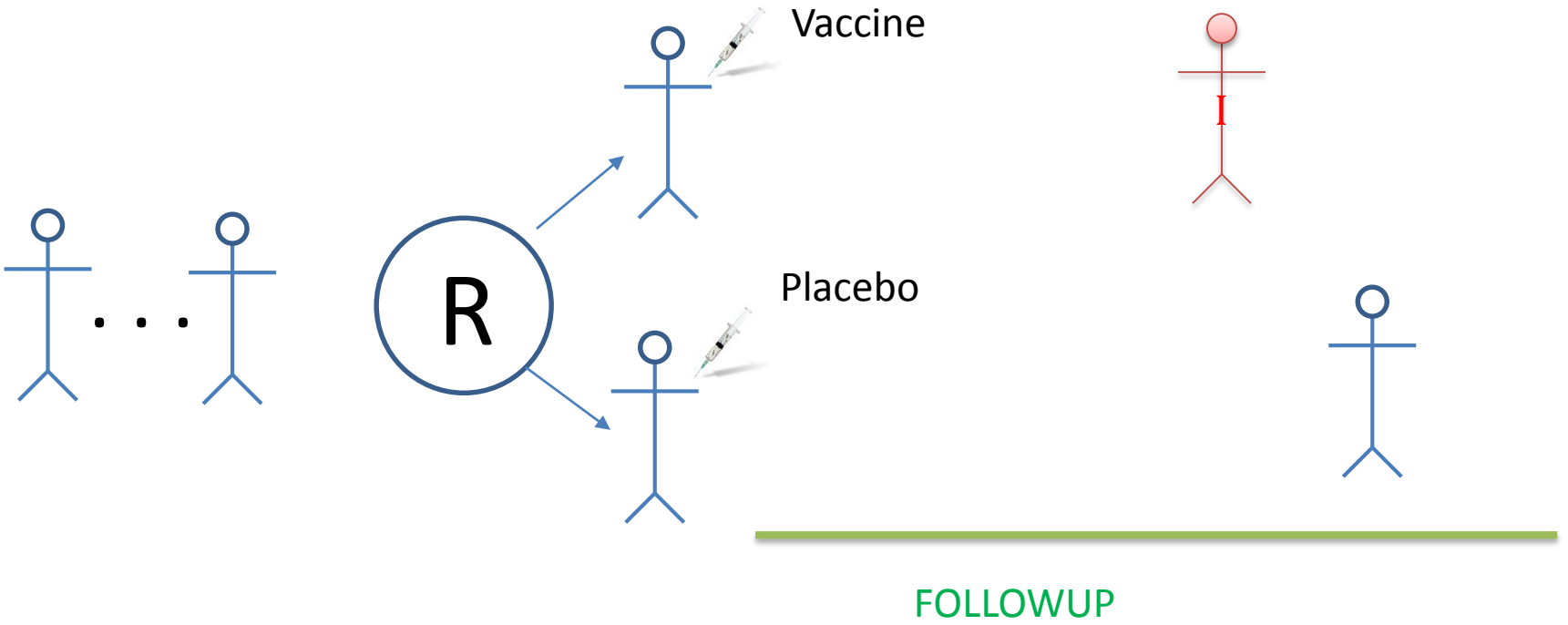
Incorporating Infecting Pathogen Counts In Vaccine Trials

Dean Follmann

National Institute of Allergy and Infectious Diseases

Vaccine Trial

- Randomize at risk healthy volunteers to vaccine or placebo
- Follow them & count significant infections

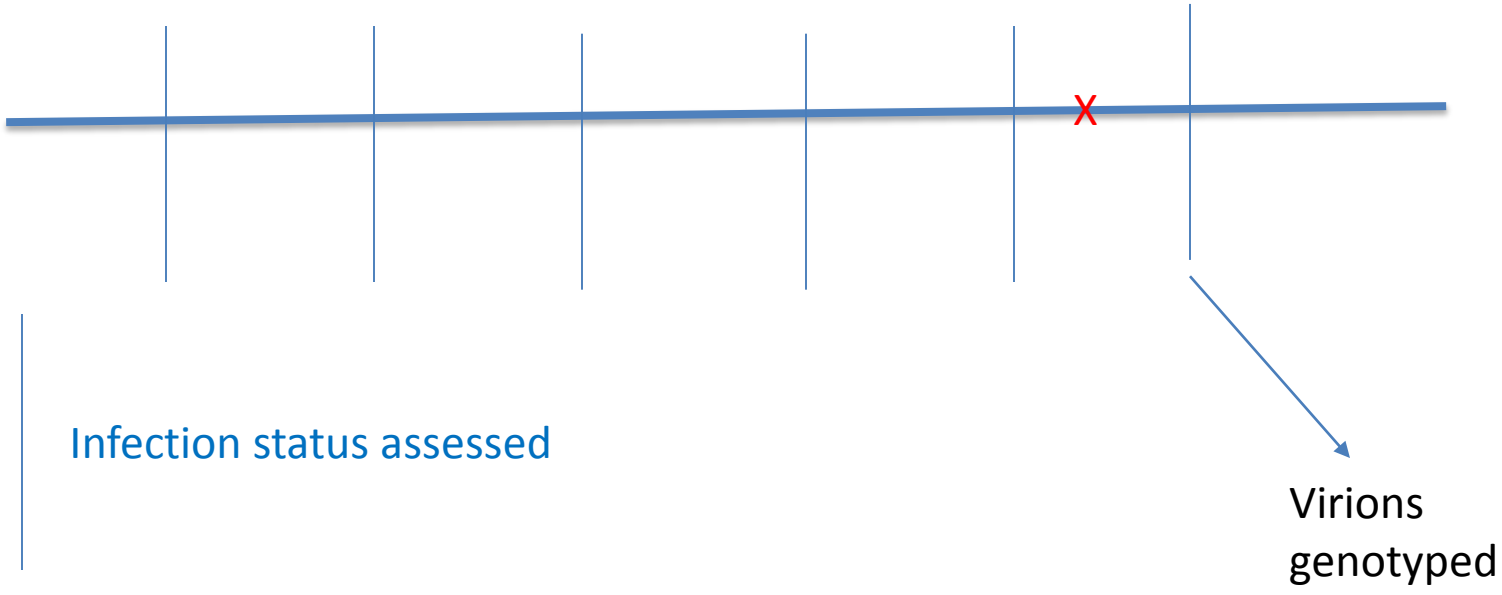


Vaccine Efficacy (VE)

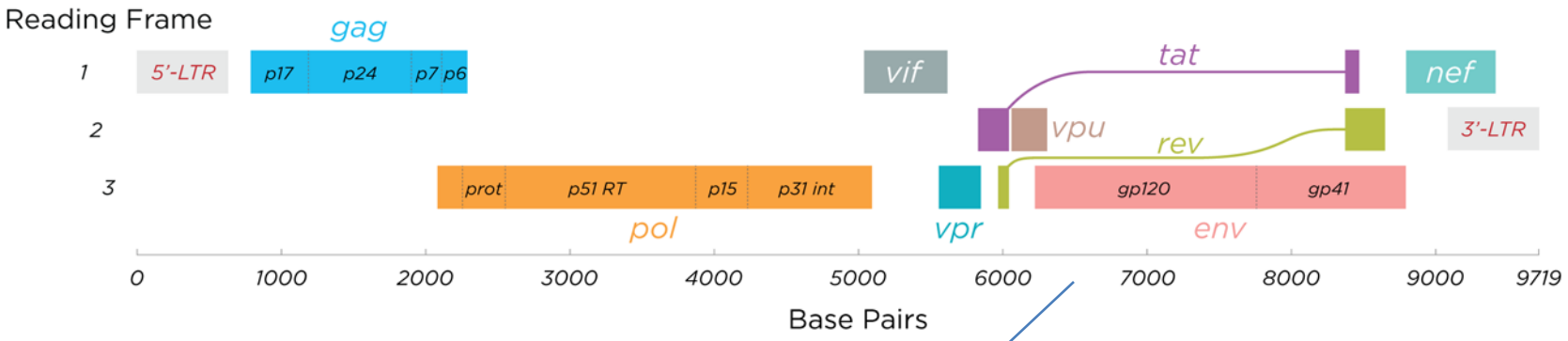
- What is the proportion reduction in some outcome on vaccine compared to placebo?
- $VE = 1 - \frac{Infection\ Rate\ on\ Vaccine}{Infection\ Rate\ on\ Placebo}$
- $VE = 1 - \frac{hazard\ rate\ on\ vaccine}{hazard\ rate\ on\ palcebo}$
- Based on human infection yes/no . . .

