

## Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal

Peter B. Gilbert<sup>1,\*</sup>,†, Ian W. McKeague<sup>2</sup>, Geoffrey Eisen<sup>3</sup>, Christopher Mullins<sup>3</sup>,  
Aissatou Guéye-NDiaye<sup>4</sup>, Souleymane Mboup<sup>4</sup> and Phyllis J. Kanki<sup>3</sup>

<sup>1</sup>*Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center and Department of Biostatistics, University of Washington, Seattle, WA 98105, U.S.A.*

<sup>2</sup>*Department of Statistics, Florida State University, Tallahassee, FL 32306, U.S.A.*

<sup>3</sup>*Harvard AIDS Institute and Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, U.S.A.*

<sup>4</sup>*Laboratoire de Bactériologie-Virologie, Université Cheikh Anta Diop, Dakar, Senegal*

### SUMMARY

From a prospective cohort study of 1948 initially human immunodeficiency virus (HIV) uninfected female commercial sex workers followed between 1985 and 1999 in Dakar, Senegal, the authors compared the male to female per infectious sexual exposure transmission probability of HIV types one (HIV-1) and two (HIV-2). New non-parametric competing risks failure time methods were used, which minimized modelling assumptions and controlled for risk factors for HIV infection. The HIV-1 versus HIV-2 infectivity ratio over time was estimated by the ratio of smoothed non-parametric kernel estimates of the HIV-1 and HIV-2 infection hazard functions in sex workers, adjusted by an estimate of the relative HIV-1 versus HIV-2 prevalence in the partner population. HIV-1 was found to be significantly more infectious than HIV-2 throughout the follow-up period ( $P < 0.001$ ). The HIV-1/HIV-2 infectivity ratio was inferred to be approximately constant over time, with estimated common value 3.55. The finding of greater HIV-1 infectivity persisted in sensitivity analyses and in covariate-adjusted analyses, with adjusted infectivity ratio estimates ranging between 3.40 and 3.86. Understanding the mechanisms by which HIV-1 infects more efficiently than HIV-2 may be useful in the development of HIV-1 vaccines. Additionally, the methodology developed here may be useful for analysing other data sets. Copyright © 2003 John Wiley & Sons, Ltd.

**KEY WORDS:** cause specific hazard rates; competing risks; HIV transmission; infectious diseases; sexually transmitted diseases; survival analysis

\* Correspondence to: Peter Gilbert, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, MW-500, P.O. Box 19024, Seattle, WA 98109, U.S.A.

† E-mail: pgilbert@ssharp.org

Contract/grant sponsor: National Institutes of Health; contract/grant numbers: AI24643, AI46703-01.

Contract/grant sponsor: U.S. Army Medical Research; contract/grant numbers: C-7072, 17-C-0138, 17-95-5005, PO1 A1 30795.

Contract/grant sponsor: National Science Foundation; contract/grant number: DMS-9971784.

## 1. INTRODUCTION

The human immunodeficiency virus (HIV) can be classified as types one (HIV-1) and two (HIV-2). HIV-1 accounts for the vast majority of HIV in the world, with HIV-2 present mainly in Western Africa. Laboratory studies and cohort studies have shown that the HIV types likely differ in various *in vivo* and *in vitro* phenotypic properties including replicative capacity [1–3], cytopathicity [4, 5], pathogenicity [6–8], perinatal infectivity [9–11] and heterosexual infectivity [12]. Here, we are interested in evaluating differential male to female infectivity, where infectivity is defined as the per sexual contact probability of transmission from an HIV infected male to an HIV uninfected female. To address this, we considered long-term follow-up data from a cohort of female commercial sex workers in Dakar, Senegal, where both HIV-1 and HIV-2 have been circulating since at least the mid-eighties.

Donnelly *et al.* [12] evaluated the question of differential male to female infectivity by analysing the Dakar cohort of 780 initially HIV seronegative sex workers followed between February 1985 and December 1989. Their conclusion, at about the one per cent significance level, was that HIV-1 infectivity was greater than HIV-2 infectivity. We re-evaluated this question by analysing the most recent data set of 1948 initially HIV uninfected sex workers followed through November 1999. The time is ripe for a reanalysis, since the number of evaluable participants with an HIV infection event has matured to 196, compared to 29 for the earlier analysis. We applied new non-parametric competing risks failure time methods, which rely on fewer modelling assumptions than the methods used by Donnelly *et al.* [12]. Unadjusted and covariate-adjusted analyses provide evidence that HIV-1 is more infectious than HIV-2 ( $P < 0.001$  in each analysis), with HIV-1 estimated to be 3.55-fold more infectious than HIV-2 in the unadjusted analysis, and 3.40–3.86-fold more infectious in the covariate-adjusted analyses.

