Epidemiol. Infect. (1999), **123**, 65–88. Printed in the United Kingdom

Simulation studies of phase III clinical trials to test the efficacy of a candidate HIV-1 vaccine

K. N. DESAI[1,2], M.-C. BOILY[1,2,*], B. R. MASSE[1,2], M. ALARY[1,2] AND R. M. ANDERSON[3]

[1] *CHA – Pavillon St-Sacrement, Groupe de recherche en épidémiologie, 1050 chemin Ste-Foy, Québec (Québec), Canada, G1S 4L8*
[2] *Département de médecine sociale et préventive, Université Laval, Québec, Canada*
[3] *Wellcome Trust Centre for the Epidemiology of Infectious Diseases, University of Oxford, South Parks Road, Oxford, UK*

(*Accepted 25 March 1999*)

SUMMARY

One question of particular importance in phase III HIV vaccine trials is the choice of efficacy measure (EM) to validly and precisely estimate the true vaccinal efficacy. Traditional EMs, based on hazard rate ratio (HRR) or cumulative incidence ratio (CIR) are time-sensitive to mode of vaccine action and population heterogeneities. Through Monte-Carlo simulation, the performance of HRR and CIR based EMs are examined across different trial designs and vaccine and population characteristics. A new EM based on log-spline hazard regression (HARE) is proposed. Given that vaccinal properties (mode of action, time-lag, waning) are unknown *a priori*, appropriate selection of EM is problematic, and HRR and CIR can be unreliable to estimate the true maximum efficacy of candidate products. Non-random sexual mixing can exacerbate biases in HRR and CIR. HARE can offer valid estimation across different modes of vaccine action and in presence of frailty effects, contrary to its traditional counterparts. Our simulation studies highlight the weaknesses of widely used EMs while offering guidelines for trial design and suggesting new avenues for statistical analysis.

INTRODUCTION

The planning of clinical trials of candidate HIV-1 vaccines presents many methodological, ethical and financial problems. Because of the need for vaccines to slow the spread of HIV-1 in most countries of the world, governmental and international agencies have been preparing for phase III trials for candidate products that are in the phase I (safety) or phase II (immunogenicity) stages of evaluation [1]. The purpose of a phase III trial is to test the efficacy of the product under natural conditions of exposure to infection. In the design of such trials, two key issues are the choice of a statistical measure of efficacy (EM) that will estimate in an unbiased manner the true efficacy of the vaccine (VE), and the appropriate sample size to achieve a desired degree of precision in this measurement. The question of sample size can only be addressed after making a decision on the precise form of the EM.

Unfortunately, the method widely used in studies of vaccine efficacy, namely, the randomized double-blind treatment and placebo clinical trial, does not guarantee validity and precision in efficacy measurement because of many heterogeneities that may influence exposure and infection in a differential manner in the treated and untreated arms during the course of the trial. For HIV-1, these include variability in the degree and duration of immunity post vaccination (due to genetic heterogeneity in both host and virus populations), heterogeneity in exposure to infection both

* Author for correspondence.

over time (the stage of the epidemic) and between patients (e.g. due to different sexual behaviours), and viral evolution during the conduct of the trial in the study population.

The validity of traditional EMs, based on hazard rate ratios (HRR) or cumulative incidence ratios (CIR), depends whether key assumptions on which these statistics are based are indeed satisfied in any given trial. These are independence of infection events, proportionality of risks and equality in exposure within the control and treated groups. Studies have suggested that such key assumptions may present difficulties in HIV-1 vaccine trials due to a lack of independence in events [2–4], biases within the study population arising from heterogeneity in sexual behaviour [5] and the mode of action of candidate vaccines [6–8]. These issues are considered in more detail in the context of HIV-1 in the following section before examining their effect on reliability of the traditional EMs in double-blind randomized controlled trial designs.

The measurement of efficacy

A key problem in the measurement of efficacy is that of dependency between exposure and infection events within the study population. In populations in which HIV-1 is transmitting between infected and susceptible individuals, the rate at which new infections arise is dependent on the numbers of infected and susceptible people in previous time intervals. As such, observations on the rate of infection (number of new infections in a given time interval) of different members of the study population will be interdependent. This process often produces a pattern of spread where the incidence rises rapidly, reaches a peak and then falls to a stable level if the infection becomes endemic [9].

Heterogeneity in sexual behaviour is a further complication [10, 11]. How it acts on EMs in randomized trials is not immediately clear, but it will give rise to frailty selection which may undermine the benefit of the initial randomization. More precisely, individuals in high sexual activity classes will, on average, acquire infection earlier in the trial than individuals with low activity. Thus, as the trial progresses, the uninfected controls will tend to belong to the lower sexual activity classes, by comparison with the uninfected vaccinees (given that the vaccine provides some protection against infection), eventually inducing differential exposure between control and treated groups.

The properties of the vaccine will perhaps have the greatest effect on the choice of an EM. At one extreme, a vaccine may act to reduce the susceptibility of all vaccinees to infection upon exposure (of a defined type) with an infectious person by a constant fraction, δ. We refer to this mode of action as Model 1 ($\alpha = 1, 0 \leqslant \delta \leqslant 1$). At the other extreme the vaccine may confer complete protection to only a fixed proportion of the vaccinees, α. We refer to this case as Model 2 ($0 \leqslant \alpha \leqslant 1,\ \delta = 1$). Previous studies have suggested that the hazard based EMs (e.g. HRR) are more appropriate for Model 1 while proportion based measures (e.g. CIR) are better for Model 2 [6]. In practice, the properties of a vaccine may lie between these extremes and we refer to this situation as the Mixed Model ($0 \leqslant \alpha \leqslant 1,\ 0 \leqslant \delta \leqslant 1$). Alternatively, the reduction in susceptibility in those who respond to immunization (i.e. it provides some protection) may vary from one individual to another. We call this the Distributional Model ($0 \leqslant \alpha \leqslant 1, 0 \leqslant \delta_i \leqslant 1$, for individual i). Further complications are introduced if the action of the vaccine alters over time due either to a waning in protection, or to a delay in the generation of a protective response post immunization (time-lag). In absence of prior knowledge on the HIV-1 vaccine and certain population characteristics, all these factors considered together highlight the need for a robust statistical EM that is reliable for a diverse array of circumstances.

Statistical and mathematical studies of vaccine efficacy

In an early study of vaccine efficacy measurement, Smith and colleagues [6] noted the time-dependency both of incidence based EMs under the assumption of Model 2, and simple proportions under the assumption of Model 1. This problem was also highlighted by Greenland and colleagues [7] who demonstrated time-dependence in both measures under a Mixed Model. More recently, an increasing number of researchers have used mathematical models of infectious agent transmission to consider the adequacy of trial designs and EMs [2, 8, 12–23]. For example, studies by Svenssen [17] demonstrated that in a stable closed population with heterogeneity in susceptibility to infection (either due to behaviour or genetic background) their Model 1 estimator (a logarithmic

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

transformation of CIR or attack rates) will be negatively biased. On the other hand, positive bias in the estimator may occur when the mode of vaccine action is heterogeneous within the trial population. Haber and colleagues [12] showed in the context of acute outbreaks in a closed homogeneous randomly mixing population that validity of vaccinal efficacy measurement using their Model 1 estimator depended on the fraction vaccinated even under a Model 1 mode of vaccine action. However, under a Model 2 vaccine an EM based on simple proportions of attack rates behaved adequately. They generalize the results to consider heterogeneity in transmission rates and demonstrate that the traditional EMs can have considerable bias [13]. Extending results to consider non-random mixing, analyses revealed further that the EMs that presuppose random mixing can be very biased [14]. On the basis of these results Haber and colleagues [14] proposed an EM that takes account of contact patterns between members of the population. Although bias improved, use of the proposed new estimator in HIV vaccine trials would depend on knowledge of sexual contact patterns between sexual activity classes which is rarely available in practice.

Halloran and colleagues [8] employed a stochastic simulation model to examine the effect of heterogeneity in vaccinal protection. They showed that their Model 1 and Model 2 estimators can provide confidence bounds for the mean vaccinal efficacy estimate. Halloran and colleagues [15] also examined the performance of EMs based on transmission probabilities, hazard ratios and attack rates under Models 1 and 2 with the assumption that vaccinees increase exposure by adopting high risk behaviours in the belief that they are fully protected against infection. They showed that EMs are very sensitive to the assumption of increased exposure and may even adopt negative values interpreted as an immunosuppressive effect of vaccination. To palliate for this difficulty, they proposed a measure based on transmission probabilities. Although less sensitive, this parameter is difficult to measure in practice. Halloran and colleagues [16] also use frailty mixing models to conclude that heterogeneities in susceptibility or exposure generally lead to underestimations in efficacy estimates.

Other studies reveal the sensitivity of vaccine efficacy measures to various forms of heterogeneity but in the majority of cases the model framework is not geared specifically for phase III trials for HIV candidate vaccines [18–20]. Some discuss trials designed to measure aspects of vaccination other than direct effects on vaccinees, such as measurement of reduction in infectiousness or population effectiveness [21, 22]. Others yet do not embark directly on the question of EMs, but on other questions in the context of HIV dynamics and control. These include deterministic studies of eradication criteria in defined populations [24], the influence of the phase of the epidemic on biases in the recruitment of patients for trials [25], the significance of time varying effects on estimation during the trial [23], as well as others [26–28].

These studies highlight many problematic sources of heterogeneity that can act within study populations exposed to HIV infection. To date most of the available results have been derived under several assumptions such as closed or small populations, observational studies, sexually homogeneous populations, random mixing, equal exposure to infection or constant baseline hazard or incidence. To our knowledge, few or no studies investigating EMs consider specifics such as sex, cohort type or staggered entry in a randomized double-blinded controlled vaccine trials embedded in a much larger population in which HIV is not at equilibrium. Few (or none) account for biological specifics of HIV such as transmission probabilities that vary within the three phases of HIV infectivity necessary to better represent HIV dynamics, incidence and prevalence. Furthermore, few (or none) investigate the importance of the Distributional Model, time-lag or waning properties of vaccination on EMs.

In an effort to assist in the planning of phase III trials and in the interpretation of trial results, we study how different sources of heterogeneity influence the traditional EMs based on hazard rates (HRR) on simple proportions (CIR) in large double-blind controlled trials. The aim is to find which measure has the greatest precision and the least bias, given the complex nonlinear dynamics of HIV transmission. While we may not be able to control for all the different influences, it will be important (for clinicians, public health specialists, etc.) to know their effects in order to interpret well the results and to know when to be confident or reserved with respect to conclusions. Lastly, in an endeavour to improve estimation, EMs based on hazard estimation (HARE) and hazard regression (HEFT) [29, 30], an adaptive log-spline technique for estimation of conditional hazard functions apt to deal with time dependent data, is introduced.

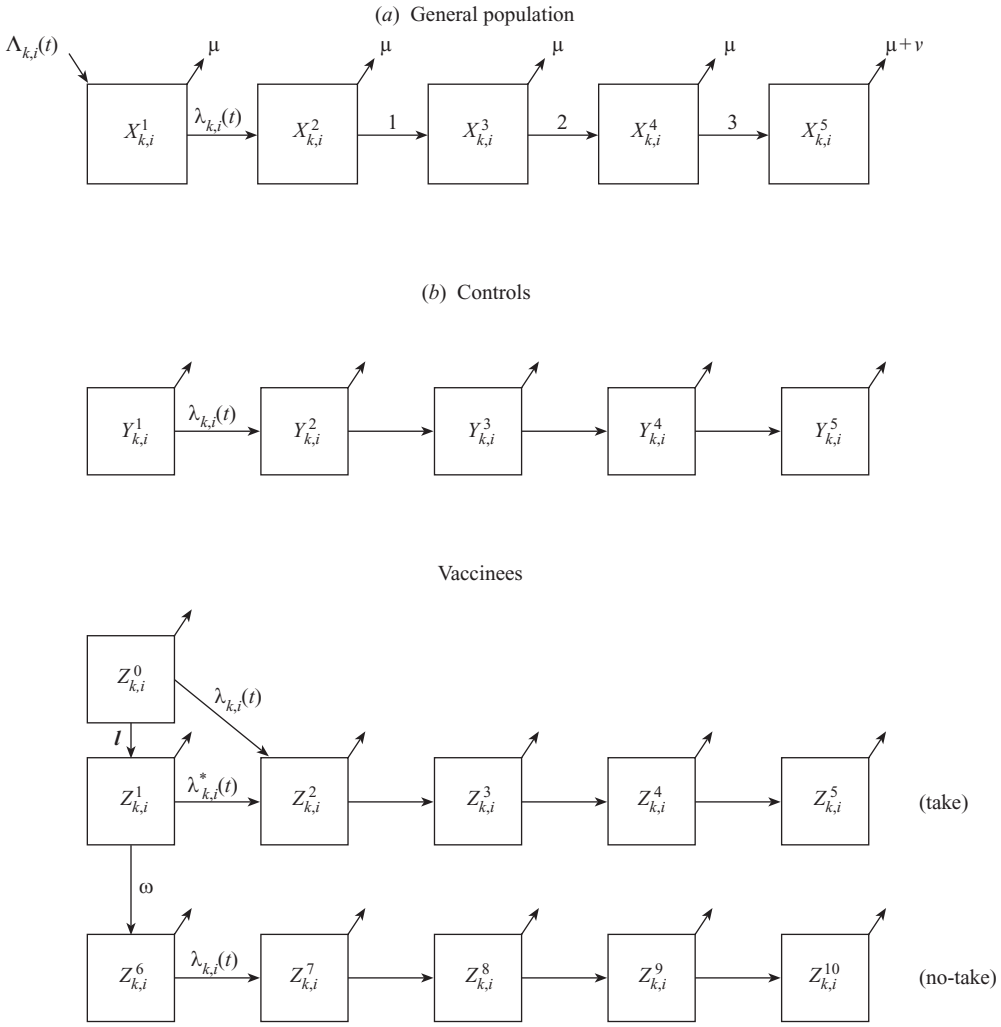

Fig. 1. (*a*) Stochastic Mathematical Model representing birth, death, HIV infection and passage between the three different stages of HIV infectivity in a sexually active population before commencement of a clinical trial. (*b*) Stochastic Mathematical Model representing the general population, controls, vaccinees in whom the vaccine provides protection (i.e. takes) and vaccinees who are not protected (i.e. does not take) in an HIV clinical trial. Variables are defined as follows: $\lambda(t)$ is the force of HIV infection general population, controls and vaccinees without developed immunity. $\lambda^*(t)$ is the force of infection in vaccinees with developed immunity. $X^1_{k,i}$, $Y^1_{k,i}$, $Z^1_{k,i}$, $Z^0_{k,i}$ and $Z^6_{k,i}$ are HIV susceptibles respectively in the general population, controls, vaccinees with developed immunity, vaccinees with undeveloped immunity, and vaccinees who do not take (vaccine failures). Similarly $X^j_{k,i}$, $Y^j_{k,i}$, $Z^j_{k,i}$, $Z^{j+5}_{k,i}$ for $j = 2, 3, 4$ are HIV infecteds. $X^5_{k,i}$, $Y^5_{k,i}$, $Z^5_{k,i}$, $Z^{10}_{k,i}$ are individuals afflicted with full blown AIDS. Description of other parameters and their values are given in Table 2. Forces of infection $\lambda(t)$ and $\lambda^*(t)$ are detailed in Appendix 1.

