time.

A Public Health Necessity

Tracking and surveillance of the epidemic

- Resource allocation

Prevention efforts

- Sample size calculations
- Selecting a target population
- Evaluation

Approaches for Estimating Incidence

- Longitudinal studies
- Changes in HIV prevalence
- Cross-sectional studies

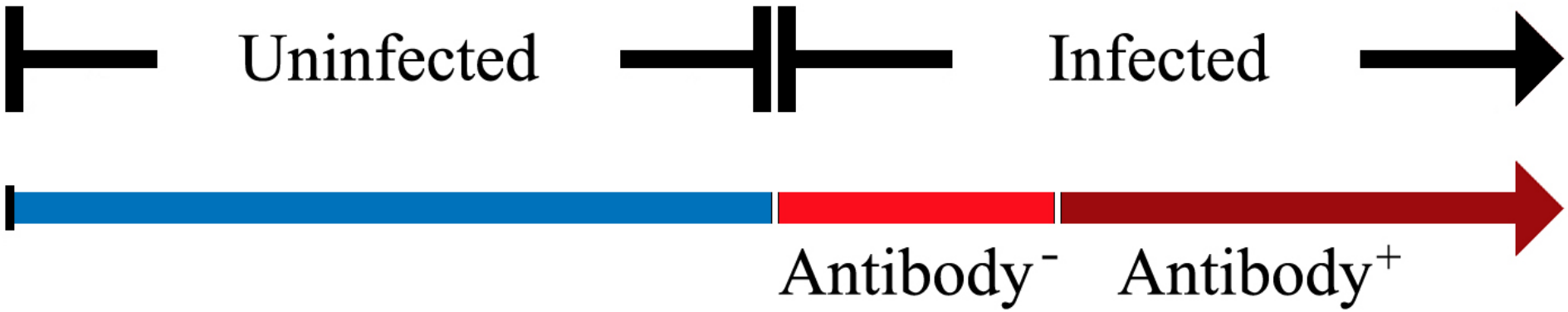
Longitudinal Studies

- Selection bias
- Loss to follow-up
- Counseling

Changes in HIV Prevalence

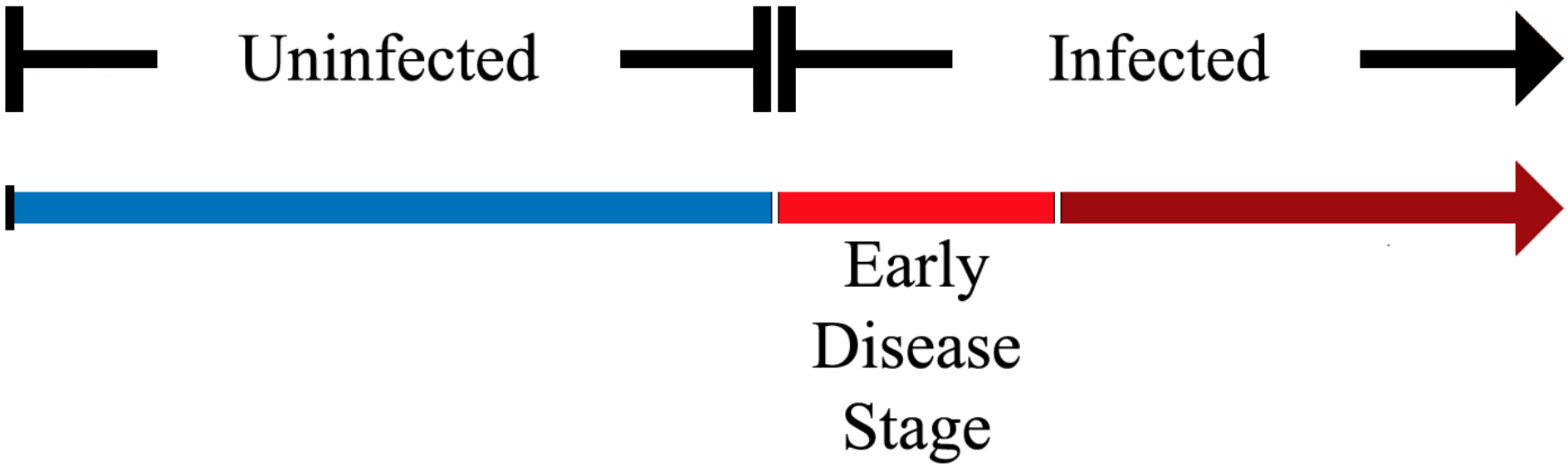
- Depends on immigration and emigration
- Depends on the relative survival rate

The Cross-Sectional Approach



Seroconversion: When antibodies become detectable in the blood

The Cross-Sectional Approach



Seroconversion: When antibodies become detectable in the blood

Estimating Incidence

X = the number of cases in the early disease stage

N_u = the number of uninfected samples

μ = the average duration individuals spend in the early disease stage

Prevalence = Total Incidence $\times$ Mean Duration

$$X = N_u I \times \mu$$

The Cross-Sectional Incidence Estimator

$$X = N_u I \times \mu$$

The Cross-Sectional Incidence Estimator

$$\widehat{I} = \frac{X}{N_u \mu}$$

Example

$$\widehat{I} = \frac{X}{N_u \mu}$$

Example: In the cross-sectional survey we find 1000 uninfected people and 50 people in the early disease stage.

People spend an average of 1/2 a year in the early disease stage.

So we have $\frac{50}{1000}$ infections per half a year

Example

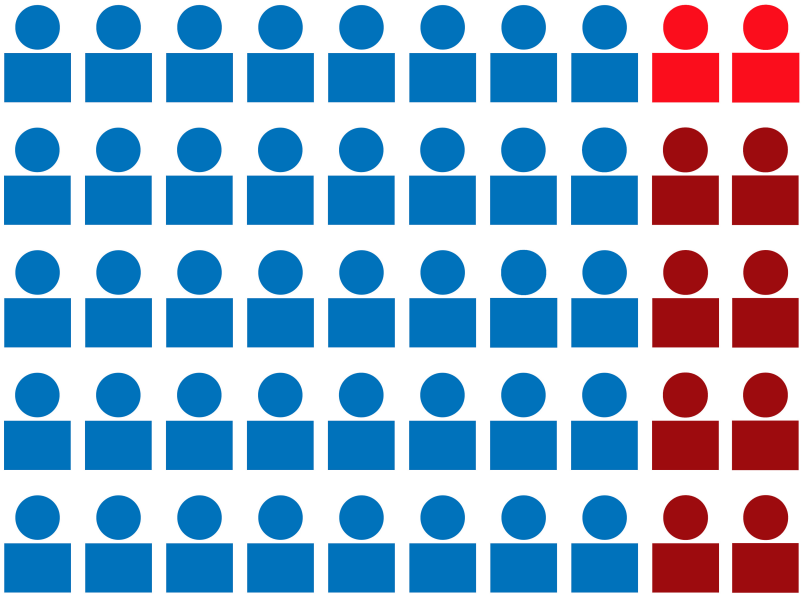
$$\widehat{I} = \frac{X}{N_u \mu}$$

Example: In the cross-sectional survey we find 1000 uninfected people and 50 people in the early disease stage.

People spend an average of 1/2 a year in the early disease stage.

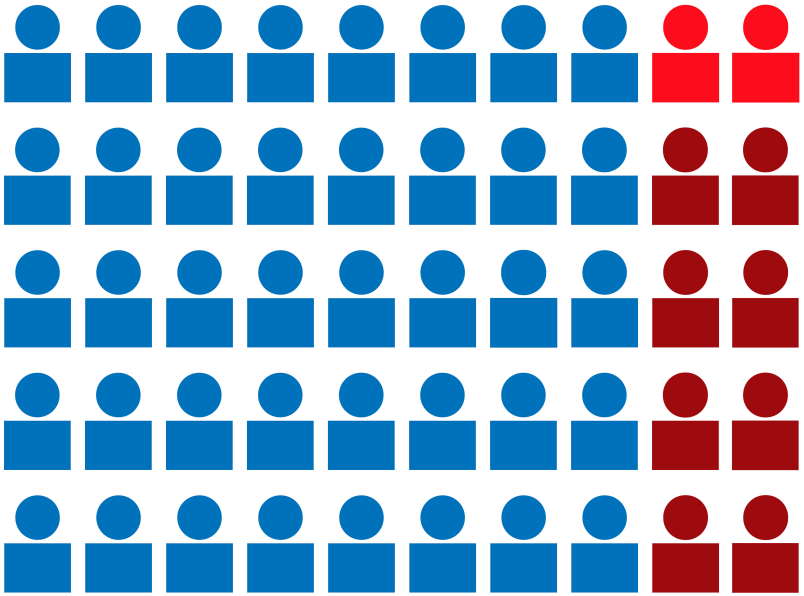
Or $\frac{50}{1000 \times 1/2} = \frac{100}{1000}$ infections per year

The Cross-Sectional Survey



$$\hat{I} = \frac{X}{N_u \mu}$$

The Cross-Sectional Survey

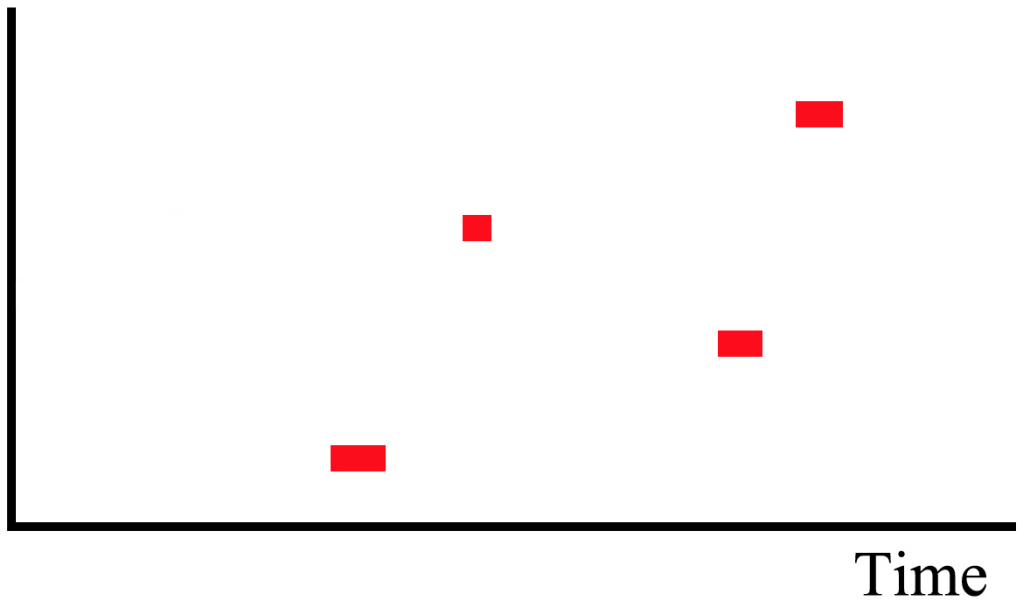


$$\hat{I} = \frac{X}{N_u \mu}$$

Average Time in the Early Disease Stage

$\mu \approx 1$ month

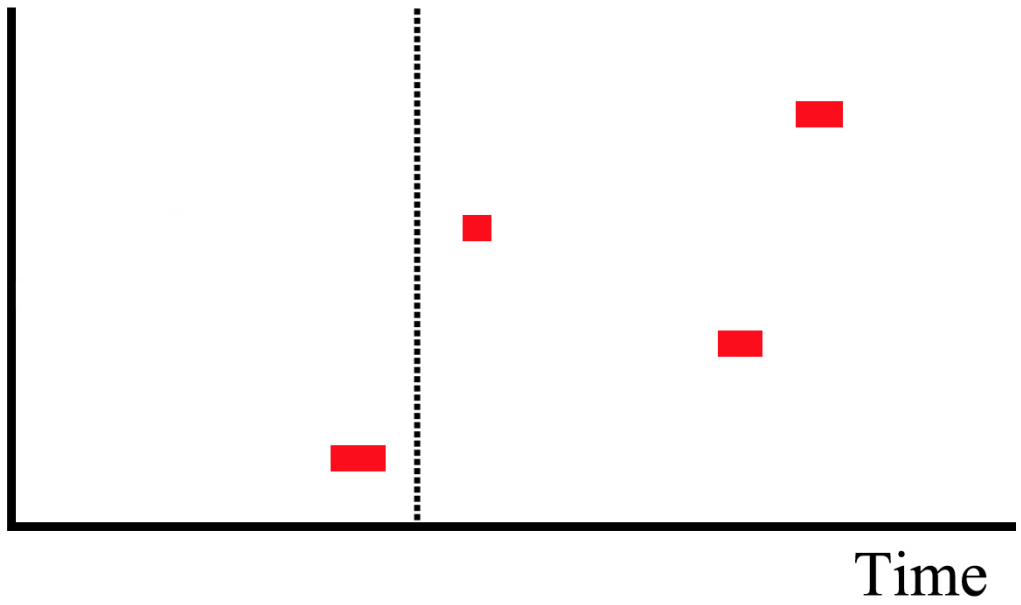
$\mu =$ the average amount of time individuals spend in the early disease stage



Average Time in the Early Disease Stage

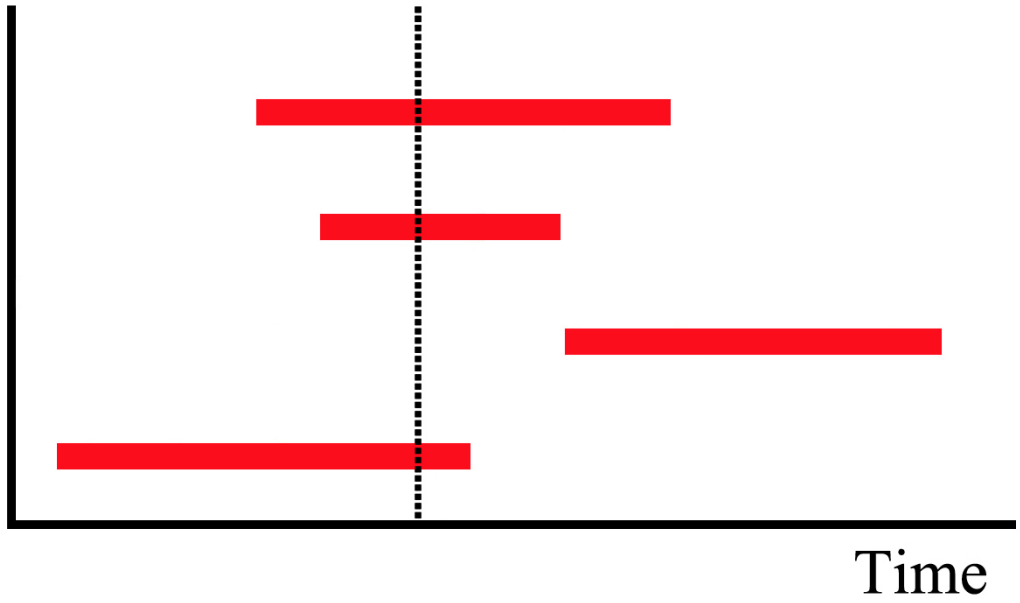
$\mu \approx 1$ month

$\mu =$ the average amount of time individuals spend in the early disease stage

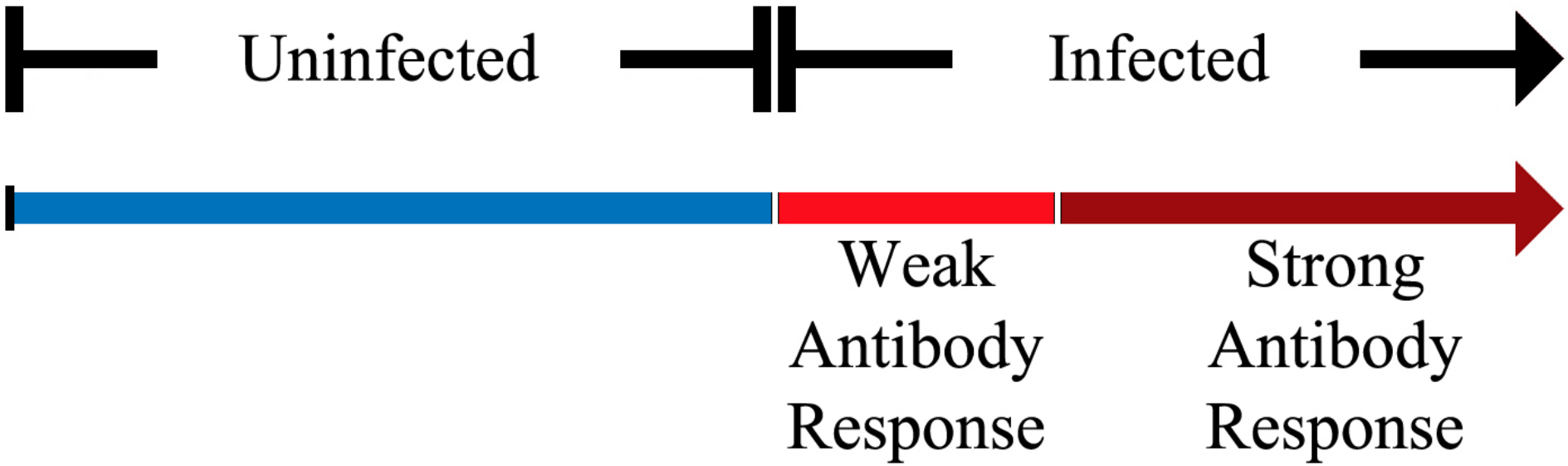


Average Time in the Early Disease Stage

μ = the average amount of time individuals spend in the early disease stage



Stages of Disease Progression

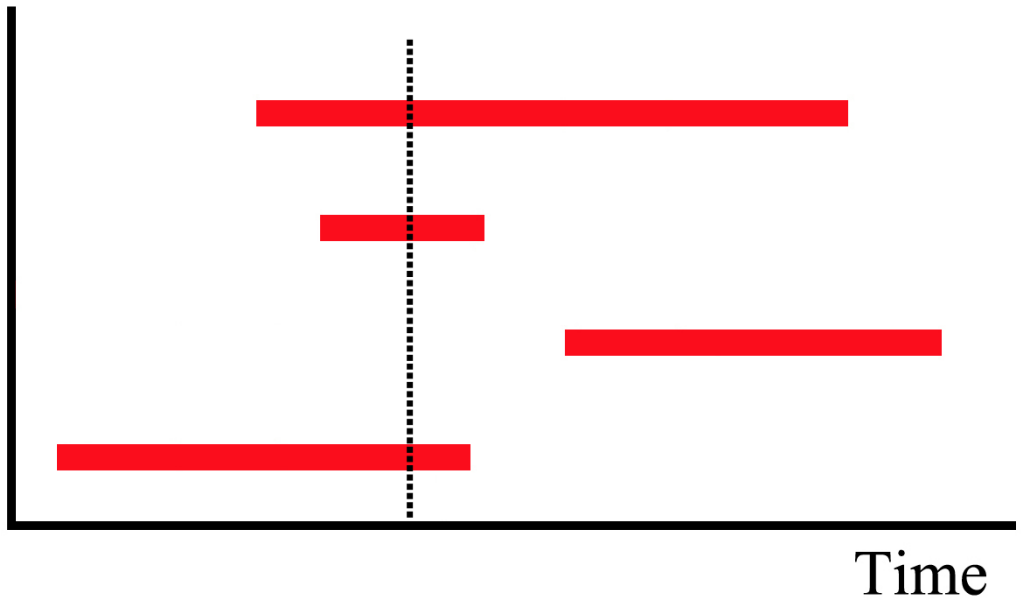


Example: BED IgG capture enzyme immunoassay result < 0.8 OD-n

Estimating Historical Incidence

$\hat{I} = \frac{X}{N_u \mu}$ is estimating past, not current, incidence ψ years before the cross-sectional survey