## 2. DATA

In 1970, the Senegalese government established a public health programme whereby self-identified female sex workers were required to register and regularly attend a health clinic, which provides regular medical evaluation and free treatment for sexually transmitted diseases. In 1985, the Inter-University Convention for the Prevention of AIDS began a prospective natural history study that involved regular HIV testing from consenting sex workers [13, 14]. For the present study, the population consisted of registered sex workers in Dakar who agreed to participate and were initially HIV seronegative. Sex workers were followed for varying time intervals between 7 February 1985 and 1 November 1999, with clinic visits scheduled every six months. At each clinic visit, women were tested for HIV-1 positivity and for HIV-2 positivity, using immunoblot antibody assays, HIV specific peptides, and HIV specific PCR [14, 15]. Seroconversions were confirmed using all available samples from individuals [14]. The time of seroconversion was estimated as the midpoint between the last seronegative visit date and the first seropositive visit date. Data on potential risk factors were also collected.

HIV-1 and HIV-2 serostatus data at each clinic visit were available from all sex workers. Information on nationality, age, date of cohort entry, and years of registered prostitution were available from greater than 99 per cent of the sex workers, and information on the average number of sexual partners per week was available from 52.6 per cent of the cohort. In total,

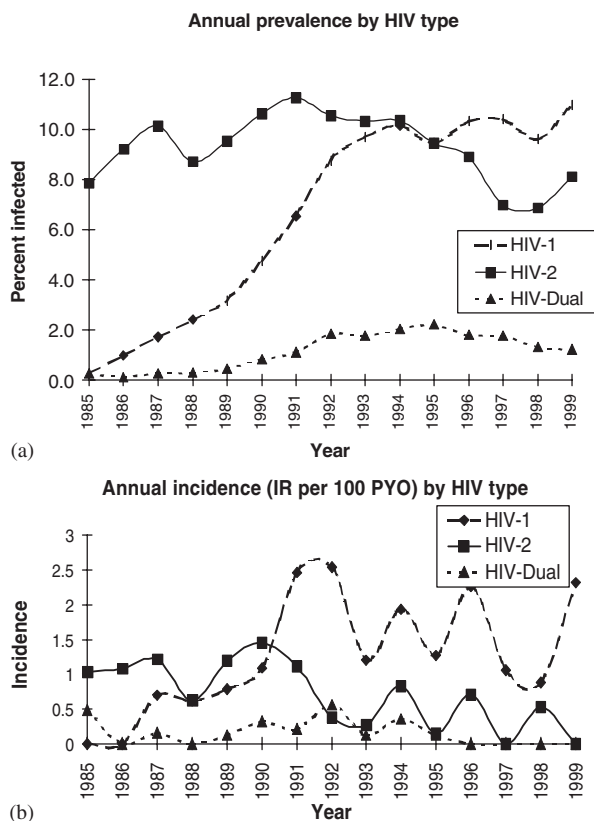


Figure 1. (a) Annual prevalence and (b) annual incidence of HIV-1 and HIV-2 infection among sex workers participating in the Dakar, Senegal, prospective cohort study. Prevalence rates were calculated from all registered sex workers visiting the clinic in the index year regardless of HIV serostatus at cohort entry (data from 3141 sex workers used in the calculations). Incidence rates were calculated from the cohort of registered sex workers HIV seronegative at the beginning of the index year (data from 1951 sex workers used in the calculations). Incident dual infection events included transitions from HIV uninfected to dual infected and from infection with one HIV type to dual infected.

199 of the 1951 initially HIV uninfected sex workers became HIV infected during the 15 year observation period, 127 with HIV-1 only, 66 with HIV-2 only, and six with both types. Among dual-infected women, three had the first seropositive test reactive for both viruses; these women were assumed to have simultaneous HIV-1 and HIV-2 seroconversion dates. For reasons given in the Methodology section, we removed these three subjects from the analysis; thus the analysed cohort consisted of 1948 sex workers, of whom 196 became HIV infected.