METHODS

The mathematical model

A compartmental stochastic model serves as the template to mirror different features of phase III trial design and HIV transmission. The model consists of five disease states ($h = 1, \ldots, 5$) representing the different stages of progression from susceptibility, to infection, to AIDS (Fig. 1*a*). Susceptibles are labelled by the superscript $h = 1$, full blown AIDS patients by $h = 5$, and three stages of progression from HIV infection (with different degrees of infectiousness) as $h = 2$, 3 and 4. The sexually active population is stratified by sex ($k = 1$ for women, and $k = 2$ for men) and by sexual activity defined by the rate of sexual partner acquisition. Six activity classes are defined ($i = 1, \ldots, 6$) where at one extreme are individuals of low sexual activity ($i = 1, 2$) and at the other are high activity individuals such as female commercial sex workers (CSW) and male clients of sexually transmitted disease clinics (MCSC) ($i = 3, \ldots, 6$). The number of individuals in disease state h, activity class i and of sex k, is given by $X^h_{k,i}(t)$. Transition between states occurs by infection, progression to disease,

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

Table 1. *Events and rates in general population, controls, and vaccinees*

(r) Description of event	Rate ($R_{r,k,i}$)
General population	
1 Recruitment into sexually active population	$R_{1,k,i} = \Lambda_{k,i}(t)$†
2 Natural death rate (NDR) of susceptibles	$R_{2,k,i} = \mu X^1_{k,i}(t)$
3 NDR of HIV infecteds in first disease state	$R_{3,k,i} = \mu X^2_{k,i}(t)$
4 NDR of HIV infecteds in second disease state	$R_{4,k,i} = \mu X^3_{k,i}(t)$
5 NDR of HIV infecteds in third disease state	$R_{5,k,i} = \mu X^4_{k,i}(t)$
6 Death rate of AIDS patients (from natural causes or AIDS)	$R_{6,k,i} = (\mu+\nu)\, X^5_{k,i}(t)$
7 HIV infection of susceptibles	$R_{7,k,i} = \lambda_{k,i}(t)\, X^1_{k,i}(t)$†
8 Progression to second state of HIV infection	$R_{8,k,i} = \gamma_1 X^2_{k,i}(t)$
9 Progression to third state of HIV infection	$R_{9,k,i} = \gamma_2 X^3_{k,i}(t)$
10 Progression to AIDS	$R_{10,k,i} = \gamma_3 X^4_{k,i}(t)$
Control group	
11 NDR of susceptibles	$R_{11,k,i} = \mu Y^1_{k,i}(t)$
12 NDR of HIV infecteds in first disease state	$R_{12,k,i} = \mu Y^2_{k,i}(t)$
13 NDR of HIV infecteds in second disease state	$R_{13,k,i} = \mu Y^3_{k,i}(t)$
14 NDR of HIV infecteds in third disease state	$R_{14,k,i} = \mu Y^4_{k,i}(t)$
15 Death rate of AIDS patients (from natural causes or AIDS)	$R_{15,k,i} = (\mu+\nu)\ Y^5_{k,i}(t)$
16 HIV infection of susceptibles	$R_{16,k,i} = \lambda_{k,i}(t)\ Y^1_{k,i}(t)$†
17 Progression to second state of HIV infection	$R_{17,k,i} = \gamma_2\, Y^2_{k,i}(t)$
18 Progression to third state of HIV infection	$R_{18,k,i} = \gamma_2\, Y^3_{k,i}(t)$
19 Progression to AIDS	$R_{19,k,i} = \gamma_3\, Y^4_{k,i}(t)$
Vaccinees who take	
20 NDR of vaccinees before time lag	$R_{20,k,i} = \mu Z^0_{k,i}(t)$
21 NDR of vaccinees after time lag	$R_{21,k,i} = \mu Z^1_{k,i}(t)$
22 NDR of HIV infecteds in first disease state	$R_{22,k,i} = \mu Z^2_{k,i}(t)$
23 NDR of HIV infecteds in second disease state	$R_{23,k,i} = \mu Z^3_{k,i}(t)$
24 NDR of HIV infecteds in third disease state	$R_{24,k,i} = \mu Z^4_{k,i}(t)$
25 death rate of AIDS patients (from natural causes or AIDS)	$R_{25,k,i} = (\mu+\nu)\ Z^5_{k,i}(t)$
26 Development of immunity (time-lag)	$R_{26,k,i} = lZ^0_{k,i}(t)$
27 HIV infection of susceptibles not having developed immunity	$R_{27,k,i} = \lambda_{k,i}(t)\ Z^0_{k,i}(t)$†
28 HIV infection of susceptibles with developed immunity	$R_{28,k,i} = \lambda^*_{k,i}(t)\ Z^1_{k,i}(t)$†
29 Progression to second state of HIV infection	$R_{29,k,i} = \gamma_1 Z^2_{k,i}(t)$
30 Progression to third state of HIV infection	$R_{30,k,i} = \gamma_2 Z^2_{k,i}(t)$
31 Progression to AIDS	$R_{31,k,i} = \gamma_3 Z^4_{k,i}(t)$
32 Loss of vaccine protection (waning)	$R_{32,k,i} = \omega Z^1_{k,i}(t)$
Vaccinees who do not take	
33 NDR of susceptibles	$R_{33,k,i} = \mu Z^6_{k,i}(t)$
34 NDR of HIV infecteds in first disease state	$R_{34,k,i} = \mu Z^7_{k,i}(t)$
35 NDR of HIV infecteds in second disease state	$R_{35,k,i} = \mu Z^8_{k,i}(t)$
36 NDR of HIV infecteds in third disease state	$R_{36,k,i} = \mu Z^9_{k,i}(t)$
37 Death rate of AIDS patients (from natural causes or AIDS)	$R_{37,k,i} = (\mu+\nu)\ Z^{10}_{k,i}(t)$
38 HIV infection of susceptibles	$R_{38,k,i} = \lambda_{k,i}(t)\ Z^6_{k,i}(t)$†
39 Progression to second state of HIV infection	$R_{39,k,i} = \gamma_1 Z^7_{k,i}(t)$
40 Progression to third state of HIV infection	$R_{40,k,i} = \gamma_2 Z^8_{k,i}(t)$
41 Progression to AIDS	$R_{41,k,i} = \gamma_3 Z^9_{k,i}(t)$

† See Appendix 1 for formal definitions of $\Lambda_{k,i}(t)$, $\lambda_{k,i}(t)$ and $\lambda^*_{k,i}(t)$.

death or immigration into the population at rates where the time between events is assumed to be exponentially distributed. The rate at which any event occurs, $R(t)$, in a population of size $N(t) = \Sigma_{h,k,i} X^h_{k,i}(t)$, is defined by $R(t) = \Sigma_{R,k,i} R_{r,k,i}(t)$, where r counts the different events. Given a stratification of two sexes and six sexual activity classes, 120 mutually exclusive events can occur (2 sexes $\times$ 6 classes $\times$ 10 possible events) in the absence of clinical trial, as defined in Figure 1*a* and Table 1.

Upon commencement of the clinical trial in the defined population, a sample of individuals from

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

Table 2. *Parameters used in simulations*

Parameter	Value(s)	
Population parameters		
Duration of sexual active lifetime $1/\mu$	35 years	
Annual rate of partner change ($m_{k,i}$ in six activity classes at $t = 0$) under heterogeneity	Activity classes 1–6	
Females	0·8 1·5 100 200 300 400	
Males	1·0 2·0 5 10 15 20	
Under homogeneity		
Females	50 50 50 50 50 50	
Males	50 50 50 50 50 50	
Proportion of each sex per activity class at $t = 0$	Activity classes 1–6	
Females	0·66 0·33 0·0015 0·0015 0·003 0·003	
Males	0·53 0·27 0·05 0·05 0·05 0·05	
Sexual mixing preferences ($\phi_{k,i,j}(t)$)	Proportional ($W_{k,i,i} = 1$), medium ($W_{k,i,i} = 5$) and strong ($W_{k,i,i} = 10$) assortative. $W_{k,i,i} = 1\ \forall i \neq j, k = 1, 2$	
Population size at time = 0	$N = 1\,000\,000$	
Trial design parameters		
HIV incidence at sampling (λ_c) in person–years	6% (rising), 15% (rising and falling), 25% (rising and falling)	
Sample size	4000	
Cohort being sampled	Commercial sex workers (females classes 3–6), male clients of STD clinics (males classes 3–6)	
Duration of follow-up	2 and 5 years	
Accrual (recruitment) period	0, 1, 2·5 and 5 years	
Biological parameters		
Death rate due to AIDS (ν)	1 year^{-1}	
Duration of infectivity phases 1 to 3 in years ($1/\gamma_1$, $1/\gamma_2$, $1/\gamma_3$)	$1/\gamma_1 = 0{\cdot}5$, $1/\gamma_2 = 6{\cdot}5$, $1/\gamma_3 = 1$	
Per partnership probability of HIV transmission (β_{kij})	homogeneity	heterogeneity
Stage I		
Female class 1–2 to male class 1–2	0·05	0·16
Female class 1–2 to male class 3–6	0·05	0·087
Female class 3–6 to male class 1–6	0·05	0·05
Stage II		
Female class 1–2 to male class 1–2	0·001	0·0032
Female class 1–2 to male class 3–6	0·001	0·0017
Female class 3–6 to male class 1–6	0·001	0·001
Stage III		
Female class 1–2 to male class 1–2	0·0071	0·051
Female class 1–2 to male class 3–6	0·0071	0·026
Female class 3–6 to male class 1–6	0·0071	0·0071
Values are doubled for male to female transmission		
Vaccine parameters		
Waning of vaccine (half life in years)	2, 5, 10 and lifelong	
Mean reduction in susceptibility δ (Models 1, 2, Mixed), μ_δ (Distributional Model)	25%, 50%, 75%, 100%	
Standard Deviation in reduction in susceptibility (σ_δ (Distributional Model only))	0%, 8%, 16·7%, 25%	
Take (α)	25%, 50%, 75%, 100%	
Time-lag before development of immunity ($1/l$)	0, 0·5, 1 year	
Mode of vaccine action	Models 1, 2, Mixed, distributional	

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

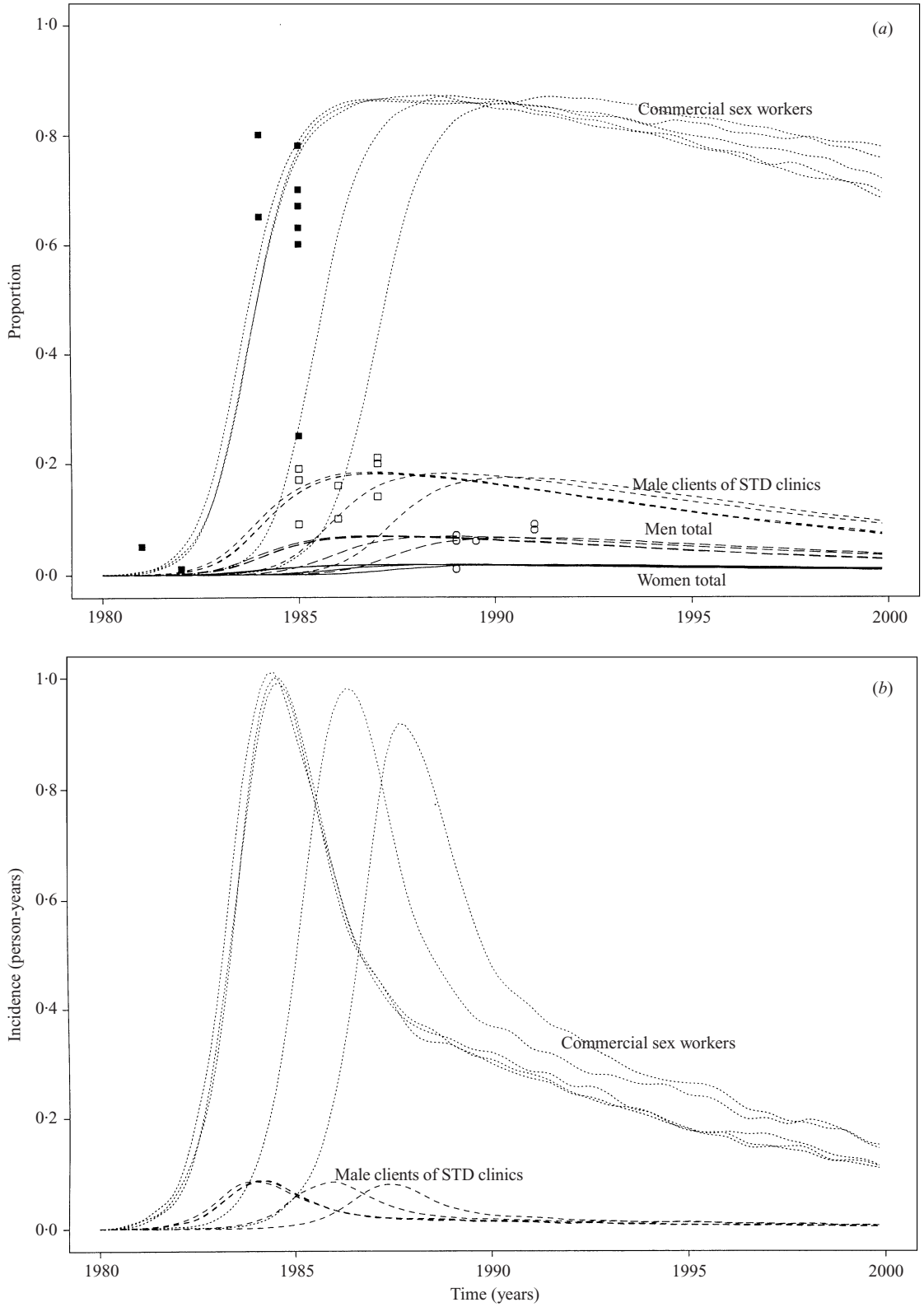

Fig. 2. (*a*) HIV prevalence curves in absence of vaccination for CSW, MCSC, total male and total female population in five repetitions of the mathematical model under parameter values characteristic of a sub-Saharan population (scenario 2). Black squares indicate HIV prevalence in prostitutes in Nairobi from field studies. Similarly, the squares indicate prevalence in male clients of STD clinics and the circles prevalence in women in general [43]. Introduction of HIV is assumed to occur in 1980. (*b*) Corresponding HIV incidence curves for the same five repetitions. HIV incidence for female-total and male-total are omitted for clarity.