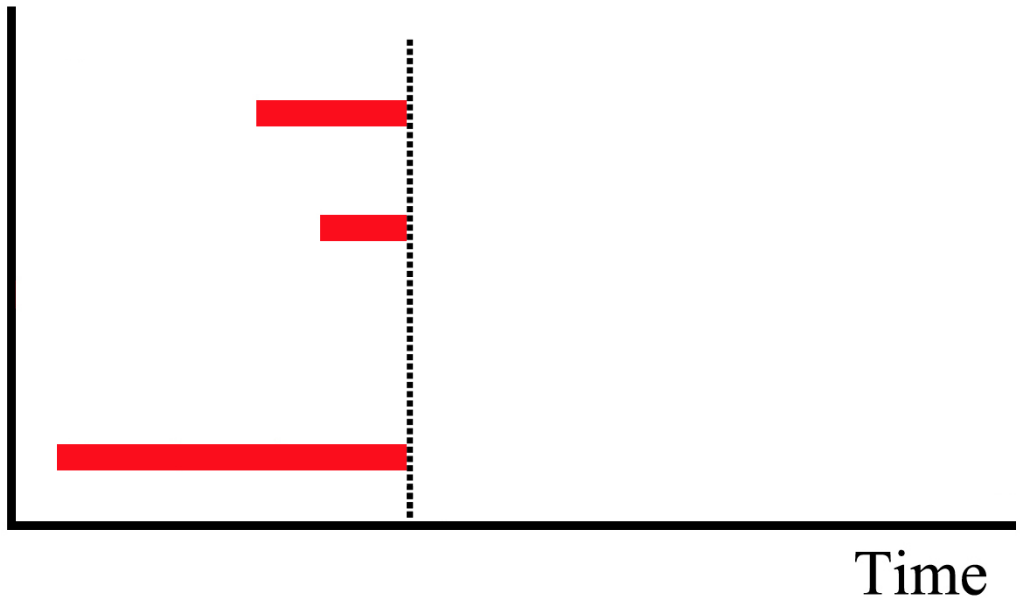
ψ is a weighted average of the amount of time prevalent early disease stage cases spent in that stage *before* the survey



Estimating Historical Incidence

$\hat{I} = \frac{X}{N_u \mu}$ is estimating past, not current, incidence ψ years before the cross-sectional survey

ψ is a weighted average of the amount of time prevalent early disease stage cases spent in that stage *before* the survey



Estimating Historical Incidence

$\hat{I} = \frac{X}{N_u \mu}$ is estimating past, not current, incidence ψ years before the cross-sectional survey

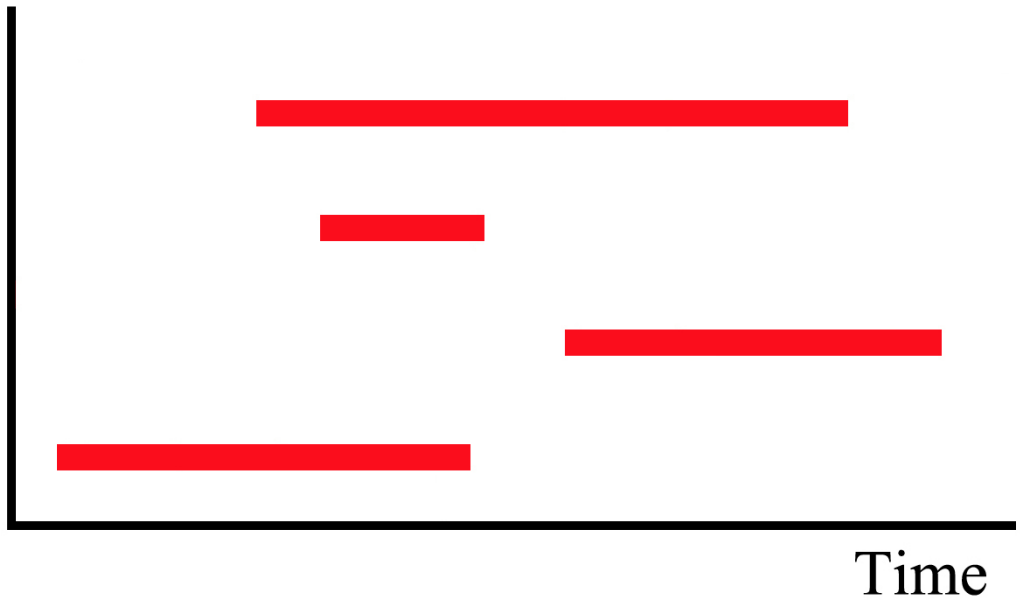
ψ is a weighted average of the amount of time prevalent early disease stage cases spent in that stage *before* the survey

$$\psi = \frac{\int_0^\infty t\phi(t)dt}{\int_0^\infty \phi(t)dt}$$

$\phi(t)$ is the probability that persons will be in the early disease stage t years after seroconversion

μ and ψ

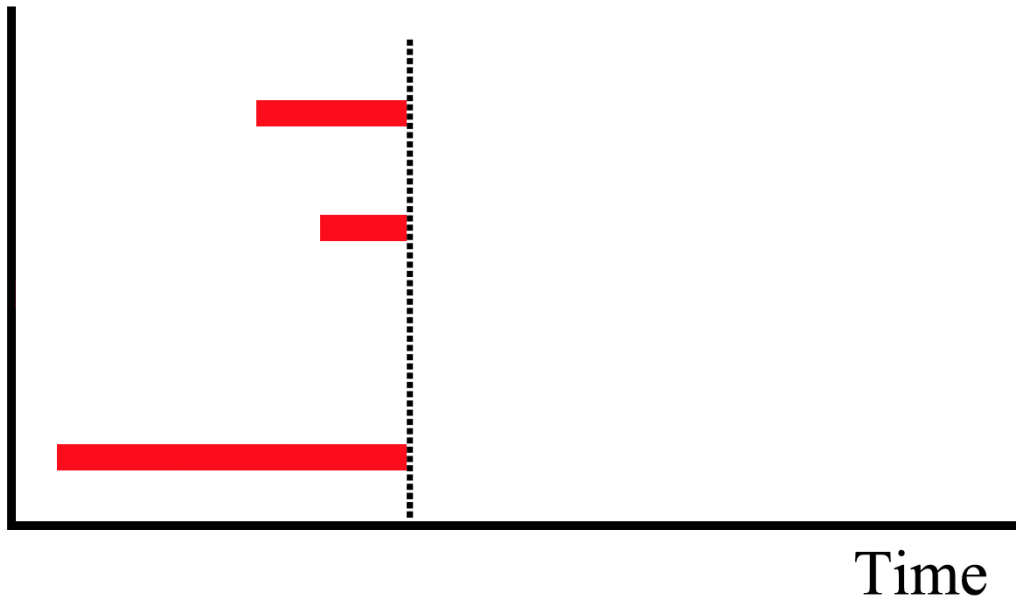
$$\mu = \int_0^\infty \phi(t)dt$$



μ and ψ

$$\mu = \int_0^\infty \phi(t)dt$$

$$\psi = \frac{1}{\mu}\int_0^\infty t\phi(t)dt$$



Key Ideas

The cross-sectional incidence estimator is

$$\hat{I}(\psi) = \frac{X}{N_u \mu}$$

μ is the average duration spent in the early disease stage

ψ is the time in the past when we estimate incidence

$\phi(t)$ is the probability that persons will be in the early disease stage t years after seroconversion

Evaluation of Biomarkers

The BED CEIA

For the BED CEIA μ has been repeatedly estimated to be about 200 days

- Hall et al. *JAMA* (2008) $\rightarrow$ United States

The BED CEIA

For the BED CEIA μ has been repeatedly estimated to be about 200 days

- Hall et al. *JAMA* (2008) $\rightarrow$ United States
- Hu et al. *AIDS Research & Human Retroviruses* (2003) $\rightarrow$ Bangkok

The BED CEIA

For the BED CEIA μ has been repeatedly estimated to be about 200 days

- Hall et al. *JAMA* (2008) $\rightarrow$ United States
- Hu et al. *AIDS Research & Human Retroviruses* (2003) $\rightarrow$ Bangkok
- Jiang et al. *AIDS* (2007) $\rightarrow$ China

The BED CEIA

For the BED CEIA μ has been repeatedly estimated to be about 200 days

- Hall et al. *JAMA* (2008) $\rightarrow$ United States
- Hu et al. *AIDS Research & Human Retroviruses* (2003) $\rightarrow$ Bangkok
- Jiang et al. *AIDS* (2007) $\rightarrow$ China
- Saphonn et al. *JAIDS* (2005) $\rightarrow$ Cambodia

The BED CEIA

For the BED CEIA μ has been repeatedly estimated to be about 200 days

- Hall et al. *JAMA* (2008) $\rightarrow$ United States
- Hu et al. *AIDS Research & Human Retroviruses* (2003) $\rightarrow$ Bangkok
- Jiang et al. *AIDS* (2007) $\rightarrow$ China
- Saphonn et al. *JAIDS* (2005) $\rightarrow$ Cambodia
- Kim et al. *AIDS Research & Human Retroviruses* (2010) $\rightarrow$ sub-Saharan Africa

Controversy

UNAIDS

"Does not recommend the BED assay for determining HIV incidence"

Controversy

UNAIDS

"Does not recommend the BED assay for determining HIV incidence"

"There is evidence that [...] the BED-assay captures not only recent infections, but also late stage HIV infection (with or without antiretroviral therapy) when the levels of antibodies fall."

Data

- Collect a repository of samples with information on seroconversion dates
- Assay samples for biomarkers

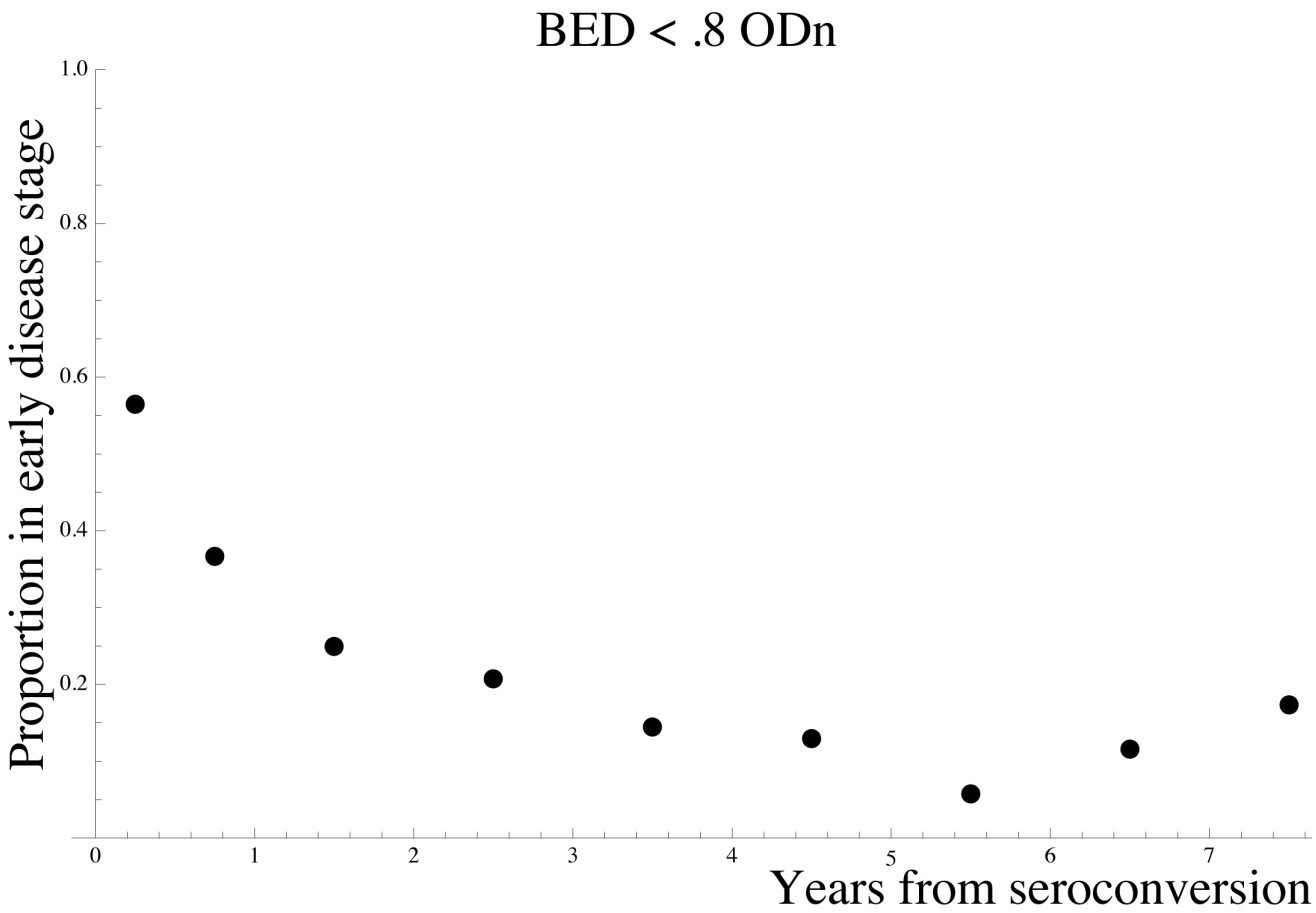
Dr. Susan Eshleman and Dr. Oliver Laeyendecker

Cohort Studies

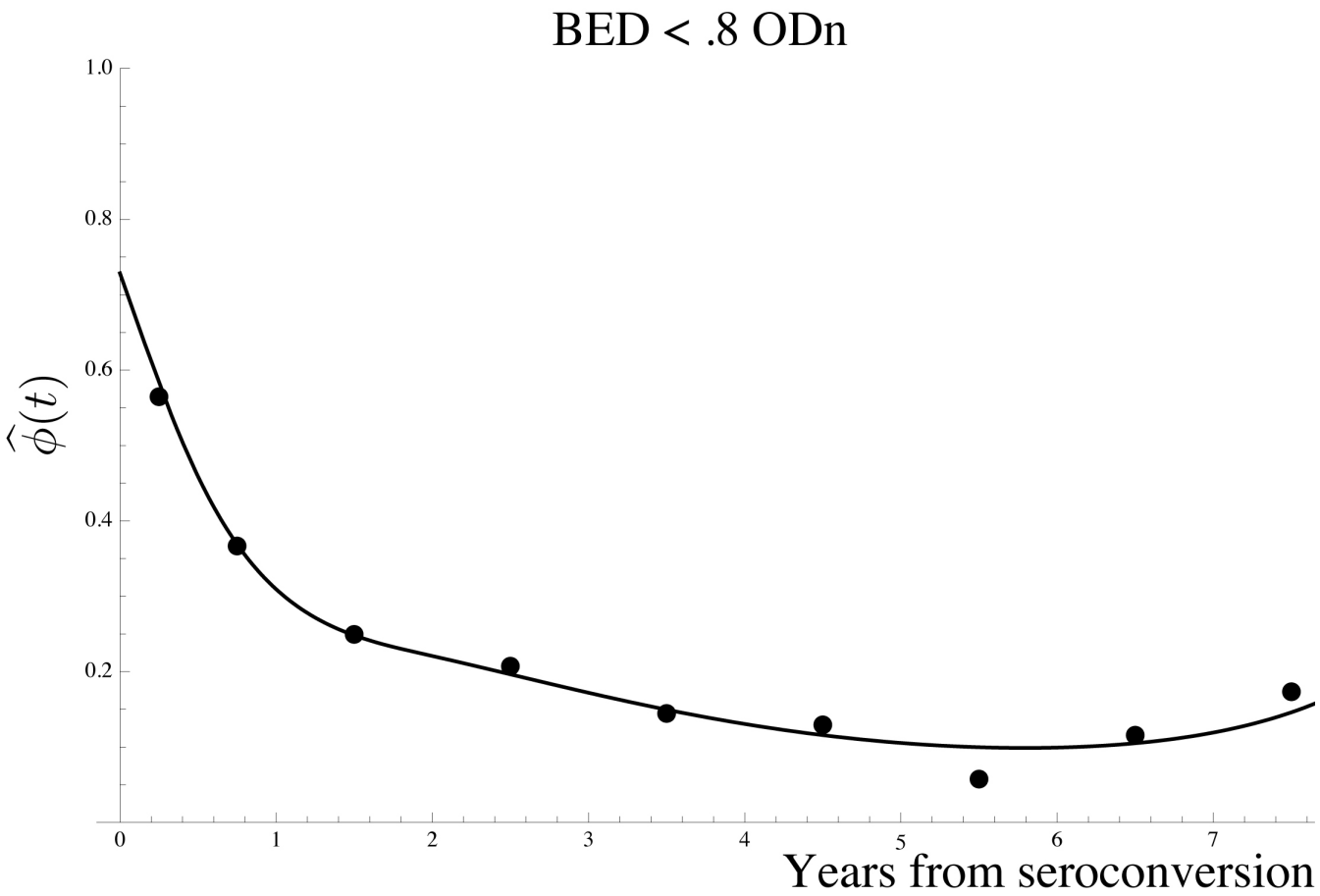
1. HIV Network for Prevention Trials (HIVNET001)
2. AIDS Link to Intravenous Experience (ALIVE)
3. Multicenter AIDS Cohort Study (MACS)