For calculating annual point prevalences of the viruses in sex workers, data were used from all sex workers registered during the year under consideration. A total of 3141 women contributed data to the prevalence calculations, including the 1951 initially uninfected women plus 1190 sex workers who entered the cohort with HIV infection. The data demonstrated a relative plateau of HIV-2 prevalence, with HIV-1 prevalence surpassing that of the more endemic virus, HIV-2, by the end of the observation period (Figure 1(a)). Incidence data

showed a steady increase in HIV-1 incidence over time, with HIV-2 incidence remaining fairly stable and then gradually decreasing after 1994 (Figure 1(b)). These data indicate that the epidemic curves for these two related viruses differ. One explanation would be greater infectivity of HIV-1 compared to HIV-2, which we evaluate here.

### 3. METHODOLOGY

Various authors have estimated the infectivity probability of HIV-1, by modelling infection risk as a function of the number of sexual contacts within sexual partnerships. Commonly the models have been formulated for studies of monogamous individuals with HIV-1 infected partners [16–18], or for studies of non-monogamous individuals with multiple partners of known HIV-1 prevalence [16, 19, 20]. Hu *et al.* [21] reviewed methodologies and challenges for comparing the infectivity of HIV variants.

Donnelly *et al.* [12] modelled the infectivity probabilities  $r_1$  for HIV-1 and  $r_2$  for HIV-2 as functions of the reported number of sexual contacts and the partner prevalences  $p_1$  of HIV-1,  $p_2$  of HIV-2, and  $p_{12}$  of dual HIV-1 and HIV-2 infection. The partner prevalence rates were assumed known and constant over time. Probabilities of becoming infected with either type, both, or neither were expressed in terms of the estimated number of sexual contacts and the parameters  $r_1, r_2, p_1, p_2, p_{12}$ . The resulting parametric likelihood was maximized under an independent competing risks assumption using standard methods to obtain point estimates and variance estimates of  $r_1$  and  $r_2$  [22]. Then, inference about differential infectivity was made by testing  $r_1 = r_2$  with a Wald statistic.

The analysis was conducted under six sets of assumptions, for two specifications of partner prevalences crossed with three ways of imputing values for the average number of sexual contacts per week for the 355 (45.5 per cent of sample) sex workers with a missing value. The infectivity ratio estimate  $\hat{r}_1/\hat{r}_2$  ranged between 5.8 and 8.9 for the six analyses, and the two-sided  $p$ -values for testing  $r_1 = r_2$  ranged between 0.0064 and 0.013.

Rather than estimating  $r_1$  and  $r_2$  separately and then comparing the estimates to evaluate differential infectivity, our approach estimated the ratio  $r_1/r_2$  directly and assessed if it significantly differed from one. Targeting inference on the infectivity ratio conceptually addresses the differential infectivity question more directly. In addition, estimating the infectivity ratio can be done with greater accuracy than estimating the type-specific infectivities separately, since the inference procedure does not rely on having accurate measurements on the number of sexual contacts (as described below; see equations (1) and (2)). There are other advantages to the new approach. First, it does not assume that the HIV-1 and HIV-2 partner prevalences are known; rather, estimates were used and their uncertainty was partially accounted for. Secondly, the HIV-1 and HIV-2 partner prevalences were allowed to vary with calendar time. This is important because the HIV-1 prevalence in sex workers varied substantially between 1985 and 1989 (Figure 1(a)), and the sex worker prevalence is closely related to the partner prevalence. Thirdly, it allows the HIV-1 and HIV-2 infectivities  $r_1(t)$  and  $r_2(t)$  to vary with time rather than assuming they are fixed constants. The infectivity ratio could vary over time, for example, if HIV-1 and HIV-2 underwent different evolutionary pathways towards more or less infectious phenotypes. Fourthly, it adjusts for several risk factor covariates. Fifthly, the testing procedure used to assess differential infectivity makes no modelling assumptions about the hazard rates of HIV-1 and HIV-2 infection over time, and requires no assumptions about

the nature of dependence between the competing risks. Not requiring independent competing risks is important because the risks likely were dependent (for example, because sex workers at high risk for infection with one type of HIV may also have been at high risk for infection with the other type). Limitations of the methods are discussed in the Discussion.