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

designated risk groups (i.e. sexual activity classes of given sex) are randomized into controls and vaccinees. As defined in Figure 1*b*, the number of controls of disease state h, sex k in the activity class i is denoted $Y^h_{k,i}(t)$ $(h = 1, \ldots, 5)$. The number of vaccinees at time t of disease state h, sex k and activity class i is defined as $Z^h_{k,i}(t)$ $(h = 0, \ldots, 5)$. Vaccinees in state $h = 0$ represent those in whom protective immunity has not as yet developed post immunization. All vaccinees start in state $h = 0$ and then pass to state $h = 1$ once protective immunity has developed after a defined average time delay. The number of vaccinees in whom immunization does not produce any protection (i.e. does not 'take') is defined as $Z^h_{k,i}(t)$ $(h = 6, \ldots, 10)$. Parameter assignments for vaccine characteristics including the mode of vaccine action, vaccine efficacy and the average time-lag $(1/l)$ are specified in each simulation. Table 2 lists the parameter assignments. The time-lag in the development of protection is mirrored by the movement of individuals between two states at the average rate of l. In the first state $(h = 0)$ individuals are fully susceptible to infection, while in the second state $(h = 1)$ susceptibility is reduced to the maximum level induced by the vaccination. Note those who have been vaccinated, but acquire infection due to imperfect vaccine efficacy, are assumed to be as equally infectious as unvaccinated infecteds. Waning of vaccine protection over time post immunization, is represented by individuals totally losing their protection at a rate of ω (i.e. they move directly to $Z^6_{k,i}(t)$). The half life of immunization is taken to be the period from the receipt of vaccination (maximal protection) to when 50% of individuals lose protection. The adaptation of the basic model to mirror a phase III vaccine trial requires the specification of an additional 31 rates $(r = 11, \ldots, 41)$. These are detailed in Table 1. Appendix 2 provides details on the Monte-Carlo procedure.

Simulation procedures

The mathematical model is used to generate infection and vaccination events within a simulated phase III trial of a candidate HIV vaccine to investigate the performance of different measures of vaccine efficacy. A variety of simulations are performed with different parameter assignments (see Table 2) for true vaccine efficacy (25, 50, 75%), the mode of vaccine action (Models 1 and 2, plus Mixed and Distributional Models), the waning of vaccine induced immunity (half-lives of 2, 5 and 10 years and lifelong protection) and the time-lag (0, 6 and 12 months). Different behavioural characteristics are examined including heterogeneity in sexual activity (compared with homogeneous activity) and various mixing patterns (proportional, moderate and strong assortativeness). The trial designs considered are cohort based (e.g. CSW, MCSC) with patient follow-up of 2 or 5 years and different periods of recruitment of patients to the trial (all enter at the same time, or over 1, 2·5 and 5 years). Different HIV incidence levels at the start of the trial (0·06, 0·15 and 0·25 person years) and under conditions of rising and falling incidence over the duration of the trial are examined. All simulation trials are assumed to be longitudinal, randomized and double-blind with two arms and that there is no loss to follow-up. Unless otherwise stated, the study population is 4000 in size with an initial incidence of 0·06 per annum in CSW. This type of design is appropriate for the evaluation of the direct effects of vaccination on the individual (efficacy). Total, indirect (i.e. herd immunity) and reduced infectiousness effects can be studied with different types of design [2, 21, 22].

Values for the epidemiological, behavioural and demographic parameters (Table 2) are chosen to represent CSW and MCSW, as well as the general population in a sub-Saharan city such as Nairobi. In the simulated trials, cohorts of CSW have been used primarily because of their high HIV incidence; however, the model has been adapted to represent cohorts of MCSC as well. The chosen parameter values give rise to the prevalence and incidence curves plotted in Figure 2 (in absence of a clinical trial).

The validation scenario considered is that of homogeneity in sexual activity within a group of CSWs, with three levels of vaccine efficacy under the mode of action defined by Models 1 and 2, plus the Mixed model, with 2 or 5 years of follow-up. However, the second scenario, incorporating heterogeneity in sexual activity (i.e. six sexual activity classes), provides the basic set of parameter assignments for further, more elaborate, trials and assumptions. The different simulations for various parameter assignments and trial designs are grouped into the nine scenarios as listed in Table 3.

Statistical analyses

For each parameter set 50 repetitions were performed. Except for the chance effects introduced into the

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

Table 3. *Trial scenarios studied*

Scenario	Factors studied	The details examined
1	Homogeneity	Models 1, 2 and Mixed of vaccine action. Cohort of female CSW, no accrual, waning, or time-lag. Initial incidence = 0·06 and rising. Annual rate of partner change = 50 for entire population. Other parameter values are as in scenario 2
2	Heterogeneity	Parameters as indicated in Table 2 with proportional mixing, cohort of prostitutes and Models 1, 2 and Mixed for vaccine action. No accrual, waning or time-lag
3	Mixing	Individuals are 5 times (medium assortative) or 10 times (strong assortative) more likely to choose a partner within the same sexual activity class. Other parameters are as in scenario 2
4	Vaccine waning	Waning half-life of 2, 5 and 10 years and lifelong immunity (i.e. no waning). Other parameters are as in scenario 2
5	Time-lag	Time-lags of 6 months and 1 year before development of immune response. Other parameters are as in scenario 2
6	Distributional mode of vaccine action	Vaccinees have individually varying reduction in susceptibilities (δ) taken from a Normal (μ_δ, σ_δ^2) distribution with $\mu_\delta = 0{\cdot}5$ and $\sigma_\delta = 8\,\%$, 16·7 % and 25 %, truncated at $\mu_\delta \pm 0{\cdot}5$. Other parameters are as in scenario 2
7	Cohort-male clients of STD clinics	Cohort formed from males of classes $i = 3, \ldots, 6$. Other parameters are as in scenario 2
8	Accrual (staggered entry)	Individuals assumed to enter uniformly at one month intervals over a period of 1, 2, 2·5 and 5 years. Sample size is 1000. Other parameter values are as in scenario 2
9	Incidence	Clinical trials begins at different incidence levels (0·15, 0·25) and at rising and falling incidence. Sample size = 300. Other parameter values are as in scenario 2

Monte-Carlo simulations (i.e. the sequences of random numbers drawn to decide which type of event occurs and when), the conditions of each repetition were identical. The number of repetitions was sufficient to assess precision and accuracy between scenarios. Here, accuracy measures the degree of systematic (non-random error) bias in the measurement of vaccine efficacy, given knowledge of true efficacy (i.e. that defined in the parameter assignments of the model). Precision defines the degree of dispersion (random error) around the expected value of the measure. The term validity implies a high degree of accuracy. For each simulation repetition the CIR is calculated by the Cochran–Mantel χ^2 statistic [31] and the HRR is determined by fitting a Cox proportional hazard model [32, 33] using exact infection times as endpoints at 2 and 5 years of follow-up. Both CIR and HRR were also calculated adjusting for sexual activity class according to standard procedures [31, 32]. Vaccine efficacy measurement based on adjusted or unadjusted HRR and CIR are denoted, respectively, VE_{HRR}^{adj}, VE_{HRR}^{unadj}, VE_{CIR}^{adj} and VE_{CIR}^{unadj}. Writing VE_{HRR} or VE_{CIR} refers to both the adjusted or unadjusted measure.

A new EM based on HARE and HEFT log-spline hazard estimation [29, 30] was applied to a limited number of scenarios. Basically, this approach consists of constructing the hazard curves for vaccinees and controls based on follow-up data. The hazard curves will be functions of time t. Estimation of VE then consists of taking the hazard ratio at a certain t. Brunet and colleagues [23] suggest choosing $t = 0$ which theoretically should provide valid estimation across different modes of action and in presence of frailty effects. We consider $t = 6$ months instead to gain precision while striving for validity. The HARE and HEFT methodology is described further in Appendix 3.

Upon calculation of HRR, CIR, HARE or HEFT, vaccine efficacy was then estimated by the relationship $VE_{CIR} = 1 - CIR$, $VE_{HRR} = 1 - HRR$. In the case of HARE (or HEFT)

$$VE_{HARE}(t) = 1 - \frac{h_v(t)}{h_c(t)},$$

where $t = 6$ months and $h_v(t)$ and $h_c(t)$ are the hazard functions of vaccinees and controls estimated by HARE (or HEFT).

Assessment of the validity and variability of the EMs was performed by the graphical examination of

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

VE_{CIR}, VE_{HRR}, VE_{HARE} and VE_{HEFT} utilizing bar plots of the relative bias

$$\left(RB = \frac{VE_{OBS} - VE_{TRUE}}{VE_{TRUE}}\right)$$

of the EMs calculated from the 50 repetitions. True vaccine efficacy is defined as $VE_{TRUE} = \delta.\alpha$ under Models 1, 2 and Mixed, and $VE_{TRUE} = \mu_{\delta}.\alpha$ under the Distributional Model, where μ_{δ} is the mean vaccinal efficacy. In the case of time-varying efficacy, as under time-lag and waning, VE_{TRUE} is defined as the maximum protection conferred by the vaccine (i.e. after development of full immunity and before any waning).

Computational details

Numerical results were generated by a program written in FORTRAN and simulations were performed on a SUN Sparc4, 5 and Ultra workstations. HRR and CIR were calculated using SAS Version 6.11, and HARE and HEFT were calculated using S-PLUS Version 3.4. HARE and HEFT software was obtained from StatLib [34].

RESULTS

Homogeneous sexual behaviour (Scenario 1)

The first scenario examined is that of a vaccine trial with 2 and 5 years of follow-up in a population with homogeneous sexual behaviour. Parameter values for this and subsequent scenarios are documented in Table 3, and summary statistics are listed in Table 4. As expected from previous studies [6, 7, 12, 15], under the Model 1 assumptions on the mode of vaccine action, the VE_{HRR}^{unadj} is the most appropriate measure to use in terms of its accuracy and precision, while the VE_{CIR}^{unadj} systematically underestimates true efficacy and its bias increases with longer periods of follow-up. Conversely, under Model 2 the VE_{CIR}^{unadj} demonstrates its validity, while VE_{HRR}^{unadj} overestimates particularly with longer follow-up. Under the Mixed Model both measures exhibit bias, the degree of which depending on follow-up duration. The VE_{HRR}^{unadj} under the Mixed Model is somewhat more precise at low and medium vaccine efficacies, while the VE_{CIR}^{unadj} performs better at high efficacy. At 5 years of follow-up both measures will be considerably biased.

Heterogeneous sexual behaviour (Scenario 2)

Heterogeneity in sexual behaviour introduces several new effects into the picture. Both VE_{HRR}^{unadj} and VE_{CIR}^{unadj} exhibit a negative bias under Model 1, which increases as the follow-up period lengthens. However, this can be rectified in the case of the VE_{HRR}^{unadj} measure by adjusting for sexual activity classes. For example, at 5 years of follow-up the VE_{HRR}^{unadj} had $-0{\cdot}3\,\%$ and $-10{\cdot}8\,\%$ relative bias under the homogeneous and heterogeneous assumptions, respectively for a vaccine of 50% efficacy while the VE_{HRR}^{adj} demonstrates respectively $-0{\cdot}3\,\%$ and $0{\cdot}6\,\%$ relative bias. The picture is similar for all vaccine efficacies and follow-up periods. Use of adjusted measures, however, implies that the trial must incorporate measurement of sexual activity via the administration of behavioural questionnaires to trial subjects.

Heterogeneity in sexual behaviour also acts to increase variability in the measures. For example, comparing the homogeneous case with the heterogeneous one for a Model 1 vaccine of 50% efficacy and with 2 years of follow-up, the standard deviations of the VE_{HRR}^{unadj} and VE_{CIR}^{unadj} increase from $1{\cdot}8$ to $2{\cdot}7\,\%$, and from $1{\cdot}8$ to $2{\cdot}6\,\%$, respectively. The pattern is similar for other models of action, efficacies and different periods of follow-up.

Under Model 2, the VE_{CIR} gives the best performance, although adjusting for sexual behaviour seems to contribute nothing to the precision or accuracy.

The measure that performs least well (i.e. VE_{CIR} under Model 1 and VE_{HRR} under Model 2) is typically more biased under heterogeneity is behaviour than homogeneity.

Under the Mixed Model the VE_{HRR}^{unadj} is subject to two competing time-dependent biases. First, is the same negative bias present in Model 1 under heterogeneous sexual activity and second, is the positive bias under the Mixed Model. We observe (see Table 4) that at low or moderate efficacy (25% or 50%), the VE_{HRR}^{unadj} increasingly underestimates with longer follow-up (the negative bias due to heterogeneity predominates), whereas at high efficacy (75%) it increasingly overestimates (the bias due to the Mixed Model is more important). However, the VE_{HRR}^{adj}, which is not subject to the negative bias under heterogeneous sexual activity, incurs only a positive bias due to the Mixed Model, and therefore overestimates at all vaccinal efficacies.