HIV Infection Detection

- Volunteers are followed at regular intervals (e.g. 6 months for infection)

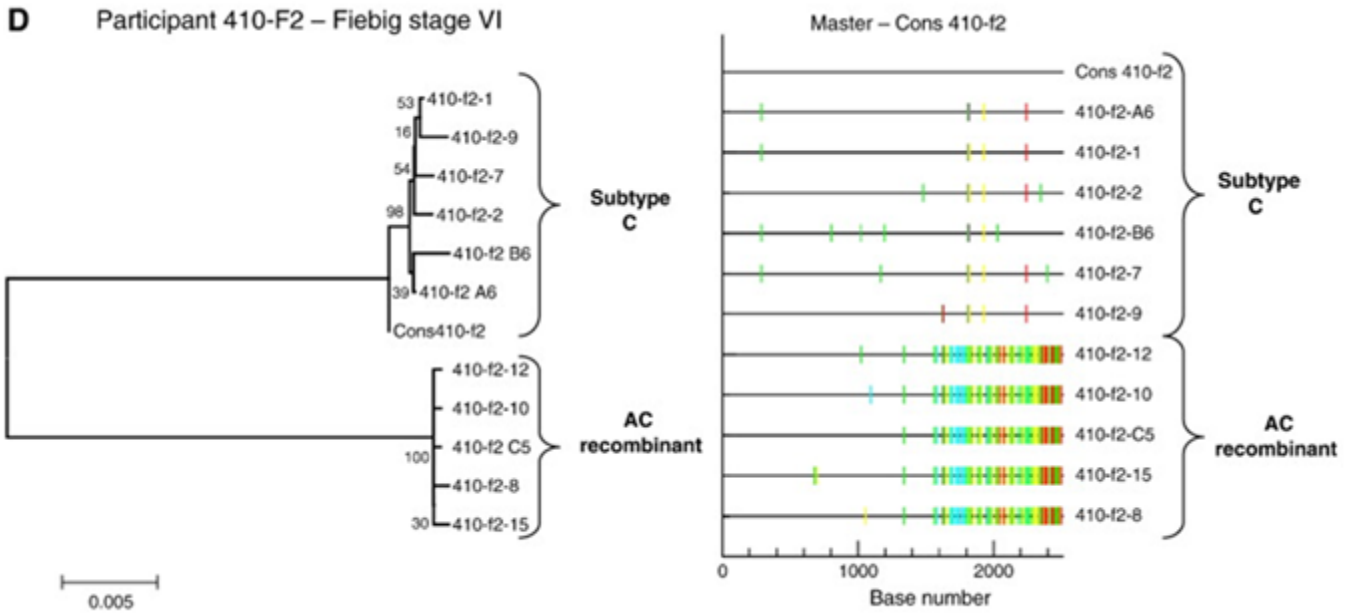


The swarm of HIV virions in an infected individual are not genetically identical



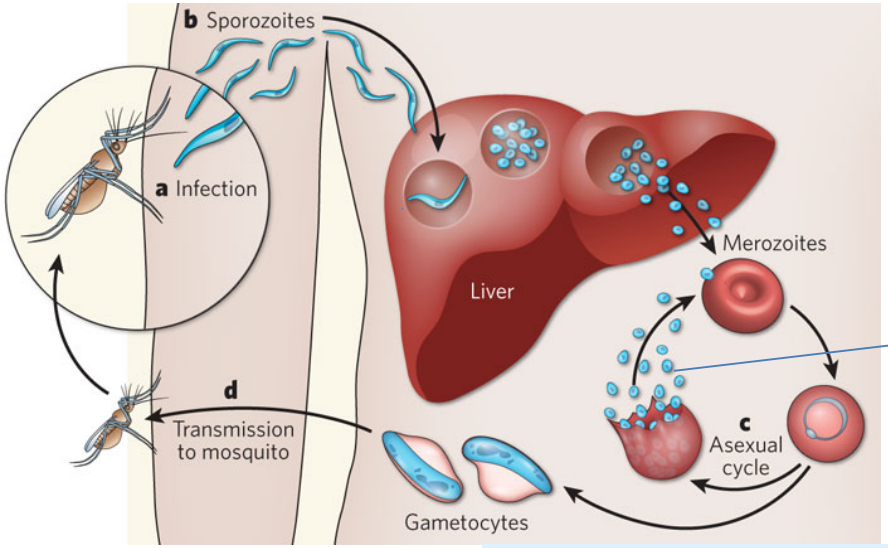
Virion 1	A T C T A T
Virion 2	A T G G C T
Virion 3	T T C T A T
CONSENSUS	A T C T A T

Founder Viruses Tell More Than Infection Yes/No



X=2

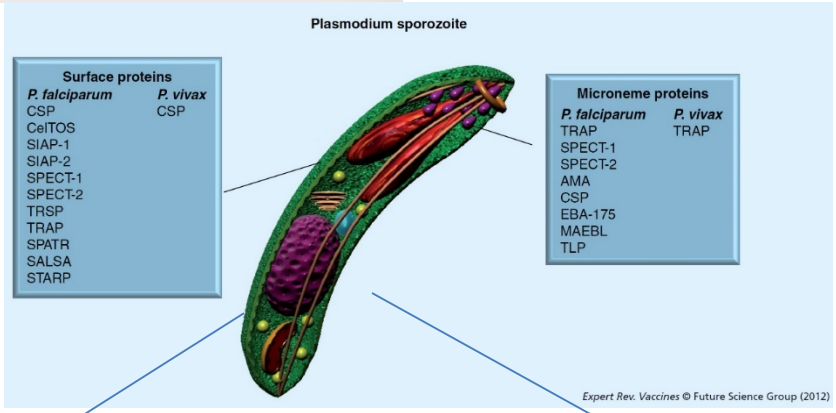
Malaria Sampling



Malaria life cycle

Sample blood stage parasites
PCR amplification of CS region
Then Next Gen sequencing.

NRNAN . . . EW
NRNEN . . . TW

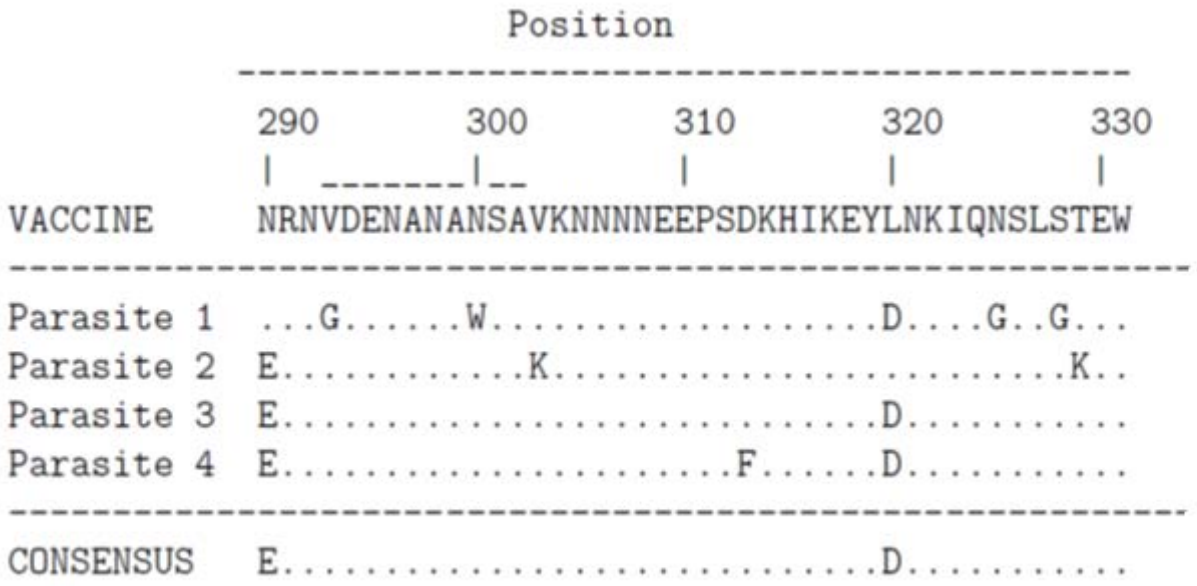


290 300 310 320 330

NRNVDENANANSAVKNNNNEEPSDKHIKEYLNKIQNSLSTEW

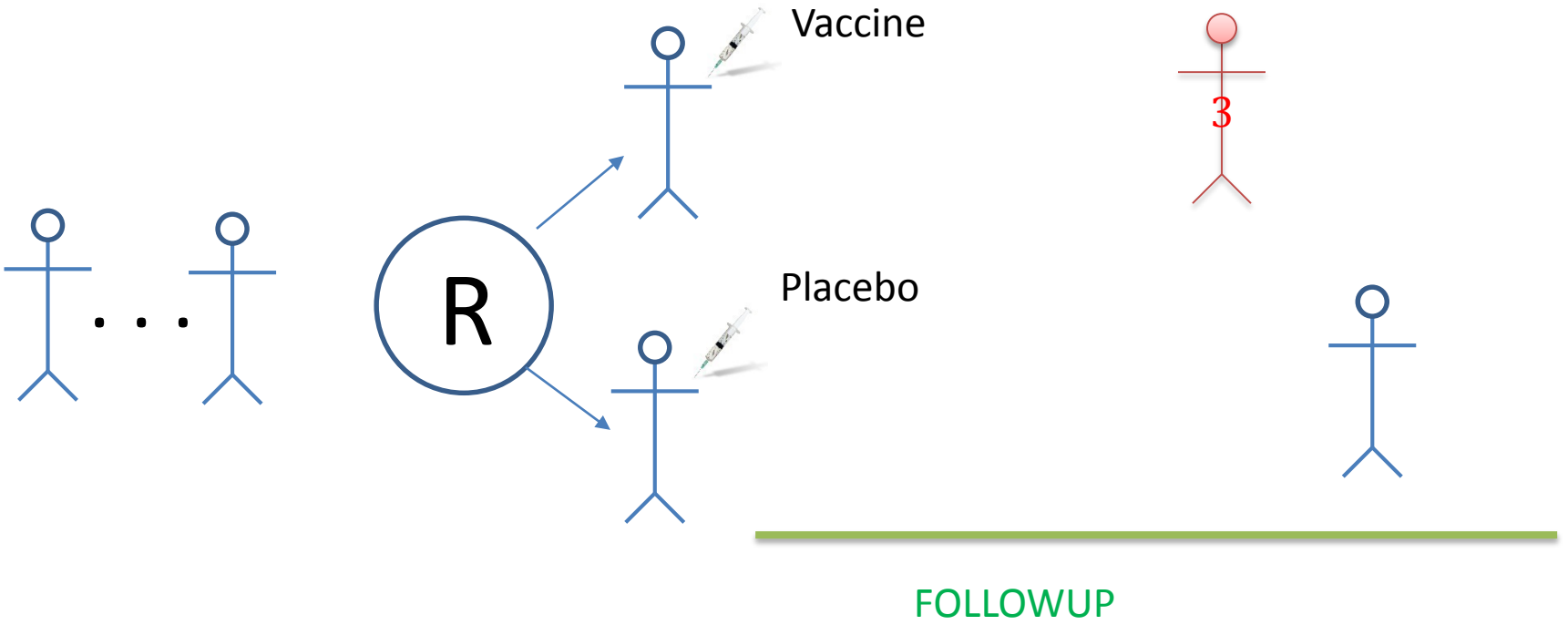
AA sequence of
Parasite used in
RTS,S/AS01 Vaccine