- 1782 samples from 709 individuals with known seroconversion windows
- Samples taken between 14 days and 8.6 years after seroconversion

Raw Data for the BED CEIA



$\widehat{\phi}(t)$ for the BED CEIA



$\hat{\mu} \approx 1.5$ years

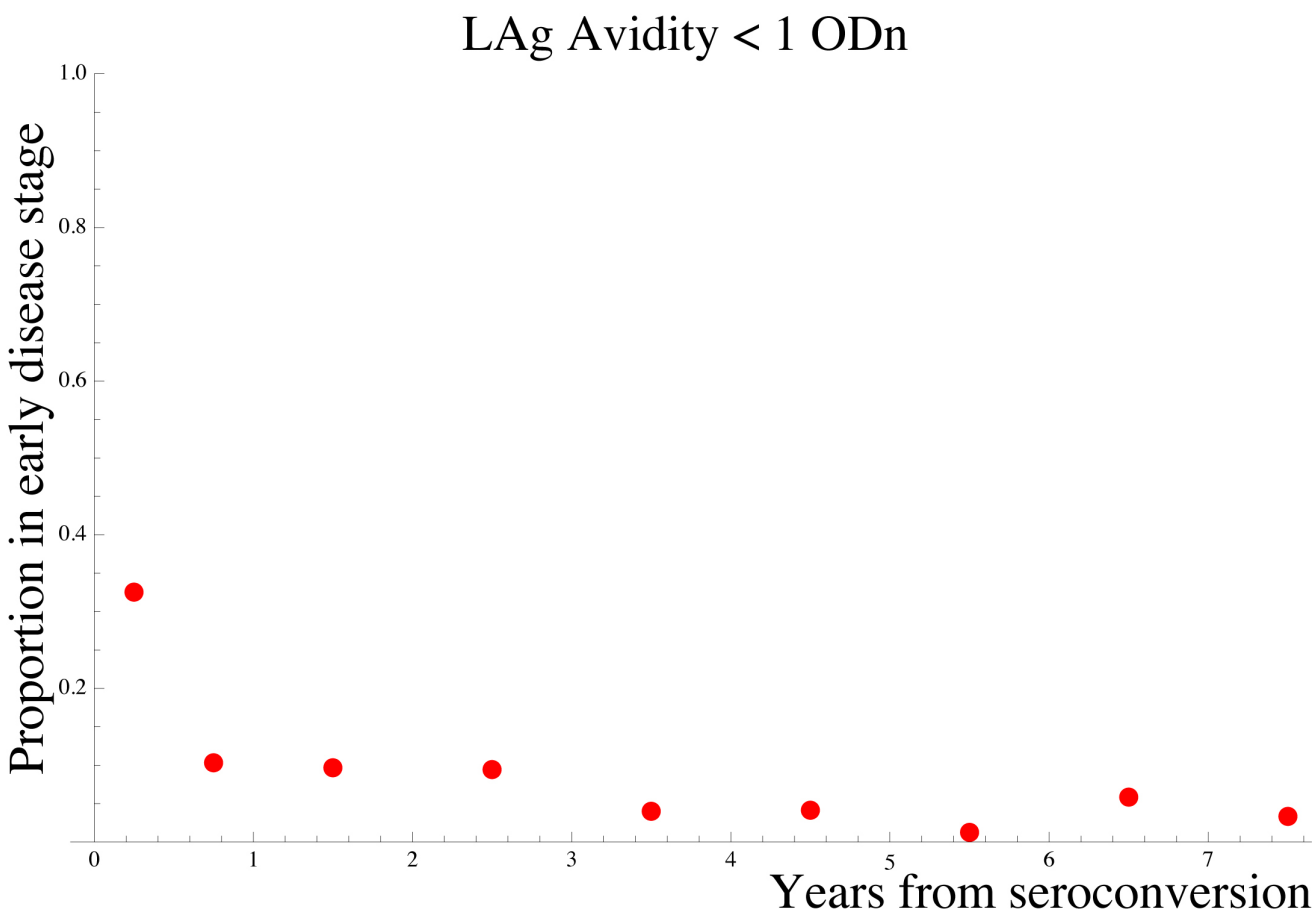
$\hat{\psi} \approx 3$ years

Better Assays

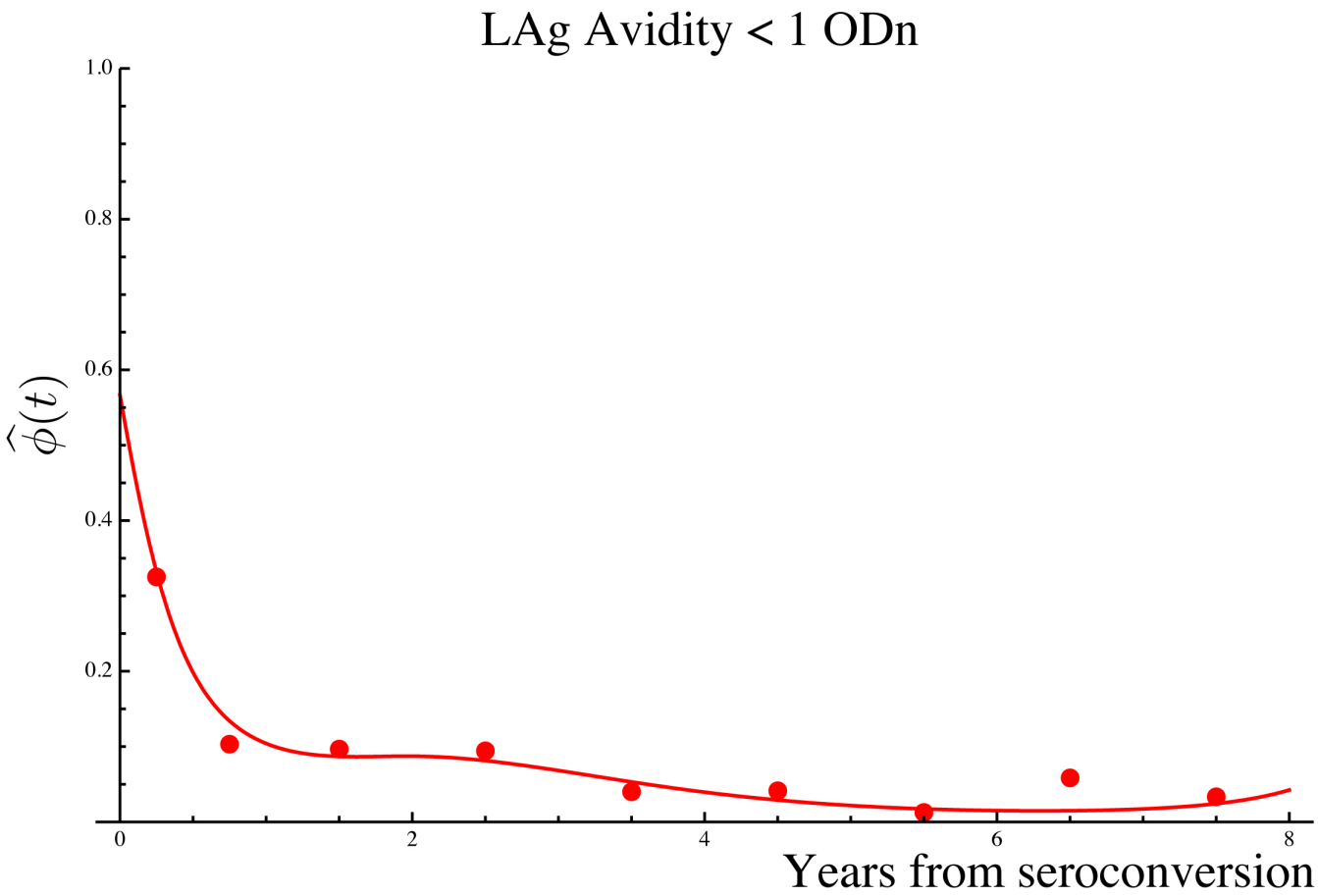
LAg-Avidity assay: measures the avidity of antibody binding to low concentrations of a multi-subtype peptide derived from an immunodominant region of gp41.

BioRad Avidity assay: measures the percentage of antigen-binding chaotropic-treated antibody relative to the antigen-binding of nontreated antibody.

Raw Data for the LAg Avidity assay



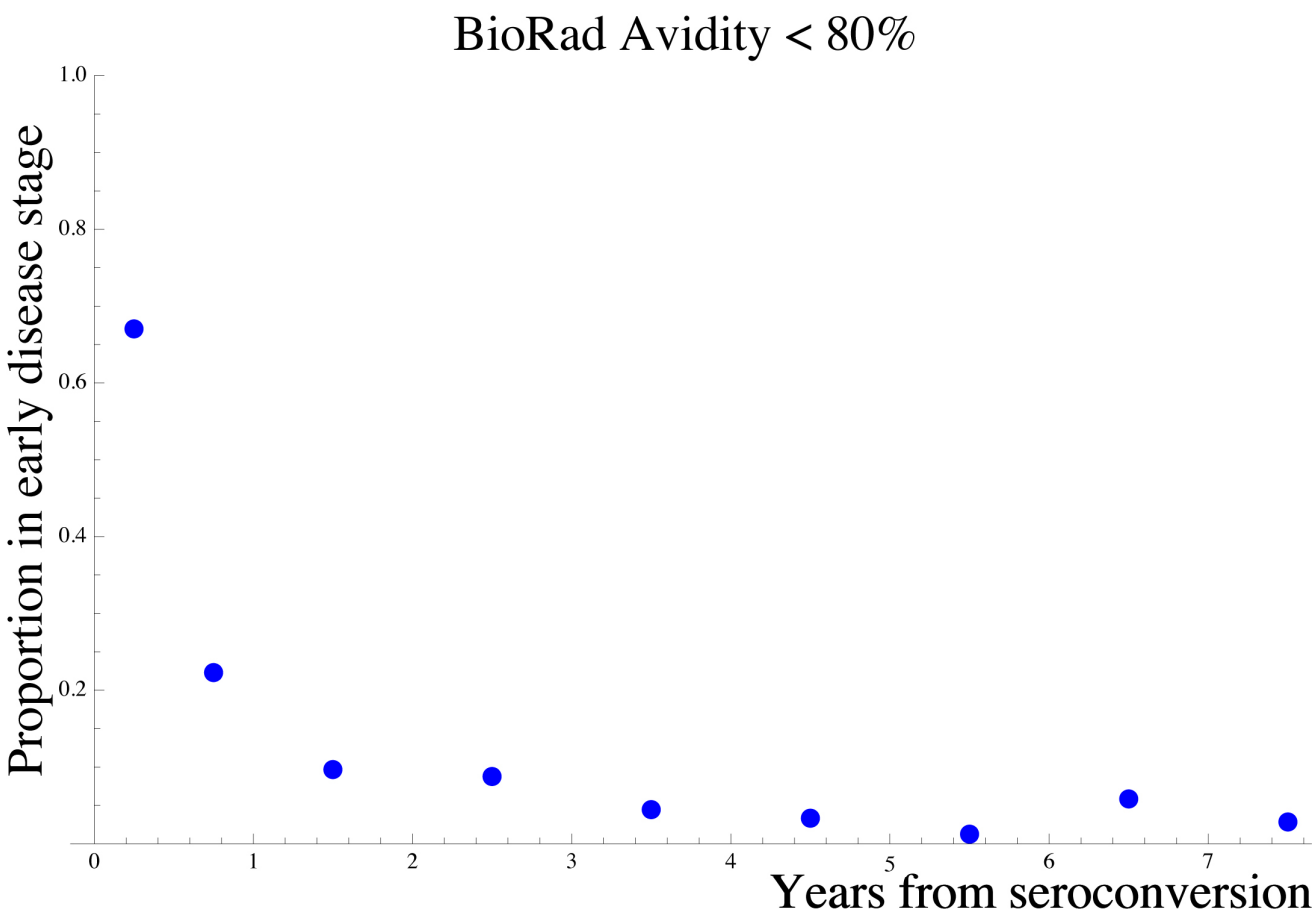
$\widehat{\phi(t)}$ for the LAg Avidity assay



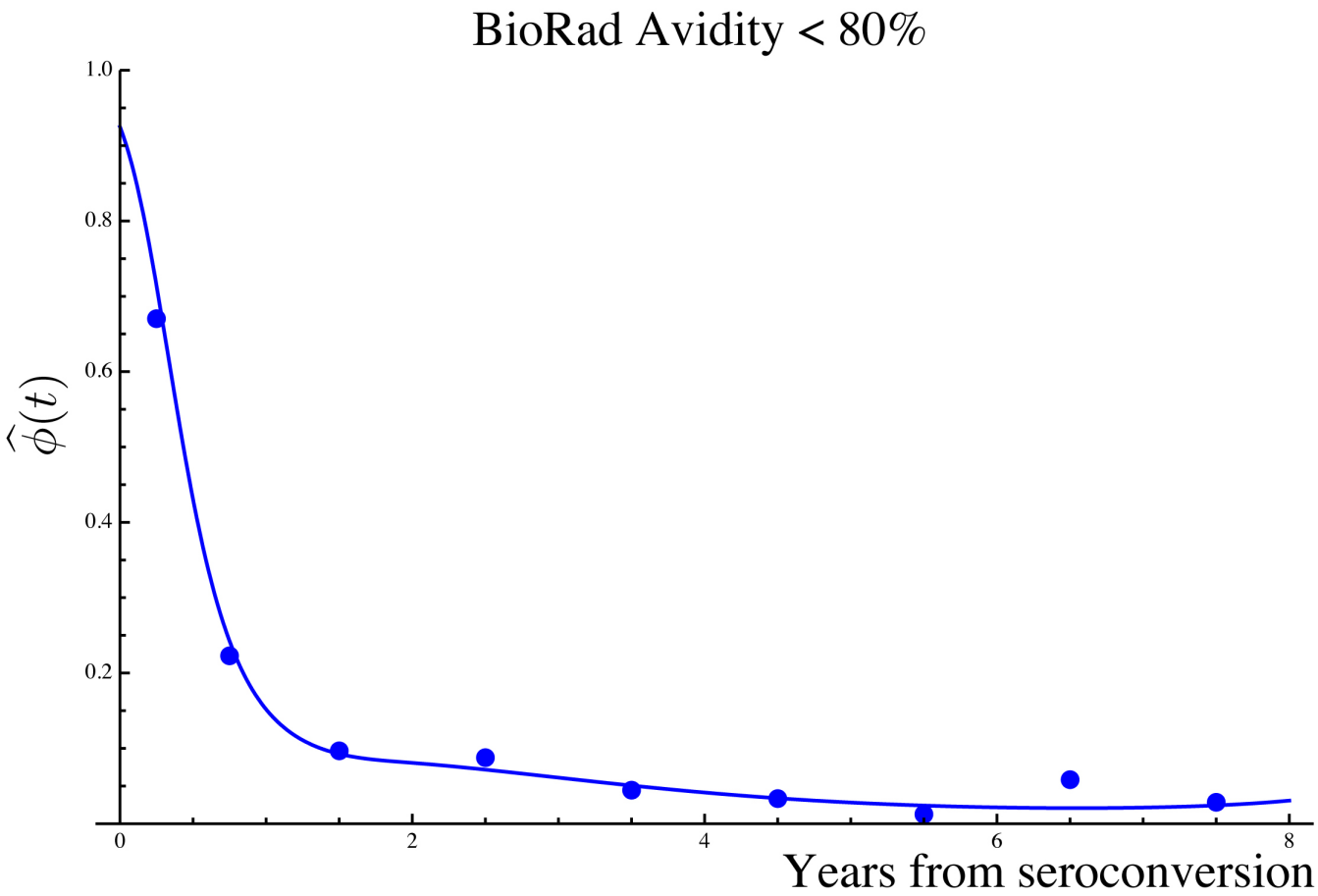
$\hat{\mu} \approx .6$ years

$\hat{\psi} \approx 2$ years

Raw Data for the BioRad Avidty



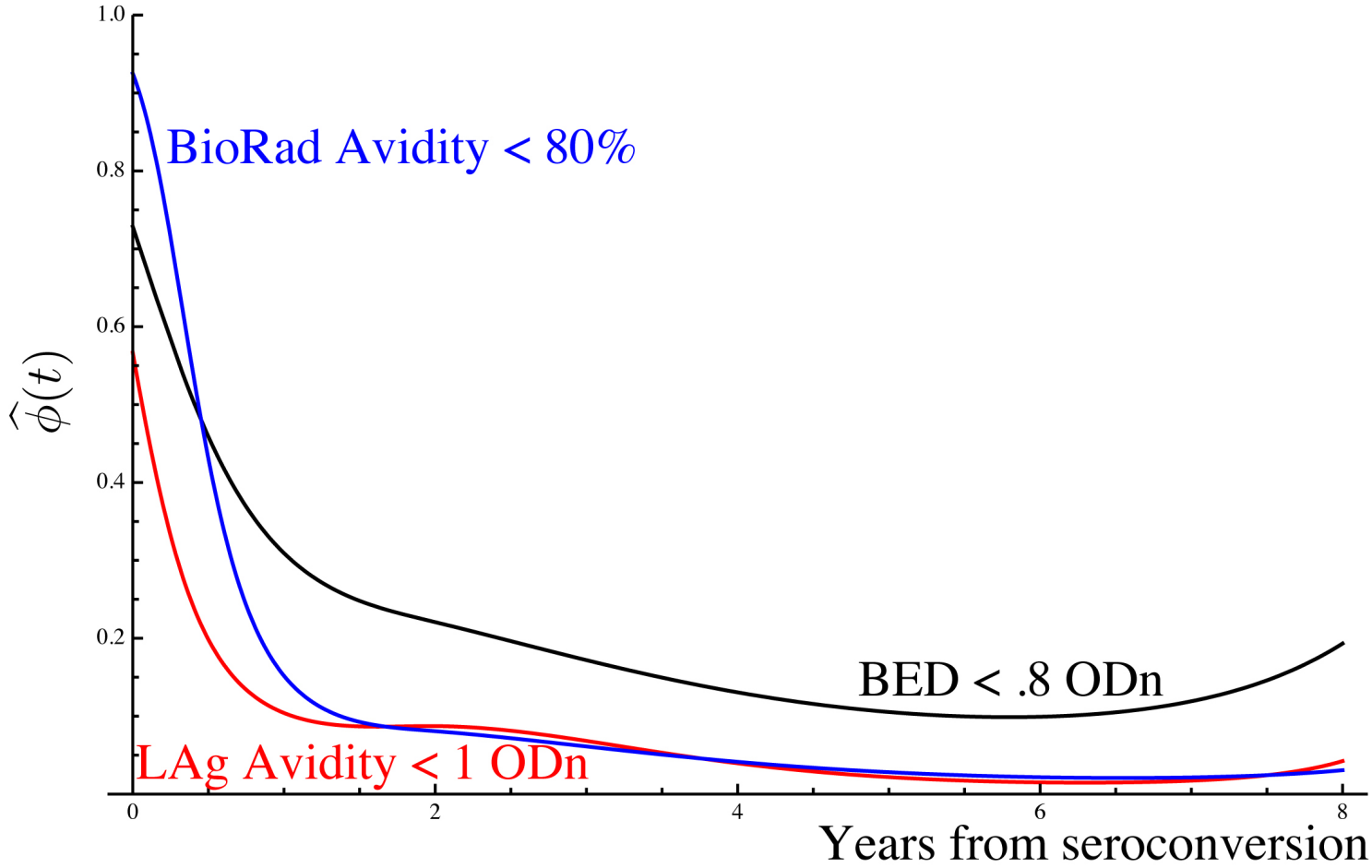
$\widehat{\phi(t)}$ for the BioRad Avidity