### 3.1. Non-parametric competing risks failure time methods

We viewed the two virus types as competing risks of infection, a framework also used (somewhat differently) by reference [23] for comparing transmissibility of HIV-1 subtypes in Thailand, and applied non-parametric statistical methods to compare the HIV-1 and HIV-2 infection hazard rates. The outcome measures on each subject were the time and type of the first HIV infection. Since a first HIV infection may modify the risk of a second HIV infection (for example, analyses of the Dakar cohort have suggested that infection with HIV-2 partially protects against subsequent superinfection with HIV-1 [15, 24]), no data on sex workers were considered beyond first infection events. Thus, HIV-1 infections censored HIV-2 infections, and vice versa. Since very few superinfection events occurred (a total of three), ignoring these events did not appreciably affect the statistical power of the analysis.

The competing risks approach does not allow for the possibility of simultaneous co-infection with competing virus types. To accommodate this, we removed the three subjects from the analysis who were simultaneously co-infected by the definition of seroconversion time we used. Since only three subjects had this endpoint, it is unlikely that an alternative analysis that retained these subjects (for example, an analysis that considers simultaneous type 1 and 2 co-infection as a third competing risk of infection) would affect the results appreciably.

The time to infection was measured as the time from entry into the cohort until seroconversion. All analyses were based on this time scale, 'study time', although the calendar time scale was also used for adjusting HIV-1 and HIV-2 hazard estimates by HIV-1 and HIV-2 partner prevalences, as described below. Sex workers who were never observed to be infected were censored with censoring time equal to the time interval of follow-up. An alternative analysis based on calendar time would accommodate the possibility that the infectivities vary more with calendar time than with study time. We chose the study time scale because it allows the use of relatively simple survival analysis methods (a calendar time scale would require that the methods account for the left truncation of survival times resulting from staggered entry), and because our approach provides a way to adjust for the effects of calendar time.

We defined  $r_i(t)$ ,  $i = 1, 2$ , as the average type  $i$  infectivity among the population of all sex workers  $t$  years into follow-up. Our goal was to estimate  $r_1(t)/r_2(t)$  non-parametrically for  $t$  ranging over the follow-up period 0 to 14.73 years. To this end, let  $\lambda_i(t)$ ,  $i = 1, 2$ , represent the hazard of type  $i$  infection for a sex worker at study time  $t$ . Each hazard function has 'crude' interpretation as the instantaneous type-specific infection risk in the presence of both circulating viruses [25]. Consider calendar times ranging between the opening and closing of the study, 7 February 1985 to 1 November 1999. We defined calendar time  $t_c$  as the number of years since 7 February 1985. At time  $t$ , the weekly risk of HIV-1 infection,  $\lambda_1(t)$ , equals the product of the type 1 infectivity probability at time  $t$ ,  $r_1(t)$ , times the number of sexual contacts during the week with a client infected with either virus type,  $c(t)$ , times the proportion  $\pi_1$  of these infected clients who are infected with HIV-1 rather than HIV-2. A similar formula holds for the weekly risk of HIV-2 infection. This key relationship

implies that

$$\frac{\lambda_1(t)}{\lambda_2(t)} = \frac{r_1(t) \times c(t) \times \pi_1(t_{c_j} + t)}{r_2(t) \times c(t) \times \pi_2(t_{c_j} + t)} \quad (1)$$

where  $t_{c_j}$  denotes the calendar time at which the  $j$ th sex worker entered the study, so that  $\pi_1(t_{c_j} + t)$  is the proportion of infected clients who were HIV-1 seroprevalent at calendar time  $t_{c_j} + t$ . We refer to  $\pi_1(\cdot)$  as the *partner relative prevalence* of HIV-1 versus HIV-2. Since the contact rate  $c(t)$  cancels in the numerator and denominator of (1), the ratio of type-specific infection hazards  $\lambda_1(t)/\lambda_2(t)$  depends only on the infectivity ratio  $r_1(t)/r_2(t)$  and on the partner relative prevalence ratio  $\pi_1(t_{c_j} + t)/\pi_2(t_{c_j} + t)$ . Consequently, in our approach it is not necessary to estimate contact rates of sex workers, an important advantage given the difficulty of this task. For all analyses, we assumed that for any given calendar time  $t_c$ , all sex workers at risk for HIV infection at that time had the same partner relative prevalence. For initial analyses that did not incorporate covariates, we also assumed that the product of the type-specific infectivity and the contact rate,  $r_i(t)c(t)$ , was common for all sex workers,  $i = 1, 2$ , and based estimation of  $r_1(t)/r_2(t)$  on the formula resulting from expression (1):

$$\frac{r_1(t)}{r_2(t)} = \frac{\lambda_1(t)/\pi_1(t_{c_j} + t)}{\lambda_2(t)/\pi_2(t_{c_j} + t)} \quad (2)$$

For analyses that included covariates, a common infectivity ratio  $r_1(t|z)/r_2(t|z)$  was assumed for all sex workers with covariate vector  $z$ , and was estimated using model (3), as described later in this section.

