# Estimation of vaccine-attributable reduction in disease incidence

ISCB, Montpellier, 29 Aug to 2 Sept, 2010


**Yin Bun Cheung**

Singapore Clinical Research Institute, Singapore

Duke-NUS Graduate Medical School, Singapore

University of Tampere, Finland

# In collaboration with

**Tina Xu**
Singapore Clinical Research Institute, Singapore
Duke-NUS Graduate Medical School, Singapore

**Eddy Lam**
University of Hong Kong, China

**Paul Milligan**
London School of Hygiene & Tropical Medicine, UK

# Multiple disease episodes in clinical trials

- Current research on vaccines for diseases that do not confer lifetime immunity after a single disease episode
- WHO MALVAC meeting 2008
- Review of pneumococcal vaccine trials (Jahn-Eimermacher et al., *Vaccine* 2007)

# Vaccine-attributable reduction

- Vaccine efficacy (VE) = $(I_U - I_V)/I_U = 1 - \text{IRR}$,
  where Incidence $I_j$, j=U or V, is defined as (number of events/person-years)
- Vaccine- attributable reduction (VAR) = $(I_U - I_V)$ = IRD
- $\text{VAR} = \text{VE} \times I_U$
- VAR reflects disease burden and public health importance of an intervention
- 1/VAR may be seen as "Number Needed to Treat"

# VE (IRR) or VAR (IRD)?

- In the analysis of binary variable, there has been a lot of discussion on the relative merits of using OR, RR, RD.
- Much less attention has been given to IRR vs IRD
- WHO MALVAC 2008 and review in *Vaccine* 2007 both concern VE (IRR)

# How to define incidence?
# First or all episodes?

- Many clinical trials only analysed time to first episode, even if subsequent episodes were recorded.

- Incidence = $\dfrac{\text{number of first events}}{\text{person-years up to first event/censoring}}$

# How to define incidence?
# First or all episodes?

- Disadvantage of analysis of first versus all episodes:
  - Focuses on short-term effects
  - Loss of precision
  - Heterogeneity

# Loss of precision

- If there is no event dependency

- Total effect = primary + secondary (via effect dependency)

- All vs first events analysis may answer different questions

# Heterogeneity example: proportional reduction, constant incidence

- True total (number of events, person-years)

| | Low | High | Incidence |
|---|---|---|---|
| Vaccine | (2, 1) | (4, 1) | (2+4)/(1+1)=3 |
| Unvac. | (4, 1) | (8, 1) | (4+8)/(1+1)=6 |