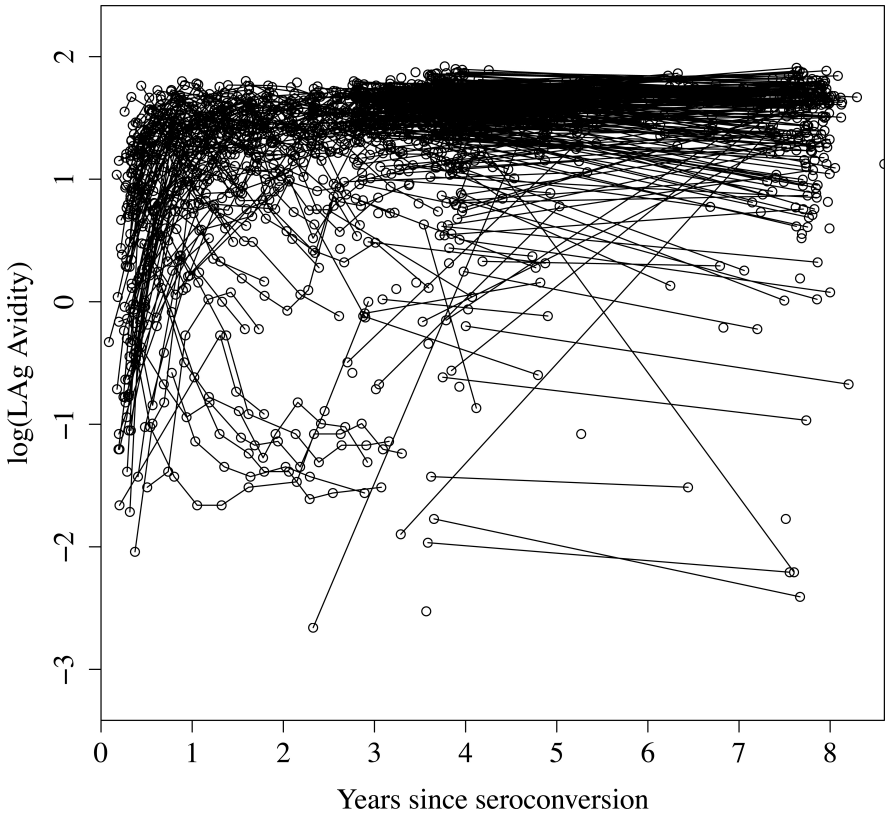
$\hat{\mu} \approx .8$ years

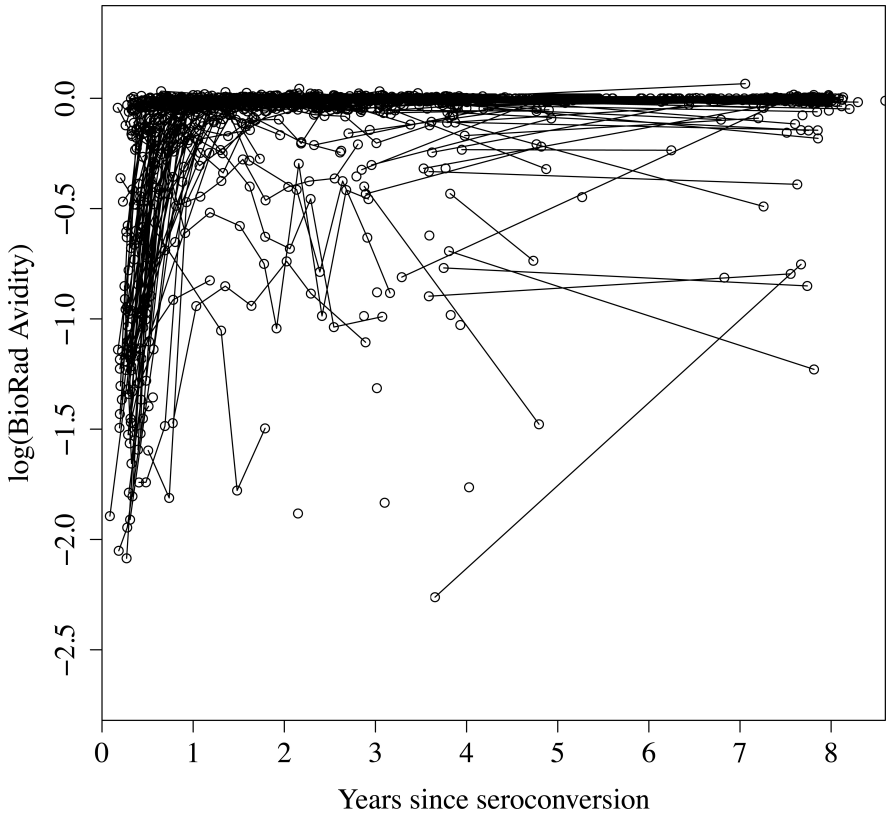
$\hat{\psi} \approx 1.6$ years

Comparison of $\widehat{\phi}(t)$ Curves

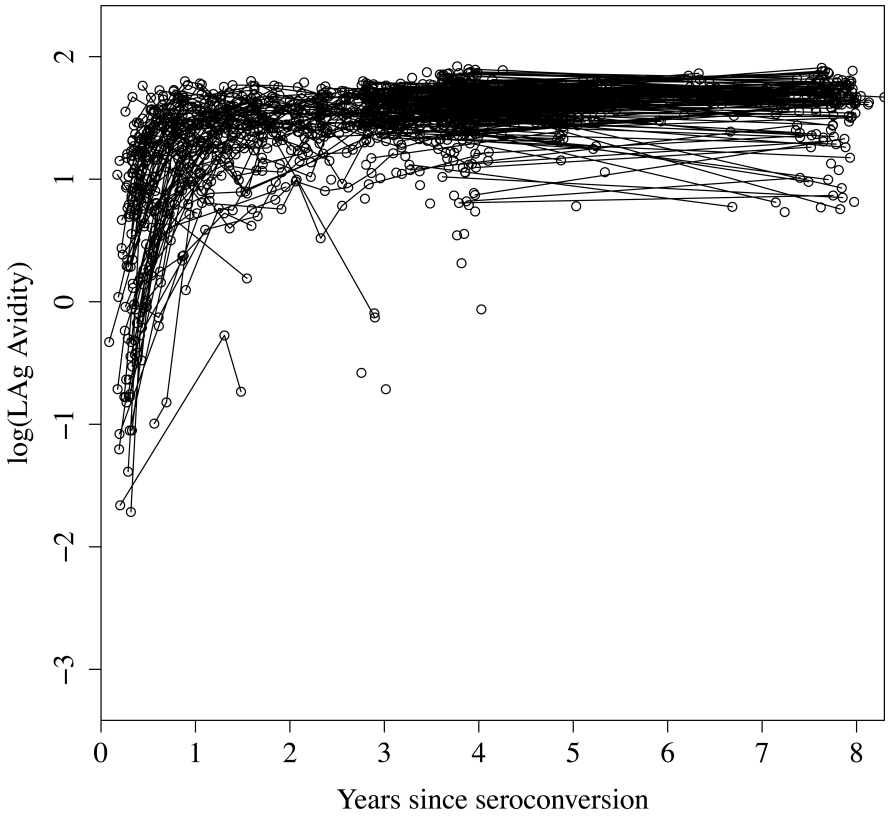


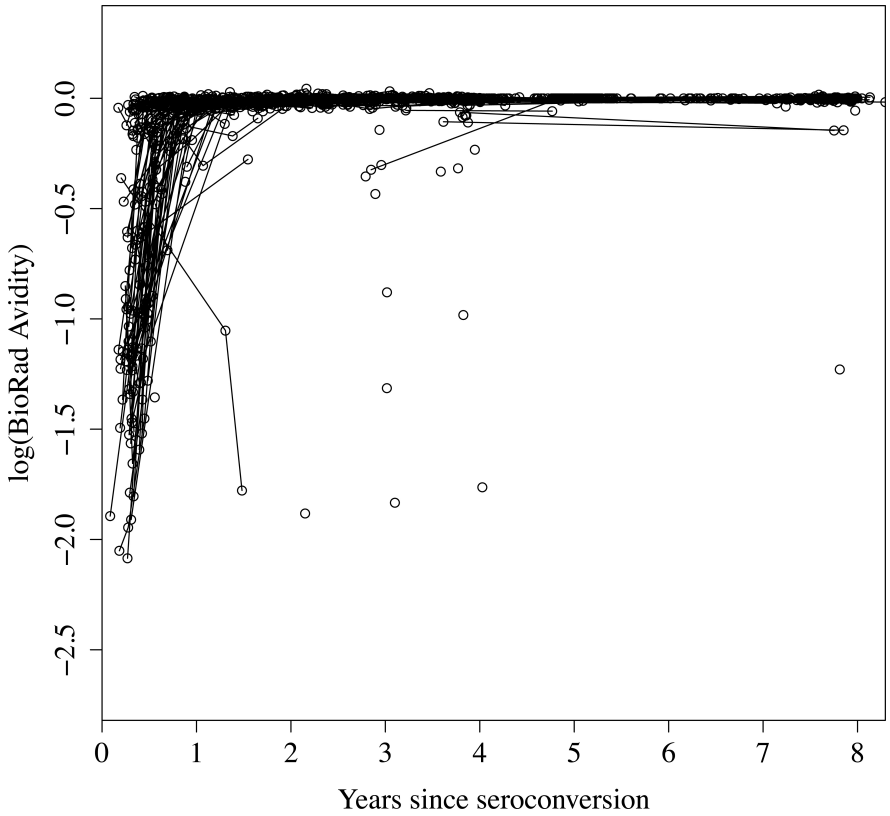
Profile Plots



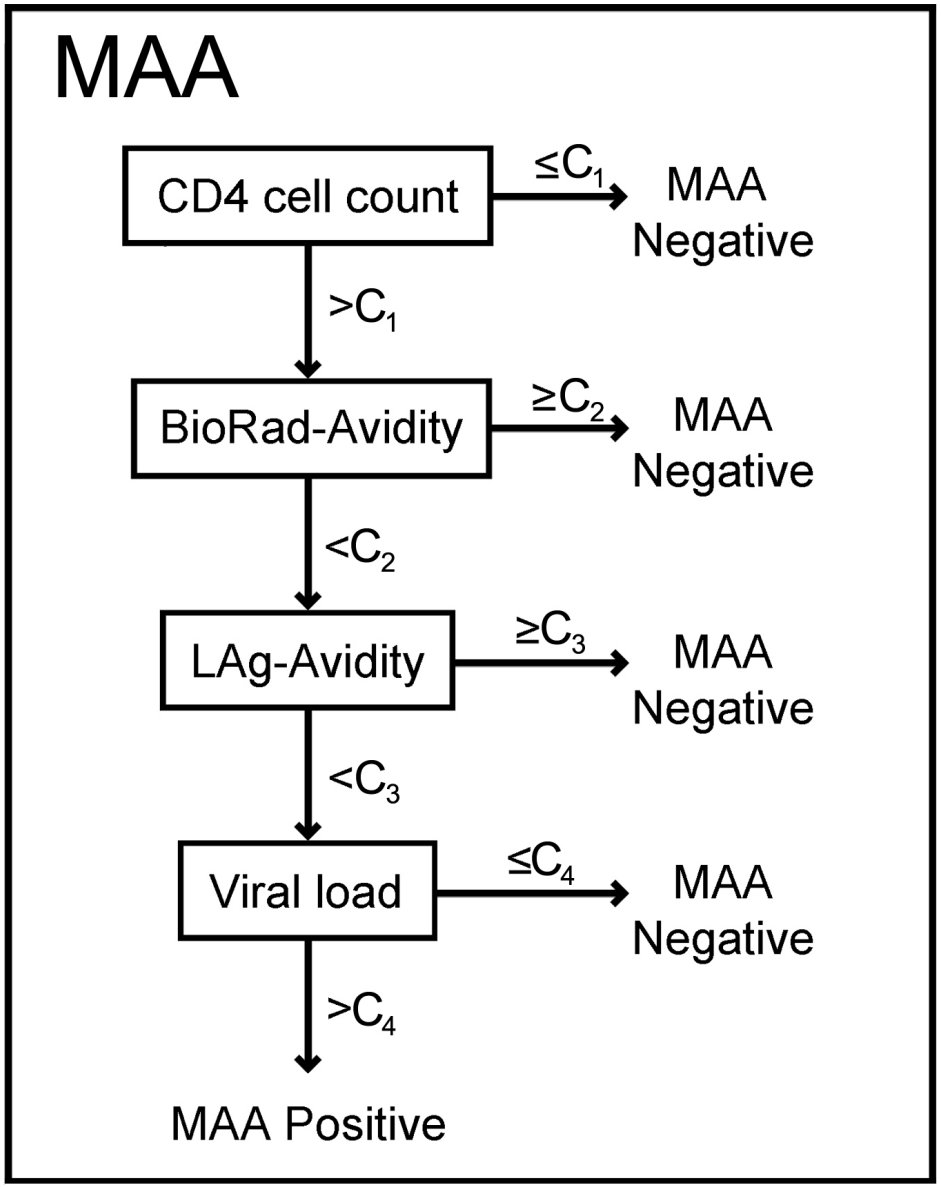


Profile Plots





Sequential Algorithms

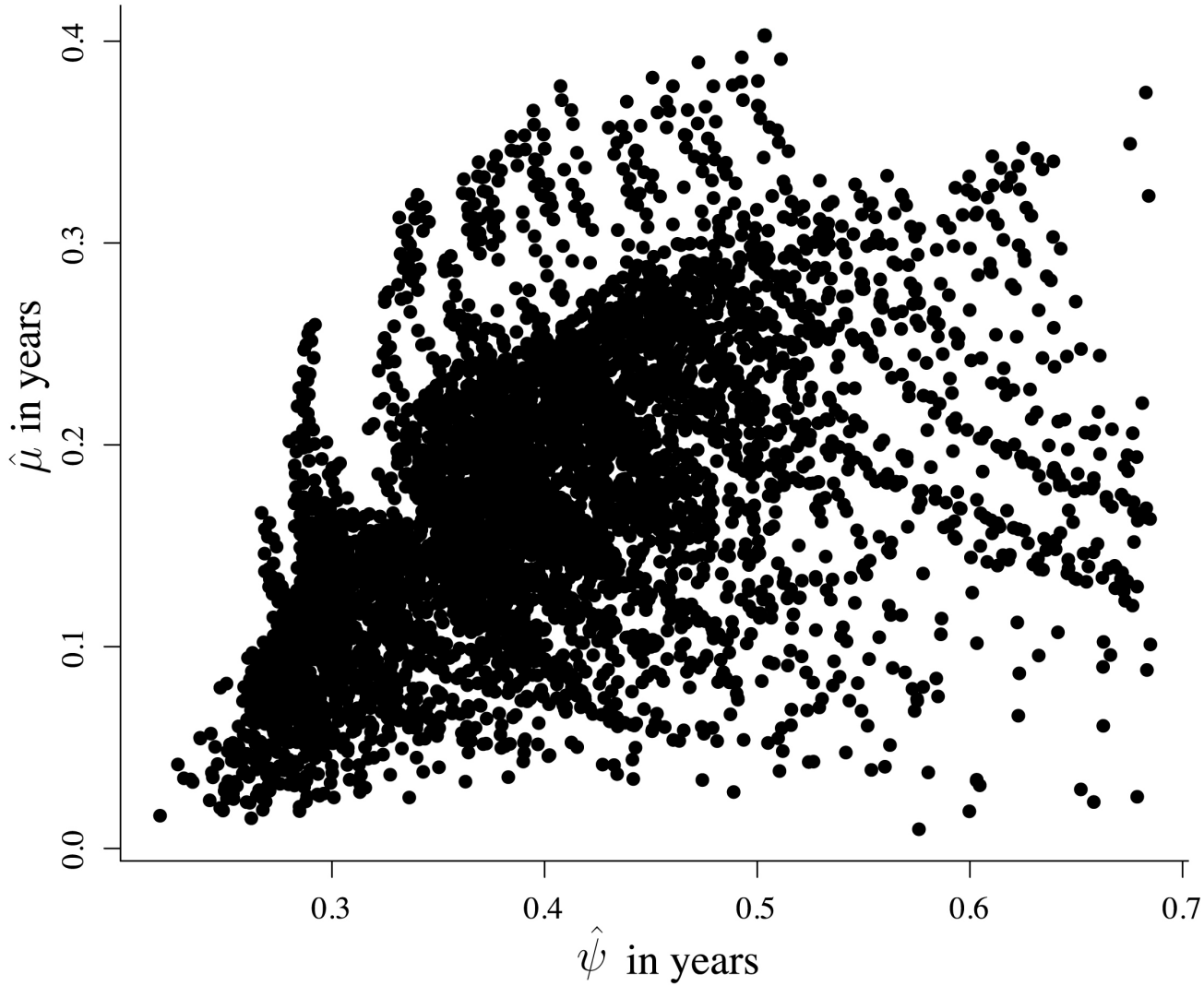


Statistical Search for MAAs

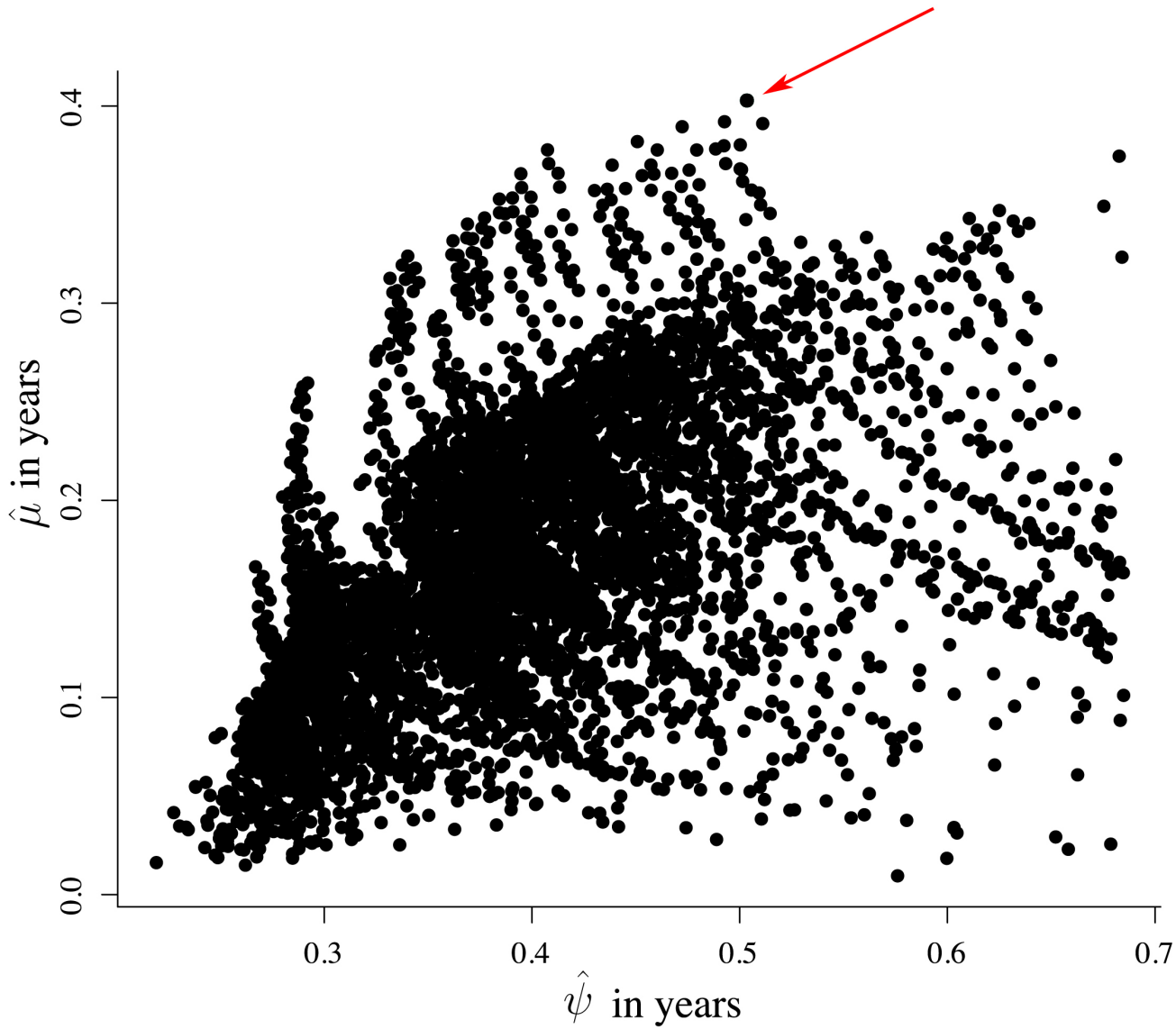
We searched for the MAA with the largest μ such that:

1. The estimated probability of being classified in the early disease stage at 8 years after seroconversion needed to be < 0.001.
2. None of the samples infected more than 8 years could be found to be in the early disease stage.
3. The upper bound of the 95% CI for the shadow needed to be less than 1 year.