4 Founding Parasites

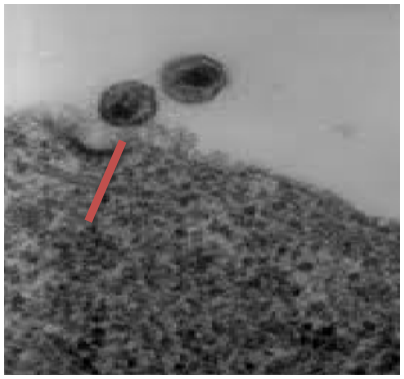


Vaccine Trial Redux

- Randomize at risk healthy volunteers to vaccine or placebo
- Follow them & count # infecting pathogens



Placebo Volunteer

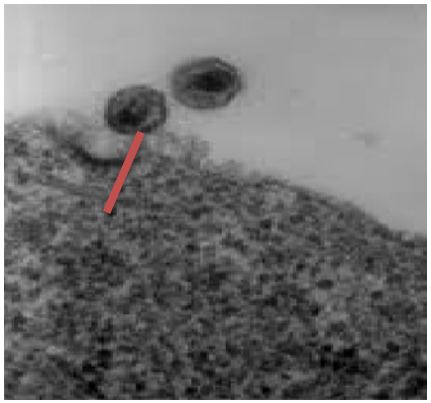


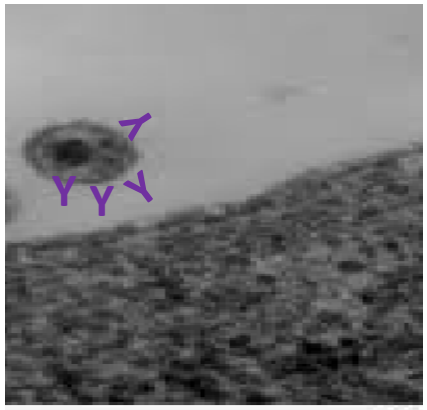
Cell infected

2 Virions infect cells

X = 2

Vaccine Volunteer





1 Virion infects a cell

Antibodies **Y** block infection

X=1

Both humans are infected, but the vaccine reduces founder viruses

Useful information that the vaccine is doing something

Mechanisms of Vaccine Protection

- All-or-none vaccine: a proportion of vaccinees are protected for *all* exposures.
- Leaky vaccine: chance of *human disease after exposure* is like flipping a coin w.p. Q
  - Q_v in vaccine arm Q_P in placebo arm
- Leaky leaky vaccine: chance of *pathogen infecting a cell* is like flipping a coin w.p. P
  - P_v in vaccine arm P_P in placebo arm

Smith et al 1984
Struchiner et al 1990
Halloran et al 1991

Vaccine Efficacy From the Virion's View

- Exposure has N virions. Each has probability p ($p\Delta$) of infecting a cell in a placebo (vaccine) recipient.
- Model X = # founder viruses
  - Vaccine E(X) = N p Δ = $\mu\,\Delta$
  - Placebo E(X) = N p = μ
- $VE_V = 1 - \frac{E(X|Z=1)}{E(X|Z=0)} = 1 - \Delta$

Per virion reduction in probability of infection
Holds for any mixture over μ

Efficiency gain using X in lieu of I(X>0)

- Suppose $X_1,\ldots,X_n \sim$ Poisson (μ)
- Dumb Method
  - Convert X to Y = I(X>0)
  - Estimate P(X>0) by avg(Y)
- Smart Method
  - Estimate $\hat{\mu}$ = avg(X)
  - Estimate P(X>0) by 1-exp(-$\hat{\mu}$)
- var (smart) /var(dumb) --- estimates of P(X>0)

μ = .25	μ= 1	μ= 3
1.1	1.7	5.8

Monkey Studies

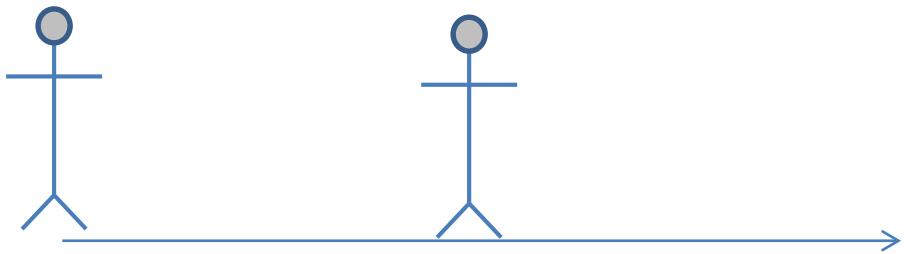
- Monkeys repeatedly *challenged* by exposing them to virus
- Assume X per challenge is Poisson($\mu\ \Delta^Z$)
- Likelihood contribution for a monkey infected on *third* challenge with 4 founder viruses.
  - P(X=0) P(X=0) P(X=4)
- Use maximum likelihood to estimate $\mu\ \Delta$
  - Form $\widehat{VE_V}$ =1-$\widehat{\Delta}$

Animal vs Human Experiments

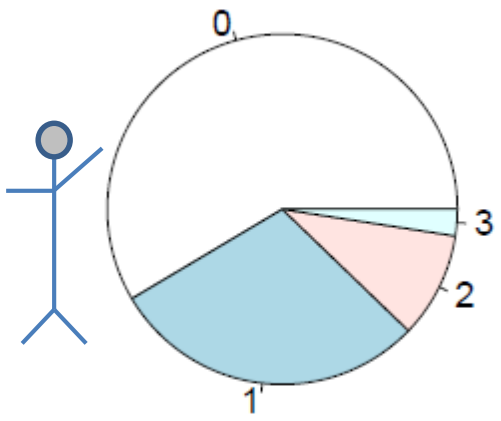
- Animal Experiments
  - Control exposure: N virions from known pool
  - Identify all Xs, even when X=0
- Human Field Trials
  - N=inoculum size uncontrolled and unknowable
  - Exposure not crisply defined
  - Exposures unknown unless infection occurs
    - *X=0 never seen*

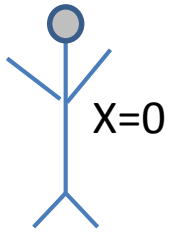
Casino Behavior

Placebo Queue



Placebo Roulette



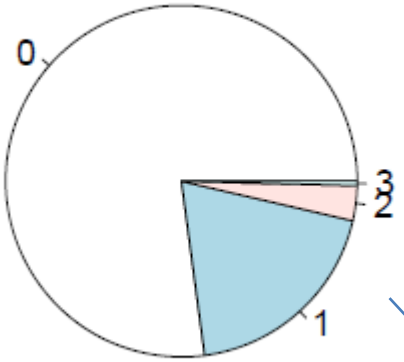


Uninfected

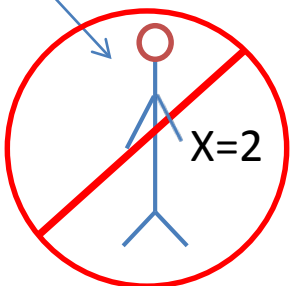
$\omega(t)$ = Instantaneous risk of gambling