### 3.2. Estimation and confidence intervals

To estimate  $r_1(t)/r_2(t)$  via equation (2), we estimated each  $\lambda_i(t)$  with adjustment at each event time for each sex worker by an estimate of the reciprocal of the HIV- $i$  partner relative prevalence. Specifically, we estimated an adjusted version of  $\lambda_i(t)$ ,  $\lambda_{i\text{adj}}(t)$ , by smoothing the non-parametric Nelson–Aalen estimate  $\hat{\Lambda}_i(t)$  of the cumulative HIV- $i$  hazard function  $\Lambda_i(t)$  [26, 27], with the integrand corresponding to the HIV- $i$  infection counting process for the  $j$ th sex worker divided by an estimate of  $\pi_i(t_{c_j} + t)$  (see Appendix). The adjustment effectively prorates the amount of person-years exposed to HIV- $i$  so that the estimated function  $\lambda_{i\text{adj}}(t)$  is proportional to  $r_i(t)$ . The adjusted hazard  $\lambda_{i\text{adj}}(t)$  has interpretation as the HIV- $i$  infection rate in women had all exposures with infected clients been to type  $i$  HIV. The ratio of the ‘relative prevalence adjusted’ estimates  $\hat{\lambda}_{1\text{adj}}(t)$  and  $\hat{\lambda}_{2\text{adj}}(t)$  was then used to estimate  $r_1(t)/r_2(t)$ .

To estimate  $\pi_1(t_c)$  over the range of calendar times of the study, we first estimated the annual partner relative prevalence for each year 1985 to 1999 by the ratio of the observed HIV-1 and HIV (HIV-1 and HIV-2 combined) prevalences in all female sex workers registered during the given year. Dual infected women were counted as infected in both the numerator and denominator prevalences. These calculations used data on sex workers HIV seronegative and seropositive at entry, totalling 3141 women. The annual relative prevalence estimates were then smoothed (using lowess in S-plus) to obtain estimates of  $\pi_1(t_c)$  for all intermediate calendar times. For the  $j$ th sex worker under observation at study time  $t$ , her partner relative prevalence of HIV-1 at that time was simply estimated by  $\hat{\pi}_1(t_{c_j} + t)$  (and her partner relative prevalence of HIV-2 was estimated by  $\hat{\pi}_2(t_{c_j} + t) = 1 - \hat{\pi}_1(t_{c_j} + t)$ ).

A 95 per cent confidence interval (CI) for  $r_1(t)/r_2(t)$  was calculated using the delta method applied to the asymptotic variances of  $\hat{\lambda}_{1\text{adj}}(t)$  and of  $\hat{\lambda}_{2\text{adj}}(t)$ . To adjust the CI for the uncertainty in the estimated partner-relative prevalence ratio, the CI was recalculated 100 times using 100 randomly sampled partner relative prevalence estimates from its asymptotic normal distribution (details provided in the Appendix). ‘Unadjusted’ and ‘adjusted’ CIs were calculated, where unadjusted CIs assume the partner-relative prevalence ratio is known and equal to the estimate, and the limits of the adjusted CIs span the 2.5th and 97.5th percentiles of the 100 recalculated confidence limits. Note that the adjusted confidence limits do not take into account uncertainty due to the assumption that the relative prevalence among clients is the same as among sex workers. Thus, though the adjusted CIs are wider than the unadjusted CIs, they are likely still too narrow.

Since the non-parametric analysis suggested a common infectivity ratio  $r_1/r_2 = r_1(t)/r_2(t)$  over time, as shown in the Results, we estimated the common ratio. The common parameter  $r_1/r_2$  was estimated by the average of the estimates of  $r_1(t_k)/r_2(t_k)$  over a discrete grid of 132 evenly spaced study times  $t_k$  spanning the follow-up period, with inverse-variance weighting.