In general, under heterogeneity in sexual activity, the VE_{HRR}^{adj} is better at low and medium vaccine efficacy, and the VE_{CIR} is better at high efficacy. With respect to direction of bias, our results are consistent with those of other studies [6, 7, 12, 15], particularly with Svenssen [17], who used a somewhat simpler

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

Table 4. *Selected simulation results. Vaccinal efficacy estimates and standard deviations* (*in parentheses*) *for adjusted and unadjusted* VE_{HRR} *and* VE_{CIR} *calculated at 2 and 5 years of follow-up in the nine scenarios investigated*

Scenario		Selected results					
(1) Homogeneity		25% efficacy		50% efficacy		75% efficacy	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	25 (2·4)	25 (1·6)	50 (1·8)	50 (1·4)	75 (1·3)	75 (0·8)
	CIR	20 (2·2)	10 (1·0)	43 (1·8)	27 (1·1)	70 (1·4)	55 (1·0)
Model 2	*HRR*	30 (3·2)	42 (2·0)	56 (2·3)	68 (1·2)	79 (1·3)	86 (0·7)
	CIR	25 (2·7)	25 (1·4)	50 (2·3)	50 (1·3)	75 (1·4)	75 (1·0)
Mixed Model	*HRR*	25 (2·5)	26 (1·8)	52 (2·5)	55 (1·6)	79 (1·3)	86 (0·7)
	CIR	20 (2·1)	11 (1·1)	45 (2·4)	33 (1·3)	75 (1·4)	75 (1·0)
(2) Heterogeneity		25% efficacy		50% efficacy		75% efficacy	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	23 (3·2)	20 (2·1)	50 (2·7)	44 (1·6)	74 (1·9)	70 (1·1)
	HRR_{adj}	25 (3·3)	25 (2·1)	51 (2·7)	50 (1·4)	75 (1·7)	75 (1·0)
	CIR	16 (2·6)	2 (1·0)	42 (2·6)	10 (1·0)	70 (2·3)	39 (2·1)
	CIR_{adj}	16 (2·7)	2 (1·0)	42 (2·6)	10 (1·1)	70 (2·2)	39 (2·0)
Model 2	*HRR*	30 (4·0)	42 (1·9)	55 (3·9)	67 (1·7)	79 (2·3)	85 (1·0)
	HRR_{adj}	31 (4·0)	45 (2·1)	56 (3·9)	69 (1·5)	79 (2·3)	86 (1·0)
	CIR	25 (3·4)	25 (1·2)	50 (3·8)	50 (1·7)	76 (2·3)	75 (1·3)
	CIR_{adj}	25 (3·3)	25 (1·2)	50 (3·8)	50 (1·6)	76 (2·4)	75 (1·3)
Mixed Model	*HRR*	24 (3·3)	23 (2·3)	52 (3·9)	51 (1·7)	79 (2·3)	85 (1·0)
	HRR_{adj}	25 (3·4)	27 (2·3)	53 (3·8)	56 (1·5)	79 (2·3)	86 (1·0)
	CIR	18 (2·8)	3 (1·0)	45 (3·9)	20 (1·5)	76 (2·3)	75 (1·3)
	CIR_{adj}	18 (2·7)	4 (1·0)	45 (3·8)	20 (1·3)	76 (2·4)	75 (1·3)
(3) Mixing medium assortative		25% efficacy		50% efficacy		75% efficacy	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	21 (3·2)	19 (2·6)	45 (2·6)	39 (1·9)	72 (1·4)	65 (1·1)
	HRR_{adj}	25 (3·0)	25 (2·6)	50 (2·6)	50 (1·7)	75 (1·3)	75 (1·0)
	CIR	11 (2·2)	2 (0·9)	31 (2·4)	8 (1·1)	63 (1·8)	27 (1·3)
	CIR_{adj}	11 (1·9)	2 (0·8)	31 (2·2)	8 (1·1)	63 (1·9)	27 (1·3)
Model 2	*HRR*	33 (2·8)	43 (2·1)	58 (2·5)	68 (1·4)	80 (1·7)	86 (1·0)
	HRR_{adj}	37 (2·6)	47 (1·8)	62 (2·5)	71 (1·2)	82 (1·6)	88 (0·9)
	CIR	25 (2·4)	24 (1·5)	50 (2·4)	50 (1·3)	75 (1·8)	75 (1·3)
	CIR_{adj}	25 (2·2)	25 (1·5)	50 (2·4)	50 (1·3)	75 (1·7)	75 (1·3)
Mixed Model	*HRR*	22 (3·0)	20 (2·7)	50 (2·7)	46 (1·8)	80 (1·7)	86 (1·0)
	HRR_{adj}	26 (3·0)	27 (2·3)	54 (2·2)	56 (1·4)	82 (1·6)	88 (0·9)
	CIR	12 (2·1)	3 (0·9)	38 (2·9)	14 (1·3)	75 (1·8)	75 (1·3)
	CIR_{adj}	13 (1·9)	3 (0·9)	38 (2·6)	14 (1·2)	75 (1·7)	75 (1·3)
(3) Mixing strong assortative		25% efficacy		50% efficacy		75% efficacy	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	19 (2·6)	17 (2·4)	43 (2·6)	37 (1·9)	70 (1·8)	64 (1·5)
	HRR_{adj}	25 (2·2)	25 (2·1)	50 (2·5)	50 (1·4)	75 (1·5)	75 (1·0)
	CIR	11 (2·2)	5 (1·3)	31 (2·4)	14 (1·8)	61 (2·3)	38 (1·5)
	CIR_{adj}	11 (1·8)	5 (1·1)	31 (2·3)	14 (1·6)	61 (2·2)	38 (1·3)
Model 2	*HRR*	34 (2·8)	43 (1·8)	60 (2·3)	68 (1·3)	81 (1·6)	86 (0·9)
	HRR_{adj}	40 (2·6)	47 (1·8)	65 (2·1)	72 (1·2)	84 (1·3)	89 (0·7)
	CIR	25 (2·2)	25 (1·4)	50 (2·1)	50 (1·4)	75 (1·8)	75 (1·2)
	CIR_{adj}	25 (1·9)	25 (1·3)	50 (2·0)	50 (1·4)	75 (1·6)	75 (1·2)

[*continued overleaf*

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

Table 4 (*cont.*)

Scenario	Selected results						
Mixed Model	*HRR*	21 (2·8)	19 (2·4)	48 (2·9)	44 (2·0)	81 (1·6)	86 (0·9)
	HRR_{adj}	26 (2·5)	26 (2·3)	55 (2·4)	56 (1·6)	84 (1·3)	89 (0·7)
	CIR	13 (2·0)	6 (1·2)	36 (2·8)	20 (1·5)	75 (1·8)	75 (1·2)
	CIR_{adj}	13 (1·7)	5 (1·1)	36 (2·5)	20 (1·4)	75 (1·6)	75 (1·2)
(4) Waning		50 % efficacy 10 year half life		50 % efficacy 5 year half life		50 % efficacy 2 year half life	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	44 (3·3)	39 (1·9)	41 (3·3)	35 (1·8)	37 (3·5)	35 (2·0)
	HRR_{adj}	46 (3·2)	45 (1·8)	42 (3·2)	41 (1·8)	38 (3·4)	39 (1·8)
	CIR	36 (3·1)	8 (1·0)	33 (3·2)	7 (1·2)	26 (2·8)	4 (1·1)
	CIR_{adj}	36 (3·0)	8 (1·1)	33 (3·1)	7 (1·1)	26 (2·6)	4 (1·0)
Model 2	*HRR*	52 (3·2)	59 (1·7)	48 (3·6)	52 (1·7)	36 (3·9)	32 (3·6)
	HRR_{adj}	52 (3·2)	61 (1·4)	49 (3·6)	55 (1·7)	37 (3·8)	36 (3·6)
	CIR	46 (3·0)	39 (1·2)	42 (3·6)	30 (1·4)	31 (3·0)	15 (1·0)
	CIR_{adj}	46 (3·0)	39 (1·2)	42 (3·5)	30 (1·3)	31 (2·8)	15 (0·9)
Mixed Model	*HRR*	47 (3·8)	46 (2·2)	43 (3·6)	40 (2·0)	34 (2·9)	28 (2·1)
	HRR_{adj}	48 (3·7)	51 (1·8)	44 (3·3)	45 (1·6)	35 (2·9)	33 (2·0)
	CIR	40 (3·7)	16 (1·8)	36 (3·4)	13 (1·4)	27 (2·7)	7 (1·1)
	CIR_{adj}	40 (3·5)	16 (1·6)	36 (3·1)	13 (1·4)	27 (2·6)	7 (1·1)
(5) Time-lag		50 % efficacy No time-lag		50 % efficacy 6 months time-lag		50 % efficacy 1 year time-lag	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	50 (2·7)	44 (1·6)	43 (2·9)	43 (1·9)	34 (3·3)	38 (1·8)
	HRR_{adj}	51 (2·7)	50 (1·4)	44 (2·9)	47 (1·6)	35 (3·2)	42 (1·7)
	CIR	42 (2·6)	10 (1·0)	37 (2·6)	20 (1·7)	29 (2·7)	18 (1·3)
	CIR_{adj}	42 (2·6)	10 (1·1)	37 (2·5)	20 (1·6)	29 (2·6)	18 (1·3)
Model 2	*HRR*	55 (3·9)	67 (1·7)	48 (3·1)	63 (1·5)	39 (3·1)	57 (1·5)
	HRR_{adj}	56 (3·9)	69 (1·5)	49 (3·0)	65 (1·4)	40 (2·7)	59 (1·2)
	CIR	50 (3·8)	50 (1·7)	43 (2·8)	46 (1·2)	34 (2·7)	40 (1·3)
	CIR_{adj}	50 (3·8)	50 (1·6)	43 (2·8)	46 (1·3)	34 (2·3)	40 (1·2)
Mixed Model	*HRR*	52 (3·9)	51 (1·7)	43 (2·7)	48 (1·9)	35 (3·0)	43 (1·8)
	HRR_{adj}	53 (3·8)	56 (1·5)	44 (2·6)	52 (1·6)	36 (3·0)	47 (1·6)
	CIR	45 (3·9)	20 (1·5)	38 (2·6)	27 (1·6)	30 (2·8)	23 (1·5)
	CIR_{adj}	45 (3·8)	20 (1·3)	38 (2·5)	26 (1·5)	30 (2·7)	23 (1·5)
(6) Distribution Take = 100 %		$\mu_\delta = 50\,\%$ $\sigma_\delta = 8\,\%$		$\mu_\delta = 50\,\%$ $\sigma_\delta = 16{\cdot}7\,\%$		$\mu_\delta = 50\,\%$ $\sigma_\delta = 25\,\%$	
		2 years	5 years	2 years	5 years	2 years	5 years
	HRR	50 (3·4)	47 (1·9)	50 (3·2)	48 (1·9)	51 (3·5)	51 (2·0)
	HRR_{adj}	51 (3·5)	51 (1·9)	51 (3·1)	52 (1·7)	52 (3·4)	54 (1·8)
	CIR	44 (3·4)	23 (1·6)	44 (3·3)	25 (1·8)	45 (3·5)	28 (1·6)
	CIR_{adj}	44 (3·4)	23 (1·6)	44 (3·3)	25 (1·8)	45 (3·4)	28 (1·5)
Distribution Take = 75 %		$\mu_\delta = 67\,\%$ $\sigma_\delta = 8\,\%$		$\mu_\delta = 67\,\%$ $\sigma_\delta = 16{\cdot}7\,\%$		$\mu_\delta = 67\,\%$ $\sigma_\delta = 25\,\%$	
		2 years	5 years	2 years	5 years	2 years	5 years
	HRR	51 (3·1)	52 (1·6)	52 (3·5)	54 (1·6)	52 (2·9)	55 (1·6)
	HRR_{adj}	52 (3·1)	56 (1·5)	53 (3·6)	58 (1·7)	53 (3·0)	58 (1·5)
	CIR	46 (3·0)	30 (1·5)	46 (3·4)	32 (1·8)	46 (2·9)	33 (1·7)
	CIR_{adj}	46 (3·0)	30 (1·5)	46 (3·4)	32 (1·7)	46 (3·0)	33 (1·7)

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

Table 4 (*cont.*)