4. The point estimate of the shadow, $\hat{\psi}$, needed to be less than 250 days.

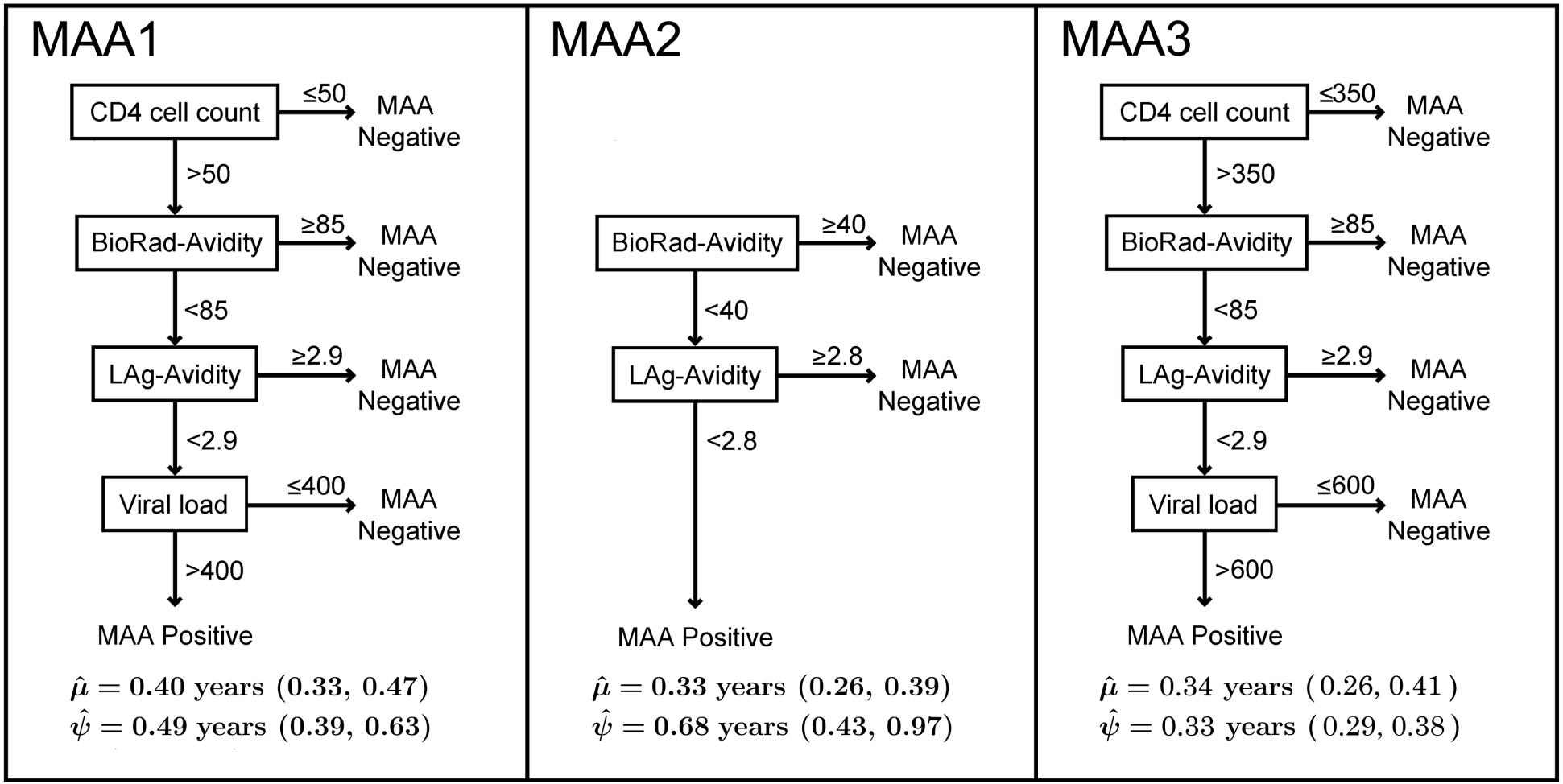
$\hat{\mu}$ vs. $\hat{\psi}$



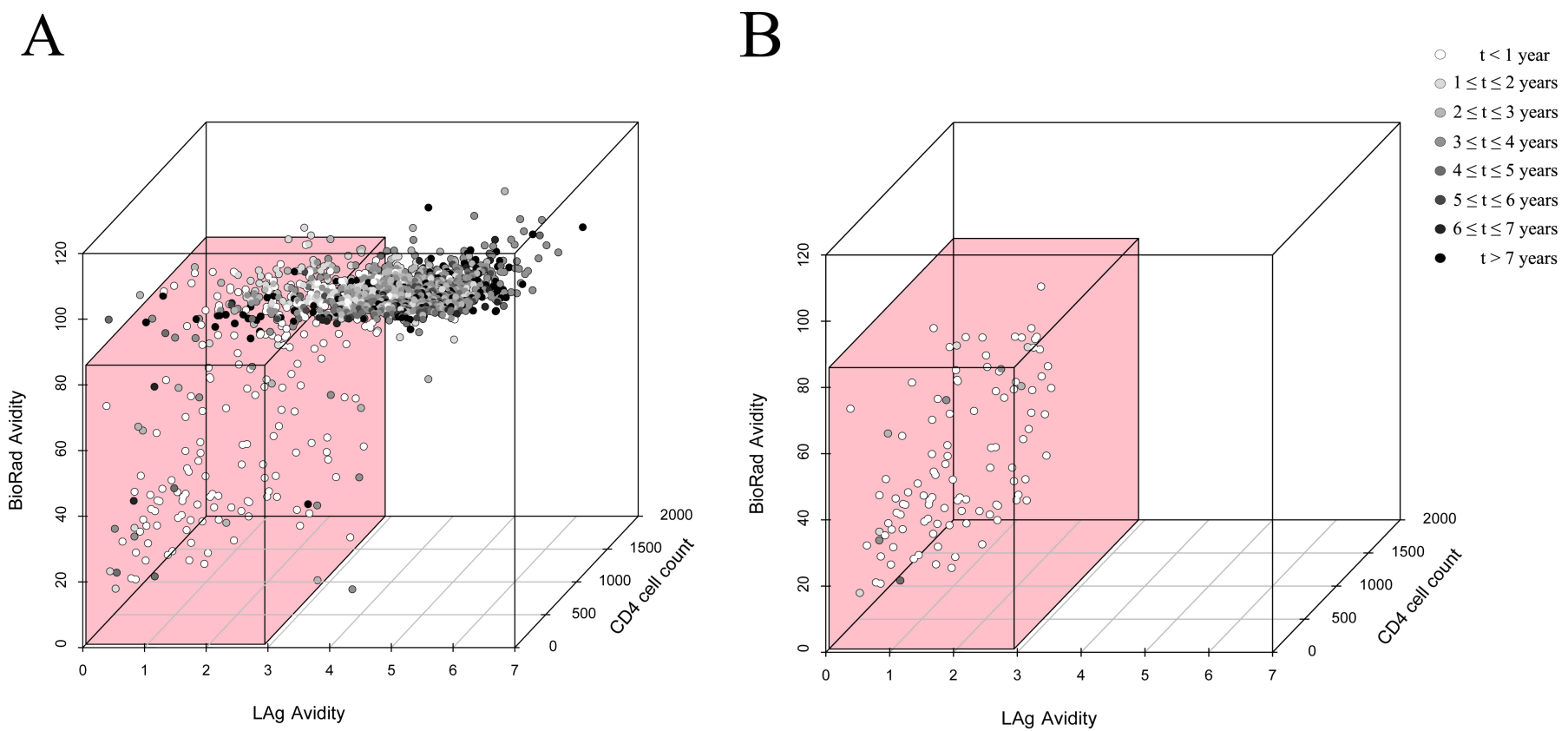
$\hat{\mu}$ vs. $\hat{\psi}$



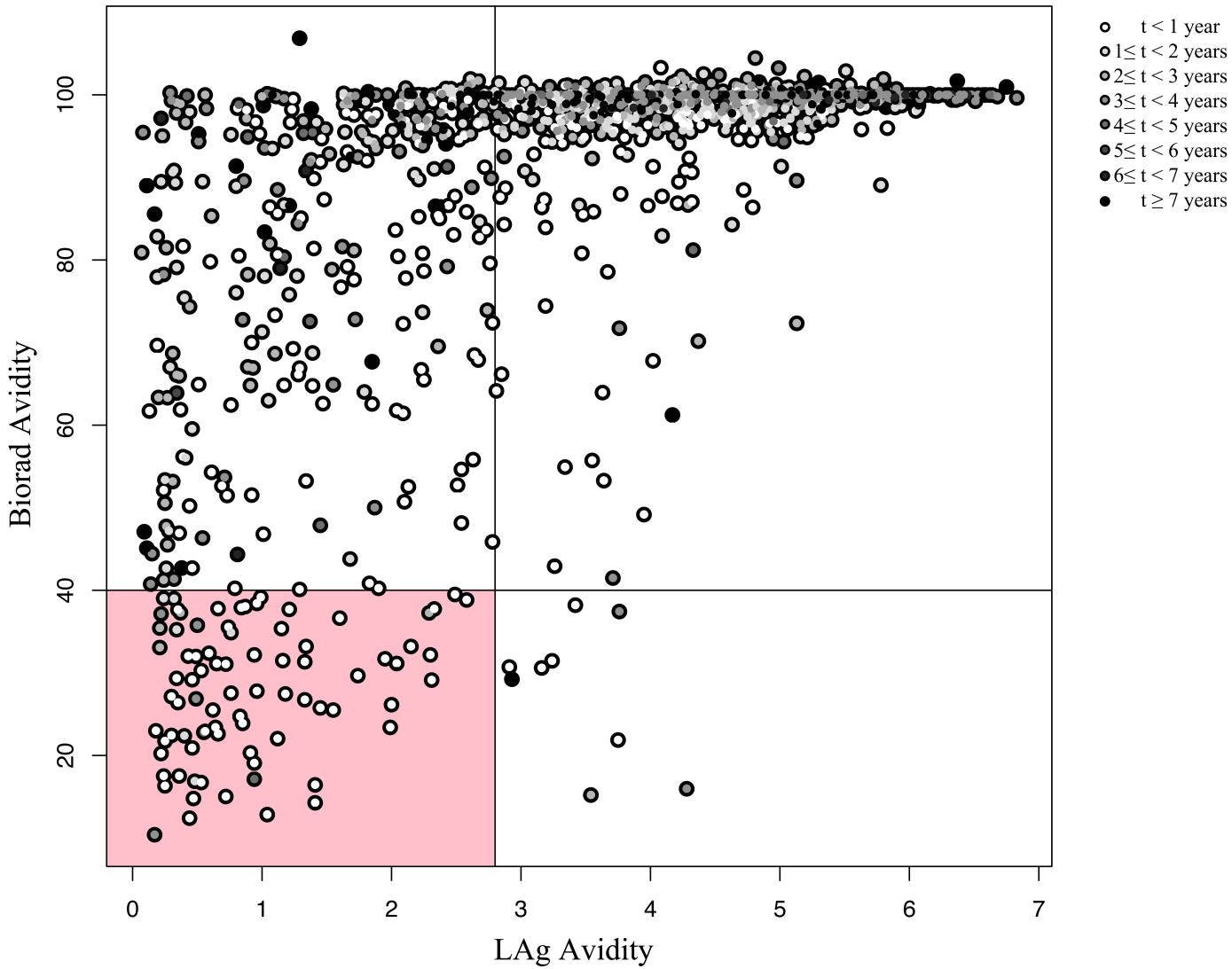
MAAs


MAA1
CD4 cell count
≤50
MAA Negative
>50
BioRad-Avidity
≥85
MAA Negative
<85
LAg-Avidity
≥2.9
MAA Negative
<2.9
Viral load
≤400
MAA Negative
>400
MAA Positive
$\hat{\mu} = 0.40$ years $(0.33, 0.47)$
$\hat{\psi} = 0.49$ years $(0.39, 0.63)$
MAA2
BioRad-Avidity
≥40
MAA Negative
<40
LAg-Avidity
≥2.8
MAA Negative
<2.8
MAA Positive
$\hat{\mu} = 0.33$ years $(0.26, 0.39)$
$\hat{\psi} = 0.68$ years $(0.43, 0.97)$
MAA3
CD4 cell count
≤350
MAA Negative
>350
BioRad-Avidity
≥85
MAA Negative
<85
LAg-Avidity
≥2.9
MAA Negative
<2.9
Viral load
≤600
MAA Negative
>600
MAA Positive
$\hat{\mu} = 0.34$ years $(0.26, 0.41)$
$\hat{\psi} = 0.33$ years $(0.29, 0.38)$