### 3.3. Estimation and confidence intervals that adjust for covariates

If the sex workers had different risk factors for HIV-1 infection than for HIV-2 infection, in either direction or magnitude, then the estimator for the infectivity ratio considered above could be biased. For example, bias could result if sex workers who became infected with HIV-1 tended to have more sexual contacts with clients than women who became infected with HIV-2. To adjust the estimate of the infectivity ratio for covariates, we used the following proportional hazards model:

$$\frac{\lambda_1(t|z)}{\lambda_2(t|z)} = \exp[\beta_0 + \beta^T z] \quad (3)$$

This version of the Cox model was studied by Lunn and McNeil [30], who showed that it can be fit using existing software programs for the ordinary Cox model without competing risks (for example, with the function `coxph` in S-plus). Since no covariates were measured on the clients of the sex workers, we were unable to adjust the infectivity ratio estimates for client covariates, and accordingly we assumed that at each calendar time the partner relative prevalence  $\pi_1(t_c)$  was common among sex workers (as stated earlier).

At each event time  $t$ , the term  $\exp[\beta_0 + \beta^T z]$  in model (3) was multiplied by an estimate of the relative prevalence ratio  $\pi_1(t_{c_j} + t)/\pi_2(t_{c_j} + t)$ , where  $j$  indicates the sex worker who had the event. This can be accomplished, for example, using the offset feature in the S-plus function `coxph`. With this adjustment, under the assumption that the infectivity ratio  $r_1/r_2$  was constant over time,  $\exp[\hat{\beta}_0]$  from model (3) estimates  $r_1/r_2$  adjusted for the covariates  $z$  at specified level  $z=0$ . Continuous covariates were centred by their mean values so that the adjustment conditions on central covariate values. The profile likelihood of model (3) was used for producing a likelihood ratio test of the null hypothesis that the adjusted  $r_1/r_2$  equalled one, and to construct a confidence interval for  $r_1/r_2$ . The validity of the proportional hazards assumption in model (3) was assessed by the method of Grambsch and Therneau [28].

### 3.4. Tests for differential infectivity

For testing for different infectivities of HIV-1 and HIV-2, we applied a test that was developed for comparing two cause-specific hazard functions with adjustment for covariate effects [29]. This test evaluates the null hypothesis  $\lambda_1(t|z_0(t)) = \lambda_2(t|z_0(t))$  for all  $t$ , for a specified covariate level  $z_0(t)$ . The test statistic is based on a weighted average of differences between non-parametric estimates of  $\lambda_1(t|z_0(t))$  and  $\lambda_2(t|z_0(t))$  under the cause specific Cox proportional hazards model over time, with weights chosen to control instability in the beginning and end of the observation period. By considering the logarithm of the estimated partner relative prevalence ratio  $\hat{\pi}_1(\cdot)/\hat{\pi}_2(\cdot)$  as a time-dependent covariate with value fixed at zero, we applied this procedure to test  $r_1(t) = r_2(t)$  for all  $t$  versus  $r_1(t) \geq r_2(t)$  with  $r_1(t) > r_2(t)$  for some  $t$ . To account for the uncertainty in the estimated partner relative prevalence ratio, the test was carried out 100 times for 100 randomly sampled partner relative prevalence ratios over time (as described in the Appendix).

### 3.5. Tests for differential infectivity that adjust for covariates

Cause specific Cox proportional hazards models were used to identify risk factors for HIV-1 infection and for HIV-2 infection, and the re-coded Cox model (3) was used to identify covariates that predicted a differential HIV-1 versus HIV-2 infection risk [30]. These analyses identified risk factors that were important to adjust for in assessing differential infectivity. The procedure of reference [29] was used for testing the null hypothesis of equal infectivities under adjustment for a covariate vector  $z$ , that is, of  $H_0: r_1(t|z) = r_2(t|z)$  for all  $t$ .

Nearly half of the sex workers in the study had no data on the covariate average number of reported sexual contacts per week. To evaluate the robustness of the results to missing values of this variable, we carried out the covariate-adjusted analyses described above using two simple techniques for handling this variable. These are: (i) a complete-case analysis; (ii) an analysis of all sex workers imputing the missing sexual contacts values to be equal to the mean.