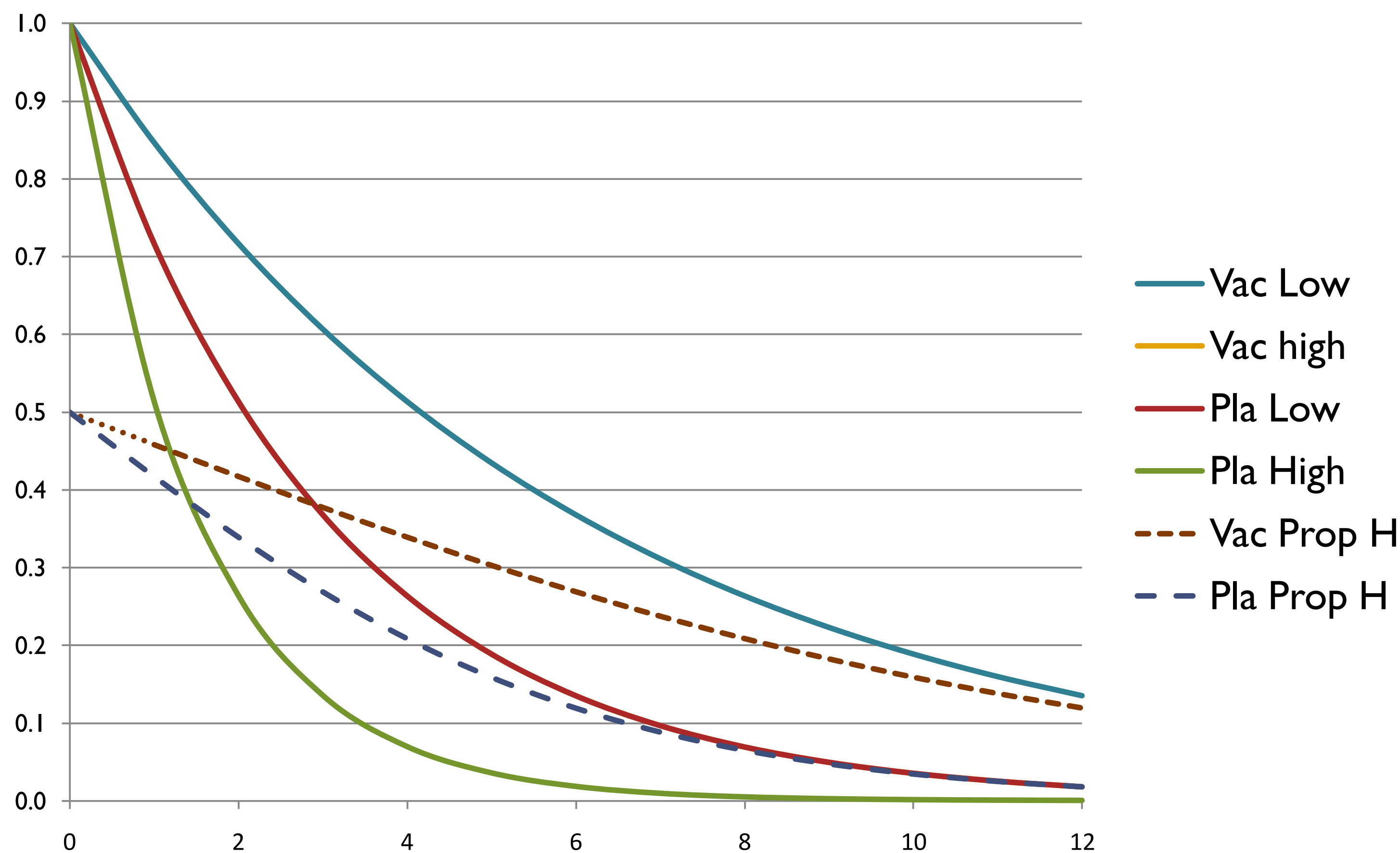
VAR = 3.0; VE = 0.50

- True (number of first events, time to first event)

| | Low | High | Incidence |
|---|---|---|---|
| Vaccine | (0.865, 0.432) | (0.982, 0.245) | (0.87+0.98)/(0.43+0.25)=2.72 |
| Unvac. | (0.982, 0.245) | (1.00, 0.125) | (0.98+1.00)/(0.25+0.13)=5.35 |

VAR = 2.63;VE = 0.49

# Heterogeneity example: proportional reduction, constant incidence



# Less discussion on absolute reduction

- No way to empirically confirm relative or absolute reduction

# Heterogeneity example: absolute reduction, constant incidence

- True total (number of events, total person-years)

| | Low | High | Incidence |
|---|---|---|---|
| Vaccine | (2, 1) | (6, 1) | (2+6)/(1+1)=4 |
| Unvac. | (4, 1) | (8, 1) | (4+8)/(1+1)=6 |