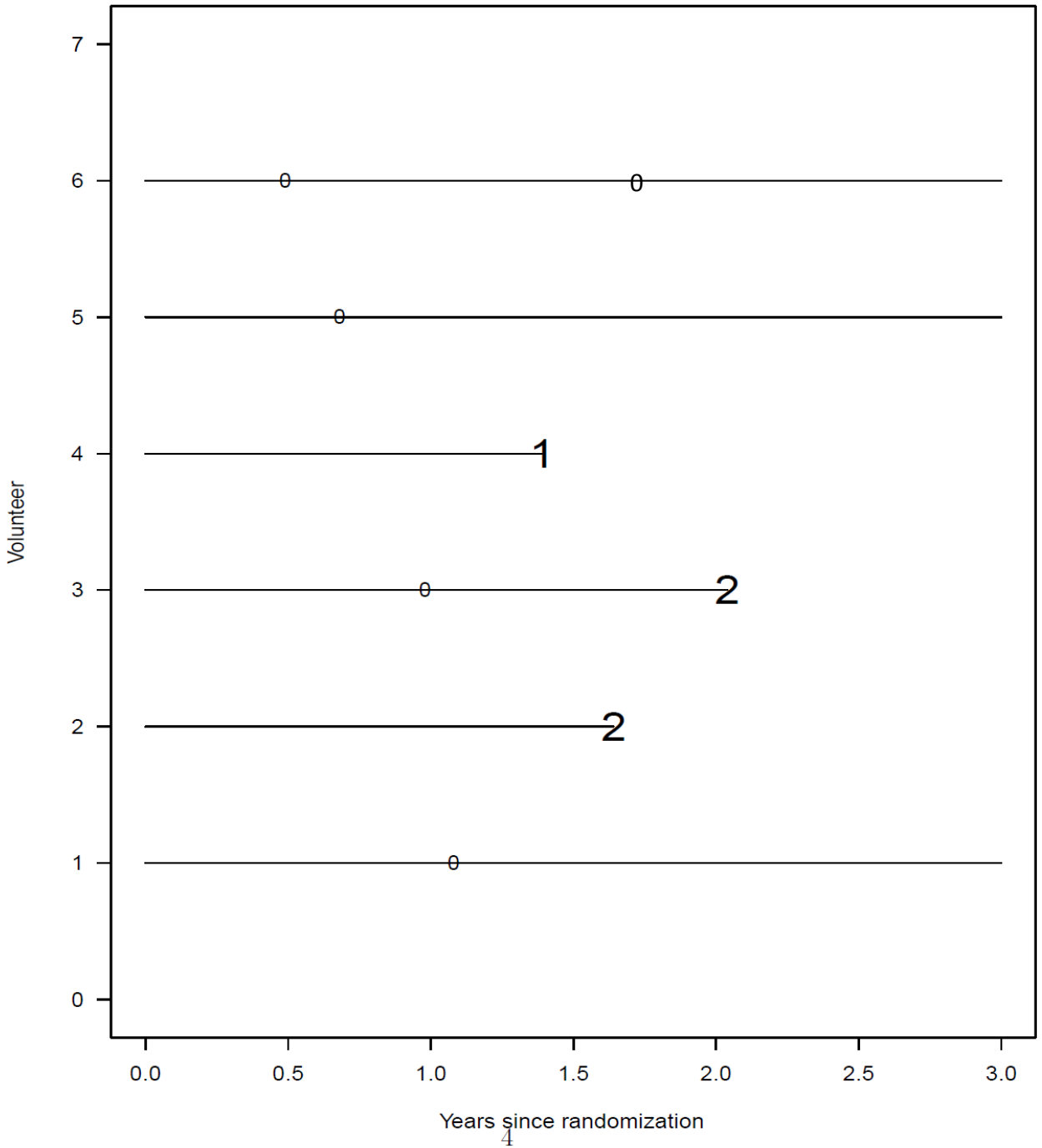
Vaccine Queue

Vaccine Roulette



Infected




Volunteer
7
6
5
4
3
2
1
0
0
0
1
2
2
0
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Years since randomization

Cox Regression For Infection

- A model for the instantaneous risk of infection

$$h(t) = \omega(t)\, P(X>0 \mid Z=0) \quad \text{in placebo group}$$

$$h(t) = \omega(t)\, P(X>0 \mid Z=1) \quad \text{in vaccine group}$$

Risk of INFECTION

Risk of EXPOSURE
Same in both groups

Probability of infection, given exposure

Cox Regression 2

- No matter the distribution of X

$$h(t) = \omega(t)\ \{P_0\ (X>0)\}\ \exp\{\ \log\left(\frac{P_1(X>0)\}}{P_0(X>0)\}}\right) Z\ \}$$

$$= h_0(t)\ \exp\{\beta\ Z\}$$

- $\beta = \log\left(\frac{P_1(X>0)\}}{P_0(X>0)\}}\right)$

- $\exp(\beta)$ is the *per-exposure* reduction in the risk of infection

Truncated mean proportional to Untruncated mean

- $E(X) = \sum_{x=0}^{\infty} xP(X = x) = \sum_{x=1}^{\infty} xP(X = x)$

$$= \sum_{x=1}^{\infty} xP(X = x)\frac{P(X>0)}{P(X>0)}$$

$$= E(X|X>0)\,P(X > 0)$$

- Thus

$$E(X|X>0) = \frac{E(X)}{P(X>0)}$$

Multiply

- Multiplication produces a product estimate
- $e^{\widehat{\beta}} \frac{\overline{X_1}}{\overline{X_0}} \rightarrow \frac{P(X>0|Z=1)}{P(X>0|Z=0)} \frac{\frac{E(X|Z=1)}{P(X>0|Z=1)}}{\frac{E(X|Z=0)}{P(X>0|Z=1)}}$

$\bar{X}_Z$ mean number of virions on Z among infected (i.e. X>0)

The Product Method Estimate of Δ

- Multiplication produces a product estimate
- $e^{\hat{\beta}} \frac{\overline{X_1}}{\overline{X_0}} \rightarrow \frac{P(X>0|Z=1)}{P(X>0|Z=0)} \frac{\frac{E(X|Z=1)}{P(X>0|Z=1)}}{\frac{E(X|Z=0)}{P(X>0|Z=1)}} = \frac{E(X|Z=1)}{E(X|Z=0)} = \Delta$

$\bar{X}_Z$ mean number of virions on Z among infected (i.e. X>0)

- Truncated X data gets ratio of *untruncated* X^* means.
- X distribution unspecified
- Arbitrary intensity of exposure function ω (t)

Horvitz-Thompson Estimator

- Population of N objects $Y_1, \ldots, Y_N$
- Sample the *i*th object with probability π_i

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{n} \frac{Y_i}{\pi_i}$$

- Estimator is unbiased

$$\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{n}\frac{Y_i}{\pi_i}\right] = \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i \frac{Y_i}{\pi_i}\right] = \frac{1}{N}\sum_{i=1}^{N} E(I_i)\frac{E(Y_i)}{\pi_i}$$

Easy Asymptotics for Product Method

- $\log(\hat{\Delta}) = \log\left(e^{\hat{\beta}_{Cox}} \frac{\overline{X}_1}{\overline{X}_0}\right)$

$$\log\left(e^{\hat{\beta}_{Cox}} \frac{\overline{X}_1}{\overline{X}_0}\right) = \hat{\beta}_{Cox} + \log(\overline{X}_1) - \log(\overline{X}_0)$$

- Delta-method $\log(\overline{X}_Z) \approx N\left(\log(\mu_Z), \frac{\sigma_Z^2}{I_Z \mu_Z^2}\right)$

- $\log(\hat{\Delta}) \sim \mathrm{N}(\log(\Delta)\ ,\ \widehat{\mathrm{var}}(\hat{\beta}_{Cox}) + \frac{S_1^2}{I_1 \overline{X}_1^2} + \frac{S_0^2}{I_0 \overline{X}_0^2}$

Product Method w/ Exponential Dbn

- Product estimate under exponential time to infection

$$\widehat{\Delta} = \left(\frac{I_1}{T_1} \Big/ \frac{I_0}{T_0}\right) \frac{\bar{X}_1}{\bar{X}_0} = \left(\frac{X_{1+}}{T_1} \Big/ \frac{X_{0+}}{T_0}\right)$$

where I_Z total number of infections on Z

T_Z total follow-up time on Z

X_{Z+} total number of virions on Z

$\bar{X}_Z$ mean number of virions on Z

V:P Hazard ratio estimate

Monkey Studies

- 10 on placebo: 1, 2, … ,10

$$\hat{\mu} = \frac{8 + 0+0+2 + \dots + 0+0+7}{1+3+\dots 3} = \frac{179}{57} = \frac{X_{0+}}{N_0}$$

- 10 on vaccine 1, 2, … ,10

$$\widehat{\mu\,\Delta} = \frac{0+0+4 + 0+\dots+0 + \dots + 0+1}{3+8+\dots 2} = \frac{75}{113} = \frac{X_{1+}}{N_1}$$

- $\hat{\Delta} = \left(\frac{X_{1+}}{N_1} / \frac{X_{0+}}{N_0}\right)$

Product Method Analogous to Estimator from Monkey Studies

- Product estimate under exponential time to infection

$$\widehat{\Delta} = \left(\frac{I_1}{T_1} \Big/ \frac{I_0}{T_0}\right) \frac{\bar{X}_1}{\bar{X}_0} = \left(\frac{X_{1+}}{T_1} \Big/ \frac{X_{0+}}{T_0}\right)$$

N_1 N_0

where N_Z total number of challenges on Z

Product method replaces total number of challenges with total time at risk

V:P Hazard ratio estimate

Concerns

- Same ω (t) for all
  - Some may have more frequent exposures
- One dbn of X for all in same group
  - Some individuals have poorer mucosal barriers...more virions get in.
- Measured covariates can address concerns

Incorporating Covariates

- Covariates for time to exposure: W^E
  - e.g. I(>3 sexual partners last month at baseline)
  - $h(t) = h_0(t) \exp(Z \beta + \theta W^E)$ . . . *product method*

- Covariates that impact X: W^X
  - e.g. damaged cells, immune response to vaccine, closeness of infecting virus to vaccine insert
  - Natural to have $E(X^*) = e^{\varphi_0+\varphi_1 Z+\varphi_2 W+\varphi_2 WZ}$

X-weighted Cox Regression

- X-weighted Cox score equation

$$\sum_{i=1}^{n} \int_0^{\infty} X_i \left\{ Z_i - \frac{\sum_{i=1}^{n} Z_i \Delta^{Z_i} I(Y_i \geq t)}{\sum_{i=1}^{n} \Delta^{Z_i} I(Y_i \geq t)} \right\} dN_i(t)$$

- *Virtually identical to product method*
- Above a functional of empirical processes. Asymptotics for $\hat{\Delta}$ from functional delta method.
- . . . but generalizes to handle both W^E & W^X.

Example HIV

- VAX003 randomized 2,546 Thai IDUs to HIV vaccine AIDSVAXB/E or placebo
  - 211 infections reported 105:106 V:P
- $VE_H = 1 - e^{-.00245} = .002$

Product Method Estimate of VE_V

- 39 volunteers, # founder viruses determined
  - High risk (IDU) volunteers
  - Infection detection within 100 days
- Mean X in vaccine 1.33, placebo 1.67

$$VE_V = 1 - e^{-.00245} \frac{1.33}{1.67} = .21$$

95% delta-method CI(-.33, .52)

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 NOVEMBER 17, 2011 VOL. 365 NO. 20

First Results of Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Children

The RTS,S Clinical Trials Partnership*

Malaria Trial

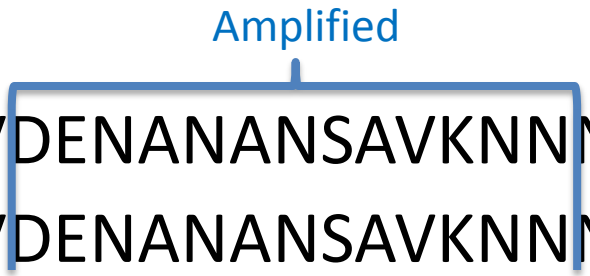
- 15,460 children randomized to malaria vaccine versus control. Focus on 5-17 months
- Primary Analysis
  - Time to clinical malaria

    VE_H = .542 95% CI (.503,.578)
- Secondary Analysis
  - Number of infecting parasites following exposure

    VE_V = .612 95% CI (.574,.612)

Undercounting

- Two nearly identical infecting pathogens may be counted as a single infecting pathogen



  - e.g. NRNVDENANANSAVKNNNNEEP
  - e.g. NRNVDENANANSAVKNNNEEEP
- Truly 2 founders but we only count 1
- Can show that VE_V is *conservative* if the undercounting process is the same in the vaccine and placebo groups.