Scenario	Selected results						
(7) Male clients		25% efficacy		50% efficacy		75% efficacy	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	24 (5·8)	24 (4·8)	49 (4·5)	49 (3·3)	75 (3·0)	75 (2·3)
	HRR_{adj}	25 (5·8)	24 (4·8)	50 (4·5)	49 (3·2)	75 (2·9)	75 (2·2)
	CIR	23 (5·6)	22 (4·5)	48 (4·5)	46 (3·3)	73 (3·1)	72 (2·4)
	CIR_{adj}	23 (5·6)	22 (4·5)	48 (4·4)	46 (3·1)	73 (3·0)	72 (2·4)
Model 2	*HRR*	26 (6·0)	27 (4·8)	51 (4·7)	52 (4·1)	77 (3·3)	78 (2·8)
	HRR_{adj}	27 (5·9)	27 (4·5)	52 (4·5)	54 (4·0)	77 (3·3)	78 (2·7)
	CIR	25 (5·8)	25 (4·5)	49 (4·7)	50 (4·1)	76 (3·4)	76 (2·9)
	CIR_{adj}	25 (5·7)	25 (4·2)	50 (4·5)	50 (4·0)	76 (3·4)	76 (2·9)
Model Mixed	*HRR*	27 (5·0)	26 (4·1)	50 (4·6)	50 (3·9)	77 (3·3)	78 (2·8)
	HRR_{adj}	27 (4·7)	26 (3·7)	50 (4·6)	51 (3·9)	77 (3·3)	78 (2·7)
	CIR	25 (4·8)	24 (3·9)	48 (4·6)	48 (4·0)	76 (3·4)	76 (2·9)
	CIR_{adj}	25 (4·6)	24 (3·6)	48 (4·6)	48 (3·9)	76 (3·4)	76 (2·9)
(8) Accrual		50% efficacy 1 year recruitment		50% efficacy 2·5 years recruitment		50% efficacy 5 years recruitment	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	48 (4·2)	44 (3·6)	47 (4·4)	45 (3·6)	47 (3·9)	44 (3·6)
	HRR_{adj}	50 (4·2)	50 (3·5)	49 (4·3)	49 (3·4)	50 (4·3)	49 (3·4)
	CIR	37 (4·1)	19 (2·6)	34 (3·8)	20 (2·7)	36 (3·5)	24 (3·1)
	CIR_{adj}	38 (4·0)	19 (2·5)	33 (3·5)	20 (2·5)	36 (3·5)	24 (2·8)
Model 2	*HRR*	59 (3·8)	69 (2·5)	62 (3·3)	69 (2·5)	59 (3·8)	65 (3·0)
	HRR_{adj}	61 (3·7)	71 (2·4)	64 (3·1)	71 (2·3)	62 (3·7)	68 (2·9)
	CIR	50 (3·7)	50 (2·6)	50 (3·4)	50 (2·6)	50 (3·7)	50 (3·0)
	CIR_{adj}	50 (3·6)	50 (2·7)	50 (3·3)	50 (2·6)	50 (3·7)	50 (3·0)
Model Mixed	*HRR*	51 (4·4)	51 (3·3)	51 (3·6)	51 (2·9)	50 (3·8)	49 (3·2)
	HRR_{adj}	52 (4·3)	55 (2·8)	53 (3·5)	55 (2·7)	53 (3·7)	55 (3·0)
	CIR	40 (4·2)	25 (2·5)	38 (3·5)	26 (2·6)	39 (3·5)	30 (2·7)
	CIR_{adj}	40 (3·9)	25 (2·2)	38 (3·5)	26 (2·6)	40 (3·3)	30 (2·6)
(9) Incidence		50% efficacy 0·15 person-years and rising		50% efficacy 0·25 person-years and rising		50% efficacy 0·25 person-years and falling	
		2 years	5 years	2 years	5 years	2 years	5 years
Model 1	*HRR*	49	50	47	45	47	45
	HRR_{adj}	49	51	50	50	48	49
	CIR	46	44	33	19	32	19
	CIR_{adj}	46	43	33	19	32	19
Model 2	*HRR*	51	54	62	69	62	69
	HRR_{adj}	52	56	64	71	64	71
	CIR	48	49	49	49	50	50
	CIR_{adj}	49	49	50	49	50	50
Model Mixed	*HRR*	46	48	52	52	51	50
	HRR_{adj}	46	49	54	56	55	57
	CIR	43	41	38	26	37	24
	CIR_{adj}	43	41	37	25	38	25

model framework, and Halloran and colleagues [16], who used frailty mixing models.

Patterns of mixing between sexual activity classes (Scenario 3)

Non-random mixing between different sexual activity classes adds a further source of heterogeneity. We examine the effects on EMs under the assumption of proportional mixing (all individuals have equal preference where couple formation is only restricted by supply and demand [11]), of moderate assortative mixing (individuals have a five times greater prefer-

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

ence for forming sexual partnerships with people in their own sexual activity class) and strong assortative mixing (preference is ten times). Mixing details are given in Appendix 1.

As summarized in Table 4, under Model 1 all EMs with the exception of VE_{HRR}^{adj} incur a further negative bias compared to scenario 2 in presence of assortative mixing. The negative bias associated with the VE_{HRR}^{unadj} increases with the degree of assortativity, while that of VE_{CIR}^{unadj} increases dramatically when passing from proportional to medium assortative but not significantly from medium to strong assortative mixing. Interestingly, the validity of VE_{CIR}^{unadj} improves slightly at 5 years follow-up under strong assortative mixing when compared to 5 years under medium assortative mixing. This is because under strong assortativity the sampled sexual activity classes tend towards being a homogeneous group as they mix less and less with other classes.

As expected, under Model 2 the VE_{CIR} is preferred, while an increasing positive bias is incurred for both VE_{HRR}^{adj} and VE_{HRR}^{unadj} with the degree of assortativity. In this case, the VE_{HRR}^{adj} is considerably less accurate than VE_{HRR}^{unadj}. Adjusting the VE_{CIR} once again improves nothing with respect to bias.

Under the Mixed Model, it is difficult to claim the best measure; however, the VE_{HRR}^{adj} is probably preferable for weak, moderate and strong vaccines and throughout mixing types, even though it will likely overestimate the true vaccinal efficacy. The performance of other EMs is more variable on mixing type and vaccinal efficacy under the Mixed Model.

In general, non-random mixing exacerbates any biases, whether the direction is positive or negative, that may already be present in other scenarios. In some instances under assortative mixing, precision may increase compared with proportional mixing. This is because the epidemic is more concentrated within the highest sexual activity classes of the population and of the cohort of CSW sampled under assortative mixing, producing a greater proportion of realized endpoints during the clinical trial. These results are in broad agreement with Haber and colleagues [13, 14] who concur that biases for their estimators are more pronounced under non-random than random mixing.

Waning of vaccinal protection over time (Scenario 4)

The effect of waning is investigated in order to determine whether valid estimation of the maximum efficacy is possible with traditional measures. Ideally, we would be able to disentangle the maximum efficacy and the rate of waning in a controlled vaccine trial since both of these vaccine characteristics are meaningful at the public health level. Limiting ourselves to the use of VE_{HRR} and VE_{CIR} without assuming prior knowledge of the waning process, we examine whether the EMs reflect the maximum efficacy.

We find under all three models of vaccine action, a waning in efficacy over time post vaccination induces a negative bias in all of the EMs compared to the true maximum efficacy, even for VE_{HRR} under Model 1 and VE_{CIR} under Model 2. This effect becomes more severe as the half-life of protection decreases (Fig. 3). For example, for a vaccine with 50% maximum efficacy and a half-life of 10 years, an extra 11% increase in bias (relative to the maximum efficacy) in the VE_{HRR}^{adj} at only 2 years of follow-up is observed under Model 1, by comparison with lifelong protection. Under the same conditions an extra 9% increase in bias occurs in the VE_{CIR}^{unadj} under Model 2. These results could produce a false conclusion that a moderate to good vaccine, but with a short duration of protection, has low efficacy. For instance, a vaccine of 50% maximal efficacy under Model 1 and with a protection half-life of 2 years, would be recorded as having a 38% and 39% efficacy by the VE_{HRR}^{adj} at, respectively, 2 and 5 years of follow-up. Under Model 2 the same vaccine would be estimated to have an efficacy of 31% and 15% at respectively 2 and 5 years of follow-up. That negative bias is induced upon the EMs (which assume that vaccinal protection is instantaneous and constant) by waning is to be expected since some vaccinees who lose protection during the study will be infected more quickly than if protection were lifelong, thereby causing a considerable underestimation of efficacy. With respect to the use of the wrong measure under a given mode of action, the VE_{CIR}, which is already negatively biased in absence of waning under Model 1, develops a further negative bias as half-life of waning shortens. Also, VE_{HRR} which begins positively biased in absence of waning under Model 2 reverses the direction of its bias with decreasing half-life. Furthermore, note that the positive bias of the VE_{HRR} under Model 2 that is offset by the roughly equivalent negative bias induced by a 5 year half-life of waning can deceptively produce a seemingly accurate estimation. The observed trends are qualitatively the same at different vaccinal efficacies (25%, 75%).

The degree of bias observed in some of these results

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

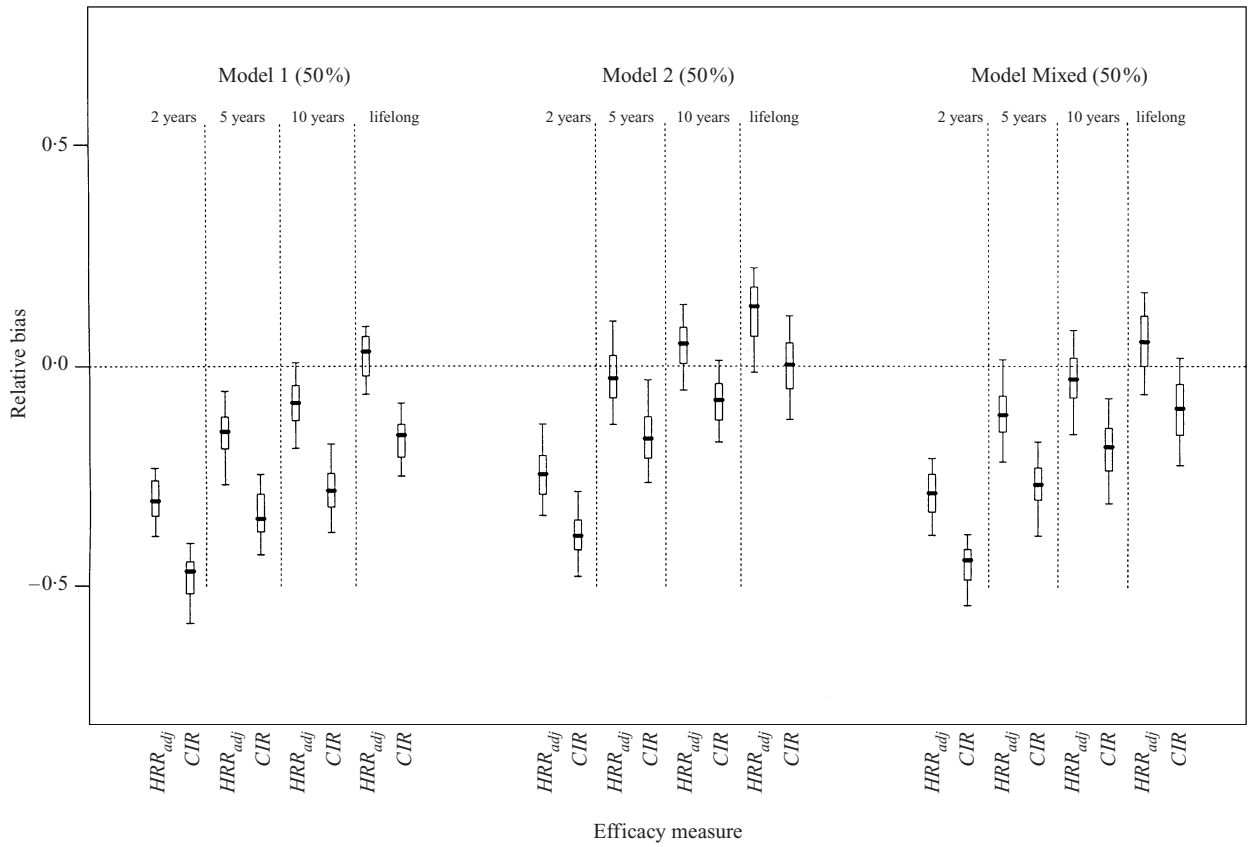

Fig. 3. Performance of VE^{adj}_{HRR} and VE^{unadj}_{CIR} in a proportionally mixing sexually heterogeneous population under the influence of waning of vaccine protection at 50% maximum efficacy under Models 1, 2 and Mixed. Waning with half-life of 2, 5 and 10 years is compared to lifelong immunity. Follow-up duration is 2 years. Other parameters are as explained in Table 3. Barplots indicate the 5th, 25th, 50th, 75th and 95th percentile of the efficacy measures calculated over the 50 repetitions. Performance of VE^{unadj}_{HRR} and VE^{adj}_{CIR} is qualitatively similar (not shown).

can be explained mathematically. Recalling that our model of waning is reflected by total individual loss in protection from the maximum level to zero (rather than gradual decline), the proportion protected is a function of time ($\alpha = \alpha(t)$), while the reduction in susceptibility drops from its maximum value to zero ($\delta =$ maximum value before waning, $\delta = 0$ after waning). To explain the bias, consider a vaccine with a duration of protection assumed to be exponentially distributed with a mean of 2 years. Then we calculate over the course of an n-year long trial that the average proportion who are protected can be given by

$$\int_0^n \exp(-\tfrac{1}{2}t)\, dt/n.$$

For a 2 year follow-up, this integral gives 63·2%, which when multiplied by the maximum efficacy of 50% gives 31·6%. By comparison, the VE^{unadj}_{CIR} and VE^{adj}_{HRR} estimate 31% and 37% efficacy under a Model 2 vaccine, and 27% and 35% under the Mixed model. Thus, while both measures will underestimate the maximum efficacy under this type of waning, the VE^{unadj}_{CIR} may be expected to accurately reflect the mean efficacy of only a Model 2 vaccine.

Time-lags in protection post immunization (Scenario 5)

As under waning, we examine the ability of the traditional EMs to estimate to maximum vaccinal efficacy under time-lags in the generation of immunity. The effect of this temporal change in efficacy over time is illustrated in Figure 4 for a vaccine of 50% efficacy. Under Models 1 and 2, plus the Mixed Model, the underestimation of vaccine efficacy by the VE_{HRR} and VE_{CIR} measures increases as the duration of the delay rises. This is to be expected, since the window of susceptibility post immunization will give rise to some cases of infection in vaccinated individuals. Some suggest that this time-lag could last until the third booster, after which maximum protection might be achieved.