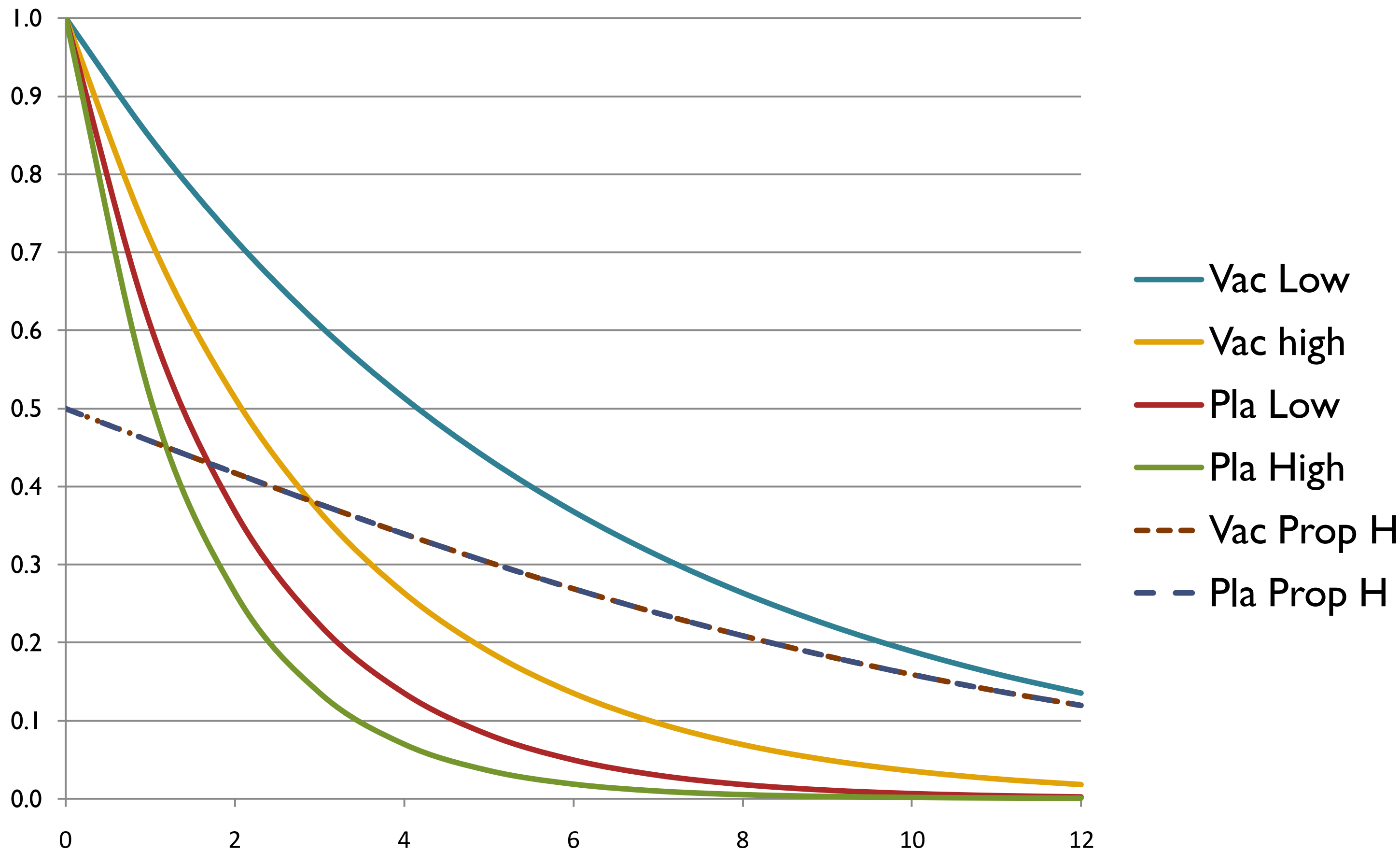
VAR = 2.0; VE = 0.33

- True (number of first events, time to first event)

| | Low | High | Incidence |
|---|---|---|---|
| Vaccine | (0.865, 0.432) | (0.998, 0.166) | (0.87+0.998)/(0.43+0.17)=3.11 |
| Unvac. | (0.982, 0.245) | (1.000, 0.125) | (0.98+1.00)/(0.25+0.13)=5.35 |

VAR = 2.24;VE = 0.42

# Heterogeneity example: absolute reduction, constant incidence



# Absolute reduction, constant incidence model, with equal stratum size

- True incidence per unit time

| | Low | High |
|---|---|---|
| Vaccine | $\mu$-$\theta$ | $(\mu+\delta)$-$\theta$ |
| Unvac. | $\mu$ | $\mu+\delta$ |

- Proportion $\text{stratum}_{\text{Low}}$ at time t = $\dfrac{1}{1+e^{-\delta t}}$