Summary

- Discussed a way to incorporate Founder virus information into vaccine trials
  - $VE_V = 1 - \frac{E(X|Z=1)}{E(X|Z=0)} = 1 - \Delta$
- Ratio of untruncated means from truncated data.
  - Product : simple, minimal assumptions
  - Martingale: good for covariates that impact X
- VE_V can complement not supplant VE_H
- Extensions and connections are interesting

Incorporating Infecting Pathogen Counts In Sieve Analysis

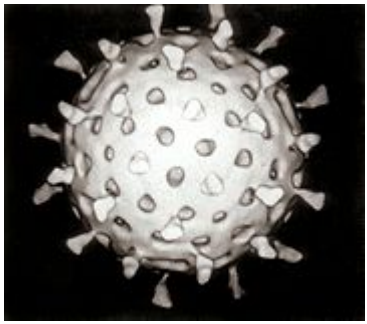
Dean Follmann

National Institute of Allergy and Infectious Diseases

Pathogens are diverse

- A pathogen species can have distinct strains
  - Serotypes ---- different surface antigens
  - Genetics ---- different DNA or RNA
- Vaccines may protect differentially against the different strains
  - Vaccine induced antibodies may protect well against some strains but not others.
  - Vaccines may induce CD4 & CD8 T-cells with differential protection
    - HIV, malaria, Ebola