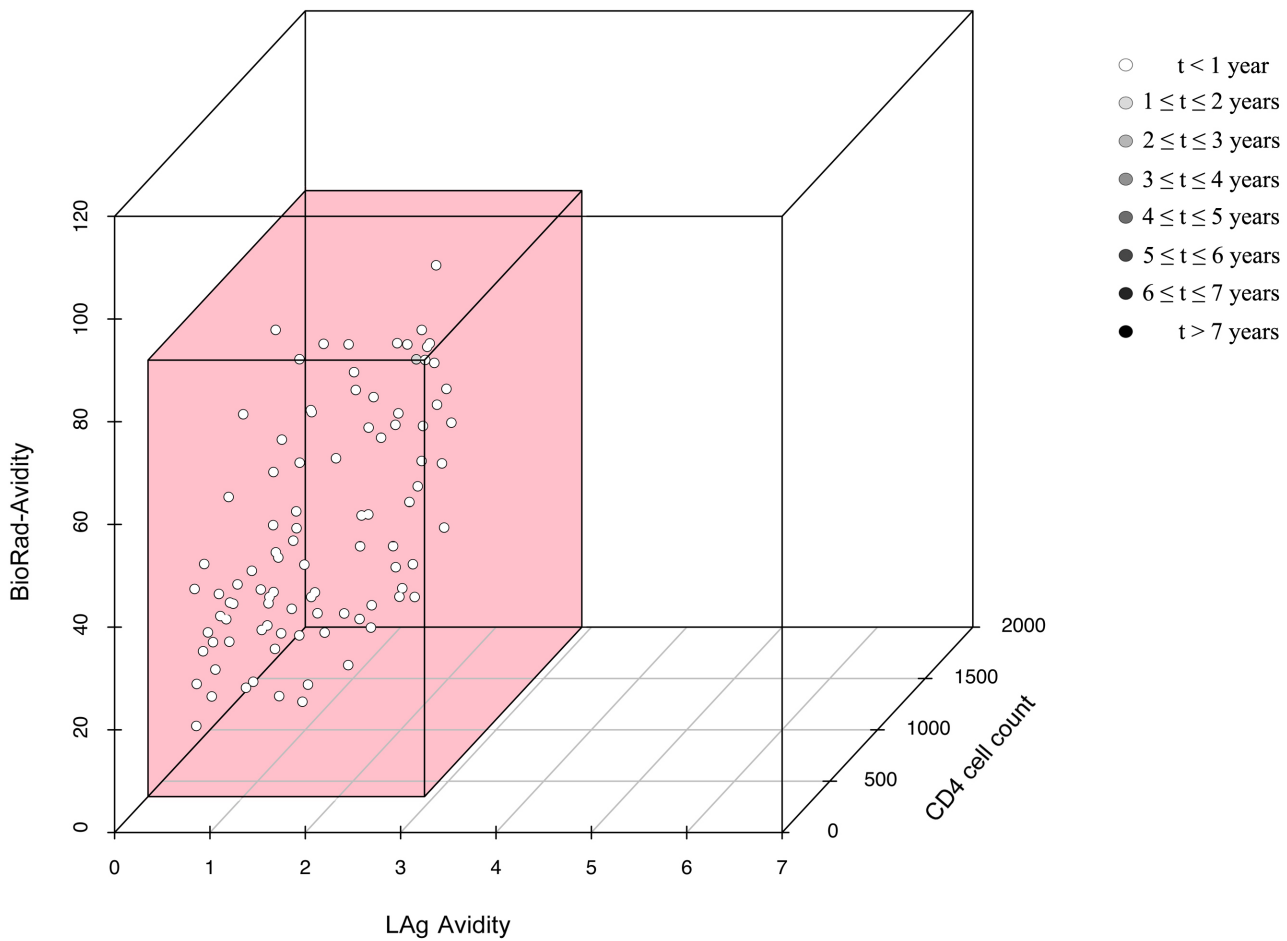
MAA1 Early Disease Stage



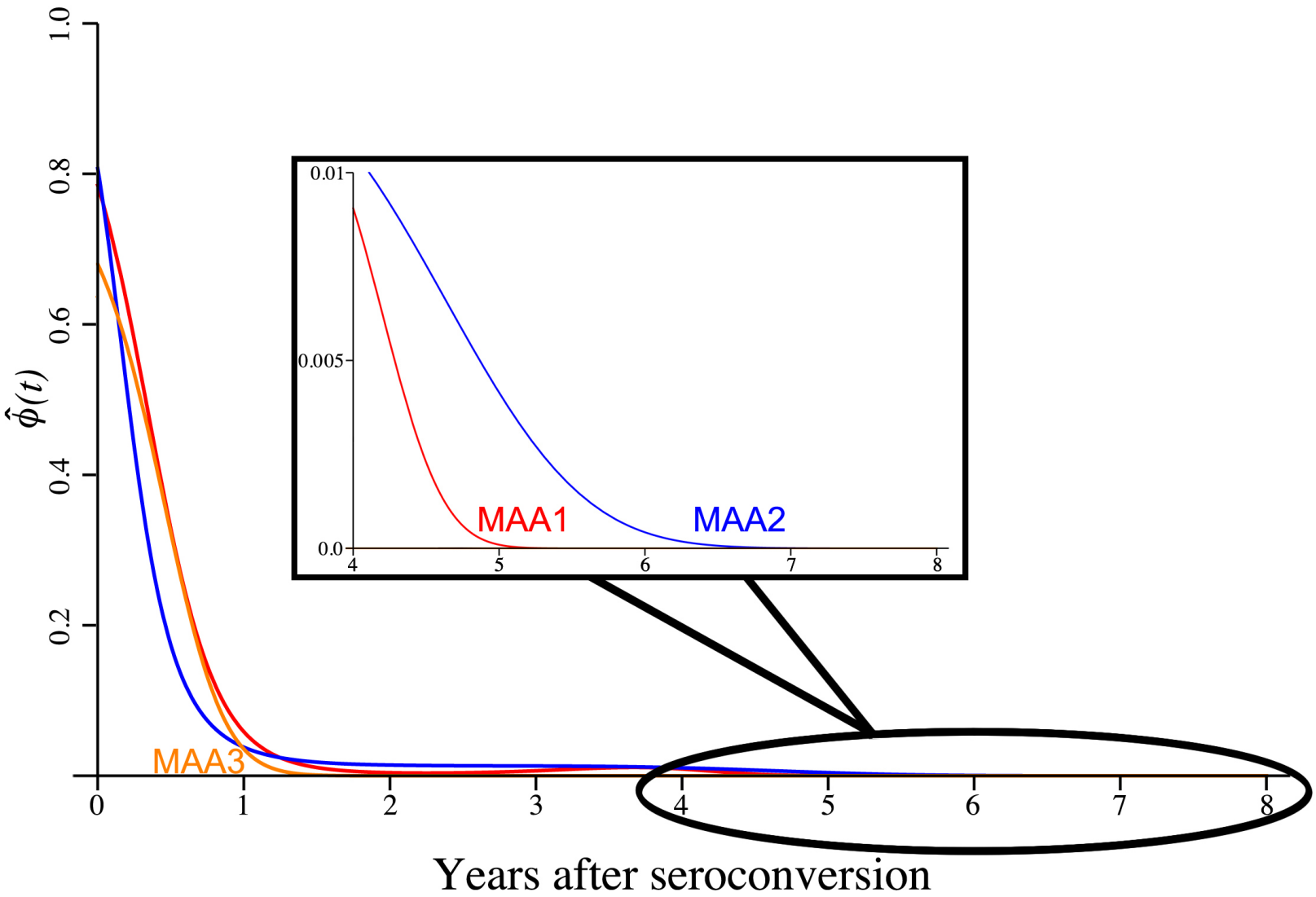
MAA2 Early Disease Stage



MAA3 Early Disease Stage



Comparison of Estimated $\phi(t)$ Curves



Supplemental Dataset

4. Johns Hopkins HIV Clinical Practice Cohort

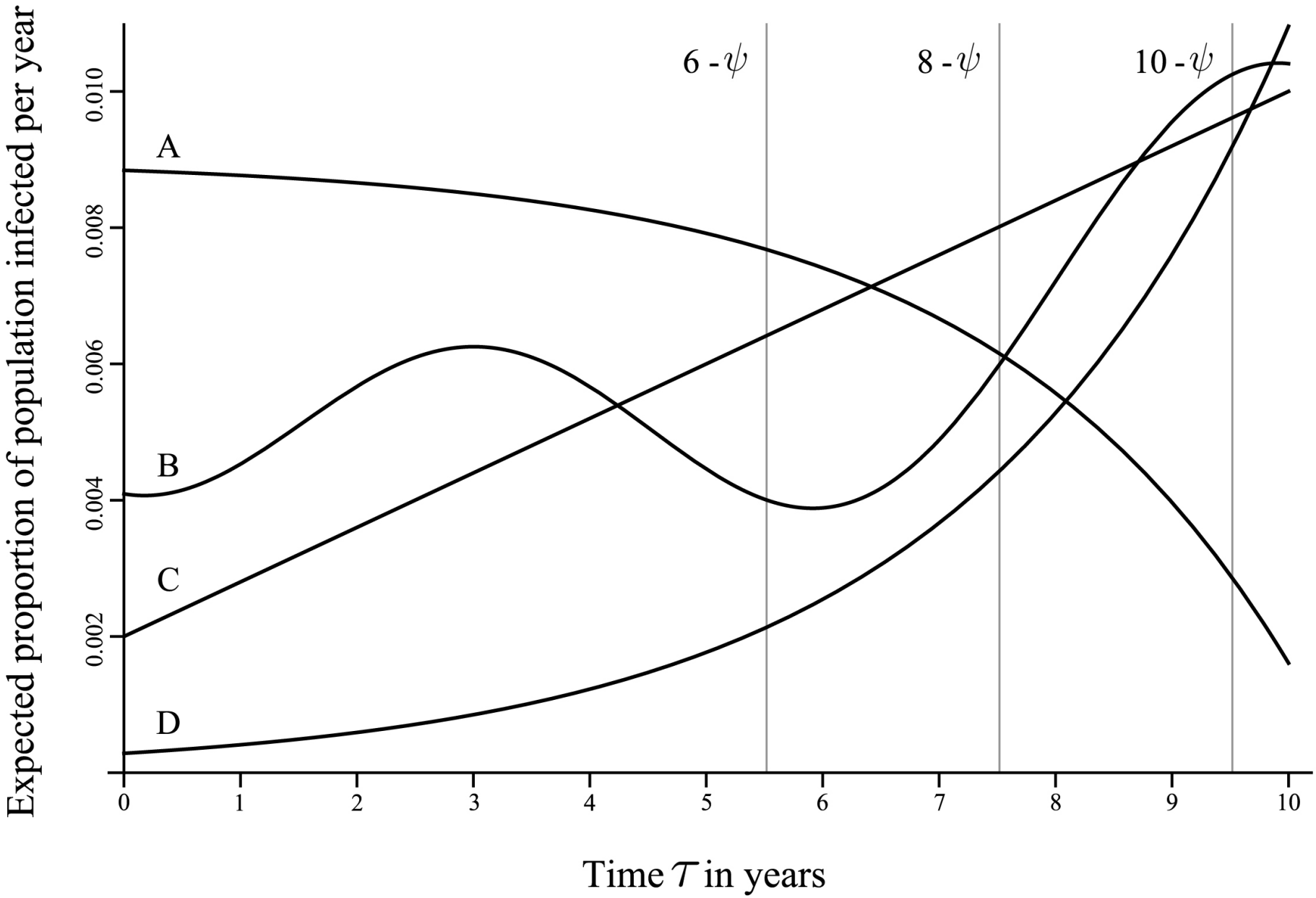
- 500 samples from individuals infected more than 8 years
- Seroconversion times unknown