# Overall incidence, simple case of two strata

- Incidence in $\text{stratum}_{\text{Low}} = \dfrac{Events_{Low}}{Time_{Low}}$

- Overall incidence $= \dfrac{Events_{Low} + Events_{High}}{Time_{Low} + Time_{High}}$

$$= \frac{\left(\dfrac{Events_{Low}}{Time_{Low}}\right) Time_{Low} + \left(\dfrac{Events_{High}}{Time_{High}}\right) Time_{High}}{Time_{Low} + Time_{High}}$$

# Constant incidence; analyse first events only

- True incidence per unit time: $\lambda$
- Expected # first event by time t: $1-e^{-\lambda t}$
- Follow-up time up to t:

$$E(T \mid T <= t)\Pr(T <= t) + t\Pr(T > t) = \frac{1-e^{-\lambda t}}{\lambda}$$

  - First derivative w.r.t. $\lambda$: $\dfrac{(1+\lambda t)e^{-\lambda t} - 1}{\lambda^2} < 0$
  - Second derivative w.r.t. $\lambda$: $\dfrac{2-e^{-\lambda t}\{(1+\lambda t)^2 + 1\}}{\lambda^3} > 0$

# How to analyse all episodes?

- Incidence = $\dfrac{\text{Total number of events observed}}{\text{Total person-years observed}}$

- $Y_i$ and $Z_i$ = total number of events and follow-up duration (person-years) in subject i

$$I_U = \frac{\sum_{i:X_{i1=0}} Y_i}{\sum_{i:X_{i1=0}} Z_i}$$

$$VAR = \frac{\sum_{i:X_{i1=0}} Y_i}{\sum_{i:X_{i1=0}} Z_i} - \frac{\sum_{i:X_{i1=1}} Y_i}{\sum_{i:X_{i1=1}} Z_i}$$

# How to analyse all episodes?

- Poisson: wrong inference; offset term needed on the log-scale
- Negative binomial: offset term needed on the log-scale; estimate may disagree with observed VAR
- Stukel et al (1994): empirical estimate of variance
- Covariate adjustment

# Research purpose

- To provide a means to estimate VAR that
  - would agree with observed VAR
  - allows differences in follow-up time
  - is convenient to adjust for covariates
  - is robust to unobserved heterogeneity
  - simple, easy to implement

# Weighted least squares approach

- Incidence in subject i, $I_i = Y_i / Z_i$
- Observed incidence

$$= \frac{\Sigma Y_i}{\Sigma Z_i} \neq \frac{1}{N} \Sigma I_i$$

$$= \frac{\Sigma Y_i}{\Sigma Z_i} = \frac{1}{\Sigma Z_i} \Sigma \left( I_i Z_i \right)$$

# Weighted least squares approach

$$Y^* = Y / \sqrt{Z}$$

$$X_0^* = \sqrt{Z}$$

$$X_1^* = X_1\sqrt{Z}$$

$$Y^* = \beta_0 X_0^* + \beta_1 X_1^*$$

$$\widehat{\beta}_0 = \frac{\sum_{i:X_{i1=0}} Y_i}{\sum_{i:X_{i1=0}} Z_i}$$

$$\widehat{\beta}_1 = \frac{\sum_{i:X_{i1=1}} Y_i}{\sum_{i:X_{i1=1}} Z_i} - \frac{\sum_{i:X_{i1=0}} Y_i}{\sum_{i:X_{i1=0}} Z_i}$$

- $X_1$ = 0 (vaccine) or 1 (unvac.), $\beta_1$ = VAR, or
- $X_1$ = 1 (vaccine) or 0 (unvac.), $\beta_1$ = -VAR

# Weighted least squares approach

$$X_i = (1, X_{i1}, \ldots, X_{ik})$$

$$\beta = (\beta_0, \beta_1, \ldots, \beta_k)$$

Disease incidence for subject i modelled as

$$\left( \frac{Y_i}{Z_i} \mid X_i, Z_i, \nu_i \right) = X_i \beta + \nu_i + e_i$$

where $\nu_i$ represents unobserved heterogeneity

$$E(e_i) = 0, Var(e_i) = \frac{\sigma^2}{Z_i}$$

$$E(\nu_i) = \upsilon, Var(\nu_i) = \rho$$

$\nu_i$ is uncorrelated with $X_i$, $Z_i$ or $e_i$.