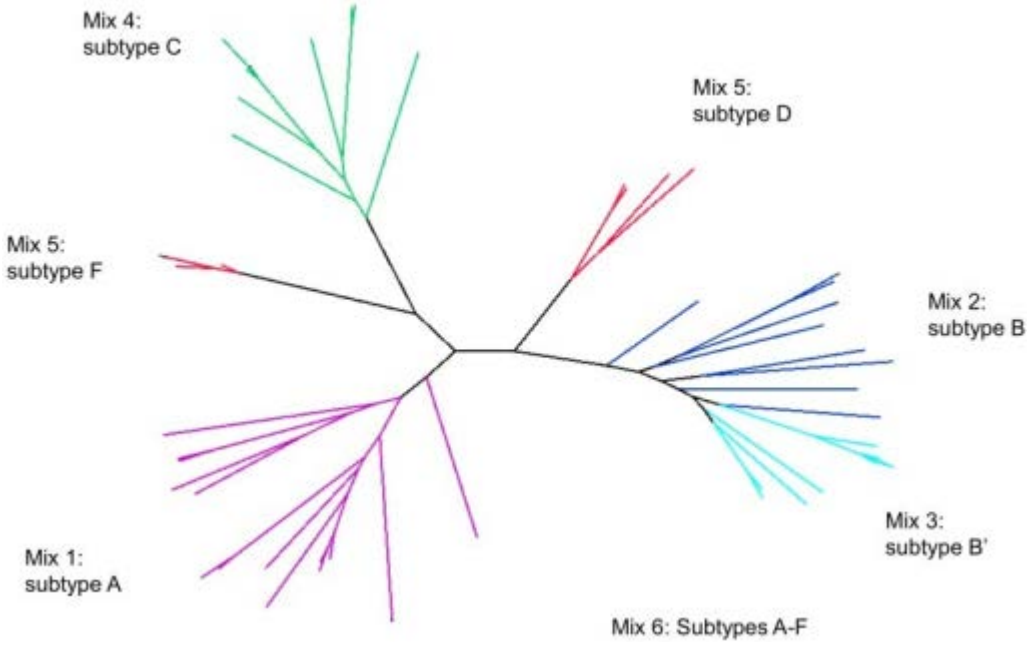
Pathogens are diverse



Rotavirus

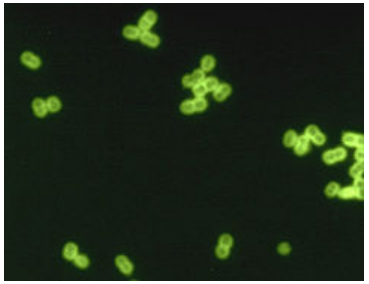
5 major serotypes

HIV multiple genotypes



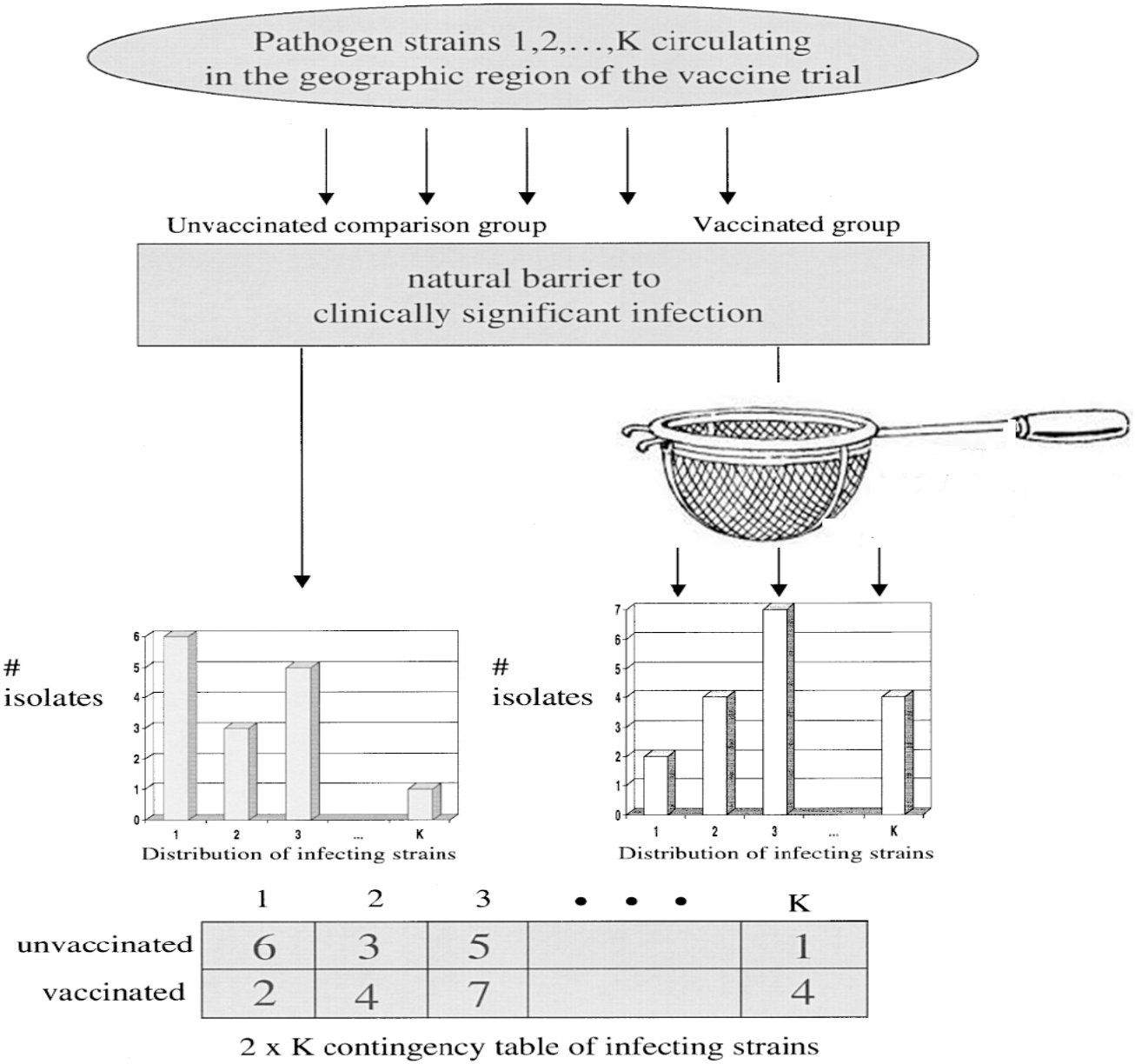
Bowles et al PLoS One 2014

Streptococcus pneumoniae



90+ serotypes


THE
SIEVE

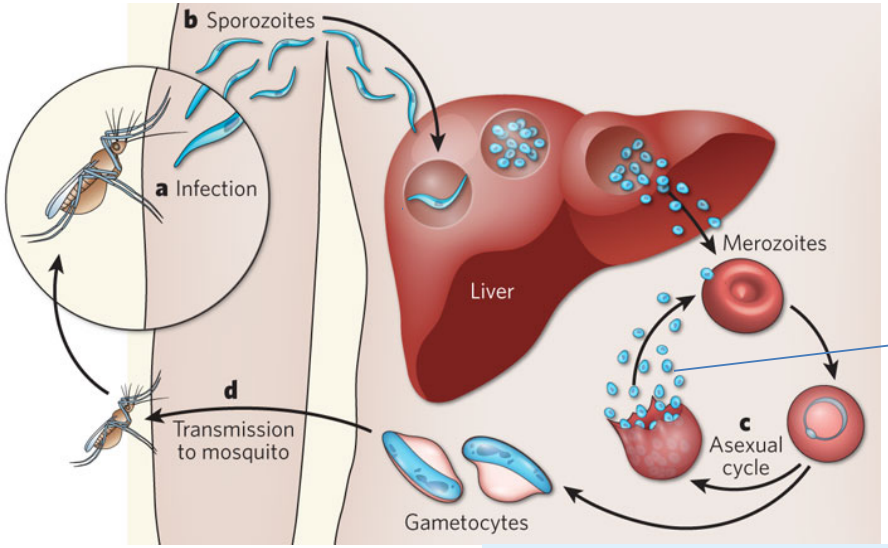


	1	2	3	• • •	K
unvaccinated	6	3	5		1
vaccinated	2	4	7		4

2 x K contingency table of infecting strains

Ref: Gilbert et al 2001

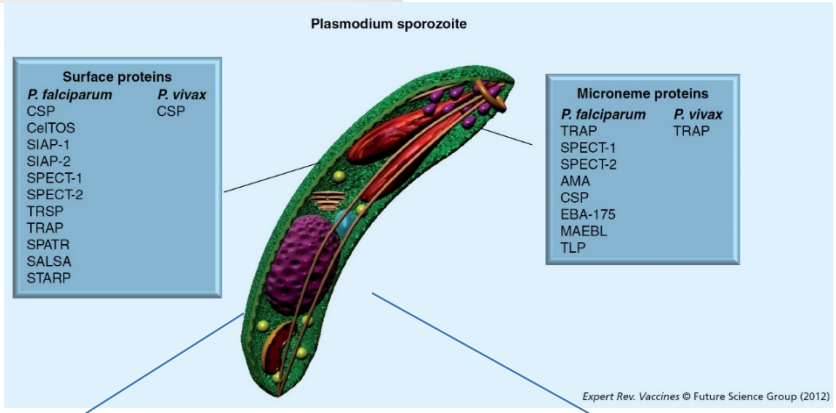
Malaria Sampling



Malaria life cycle

Sample blood stage parasites
PCR amplification of CS region
Then Next Gen sequencing.

NRNAN . . . EW
NRNEN . . . TW



290 300 310 320 330

NRNVDENANANSAVKNNNNEEPSDKHIKEYLNKIQNSLSTEW

AA sequence of
Parasite used in
RTS,S/AS01 Vaccine

of Founding Parasites

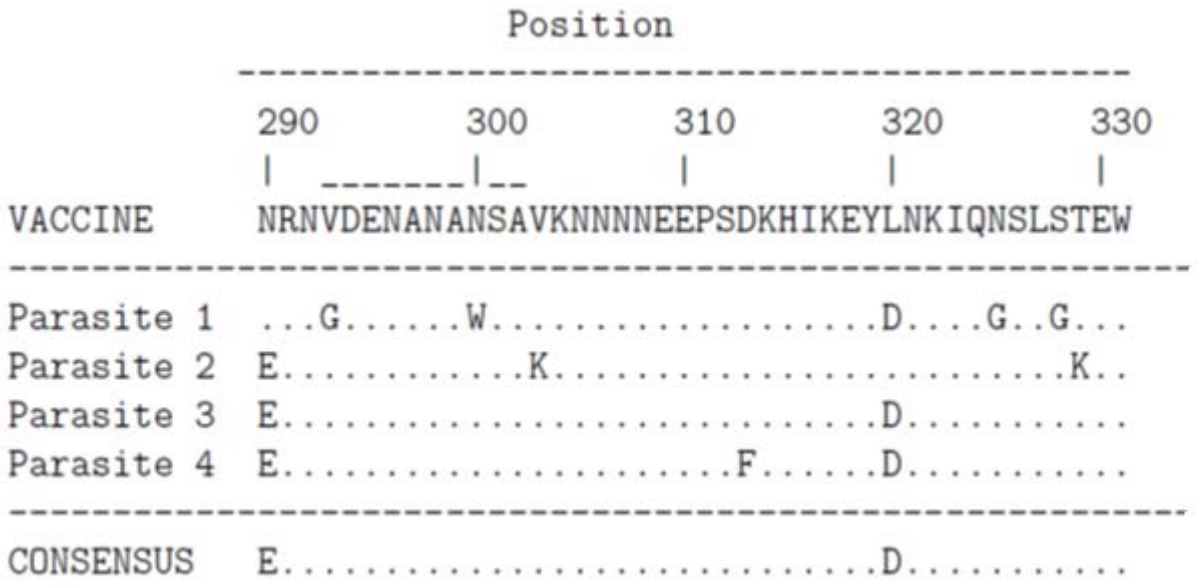


Table 1

Partial amino acid sequence of the RTS,S vaccine immunogen along with an illustration of 4 infecting parasites that could have been recovered from an infected volunteer in a vaccine trial along with the position-wise most popular sequence, i.e. the consensus sequence.

	Position							
	290	300	310	320	330	match at	match in	total mismatches
VACCINE	NRNVDENANANSAVKNNNNEEPSDKHIKEYLNKIQNSLSTEW					320	293-302	290-331
Parasite 1	...G.....W.....D....G..G...					0	0	5
Parasite 2	E.....K.....K..					1	1	3
Parasite 3	E.....D.....					0	1	2
Parasite 4	E.....F.....D.....					0	1	3
CONSENSUS	E.....D.....					0	1	2

X_1, X_2=(# match at 320, # mismatch at 320) = (1,3)

Table 1

Partial amino acid sequence of the RTS,S vaccine immunogen along with an illustration of 4 infecting parasites that could have been recovered from an infected volunteer in a vaccine trial along with the position-wise most popular sequence, i.e. the consensus sequence.

	Position							
	290	300	310	320	330	match at	match in	total mismatches
VACCINE	NRNVDENANANSAVKNNNNEEPSDKHIKEYLNKIQNSLSTEW					320	293-302	290-331
Parasite 1	...G.....W.....D....G..G...					0	0	5
Parasite 2	E.....K.....K..					1	1	3
Parasite 3	E.....D.....					0	1	2
Parasite 4	E.....F.....D.....					0	1	3
CONSENSUS	E.....D.....					0	1	2

X_a = # of infecting pathogens with 'a' total mismatches in 290-331

$X_0, X_1, X_2, X_3, X_4, X_5, \ldots$ = (0,0,1,2,0,1,0,0,0

0 1 2 3 4 5 . . . ..

Table 1

Partial amino acid sequence of the RTS,S vaccine immunogen along with an illustration of 4 infecting parasites that could have been recovered from an infected volunteer in a vaccine trial along with the position-wise most popular sequence, i.e. the consensus sequence.

	Position							
	290	300	310	320	330	match at	match in	total mismatches
	DV10 Region						DV10 Region	
VACCINE	NRNVDENANANSAVKNNNNEEPSDKHIKEYLNKIQNSLSTEW					320	293-302	290-331
Parasite 1	...G.....W.....D....G..G...					0	0	5
Parasite 2	E.....K.....K..					1	1	3
Parasite 3	E.....D.....					0	1	2
Parasite 4	E.....F.....D.....					0	1	3
CONSENSUS	E.....D.....					0	1	2

DV10 Region

$X_1 X_2$=(# match DV10 region, # mismatch DV10 region) = (3,1)

New type of data

- Before, used the consensus strain
  - Y_a = 1 if infected by `strain` a, else 0
  - e.g. (Y_1, Y_2) = (1,0) or (0,1)

- Now, get # infecting pathogens of each type
  - X_a = number of `strains` of type a

    e.g. (X_1, X_2) = (2,0) or (3,1)

Analysis of New Data

- Can we *shoehorn* this data with multiple infecting strains into existing methods for a single infecting strain?
- Can we *develop* new methods that explicitly account for multiple infecting strains?

Shoehorn: Within Cluster Resampling aka Multiple Outputation

1) Randomly pick a single pathogen for each infected person
   - Fred 4 unique strains: 1 match 3 mismatch
   - Pick a strain at random e.g. mismatch
2) Run a standard sieve analysis
   - VE(match) = .65 VE(mismatch) = .51
3) Repeat many many many times and average.

Within Cluster Resampling Schematic

Resample #	Dataset	VE(match)	VE(mismatch)
1	D_1 →	65.1	42.1
2	D_2 →	51.2	53.4
3	D_3 →	71.3	38.1
4	D_4 →	61.3	47.8
9999	D_{9999} →	52.1	38.9
10000	D_{10000} →	63.2	54.1
AVERAGE		63.1	53.9

There is an easy way to get a p-value for within cluster resampling.

Easy Inference With WCR

- Each resample gives estimates of the parameter and its variance
  - $P_1\ V_1$, $P_2\ V_2$, . . . , $P_{10000}\ V_{10000}$
- Calculate 3 Statistics
  - Average the P_i, → $\bar{P}$
  - Average the V_i → $\bar{V}$
  - Sample variance of the P_i → S^2

$$\frac{\bar{P}}{\sqrt{\bar{V}-S^2}}$$ is standard normal on the null

Easy Inference With WCR

- Each resample gives estimates of the parameter and its variance
  - $P_1\ V_1$, $P_2\ V_2$, . . . , $P_{10000}\ V_{10000}$
- Calculate 3 Statistics
  - Average the P_i, → $\bar{P}$
  - Average the V_i → $\bar{V}$
  - Sample variance of the P_i → S^2

$$\frac{\bar{P}}{\sqrt{\bar{V}-S^2}}$$ is standard normal on the null

WCR

- WCR can be used whenever you have a statistical procedure P that requires 1 outcome per person, but you have multiple outcomes.
- Can be used in lieu of GEE
  - Like exchangeable with rho -> 1
    - One person, one vote
  - Opposite of working independence rho=0
    - One pathogen, one vote

WCR = t-test on cluster means

- Test means of two groups X vs Y

$x_{11}\ x_{12}\ x_{13}$ → $\bar{x}_1$

$x_{21}\ x_{22}$ → $\bar{x}_2$

x_{31} → $\bar{x}_3$

$y_{11}\ y_{12}\ y_{13}\ y_{14}$ → $\bar{y}_1$

$y_{21}\ y_{22}$ → $\bar{y}_2$

Sieving at DV10 Region

DV10 Region	RTS,S Vaccine # Events	Control Vaccine (% Incidence)	VE
Match	90 (2.5)	86 (5.6)	63.1
Mismatch	1091 (30.8)	822 (53.7)	53.9