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

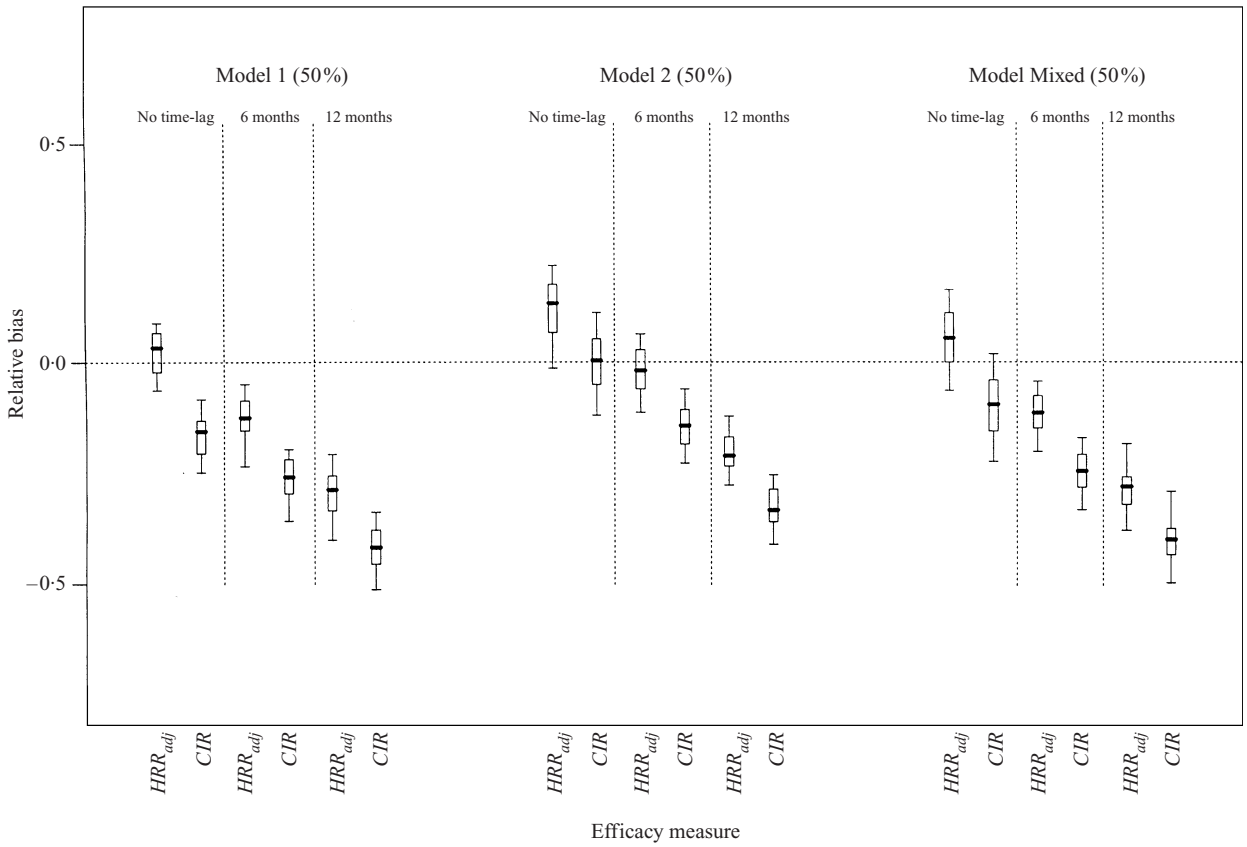

Fig. 4. Performance VE^{adj}_{HRR} and VE^{unadj}_{CIR} under the influence of 6 and 12 month time-lag of a vaccine of 50% maximum efficacy under Models 1, 2 and Mixed. Follow-up period is 2 years. Other parameters are as explained in Table 3. Barplots indicate the 5th, 25th, 50th, 75th and 95th percentile of the efficacy measures calculated over the 50 repetitions. Performance of VE^{unadj}_{HRR} and VE^{adj}_{CIR} is qualitatively similar (not shown).

For Model 1 and the Mixed Model the negative bias in the VE_{HRR} may be reduced by long periods of follow-up. However, the bias is exacerbated (expectedly so) by long follow-up with the VE_{CIR}. Under the Mixed Model the VE_{HRR} performs better than the VE_{CIR} at moderate efficacy. More generally, neither the VE_{HRR} nor the VE_{CIR} can offer reliable and unbiased estimation for any mode of vaccine action without prior information on time-lag. For instance, under a 6 month time-lag and 50% efficacy under the Mixed Model, the VE^{adj}_{HRR} would recognize 44% efficacy at 2 years of follow-up, while the VE^{unadj}_{CIR} would detect 38%.

Heterogeneity in the response to vaccination (Scenario 6)

Heterogeneity in response to vaccination was investigated assuming all vaccinees (100% take) benefited from some individual level of protection selected from a Normal distribution with mean of $\mu_\delta = 50\%$ and a standard deviation (S.D.) of $\sigma_\delta = 8\%$, 16·7% or 25%. An S.D. of 8% corresponds to a normal distribution with $\mu_\delta \pm 3\sigma_\delta$ contained within the interval [0·25, 0·75]. Similarly, an S.D. of 16·7% corresponds to $\mu_\sigma \pm 3\sigma_\delta$ contained within [0, 1] and an S.D. of 25% to $\mu_\delta \pm 2\sigma_\delta$ contained within [0, 1]. Also, considered was a normal distribution with mean $\mu_\delta = 66\%$ and 75% take. As illustrated in Figure 5, results at 5 years of follow-up show the relative bias of the VE^{adj}_{HRR} and VE^{unadj}_{HRR} to increase gradually under the Distributional Model at 75% and 100% take. On the other hand, the negative bias of VE_{CIR} under Model 1 is reduced substantially in passing from Model 1 (S.D. = 0) to the Distributional Model, and continues to do so with increasing variance σ^2_δ. The VE_{CIR} still underestimates the true vaccinal efficacy, however. The VE^{adj}_{HRR} is the best estimator under this mode of vaccine action with relative bias increasing gradually with increasing variance in vaccinal efficacy. Our results concur with those of Svenssen [17] who also report positive bias under heterogeneity in vaccinal response.

For a sufficiently long follow-up, it is natural that bias in estimates rise with increasing σ^2_δ since

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

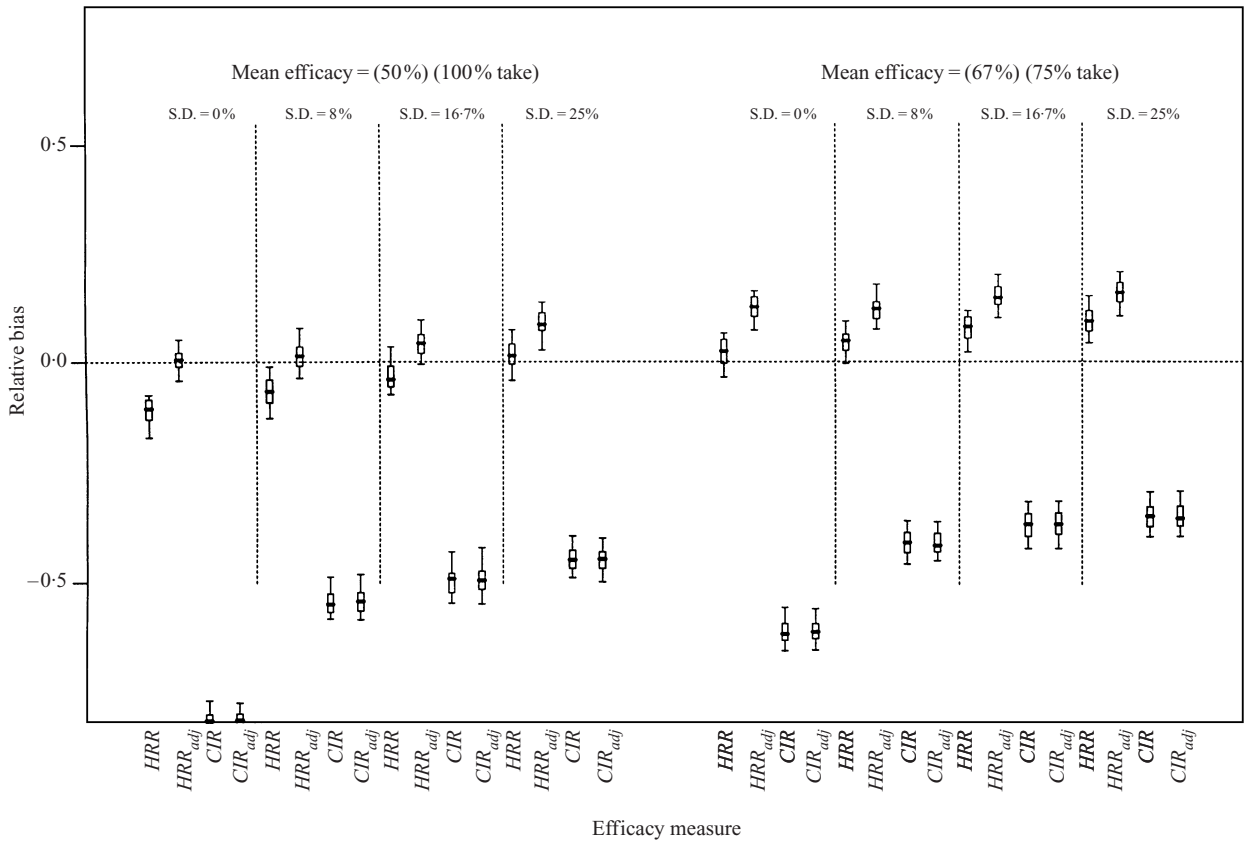

Fig. 5. Performance of adjusted and unadjusted VE_{HRR} and VE_{CIR} in a sexually heterogeneous population under the Distributional Model for two vaccines each with 50% true efficacy and three different variances ($\sigma_\delta = 8\%$, 16·7% and 25%). Take and mean efficacy of first vaccine are set at $\alpha = 1{\cdot}0$, $\mu_\delta = 50\%$ and that of the second vaccine are set at $\alpha = 0{\cdot}75$, $\mu_\delta = 67\%$. Follow-up period is 5 years. Other parameter values are as explained in Table 3. Barplots indicate the 5th, 25th, 50th, 75th and 95th percentile of the efficacy measures calculated over the 50 repetitions.

individuals at the low end of the distribution will approach a zero reduction in susceptibility almost as total vaccine failures as in Model 2 non-takers. Thus the Distributional Model can take on properties of Model 2. Furthermore, with larger variance in vaccinal efficacy, greater proportions of vaccinees will be infected earlier on average leaving the remaining vaccinees with higher average levels of protection than under small variances. This is reflected in the VE^{adj}_{HRR} and VE^{unadj}_{CIR} which are seen to increase with increasing variance in vaccinal efficacy when the follow-up period is sufficiently long.

Understandably, the precision of the measures may decrease with increasing variance of the distribution in reduction in susceptibility. For example, at 2 years of follow-up the S.D. increases from 2·7% to 3·5% and 2·6% to 3·4% of VE^{adj}_{HRR} and VE^{unadj}_{CIR} respectively in going from a distribution with $\sigma_\delta = 0\%$ to one with $\sigma_\delta = 8\%$. Thus, potentially the Distributional Model may have a non-negligible effect on variance of EMs which may play an important role during sample size considerations, at least for short follow-up periods.

Study population (Scenario 7)

A comparison was made between two types of study population, namely, MCSC and female CSW under identical conditions of sample size, vaccine efficacy, trial design and baseline incidence. In these settings the VE_{HRR} and VE_{CIR} are less prone to biases caused by the mode of vaccine action, mixing and efficacy levels in cohorts of MCSC. However, all measures show a high degree of variability in MCSC than in CSW, particularly with low efficacy vaccines. This is due to the faster rate of spread of HIV in the latter group (i.e. more events occur). Since the precision of the VE_{CIR} is time dependent under Model 1, the measure develops bias more quickly in a CSW cohort than in the MCSC cohort due to the more rapid accumulation of events in the former by comparison

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

with the latter. Similarly, the VE_{HRR} under Model 2 develops a positive bias more quickly in a CSW cohort than the equivalent MCSC cohort. The rise in variability in the MCSC setting is due to the smaller number of infection endpoints. There thus arises a conflict between improving the statistical precision or validity. Validity is enhanced within the MCSC setting, while precision is greater within the CSW setting.

Accrual of patients during the conduct of the trial (Scenario 8)

The question of interest in this scenario is whether the strategy of recruiting subjects during the conduct of the trial can act to counter frailty selection. Recruitment permits replacement of the most vulnerable trial participants. However, accrual periods of 1, 2·5 and 5 years, in combination with 2 and 5 years of follow-up of every patient recruited, did not significantly improve the performance of the EMs in either cohorts of CSWs or MCSCs. Although the biases were reduced for long follow-up periods in the VE_{CIR} under Model 1 and the Mixed Model, and in the VE_{HRR} under Model 2, this was largely due to the fact that the epidemic moved to a more stable endemic level when both follow-up and recruitment period are lengthy, rather than due to any special merit of patient accrual.

Changing HIV incidence over the conduct of the trial (Scenario 9)

In this scenario we examine varying levels of HIV incidence at the start of the trial and changes therein over its conduct. Because at different phases in the epidemic curve it was not possible to recruit 4000 uninfected prostitutes in our population, sample size was fixed at 300. Simulation results for incidence of 15% (rising), 25% (rising) and 25% (falling) are given in Table 4. The S.D. of the measures are also excluded because of the small sample size. The VE_{HRR} in Model 1 and VE_{CIR} in Model 2 do not incur appreciably any new biases due to the different incidence levels. However, the bias present for the VE_{CIR} in Model 1 and VE_{HRR} in Model 2 is larger with higher initial incidence in the rising phase. This bias, however, declines with decreasing initial incidence in the dropping phase. In fact, at incidence of 15% and falling, the wrong measures (VE_{CIR} under Model 1 and VE_{HRR} under Model 2) perform almost as well as the right measure in terms of bias (not shown in Table 4). Similarly, under the Mixed Model the EMs are least biased at incidence of 15% and falling compared to the other three incidence levels.

At incidence of 15% and falling, the change in HIV incidence is very slow and almost constant, in contrast with the other three phases of incidence. That the EMs are less biased at 15% and falling follows the same reasons given for the reduced bias in EMs under cohorts of MCSC (scenario 7). In general, these analyses reveal that biases are generated by rapidly rising incidence and that there may be practical problems in recruiting sufficient uninfecteds at certain times in the epidemic's development.

Hazard Regression (HARE) and Hazard Estimation (HEFT)

Given the numerous potential biases in traditional EMs we explore the use of log-spline models to obtain valid estimation of vaccinal efficacy across different modes of vaccine action and possible frailty effects. This approach is in part motivated by Brunet and colleagues [23]. To assess the performance of HARE and HEFT, VE_{HARE} and VE_{HEFT} were compared with VE_{CIR}^{unadj} and VE_{HRR}^{adj} under Models 1, 2 and the Mixed Model in the setting of scenario 2, a moderate degree of assortative mixing (scenario 3) and a vaccine with an average protection duration of 10 years (scenario 4). As demonstrated in Figure 6, VE_{HARE} and VE_{HEFT} demonstrated accuracy more consistently after 2 years of follow-up for all modes of vaccine action. In contrast, VE_{HRR}^{adj} and VE_{CIR}^{unadj} fluctuated in accuracy depending on the mode of vaccine action. Both VE_{HARE} and VE_{HEFT} are much less influenced by mixing patterns and waning than are VE_{HRR}^{adj} and VE_{CIR}^{unadj}. However, this increased robustness is achieved at the cost of greater variability. VE_{HARE} and VE_{HEFT} were obtained using $t = 6$ months after calculating hazard curves based on 2 years of follow-up data (see Statistical Analyses in Methods section regarding need to choose t). This choice of t was suitable for striking a balance between validity and precision, contrary to Brunet's suggestion of $t = 0$, where precision is low. How VE_{HARE} and VE_{HEFT} perform under a broader range of conditions is the subject of another study.