# Weighted least squares approach

$$E\left(\frac{Y_i}{Z_i} \mid X_i, Z_i\right) = X_i\beta + \upsilon = X_i\beta^*$$

$$\beta^* = (\beta_0 + \upsilon, \beta_1, \ldots, \beta_k)$$

$$Var\left(\frac{Y_i}{Z_i} \mid X_i, Z_i\right) = \frac{\sigma^2}{Z_i} + \rho$$

$$\text{Working variance} = \frac{\sigma^2}{Z_i}$$

# Weighted least squares approach

$$S\left(\beta^*\right) = \sum_{i=1}^{n} Z_i X_i^T \left( \frac{Y_i}{Z_i} - X_i \beta^* \right) = 0$$

Let $Y = (Y_1, \ldots, Y_n), Z = (Z_1, \ldots, Z_n) X = (X_1^T, \ldots, X_2^T)$

and $W = diag(Z_1, \ldots, Z_n)$

$$\beta^* = \left(X^T W X\right)^{-1} X^T Y$$

(Xu et al. *AJE* 2010)

# Weighted least squares approach

- Huber-White robust SE estimator

$$= \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{n} Z_i^{\,2}\hat{e}_i^2\mathbf{X}_i^T\mathbf{X}_i\right)\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}$$

- Consistent with Stukel et al (1994) in simple case

- HC1, HC2 and HC3 correction factors
  - n/(n-k),
  - $1/(1-h_i)$ and
  - $1/(1-h_i)^2$,
  - where $h_i$ is the i-th diagonal element in the hat matrix