Choosing an MAA

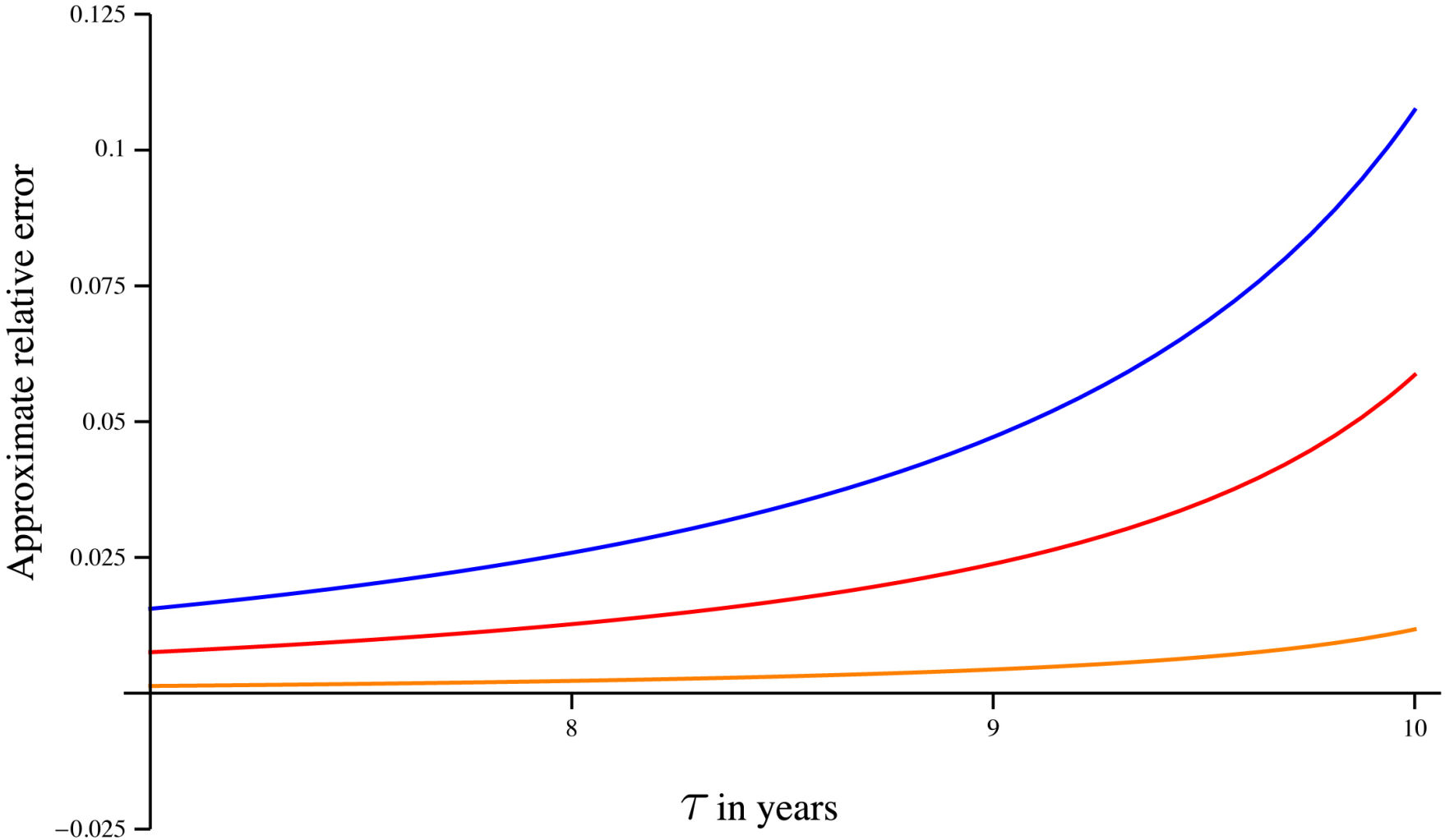
Relative Bias

$$\frac{\hat{I}(\psi) - I(\psi)}{I(\psi)}$$

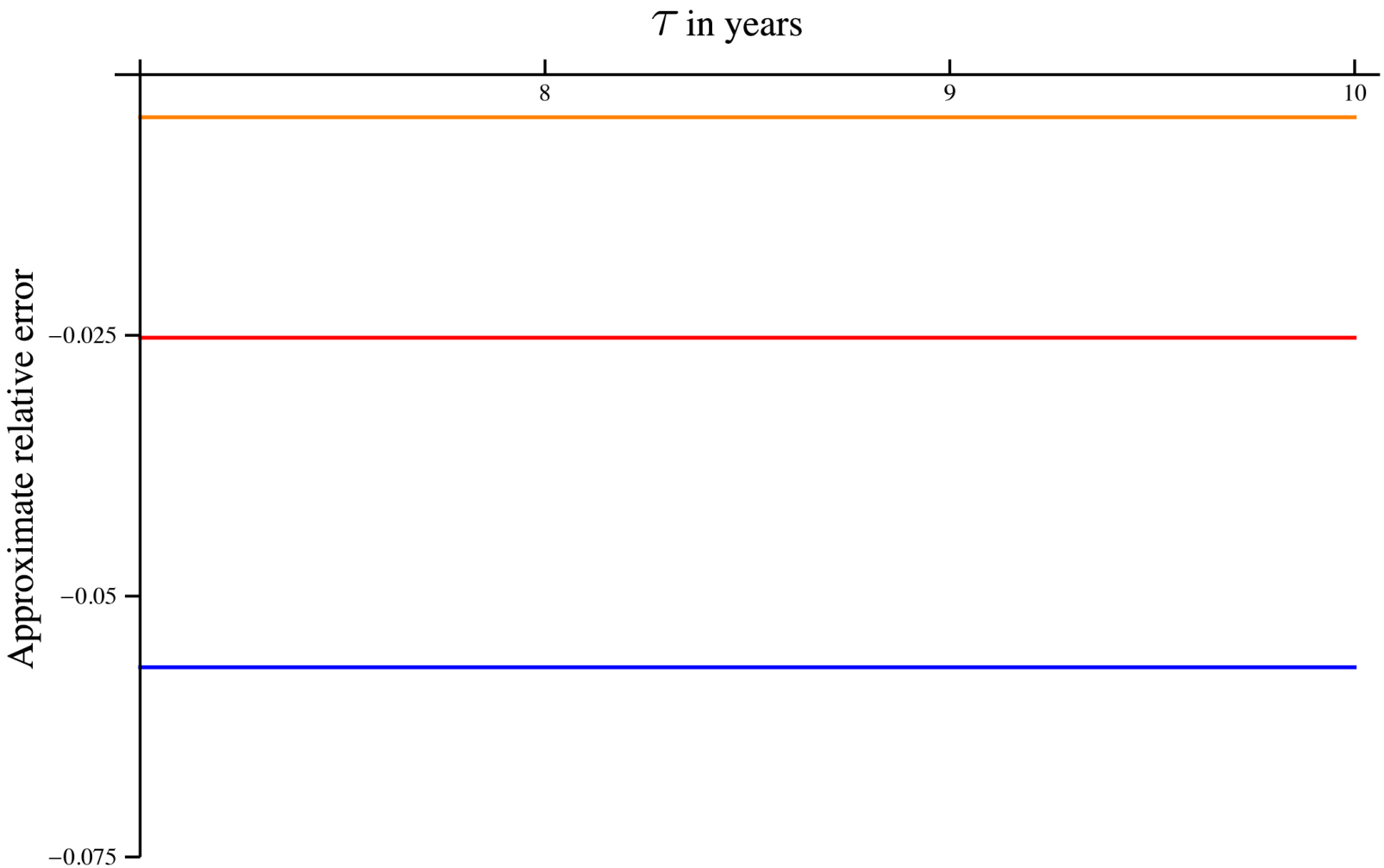
Simulated Epidemics



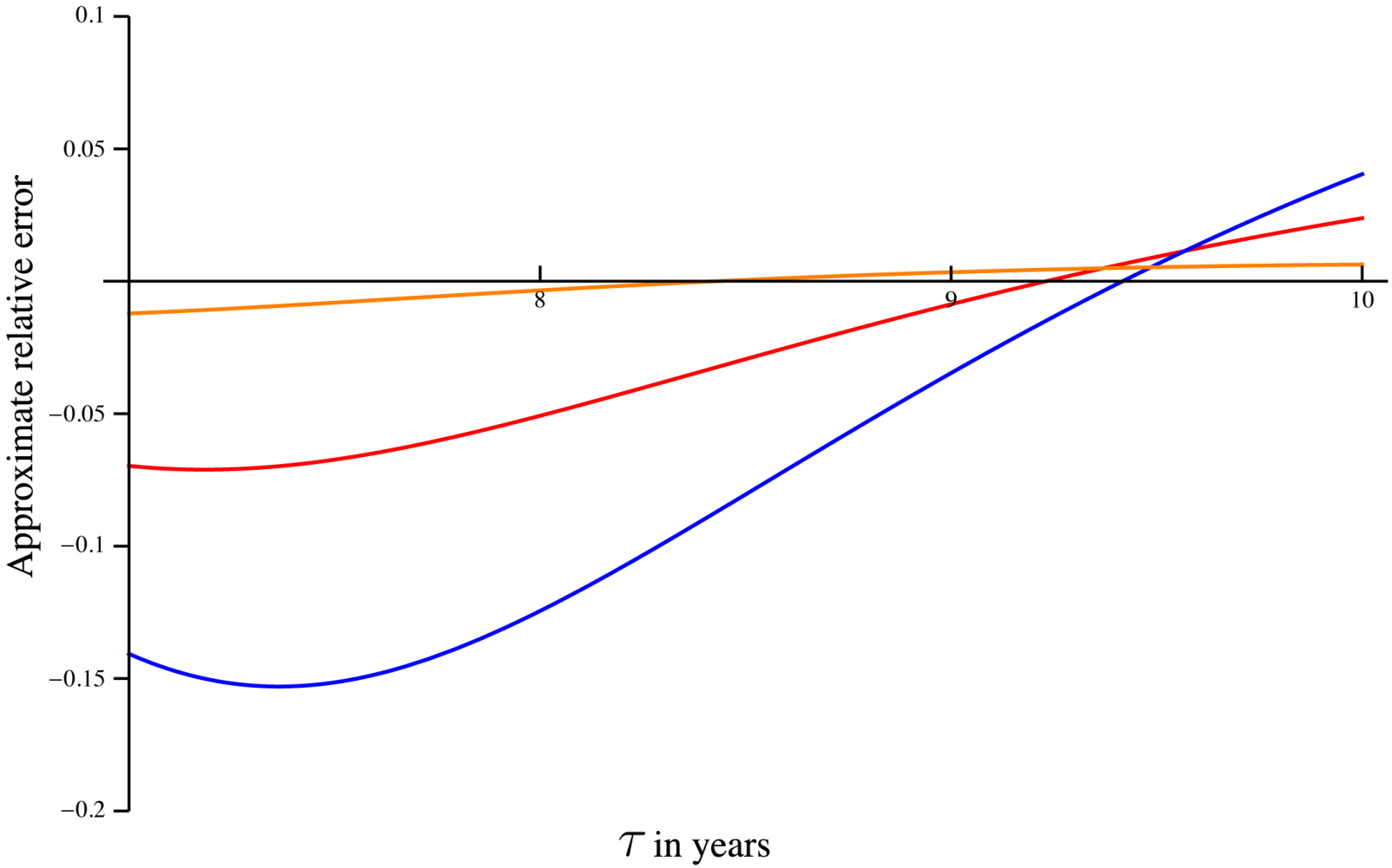
Bias in Epidemic A



Bias in Epidemic B



Bias in Epidemic D



Sample Sizes

Sample Sizes for a Single Survey

Goal: Estimate I with a desired level of precision

Random sample of n individuals

$\rho = W/I$

$W =$ the width of the confidence interval for I

The Distribution of X

We assume $X|\mu \sim \text{Poisson}(n(1-p)I\mu)$

Recall: Prevalence = Incidence $\times$ Duration

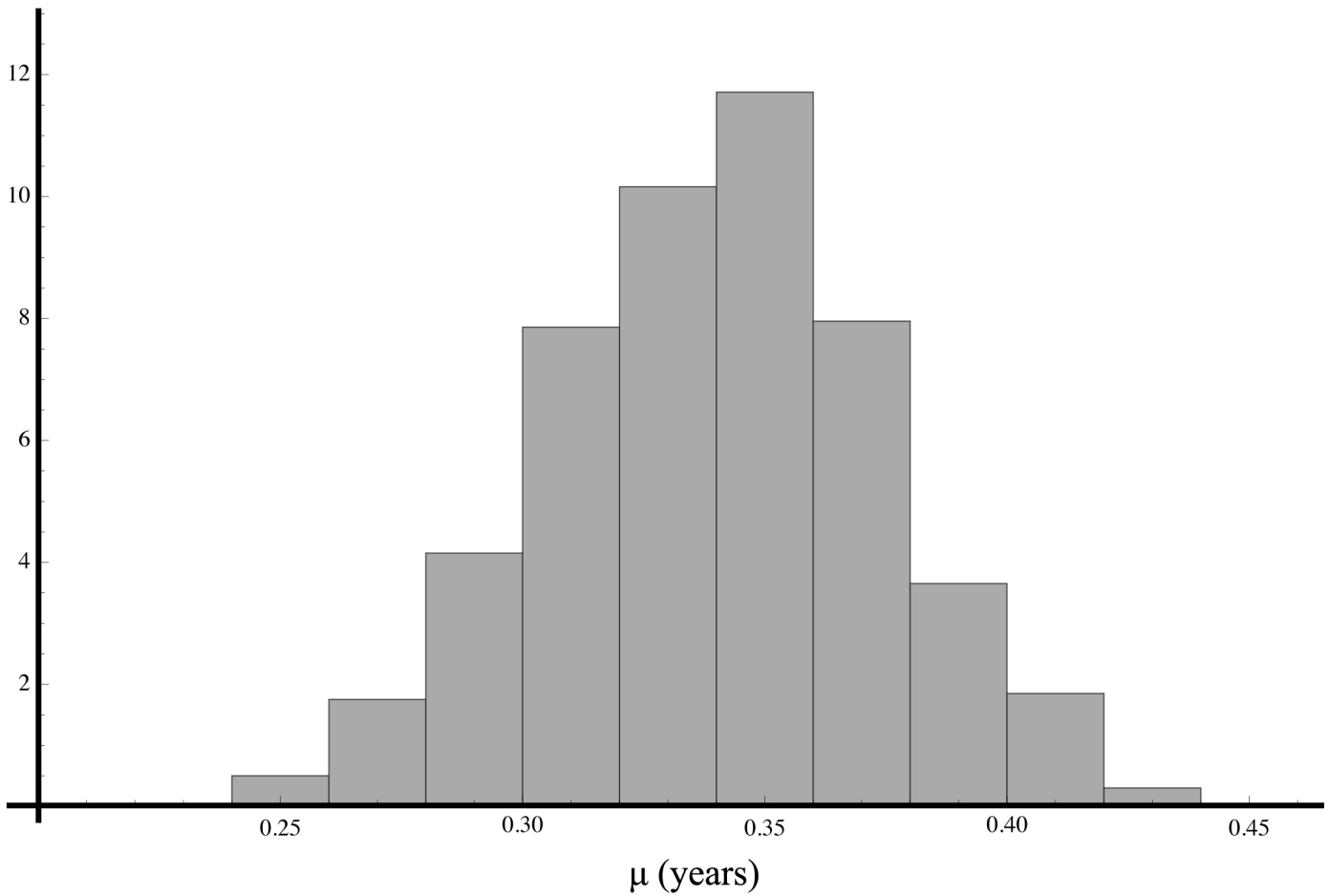
$$X = N_u I \times \mu$$

$E[N_u] = n(1-p)$

$n =$ sample size

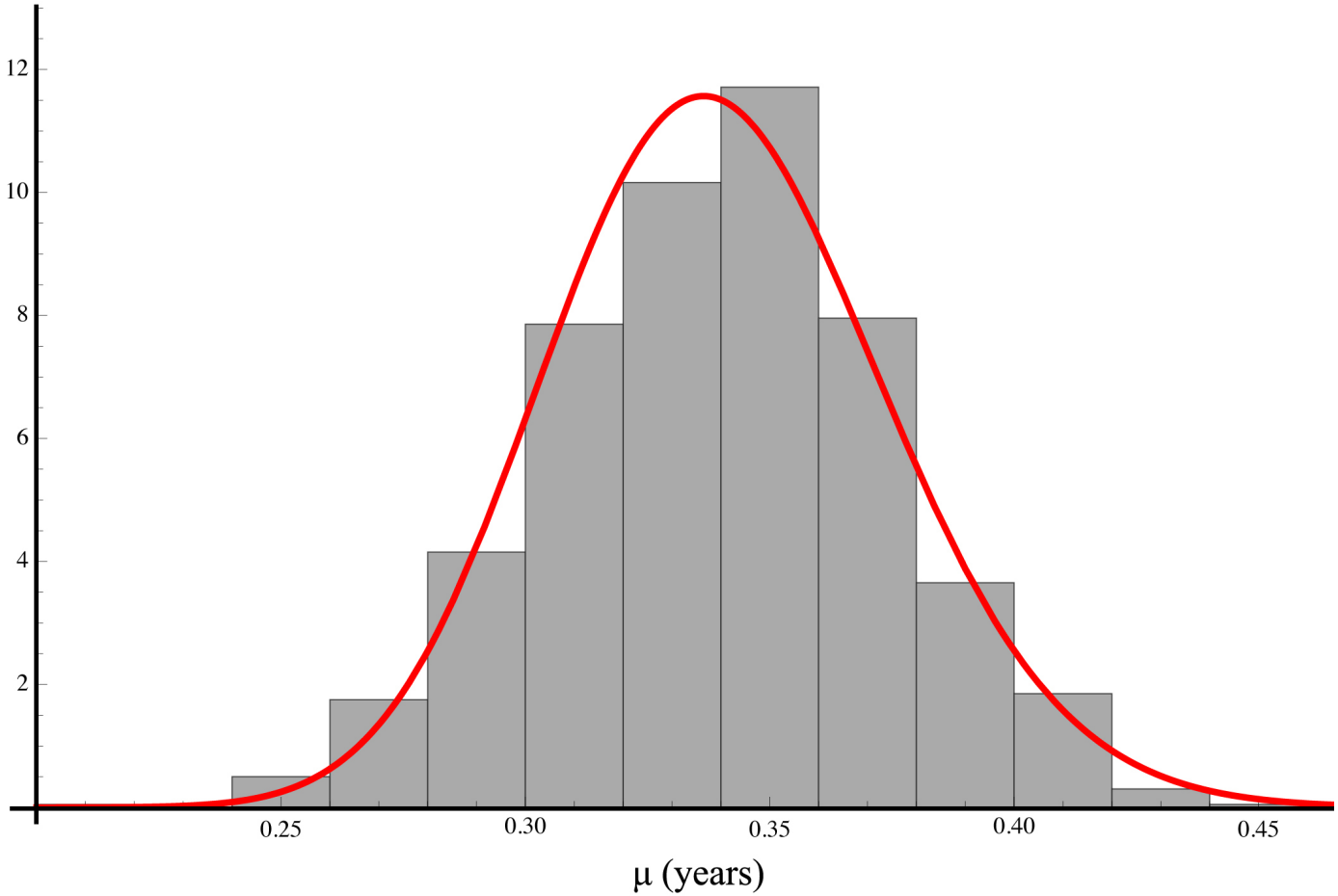
$p =$ HIV prevalence

Accounting for Uncertainty in μ



$$X|\mu \sim \text{Poisson}(n(1-p)I\mu)$$

Accounting for Uncertainty in μ



$$\mu \sim \text{Gamma}(k, \theta)$$

Solving for n

To achieve precision of $\rho = W/I$ we solve for n in

$$\rho = \frac{\left[\frac{1}{\text{Beta}(\alpha/2, k, nI^0(1-p^0)k\theta+1)] } - \frac{1}{\text{Beta}(1-\alpha/2, k, nI^0(1-p^0)k\theta)}\right]}{nI^0(1-p^0)\theta}$$

Solving for n

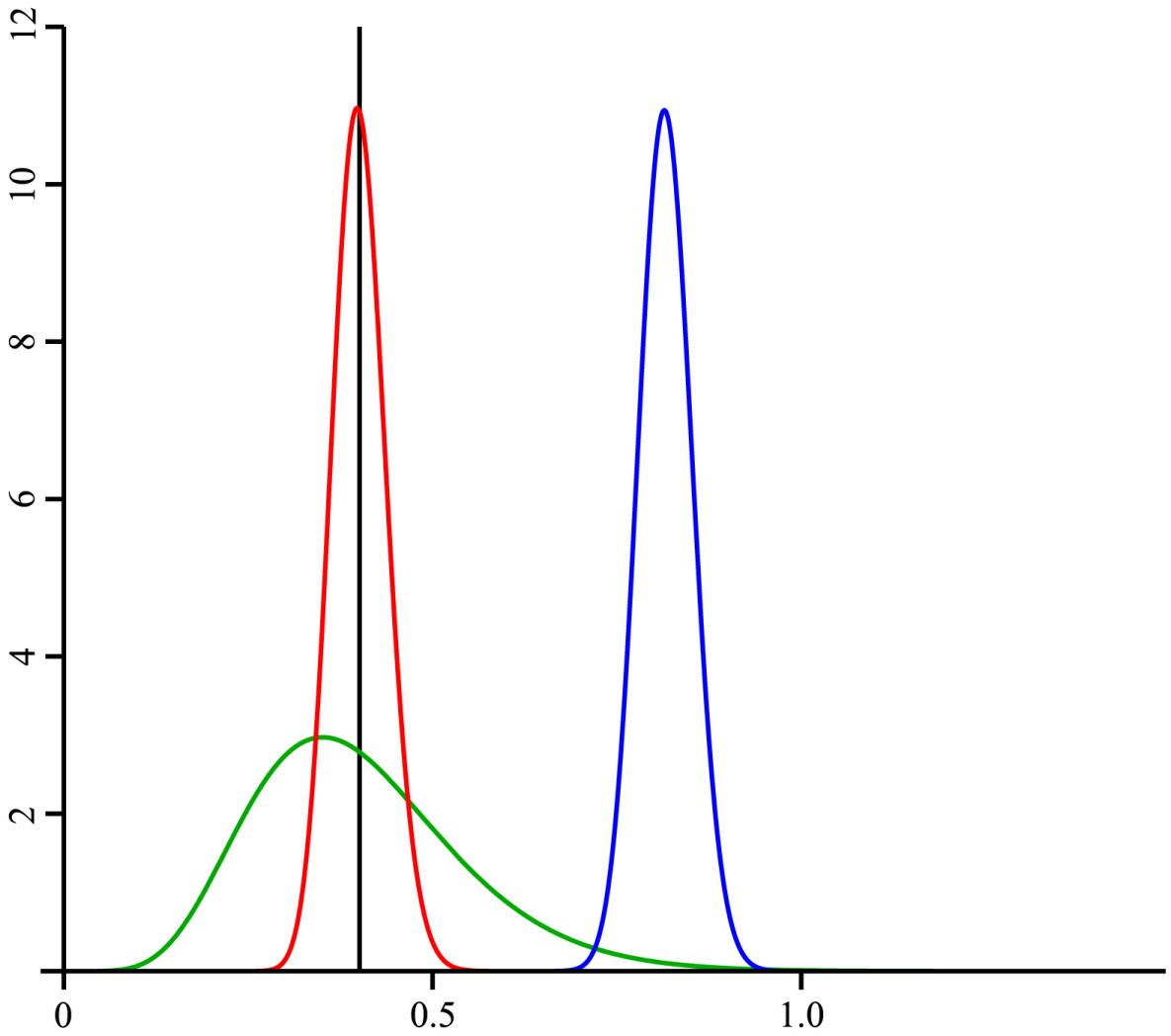
To achieve precision of $\rho = W/I$ we solve for n in

$$\rho = \frac{\left[\frac{1}{\text{Beta}(\alpha/2, k, nI^0(1-p^0)k\theta+1)] - \frac{1}{\text{Beta}(1-\alpha/2, k, nI^0(1-p^0)k\theta)}\right]}{nI^0(1-p^0)\theta}$$