DISCUSSION

At present it seems unlikely that the vaccinal properties such as mode of action, time-lag post

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

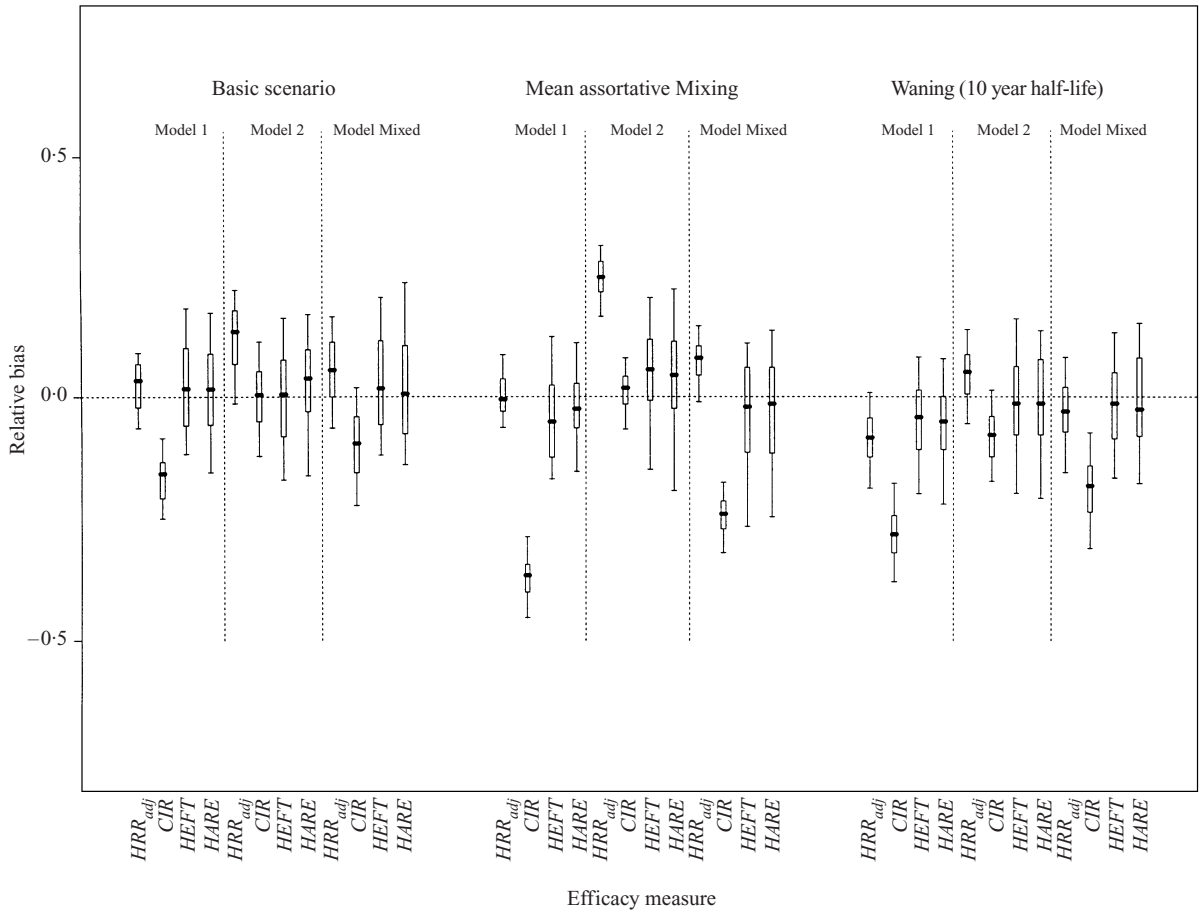

Fig. 6. Comparison of VE^{adj}_{HRR} and VE^{unadj}_{CIR} with VE_{HEFT}, VE_{HARE} under Models 1, 2 and Mixed of vaccine action and the basic scenario (scenario 2), assortative mixing (scenario 3) and 10 year vaccinal waning (scenario 4). Parameter values are as described in Table 3. Barplots indicate the 5th, 25th, 50th, 75th and 95th percentile of the efficacy measures calculated over the 50 repetitions.

immunization and waning rate of a candidate HIV vaccine will be known with precision before the conduct of a phase III trial, given current uncertainty over which immunological measures correlate best with protection against infection. It is hoped that the trials themselves will yield important information on these correlates. However, trial design raises many difficult issues which must be addressed if we are to understand how different candidate vaccines perform under conditions of natural exposure to infection. The real dilemma is that the choice of an EM, the design of the trial and concomitantly sample size determination, each depend on the precise properties of the vaccine and its mode of action. Given this somewhat unhappy state of affairs, the issue in question at present is what EM performs best both for the possible range of vaccinal properties, and in the presence of the many sources of heterogeneity that influence HIV transmission in sexually active populations.

Our simulation studies provide a series of important guidelines, although further research is required to explore new directions for EMs in a wider variety of settings. The factors influencing the performance of a defined EM for the three major possible modes of vaccine action were explored using a model framework incorporating the biological characteristics of HIV transmission and infection specifically designed to mimic the dynamics of HIV-in a population with heterogeneous sexual activity. The shortcomings of VE measures are observed despite proper randomization at the start of the trial. With time, randomization fails to keep the individual characteristics balanced between the two groups.

Given the sexual behaviour is highly heterogeneous in human communities [35], this important facet of HIV transmission was incorporated in our simulations. Our analyses first suggest that the validity of VE_{CIR} and VE_{HRR} measures is most likely unsatisfactory for all modes of vaccine action. In both

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

heterogeneous and homogenous populations accurate estimation of vaccinal efficacy will be difficult using VE_{CIR} and VE_{HRR} given the mode of vaccine action may not be known. The problem of maintaining accuracy is further exacerbated by non-random mixing, which tends to amplify any biases that already exist under random mixing. Here again, observational studies suggest that in many settings the observed pattern of mixing could be assortative [36]. Of the traditional measures the VE_{HRR}^{adj} appears to be the most reliable in these circumstances, except under the Model 2 mode of vaccine action. Adjustment of measures for sexual activity implies, of course, the importance of behavioural questionnaires for the study population. Rida [37] also has suggested that behavioural data will be essential to assess whether vaccination itself induces changes in sexual activity and therefore unequal exposure in control and treated groups.

The different possible modes of vaccine action could present yet another hurdle if more than one vaccine is tried in multi-armed trials. If the same statistical measure is used to compare two vaccines with different modes of action, then adequate evaluation of vaccine efficacy may not be achieved. For example, if the VE_{HRR} is used to compare a vaccine of Model 1 with one of Model 2 of equal efficacy, it is possible that the latter vaccine may be viewed as more effective due to the time dependent overestimation of VE_{HRR} under Model 2. In this case, differences in the survival curves (proportion of uninfected individuals at time t) between the two candidates will be due to the different modes of vaccine action since both vaccines have the same efficacy. Obviously, comparability of such candidates between trials with separate control groups is even more problematic.

However, the factors that have the greatest influence on accuracy of the EMs are waning immunity post immunization and time delays in the development of protection post vaccination. Unless additional information is available on these processes, trials will tend to underestimate true maximum efficacy. Even if average duration of time-lag and rate of waning were known, it remains to be seen whether this information would then permit correction of the traditional EMs to accurately estimate the maximum vaccine efficacy. Thus, changing efficacy over time makes trial data analysis and interpretation of vaccine efficacy estimates even more tenuous. Underestimation of true efficacy is particularly wearisome since maximum efficacy is not expected to be high for the early products that enter phase III trials [28]. Hence, ample underestimation of true efficacy could lead to the rejection of a reasonably efficacious product which may have some public health interest in certain populations [28]. This could result in a missed chance to distribute a moderately effective vaccine.

The high bias of the EMs under waning immunity and time delays in the development of protection suggest that confidence bounds derived by Halloran and colleagues [8] are unlikely to be respected. Their lower bound is based on their Model 2 estimator and upper bound based on their Model 1 estimator. Our Model 2 and Model 1 estimators are seen on average not to encompass the true maximal value under time-lag and waning influences. One way to elude the problem of waning immunity could be to administer boosters during the course of the trial in order to maintain the level of protection constant.

A distributional mode of vaccine action will also influence variability and bias. Additional bias can arise due to this mode of vaccine action, but is likely to be minor and will probably manifest only if follow-up is lengthy. The Distributional Model may play a more important role in variability of measures.

Results from the Cohort of MCSC, accrual and incidence scenarios share a common thread by demonstrating the importance of the epidemic shape on the quality of estimation. Accuracy may be improved by careful choice of a study setting (i.e. cohort type). Male attendees at sexually transmitted disease clinics have many advantages in terms of striking a sensible balance between the rate at which infection events more likely to occur and the size of the potential study population. Female CSW are very appropriate in terms of their high rates of infection, but sample size issues may restrict their potential value (i.e. they are a small proportion of the sexually active population and many are already likely to be infected in high transmission areas). A further consideration with respect to high incidence groups such as MCSC and CSW, is the high variability in sexual activity within these population strata. This can menace validity of the EMs. There is thus a dilemma between using lower activity groups to gain validity versus high transmission groups to gain precision in EMs.

Staggered entry in trial design (accrual) was in ineffective in limiting frailty or other time dependent effects. This, however, was because accrual requires lengthening the study duration in rapidly changing incidence phases in this particular cohort of CSW.

Similarly in the incidence scenario, biases were dependent on the incidence phase of the epidemic. Higher incidence rates and rising incidence at start of the study generally produced greater bias than lower incidence rates and falling incidence rates. Overall, therefore, it is difficult to eliminate bias in the complex and heterogeneous settings in which HIV spreads in a community. Furthermore, our simulations do not take account of the added complications of loss to follow-up which will probably act to further decrease precision. This is of obvious importance in the choice of a suitable cohort setting. For example, MCSC may be ideal in terms of incidence and population size, but they may be less reliable in terms of drop out in follow-up.

It is worth noting that under all the conditions investigated (except time-lag), biases of the EM actually increased with follow-up duration. While epidemiological principles suggest that the prospective cohort study involving incident cases and a short follow-up duration is the best observational design to test an aetiologic hypothesis [31], our study helps to delineate what is meant exactly by a short follow-up. Specifically, clinical trials with follow-up exceeding 2 years in an epidemic condition, such as that considered in this paper, will vary likely encompass increasingly important biases.

Playing a role in the steering of HIV clinical trials, the interim monitoring board will need to consider many factors such as time-lag post-immunization so as not to prematurely conclude from interim analyses that a vaccine is of low-efficacy. Nor should the monitoring board allow trials to continue too long in order to minimize effects of different sources of heterogeneity or waning which would cause under-estimations of the true potential of a vaccine.

Clearly, it seems reliable estimations will come only when the mode of action is known *a priori* to be either Model 1, 2 or Distributional (with 100% take) and in absence of any time-lag or waning, prerequisites which may demand a stretch of imagination. In any eventuality, we (clinicians, public health experts, immunologists and others) still want to know whether a vaccine works and how well. With this goal in mind, the factors affecting bias and variability of the VE_{CIR} and VE_{HRR} become important.

The importance of having an unbiased (valid) summary statistic of vaccine efficacy is not merely academic. As stated by Kleinbaum [31]: '[...] it is important to recognize that internal validity is the sine qua non of etiologic research [...]'. It has numerous clinical and public health implications. On the clinical side, it is important that the statistics used correctly reflect the size of the effect of vaccine [38, 39]. Moreover, with unbiased estimation of vaccine efficacy in combination with survival curves and immunological data, it may be easier to speculate on the mode of vaccine action. As clearly emphasized in past years [39], it is not enough to detect a significant difference between groups. It is also important to quantify the difference in a meaningful manner [38–40]. This is well illustrated by the controversy that followed the results of a malaria vaccine trial in Tanzania where an efficacy of 31% (VE_{IDR}) was found with a 95% CI ranging from 0% to 52% [41]. Given this level of efficacy and wide confidence interval, the decision to use the vaccine as a public health measure is not straightforward [38]. We have to know what we have on hand to make well-informed decisions. From a public health perspective it is therefore important to disentangle the issues on vaccine efficacy not only to help identify who gets infected (between vaccinees and controls) and who doesn't, but also how many. For example, without prior information on the mode of vaccine action, comparisons of survival curves in multi-armed trials may be difficult to interpret especially if the curves cross. Finally, the preventive potential of a vaccine at a population level (the effectiveness), depends not only on the efficacy of the vaccine but also on the vaccine mode of action [24]. For a fixed vaccine efficacy of 31%, McLean and Blower [24], showed that the equilibrium seroprevalence of infection is larger under Model 1 and the Mixed Model than under Model 2. This emphasizes the need for careful interpretation of vaccine efficacy estimates if we want to predict the effectiveness of future immunization programmes. After all, if we cannot deduce the true potential of a vaccine in a clinical trial, how can we expect to be able to approve and distribute a vaccine of public health interest to control the HIV/AIDS epidemic?

Despite the numerous influences that will be present in HIV vaccine trials, there is some hope that accurate efficacy estimation can be afforded by applying VE_{HEFT} and VE_{HARE} to the method originally proposed by Brunet and colleagues [23]. Their suggestion of basing the analysis on the entry point of time ($t = 0$) seems sub-optimal, since we find greater precision using the time of 6 months post entry. This time of estimation gives more robust estimates and reduces effects of frailty. In particular, HARE and HEFT

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

offer hope for valid estimation of VE across modes of vaccine action. However, more research is needed in this area both to examine the robustness of this conclusion under a wider range of settings and to derive better estimates of efficacy in heterogeneous study populations. A more general search for robust EMs is also needed. While we may not be able to eliminate bias, it is very important to understand its direction in a defined study population and with a vaccine of a given mode of action. Such an understanding will be important in the appropriate choice of sample size and study duration. Much can be learnt by the sensible use of stochastic simulation models of HIV transmission that incorporate the details of a specific trial design. Prior to implementation of phase III vaccine trials in defined study populations, such studies will help to save time and money in the planning stages, aid in the interpretation of results and in making informed conclusions in HIV vaccine clinical trials.