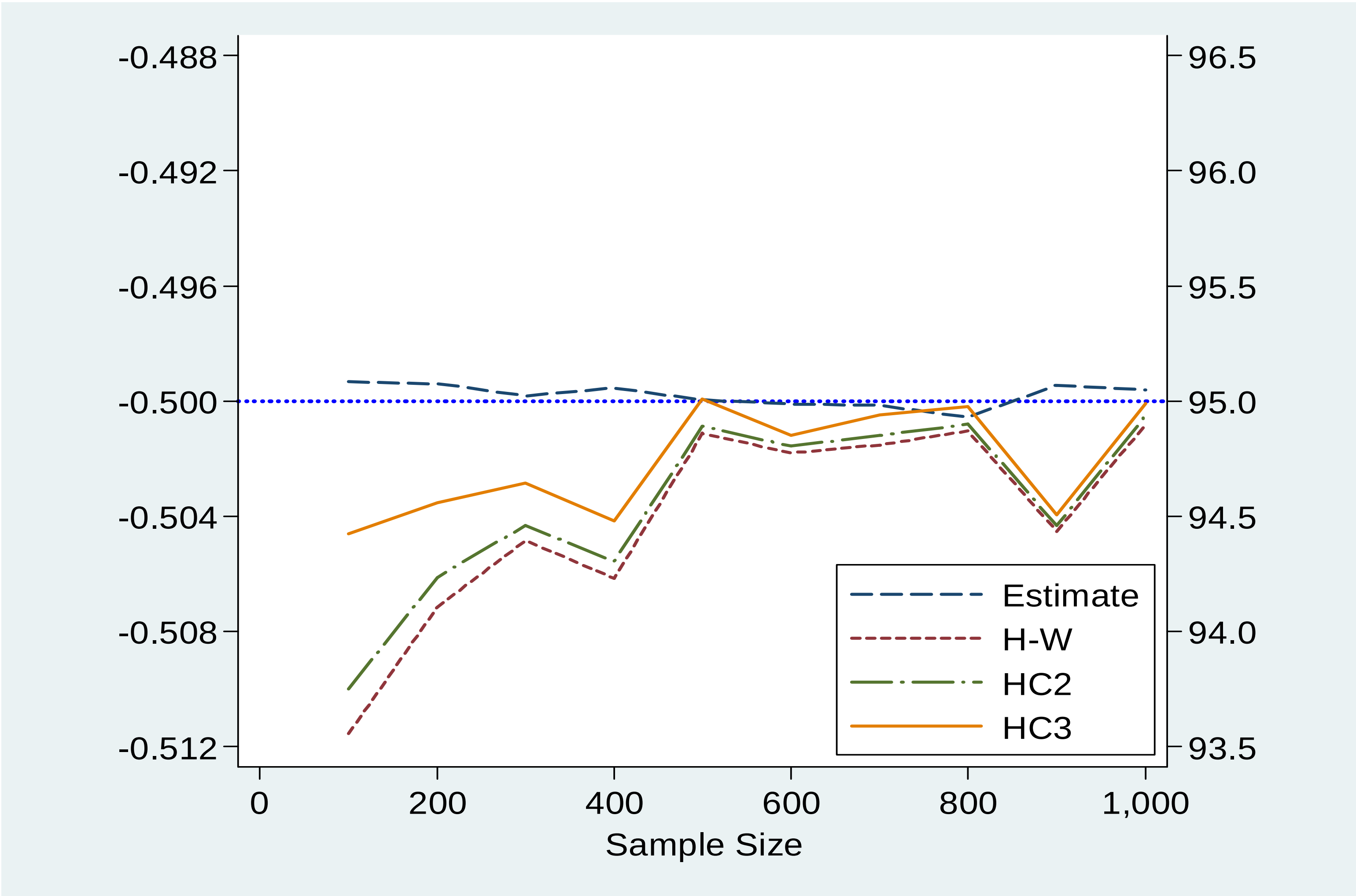
# Simulation

- $Y_i \sim \text{Poisson}(Z_i(\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\nu_i))$
- $Z_i \sim \text{uniform}(2,3)$
- $X_{i1} \sim \text{Bernouli}(p)$, $p=0.5$ or $0.7$
- $X_{i2} = 5\beta_0 \times \text{Beta}(\alpha_1+\alpha_2 X_{i1}, \alpha_3)$
- $(\alpha_1, \alpha_2, \alpha_3)$ determines collinearity and skewness