Averaged over 1000s of synthetic data sets with 1 Strain per person

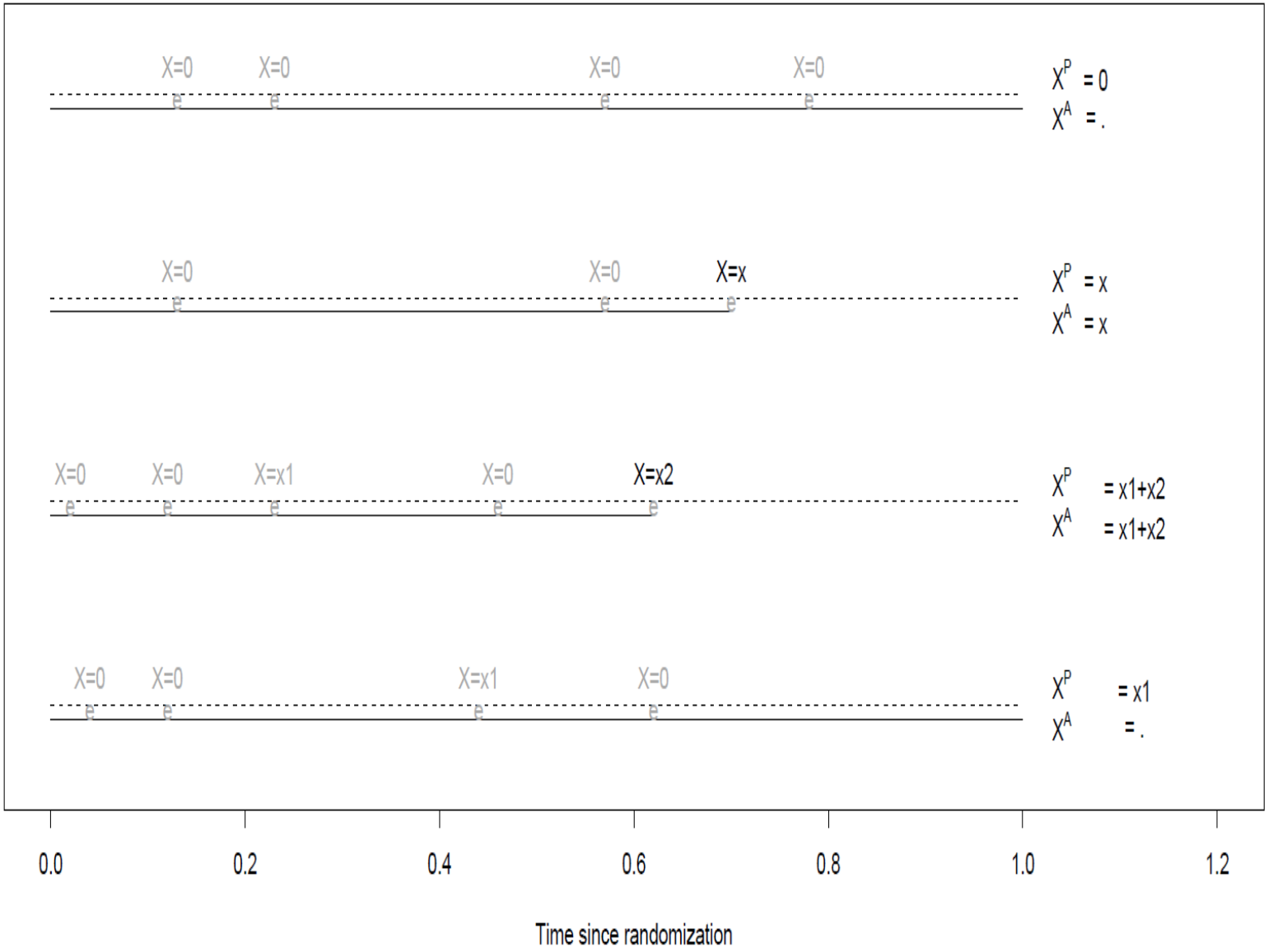
- Test of equal VE has p=.04
- Some evidence of sieving.

Neafsey et al 2015

New Methods

- Let's develop new methods that explicitly use the counts
- Passive surveillance
  - Get $(X_1, X_2) = (0,0)$ or $(3,1)$ or $(2,0)$ at end of study
- Active surveillance
  - Get time of infection detection and
  - Get $(X_1, X_2) = (0,0)$ or $(3,1)$ or $(2,0)$

Passive and active surveillance



Passive Surveillance: Modern Data & Analysis

Group	X_1	X_2
Vaccine	1	0
Vaccine	0	0
Placebo	3	0
Placebo	2	4
Vaccine	0	2
Placebo	0	0

Placebo group **5 mismatched out of 9**
Vaccine group **1 mismatched out of 3**

Passive Surveillance
Single Pathogen Data & Analysis

Group	X_1	X_2
Vaccine	1	0
Vaccine	0	0
Placebo	1	0
Placebo	0	1
Vaccine	0	1
Placebo	0	0

Placebo group **1 mismatched out of 2**
Vaccine group **1 mismatched out of 2**

Passive Surveillance: Counts

- Assume *bivariate negative binomial*
  - X_{si} Poisson exp{b_i + B0 + B1 Z + B2 I(s=1) + B3 Z I(s=1) }
  - s=1,2 i=1,…n subjects exp(b_i) ~ Gamma (μ, V)
  - Z= vaccine indicator
- Condition. $X_0 | X_0+X_1=N$ follows

$$\text{Binomial}\left(N, \frac{e^{B1}}{1+e^{B1}}\right) \quad \text{in placebo}$$

$$\text{Binomial}\left(N, \frac{e^{B1+B3}}{1+e^{B1+B3}}\right) \quad \text{in vaccine}$$

Sieve effect if B3 is nonzero

Passive Surveillance: Single Pathogen

- Identify most popular strain
  - W=1 if $X_0 > X_1$, or if $X_0 = X_1$ flip a coin
- Then W follows

$$\text{Binomial}\left(1, \frac{e^{B1}}{1+e^{B1}}\right) \quad \text{in placebo}$$

$$\text{Binomial}\left(1, \frac{e^{B1+B3}}{1+e^{B1+B3}}\right) \quad \text{in vaccine}$$

Sieve effect if B3 is nonzero

Simulation

- X~ bivariate negative binomial
  - $\exp(b_i)$ ~ Gamma(.5,v) v=0,1,2
- Counts: Binomial (= GEE-I), WCR
- Infection: Bernoulli

SIMULATION VARIANCE OF Sieve effect B3					
V	Binomial (new)	Bernoulli (old)	WCR (shoehorn)	Binomial/ Bernoulli	Binomial/ WCR
0	.066	.139	.083	2.1	1.3
1	.072	.170	.109	2.4	1.5
2	.047	.201	.090	4.2	1.9

Sweet but

- Simulations were based on an idealized model
  - Nice bivariate negative binomial model
  - Nice leaky leaky mechanism
- Can show if vaccine impacts P(X>0) but no effect on X>0, (i.e. non-leaky leaky) WCR is better
  - *Mechanism of protection important*

Active Surveillance

- Let's consider field trials
  - Time to infection as endpoint
  - Count X_1, X_2 once infected
- Only observe X_1, X_2 | $X_1+X_2 > 0$
- Do natural modification of the product method

The Product Method Estimate of Δ

- Multiplication produces a product estimate
- $e^{\hat{\beta}} \frac{\overline{X_{1s}}}{\overline{X_{0s}}} \rightarrow \frac{E(X_s|Z=1)}{E(Xs|Z=0)} = \Delta_s$
- $\bar{X}_{Zs}$ mean number of strain s virions on Z among infected (i.e. $X_{Z1} + X_{Z2} > 0$)

- Truncated X data gets ratio of *untruncated* X^* means.
- X distribution unspecified
- Arbitrary intensity of exposure function ω (t)

Sieving Effect on Counts

- Test equality of ratio of *unconditional* means
  - $\frac{E(X_1|Z=1)}{E(X_1|Z=0)} = \Delta_1 = \Delta_2 = \frac{E(X_2|Z=1)}{E(X_2|Z=0)}$

- Equivalent to testing ratio of `truncated' means.

$$\cancel{e^{\beta}}\ \frac{\mu^{t}_{11}}{\mu^{t}_{01}} = \cancel{e^{\beta}}\ \frac{\mu^{t}_{12}}{\mu^{t}_{02}}$$

$$\mu^{t}_{zs} = E(X_{ZS} | X_{Z1}+X_{Z2}>0)$$

Sieving Effect on Infections

- Let $Y_s = I(X_s>0)$
- Test equality of ratio of *unconditional* means

$$\frac{E(Y_1|Z=1)}{E(Y_1|Z=0)} = \frac{E(Y_2|Z=1)}{E(Y_2|Z=0)}$$

- Equivalent to testing ratio of `truncated' means.

$$e^{\beta}\frac{\mu^t_{11}}{\mu^t_{01}} = e^{\beta}\frac{\mu^t_{12}}{\mu^t_{02}}$$

$$\mu^t_{zs} = E(Y_{ZS}|Y_{Z1}+Y_{Z2}>0)$$

Simulation Setup

- Exponential gap times to exposures
- Bivariate negative binomial at each exposure.
  - Infected if $X_1+X_2> 0$
- Evaluate product estimate
- Compare to WCR where we pick a pathogen at random

Results

WCR **VE on *Infection*** **VE = 1- P(X>0|Z=1)/P(X>0|Z=0)**

Var	Mean(X)	% infected	ln(1-VE_1)	Ln(1-VE_2)	Ln(1-VE_1)/(1-VE_2)
1	9.4	.29	-.512	-1.030	.514
			(.042)	(.089)	(.100)
0	2.1	.30	-.449	-.981	.532
			(.036)	(.094)	(.106)

Product Estimate **VE on *# pathogens*** **VE = 1- E(X|Z=1)/E(X|Z=0)**

Var	Mean(X)	% infected	ln(1-VE_1)	Ln(1-VE_2)	Ln(1-VE_1)/(1-VE_2)
1	9.4	.29	-1.660	-2.170	.508
			(.897)	(.932)	(.042)
0	2.1	.30	-1.550	-2.030	.527
			(.043)	(.091)	(.084)

New method can be more powerful

Weighted Estimating Equations

- Covariates W for active surveillance
  - Can incorporate risk factors for exposure
- Can allow pathogen distribution $F(X_1,X_2 \mid Z)$ to change over time
- Can allow sieve effect to vary with W
  - Vaccine blocks ‘1’ in older people & blocks ‘2’ in younger people
- Details forthcoming . . . someday

Beyond Mismatch

Table 1

Partial amino acid sequence of the RTS,S vaccine immunogen along with an illustration of 4 infecting parasites that could have been recovered from an infected volunteer in a vaccine trial along with the position-wise most popular sequence, i.e. the consensus sequence.