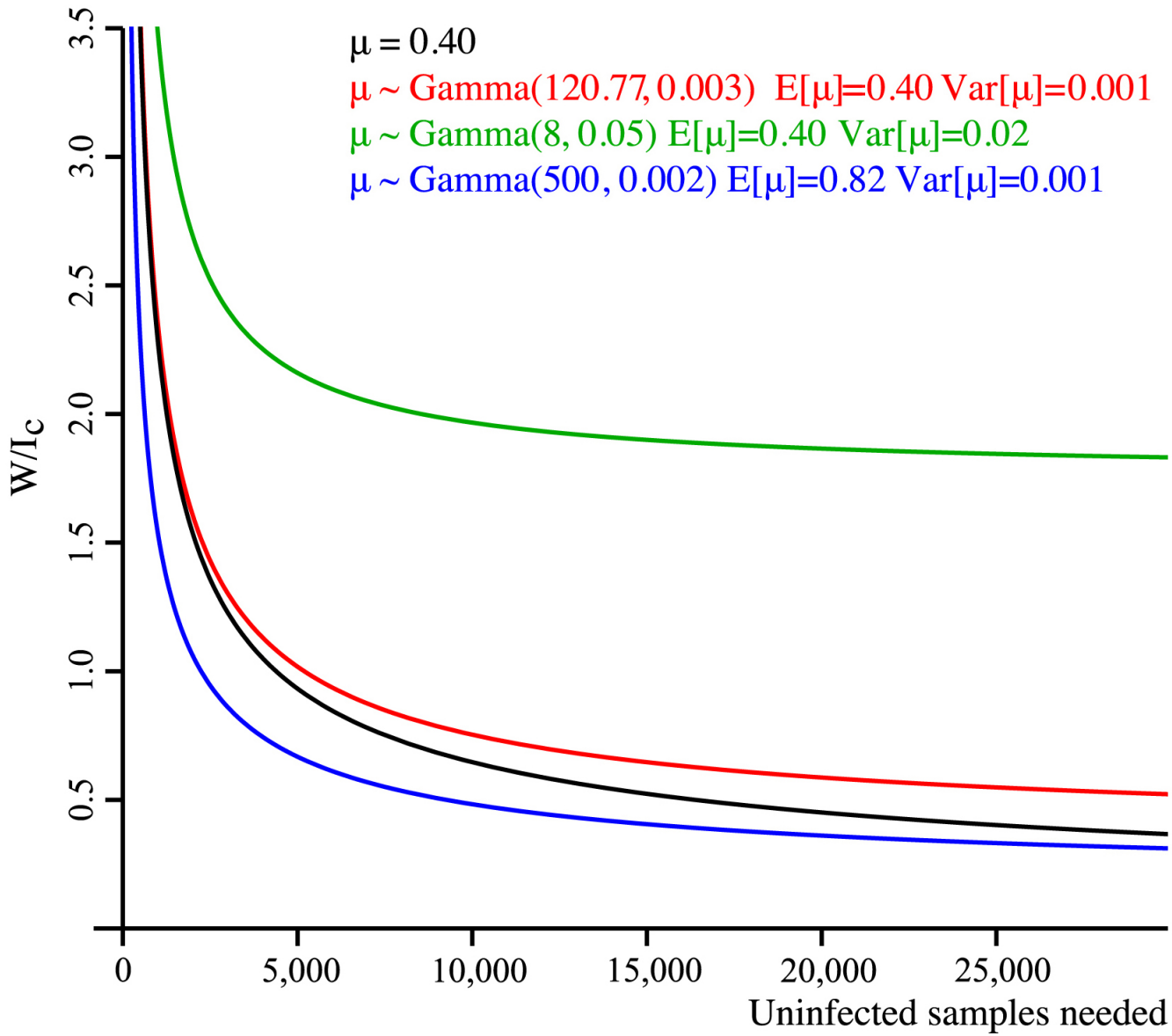
The sample size is inversely proportional to the underlying incidence

Larger sample sizes are needed to estimate smaller incidences

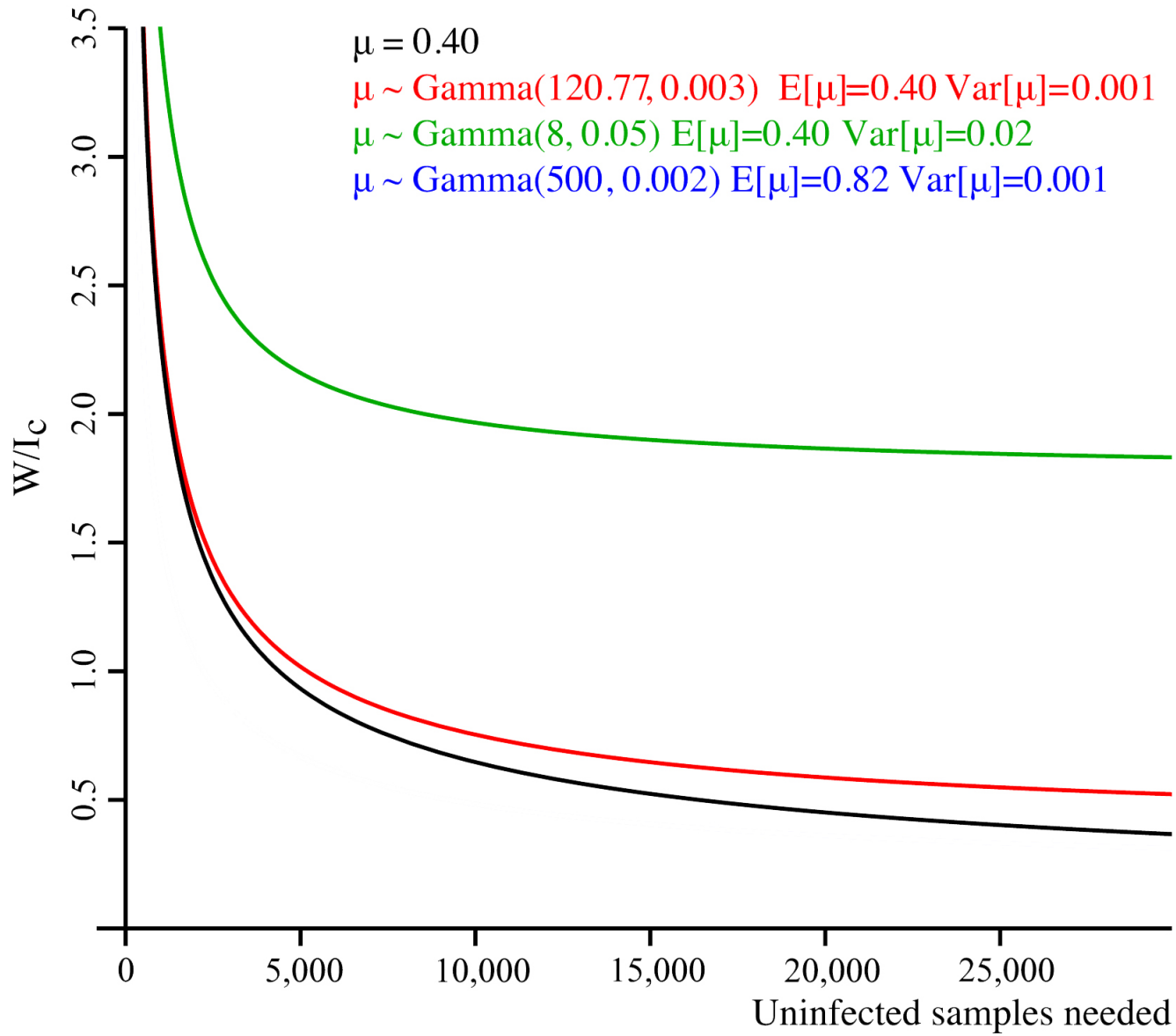
Impact of the Prior for μ



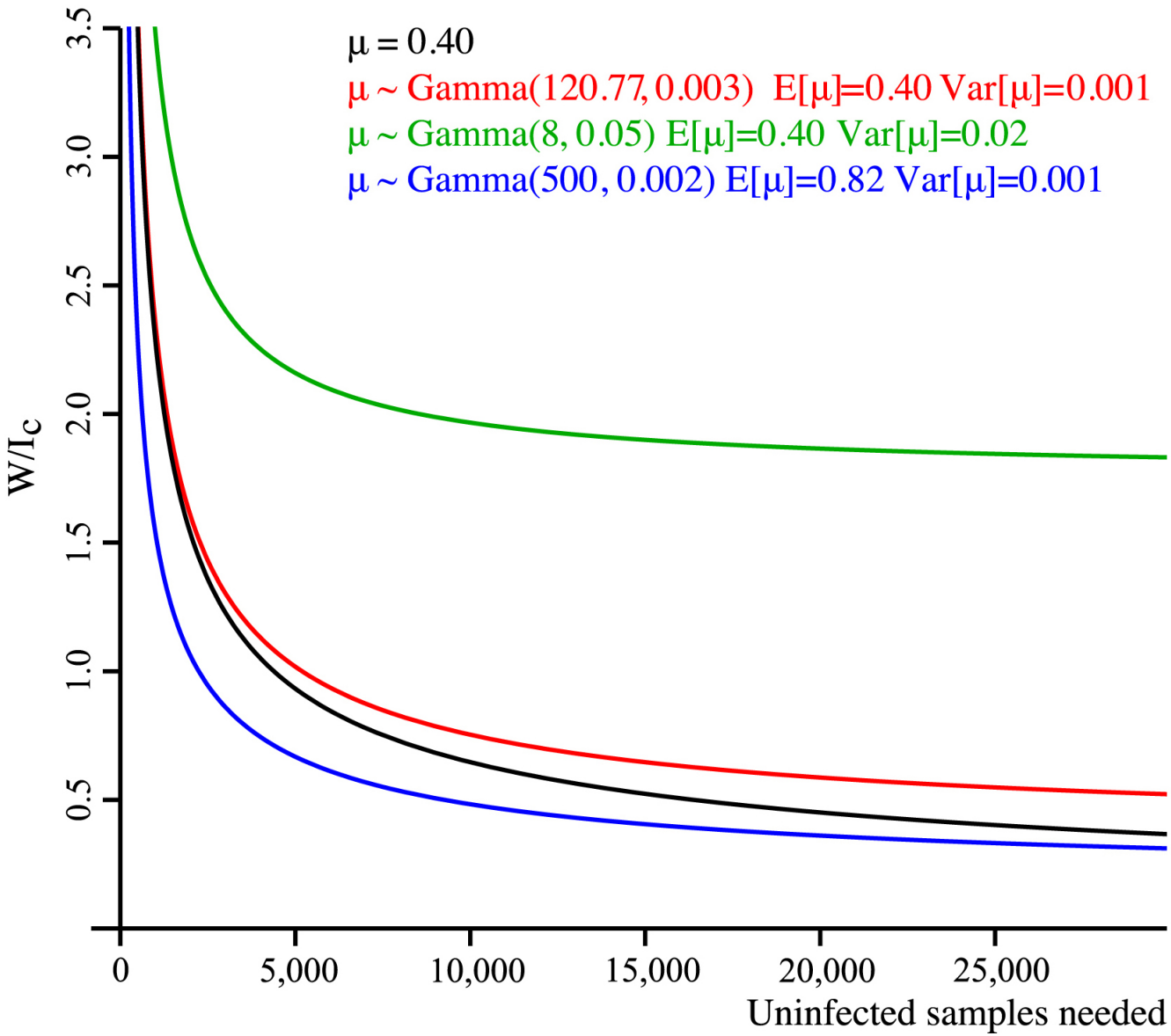
μ's Influence on Sample Size



Distributions of μ



Distributions of μ



Sample Sizes

Uninfected samples needed for a desired precision

I_c	W/I_c						
	0.50	1	1.5	2	2.5	3	3.5
0.25%	141,895	20,768	9,115	5,338	3,602	2,648	2,060
0.50%	70,948	10,384	4,558	2,669	1,801	1,324	1,030
1.00%	35,474	5,192	2,279	1,335	901	662	515
1.50%	23,650	3,462	1,520	890	601	442	344
2.00%	17,737	2,596	1,140	668	451	331	258
2.50%	14,190	2,077	912	534	361	265	206
3.00%	11,825	1,731	760	445	301	221	172
4.00%	8,869	1,298	570	334	226	166	129
5.00%	7,095	1,039	456	267	181	133	103

Consecutive Cross-Sectional Surveys

Surveys are conducted at calendar times t_1 and t_2 and estimate incidences I_1 and I_2

Test $H_0 : I_2/I_1 = 1$ against $H_A : I_2/I_1 = r > 1$

Sample sizes are a function of design parameter

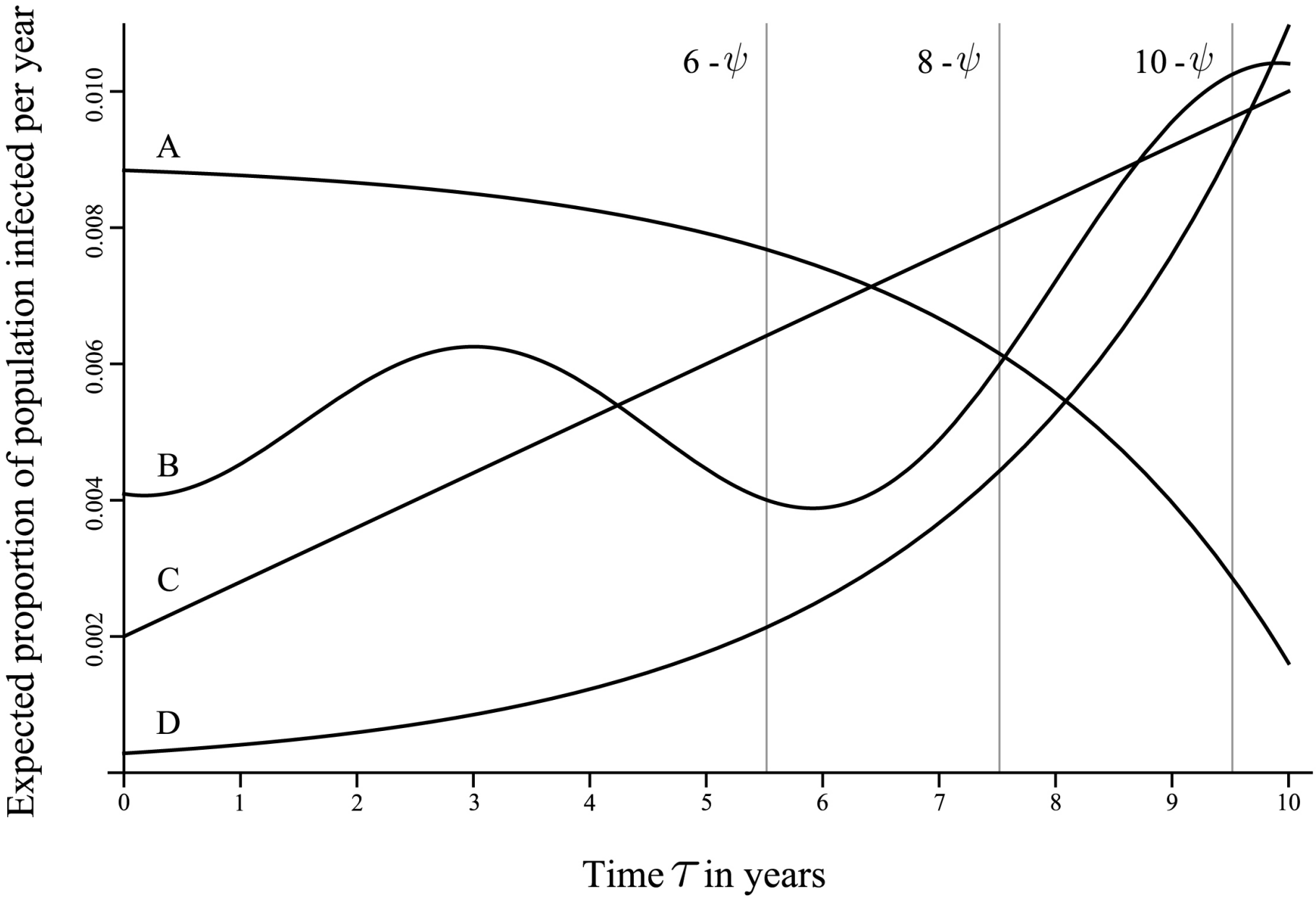
$$s \approx \frac{n_2}{n_1}$$

Sample Sizes for Successive Surveys

Needed uninfected samples at the first time point multiplied by initial incidence

	$r = 1.2$	$r = 1.5$	$r = 2$	$r = 3$	$r = 5$
$s = 2$					
i	898.7 (2696.0)	166.4 (499.0)	50.5 (151.5)	17.0 (50.8)	6.2 (18.6)
ii	907.7 (2723.1)	168.2 (504.5)	51.1 (153.2)	17.2 (51.4)	6.3 (18.8)
iii	1049.0 (3147.0)	195.2 (585.6)	59.6 (178.7)	20.2 (60.4)	7.5 (22.3)
iv	442.8 (1328.2)	82.0 (246.0)	24.9 (74.7)	8.4 (25.1)	3.1 (9.2)
$s = 1$					
i	1196.6 (2393.1)	221.2 (442.3)	67.4 (134.8)	22.2 (44.4)	8.2 (16.3)
ii	1208.8 (2417.5)	223.5 (447.0)	68.1 (136.2)	22.5 (44.9)	8.3 (16.5)
iii	1396.2 (2792.3)	259.2 (518.4)	79.0 (158.0)	26.3 (52.6)	9.8 (19.6)
iv	589.6 (1179.1)	109.0 (218.0)	33.2 (66.4)	11.0 (21.9)	4.1 (8.1)
$s = \frac{1}{2}$					
i	1793.3 (2689.9)	331.1 (496.6)	100.4 (150.6)	33.8 (50.6)	12.1 (18.1)
ii	1811.0 (2716.5)	334.3 (501.5)	101.5 (152.2)	34.1 (51.1)	12.2 (18.3)
iii	2090.3 (3135.4)	386.6 (579.8)	117.7 (176.5)	39.3 (59.0)	14.1 (21.1)
iv	883.4 (1325.1)	163.1 (244.7)	49.5 (74.2)	16.7 (25.0)	6.0 (8.9)

Simulated Epidemics



Simulation Results

	True Incidence	ASI	CI coverage	Power	u
(1)					
Epidemic A	0.80% and 0.30%	0.80% and 0.29%	0.94 and 0.92	0.92	9,984
Epidemic B	0.41% and 1.08%	0.42% and 1.05%	0.95 and 0.95	0.87	7,595
Epidemic C	0.66% and 1.01%	0.65% and 1.01%	0.95 and 0.95	0.90	29,557
Epidemic D	0.21% and 0.94%	0.22% and 0.96%	0.91 and 0.96	0.90	4,848
(2)					
Epidemic A	0.65% and 0.30%	0.64% and 0.29%	0.95 and 0.92	0.92	17,920
Epidemic B	0.62% and 1.08%	0.65% and 1.05%	0.96 and 0.94	0.82	18,383
Epidemic C	0.83% and 1.01%	0.83% and 1.01%	0.95 and 0.95	0.90	123,545
Epidemic D	0.45% and 0.94%	0.45% and 0.96%	0.95 and 0.96	0.90	12,960

Conclusions

- Cross-sectional studies measure incidence in the past
- Need to properly calibrate μ and ψ
- Multiple biomarkers offer a viable solution
- Account for uncertainty in μ in sample size methods
- Sample sizes can be prohibitively large

Acknowledgments

- ALIVE
- HIVNET 001
- MACS
- JHHCC

Dr. Oliver Laeyendecker

Dr. Susan Eshleman

Dr. Ron Brookmeyer