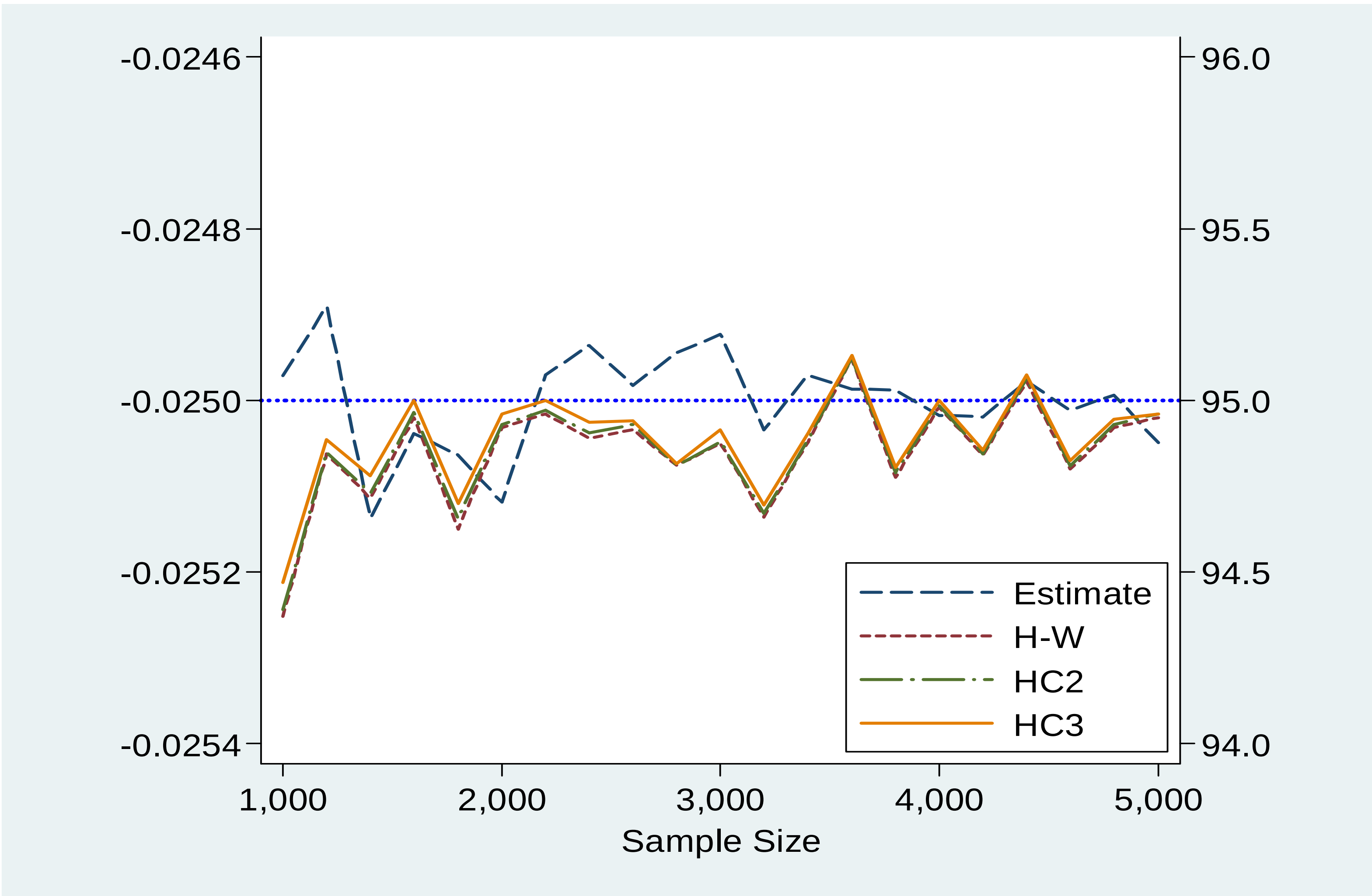
# Simulation

- Heterogeneity, $\nu_i$
  - 0
  - Discrete, $Pr(\nu_i = (\beta_0+\beta_1)/2)=p$, $Pr(\nu_i = -(\beta_0+\beta_1)/2)=(1-p)$
  - Gamma$(p(1-p), \beta_0+\beta_1)$
- $(\beta_0,\beta_1, \beta_2)=(1,-0.5,0.05)$ or $(0.05,-0.025,0.005)$
- N=200 to 5,000

# Simulation: true VAR=0.5, high incidence, $X_1$~Bernoulli(0.7), $X_1$& $X_2$ highly correlated (~0.5), X2 highly skewed (~-1.6), Gamma random effects



# Simulation: true VAR=0.025, low incidence, $X_1$~Bernoulli(0.7), $X_1$ & $X_2$ highly correlated (~0.5), X2 highly skewed (~-1.6), Gamma random effects



# Non-constant incidence and/or VAR

Regardless of baseline incidence, if true VAR, $\beta(t)$, is a function of time, it can be shown that the weighted least squares estimator converges to the weighted average of $\beta(t)$,

$$\hat{\beta} \xrightarrow{a.s.} \beta^{\tau} = \frac{\int_0^{\tau} g(t)\beta(t)dt}{\int_0^{\tau} g(t)dt}$$

where $g(t)$ is the probability of being at risk at time $t$ and $\tau$ is the length of the study duration determined by the investigators.

# Presence of event dependency

$$\hat{\beta} \xrightarrow{a.s.} \beta^{\tau} = \frac{\int_0^{\tau} g(t)\beta(t)dt}{\int_0^{\tau} g(t)dt} + \frac{\int_0^{\tau} g(t)f(Z_t)dP^1(Z_t) - \int_0^{\tau} g(t)f(Z_t)dP^0(Z_t)}{\int_0^{\tau} g(t)dt}$$

where $\tau$ is the study duration, $Z_t$ denotes the past disease history/trajectory before time *t*. *f(.)* is the effect function for the past event history. $P^0(.)$ and $P^1(.)$ are different probability measures of the event history for the placebo and intervention groups, respectively.

(similar to Cheung et al. *Stat Med* 2010)

# Thank you