	Position							
	290	300	310	320	330	match at	match in	total mismatches
VACCINE	NRNVDENANANSAVKNNNNEEPSDKHIKEYLNKIQNSLSTEW					320	293-302	290-331
Parasite 1	...G.....W.....D....G..G...					0	0	5
Parasite 2	E.....K.....K..					1	1	3
Parasite 3	E.....D.....					0	1	2
Parasite 4	E.....F.....D.....					0	1	3
CONSENSUS	E.....D.....					0	1	2

X_a = # of infecting pathogens with 'a' total mismatches in 290-331

$X_0, X_1, X_2, X_3, X_4, X_5, \ldots$ = (0,0,1,2,0,1,0,0,0

0 1 2 3 4 5 . . . . . # of mismatches

Beyond Mismatch

- Consider the region 290-331. Assume

$X_{zs} \sim \text{Poisson}\{ \exp(A_s + Z^*(B0 + B1\ s)) \}$

# mismatches	Vaccine Rate	Placebo Rate	Count
0	exp(A0+ B0 + B1 * 0)	exp(A0)	7
1	exp(A1 + B0 + B1 * 1)	exp(A1)	3
2	exp(A2 + B0 + B1 * 2)	exp(A2)	0
3	exp(A3+ B0 + B1 * 3)	exp(A3)	1
.	.	.	.
.	.	.	.
.	.	.	.
43	exp(A43 + B0 + B1 * 43)	exp(A43)	0
Count	30	55	

Sieve effect

Beyond Mismatch

- For a given subject, conditional on Z and the number of infecting pathogens, X_+

$X_1\ X_2 .. X_{43} \sim \text{Multinomial}(X_+\ p_1\ p_2 \ldots p_{43})$

$p_s = \exp(As + Z*(B0+B1\ s))$

$$p_s = \frac{\exp(As + Z*(B0+B1\ s))}{\sum_{s=1}^{43} \exp(As + Z*(B0+B1\ s))}$$

- Analogous to usual sieve methods with $X_+ = 1$
- May be hard to estimate with so many parameters
  - Redefine so there are fewer parameters
  - or

Beyond Mismatch

- Model has 43 nuisance parameters
  - Want to allow arbitrary dbn of WT viruses
- Under independence of subjects can condition on rows to eliminate them

# mismatches	Vaccine Rate	Placebo Rate	Count	Pr(Infection in vaccine | infection)
0	exp(A0+ B0 + B1 * 0)	exp(A0)	7	exp(B0+B1*0)/(1+exp(B0+B1*0)
1	exp(A1 + B0 + B1 * 1)	exp(A1)	3	exp(B0+B1*1)/(1+exp(B0+B1*1)
2	exp(A2 + B0 + B1 * 2)	exp(A2)	0	exp(B0+B1*2)/(1+exp(B0+B1*2)
3	exp(A3+ B0 + B1 * 3)	exp(A3)	1	exp(B0+B1*3)/(1+exp(B0+B1*3)
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
43	exp(A43 + B0 + B1 * 43)	exp(A43)	0	exp(B0+B1*43)/(1+exp(B0+B1*43)
Count	30	55		

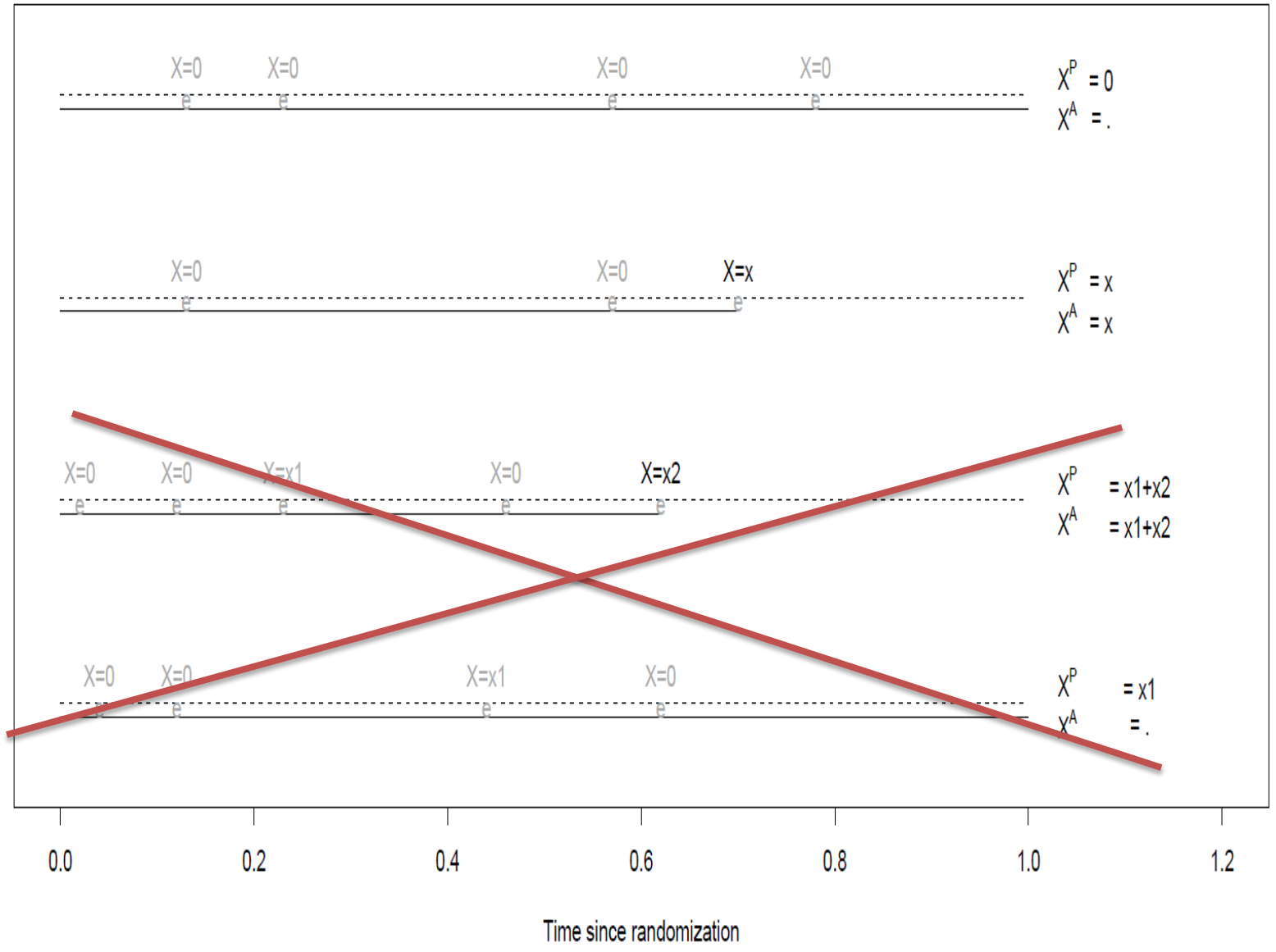
Beyond Mismatch

- Likelihood based on product of binomials (N,Y)

N	Pr(Infection in vaccine | infection)	Y = # vaccine
7	exp(B0+B1*0)/(1+exp(B0+B1*0)	3
3	exp(B0+B1*1)/(1+exp(B0+B1*1)	1
0	exp(B0+B1*2)/(1+exp(B0+B1*2)	0
1	exp(B0+B1*3)/(1+exp(B0+B1*3)	0
.	.	
.	.	
.	.	
0	exp(B0+B1*42)/(1+exp(B0+B1*42)	0

- May be able to relax independence assumption with GEE for correlated binomial data
- Analogous results obtains for active surveillance

Non-recurrent disease (e.g. HIV)



Sieve Parameter

- *per exposure* sieve effect for untruncated data

$$\theta_{a,a'} = \frac{E(X_a|Z=1)/E(X_a|Z=0)}{E(X_{a'}|Z=1)/E(X_{a'}|Z=0)}$$

- Using the contingency table, we estimate ratios based on available data
  - At end of follow-up (passive)
  - At the time of infection (active)
  - *Neither are at time of exposure*

Sieve Parameter

- Define the sieve parameters for active & passive surveillance

$$\frac{E(X_a^P|Z=1)/E(X_a^P|Z=0)}{E(X_{a'}^P|Z=1)/E(X_{a'}^P|Z=0)} \qquad \frac{E(X_a^A|Z=1)/E(X_a^A|Z=0)}{E(X_{a'}^A|Z=1)/E(X_{a'}^A|Z=0)}$$

- Can show the *per-exposure* ratio of means $\theta_{a,a'}$ equals each of the above ratios
- Analogous to work by Gilbert

Acknowledgements

- Betz Halloran, Peter Gilbert, Michael Sachs, Erin Gabriel
- Chiung-Yu Huang

  (JHU)