ACKNOWLEDGEMENTS

MCB thanks the NHRDP and MRC for financial support. MA is supported by the NHRDP and FRSQ. RMA thanks the Wellcome Trust for research grant support. The authors would like to thank Robert Poulin for his technical assistance in programming.

APPENDIX 1

Recruitment

The functional form for recruitment of new susceptibles into the sexual active population leading to a stable population in absence of AIDS is given for sex k and activity class i at time t as follows:

$$\Lambda_{k,i}(t) = \mu Z^0_{k,i}(t) + \sum_{h=1}^{4} \mu(X^h_{k,i}(t) + Y^h_{k,i}(t) + Z^h_{k,i}(t) + Z^{h+5}_{k,i}(t)). \quad (1)$$

Force of infection

The force of HIV infection for sex k and activity class i at time t in the general population, controls, vaccinees in whom the vaccine does not take, and vaccines who take but have not developed immunity is given by:

$$\lambda_{k,i}(t) = m_{k,i}(t) \sum_{j=1}^{6} \beta_{k',i,j} \phi_{k,i,j}(t) \times \left(\frac{\Sigma^4_{h=2}(X^h_{k',j}(t) + Y^h_{k',j}(t) + Z^h_{k',j}(t) + Z^{h+5}_{k',j}(t))}{NA_{k',j}(t)} \right). \quad (2)$$

Here $m_{k,i}(t)$ is the annual rate of partner acquisition of persons of sex k and class i, $\beta_{k',i,j}$ is the per partnership HIV transmission probability from sex k' and class j to sex k class i. The term $\phi_{k,i,j}(t)$ describes the mixing matrix elements (described in more detail below), and $NA_{k',j}(t)$ is the total sexually active population of sex k' and class j. Thus, $\lambda_{k,i}(t)$ is a function of the rate of sexual partner change, the HIV transmission probability and HIV prevalence. Individuals with AIDS are assumed not to contribute to the sexually active population due to the severity of their condition.

The force of HIV infection for sex k and activity class i at time t for successful vaccinees with protective immunity is given by:

$$\lambda^*_{k,i}(t) = (1 - \delta_{k,i}) \lambda_{k,i}(t), \quad (3)$$

where $\delta_{k,i}$ is the proportional reduction in susceptibility to HIV infection of sex k class i due to vaccination. In the case of the Distributional Model, $\delta_{k,i}$ is taken as the individual reduction in susceptibility for each successful vaccinee of sex k and class i.

Mixing elements

The inclusion of the elements of the mixing matrix $\phi_{k,i,j}(t)$ in the model is essential since the manner in which sexual partner formation occurs is a critical component in the spread of HIV and STDs [10, 11]. As described in Boily and Anderson (1991) [11], the mixing elements are defined as:

$$\phi_{k,i,j}(t) = \frac{W_{k,i,j} NA_{k',j}(t) m_{k',j}(0)}{\Sigma_j W_{k,i,j} NA_{k',j}(t) m_{k',j}(0)}, \quad (4)$$

subject to the constraints:

$$0 \leqslant \phi_{k,i,j}(t) \leqslant 1, \quad (5)$$

$$\sum_{j=1}^{6} \phi_{k,i,j}(t) = 1, \quad (6)$$

$$NA_{k,i}(t) \phi_{k,i,j}(t) m_{k,i}(t) = NA_{k',j}(t) \phi_{k',j,i}(t) m_{k',j}(t). \quad (7)$$

Here, $W_{k,i,j}$ defines a set of weights corresponding to the preference of individuals of sex k and activity class i for partners of the opposite sex in activity class j.

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

Thus, the mixing matrix elements $\phi_{k,i,j}(t)$ are the probability that a person of sex k class i chooses a partner of opposite sex k' and class j. The first two constraints are self-explanatory. The third, balancing supply with demand, indicates that the number of partners formed by individuals of sex k and activity class i who choose a member of the opposite sex k' and class j must equal the number of partners formed by sex k' and activity class j who choose a member of the opposite sex k and class i. In order for the supply and demand constraint to hold for all i, j and k at all t, the elements of the preference matrix $W_{k,i,j}$ (and hence the mixing matrix, $\phi_{k',j,i}(t)$) should satisfy for all i and j the following constraint:

$$\frac{W_{1,i,j}W_{1,j,1}}{W_{1,i,1}W_{1,j,j}} = \frac{W_{2,j,i}W_{2,1,j}}{W_{2,1,i}W_{2,j,j}}. \tag{8}$$

Purely assortative mixing (mixing exclusively within same activity class) occurs when $W_{k,i,j} = 1$ for $i = j$ and $W_{k,i,j} = 0$ for $i \neq j$. Random or proportionate mixing occurs when $W_{k,i,j} = 1$ for all values of k, i and j.

APPENDIX 2

In the stochastic simulations, a random sequence of individual events is generated where each of the 492 possible events (2 sexes × 6 classes × 41 events) occurs with probability $P_{r,k,i}(t)$ defined by $P_{r,k,i}(t) = R_{r,k,i}(t)/S(t)$, where $S(t) = \Sigma_{r,k,i} R_{r,k,i}(t)$. Using RAN2 [42], the event chosen at each step of the sequence is determined by a random number generated from a uniform distribution according to the 492 probabilities. The time a person of sex k and class i spends in a specific state r before making a transition is assumed to be exponentially distributed with mean $R^{-1}_{r,k,i}(t)$. Furthermore, the time between any two events is exponentially distributed with mean $S^{-1}(t)$. Therefore, the time of occurrence s of chosen event r can be determined by choosing a random number from a uniform distribution and setting it equal to $F(s)$ in the equation $F(s) = 1 - \exp[-S(t)\,s]$. Thus by an iterative process, a sequence of events and their time of occurrence is generated.

APPENDIX 3

HARE and HEFT is a general framework developed by Kooperberg, Stone and Truong [29, 30], to model the log-hazard function based on survival times. In the context of HIV clinical trials, we employ HARE and HEFT to obtain the hazard functions of controls and vaccinees based on the follow-up data. It involves an automatic procedure for stepwise addition and deletion of knots in time as well selection of covariates. Estimation of model parameters is based on maximum likelihood methods while model selection is based on the Bayes Information Criteria.

Given a possibly censored time, t, and a vector of covariables, $\mathbf{x} = (x_1, \ldots, x_m)$, HARE models the conditional hazard function, $h(t\,|\,\mathbf{x})$, $\alpha(t\,|\,\mathbf{x})$, via the log-hazard function, $\alpha(t\,|\,\mathbf{x}) = \log(h(t\,|\,\mathbf{x})$, according to the model:

$$\alpha(t\,|\,\mathbf{x}) = \sum_{j=1}^{p} \beta_j B_j(t\,|\,\mathbf{x}). \tag{9}$$

Here $B_1, \ldots, B_p$ are linear spline functions of time and the $\beta_1, \ldots, \beta_p$ are the parameters estimated by maximum likelihood methods. The resulting hazard is a piecewise linear continuous function of time. HARE also tests for interactions and non-proportional hazards in covariates and time, and accounts for them by admitting the appropriate product terms. Thus, the class of HARE models includes the sub-class of proportional hazard models.

HEFT is a similar to HARE and performs the estimation of the unconditional log-hazard function through cubic spline functions. Thus unlike HARE, HEFT does not admit covariates and employs cubic splines rather than linear ones.

REFERENCES

1. Balter M. Impending AIDS vaccine trial opens old wounds. Science 1998; **279**: 650.
2. Halloran ME, Struchiner CJ. Study designs for dependent happenings. Epidemiol 1991; **2**: 331–8.
3. Struchiner CJ, Halloran ME, Robins JM, Spielman A. The behaviour of common measures of association used to assess a vaccination programme under complex disease transmission patterns – a computer simulation study of malaria vaccines. Int J Epidemiol 1990; **19**: 187–96.
4. Halloran EM, Struchiner CJ. Causal inference in infectious diseases. Epidemiol 1995; **6**: 142–51.
5. Vaupel JW, Yashin AI. Heterogeneity's ruses: Some surprising effects on selection on population dynamics. Amer Statist 1985; **39**: 176–85.
6. Smith PG, Rodrigues LC, Fine PEM. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. Int J Epidemiol 1984; **13**: 87–93.
7. Greenland S, Frerichs RR. On measures and models for the effectiveness of vaccines and vaccination programmes. Int J Epidemiol 1988; **17**: 456–63.

https://doi.org/10.1017/S0950268899002642 Published online by Cambridge University Press

8. Halloran ME, Haber M, Longini IM. Interpretation and estimation of vaccine efficacy under heterogeneity. Am J Epidemiol 1992; **136**: 328–43.
9. Anderson RM, May RM. Infectious disease of humans: dynamics and control. Oxford: Oxford University Press, 1991.
10. Anderson RM, Gupta S, Ng W. The significance of sexual partner contact networks for the transmission dynamics of HIV. J Acquir Immune Defic Syndr 1990; **3**: 417–29.
11. Boily MC, Anderson RM. Sexual contact patterns between men and women and the spread of HIV-1 in urban centres in Africa. IMA J Math Appl Med Biol 1991; **8**: 221–47.
12. Haber M, Longini IM, Halloran ME. Measures of the effects of vaccination in a randomly mixing population. Int J Epidemiol 1991; **20**: 300–10.
13. Haber M, Longini IM, Halloran EM. Estimation of vaccine efficacy in outbreaks of acute infectious disease. Statist Med 1991; **10**: 1573–84.
14. Haber M, Halloran E, Longini IM, Watelet L. Estimation of vaccine efficacy in non-randomly mixing populations. Biom J 1995; **37**: 25–38.
15. Halloran ME, Longini IM Jr, Haber MJ, Struchiner CJ, Brunet RC. Exposure efficacy and change in contact rates in evaluating prophylactic HIV vaccines in the field. Statist Med 1994; **13**: 357–77.
16. Halloran EM, Longini IM, Struchiner CJ. Estimability and interpretation of vaccine efficacy using frailty mixing models. Am J Epidemiol 1996; **144**: 83–97.
17. Svenssen A. Analyzing effects of vaccines. Math Biosciences 1991; **107**: 407–12.
18. Boily M-C, Anderson, RM. HIV transmission and the role of other sexually transmitted diseases: measure of association and study design. Sex Transm Dis 1996; **23**: 312–32.
19. Haber M, Watelet L, Halloran E. On individual and population effectiveness of vaccination. Int J Epidemiol 1995; **24**: 1249–60.
20. Halloran ME, Haber M, Longini IM, Struchiner CJ. Direct and indirect effects in vaccine efficacy and effectiveness. Am J Epidemiol 1991; **133**: 323–31.
21. Halloran EM, Struchiner CJ, Longini IM. Study designs for evaluating different Efficacy and effectiveness aspects of vaccines. Am J Epidemiol 1997; **146**: 789–803.
22. Longini IM, Datta S, Halloran EM. Measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. J Acq Imm Defic Synr Hum Retrovirol 1996; **13**: 440–7.
23. Brunet RC, Struchiner CJ, Halloran ME. On the distribution of vaccine protection under heterogeneous response. Math Biosciences 1993; **116**: 111–25.
24. McLean AR, Blower SM. Imperfect vaccines and herd immunity to HIV. Phil Trans R Soc Lond B 1993; **253**: 9–13.
25. Ioannidis J, Cappelleri JC, Schmid CH, Lau J. Impact of epidemic and individual heterogeneity on the population distribution of disease progression rates. Am J Epidemiol 1996; **144**: 1074–85.
26. Stigum H, Magnus P, Bakketeig LS. Effect of changing partnership formation rates on the spread of sexually transmitted diseases and human immunodeficiency virus. Am J Epidemiol 1997; **145**: 644–52.
27. Halloran EM, Watelet L, Struchiner C. Epidemiologic effects of vaccine with complex direct effects in an age-structured population. Math Biosciences 1994; **121**: 193–225.
28. Anderson RM, Garnett GP. Low-efficacy HIV Vaccines: potential for community-based intervention programmes. Lancet 1996; **348**: 1010–3.
29. Kooperberg C, Stone CJ, Truong YK. Hazard regression. J Amer Statist Assoc 1995; **90**: 48–94.
30. Kooperberg C, Clarkson D. Hazard regression with interval-censored data. Biometrics 1997; **53**: 1485–94.
31. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research – principles and quantitative methods. New York: Van Nostrand Reinhold, 1992.
32. Collett D. Modelling survival data in medical research. London: Chapman and Hall, 1992.
33. Lawless JF. Statistical models and methods for lifetime data. New-York: Wiley, 1982.
34. StatLib. http://lib.stat.cmu, 1998.
35. Anderson RM, May RM, McLean AR. Possible demographic consequence of AIDS in developing countries. Nature 1988; **332**: 228–34.
36. Garnett GP, Hughes JP, Anderson RM, et al. Sexual mixing patterns of patients attending sexually transmitted diseases clinics. Sex Transm Dis 1996; **23**: 248–57.
37. Rida WN, Lawrence DN. Some statistical issues in HIV vaccine trials. Statist Med 1994; **13**: 2155–77.
38. Moore J, Anderson RM. The WHO and why of HIV vaccine trials. Nature 1994; **372**: 313–4.
39. Lang JM, Rothman KJ, Cann CI. That confounded P-Value. Epidemiol 1997; **9**: 7–8.
40. Boily M-C, Desai KN, Mâsse BR, Alary M, Anderson RM. Some important issues the design of Phase III HIV vaccine efficacy trials. Vaccine 1999; **17**: 989–1004.
41. Alonso PL, Smith T, Schellenberg JR, et al. Randomised trial of efficacy of SPf66 vaccine against *Plasmodium falciparum* malaria in children in southern Tanzania. Lancet 1994; **344**: 1175–81.
42. Press WH, Teukolsky SA, Vetterling WT, Flannery BF. Numerical recipes in fortran. New York: Cambridge University Press, 1986.
43. U.S. Census Bureau. HIV/AIDS Surveillance Data Base. http://www.census.gov:80/ftp/pub/ipc/www/hivaidsn.html, 1998.