### 3.6. Sensitivity analysis

A main assumption of the approach was that the relative prevalence over time in sex workers represented the relative prevalence over time in their partners. To evaluate sensitivity of the results to misrepresentation, we re-did the unadjusted and covariate-adjusted analyses with the estimated HIV-1/HIV-2 relative prevalence ratio  $\pi_1(\cdot)/\pi_2(\cdot)$  increased at all times by a multiplicative constant  $K > 1$ , which made it more difficult to infer greater HIV-1 infectivity. The analysis was repeated for several values of  $K > 1$ .

## 4. RESULTS

### 4.1. Description of the cohort

The analysed cohort included 1948 sex workers, of whom 196 had a first HIV infection event, 128 with HIV-1 and 68 with HIV-2. Figure 2 shows the HIV seronegative observation time of the sex workers who seroconverted while under follow-up. Sex workers infected with HIV-1 had a median entry date of 13 October 1988 and a median infection time of 3.30 years, while sex workers infected with HIV-2 had an earlier median entry date 30 September 1986



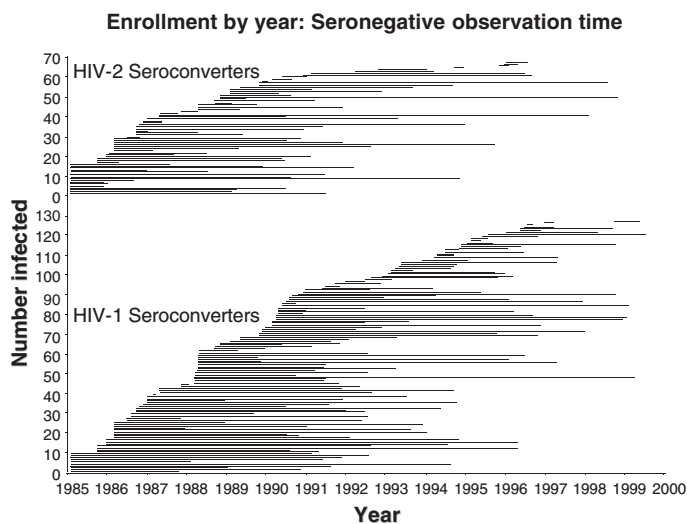


Figure 2. Timing of HIV-1 and HIV-2 infection (observation time and calendar time) for the 196 initially HIV seronegative Dakar sex worker cohort participants who became infected while enrolled, between 7 February 1985 and 1 November 1999.

and a shorter median infection time of 2.58 years. Among all enrollees, the median period of follow-up was 3.33 years, with an interquartile range of 1.33–7.42 years; 1221 of 1488 (62.7 per cent) of participants were never observed to seroconvert and were lost to follow-up prior to 1 January 1999. Participants completed a consultation form at a clinic visit at a median rate of once every 8.3 months, with an interquartile range of once every 5.2–11.9 months.

Table I summarizes participant characteristics and describes covariate specific HIV incidence rates. Briefly, most women were Senegalese (1417, 73.2 per cent), the average age at cohort entry was 30.4 years (range 19–56 years), 63.9 per cent of subjects entered before 1 January 1991, the average years of registered prostitution at entry was 2.7 years (range 0 to 26 years), and the average number of partners per week was 6.8 (interquartile range of 3.5 to 8.0 partners per week).

Figure 3 shows estimates and 95 per cent CIs of cumulative incidence and hazard rates of HIV-1 and HIV-2 infection. The gap between HIV-1 and HIV-2 cumulative incidence steadily widened during the follow-up period. The estimated HIV-1 hazard  $\hat{\lambda}_1(t)$  was considerably greater than the estimated HIV-2 hazard  $\hat{\lambda}_2(t)$  throughout the follow-up period. The HIV-2 hazard increased slightly during the first five years of follow-up and then gradually declined, while the HIV-1 hazard rose steeply for five years and then rapidly declined. This illustrates a peak in the HIV incidence among sex workers five years into follow-up followed by steady curtailment.

#### 4.2. Non-parametric estimation of $r_1(t)/r_2(t)$

Figure 4 shows estimates over time of  $r_1(t)/r_2(t)$ , using the kernel smoothing procedure described in Section 3.2. Since the adjusted lower 95 per cent confidence limit of  $r_1(t)/r_2(t)$

Table 1. Description of characteristics and incidence rates of initially HIV seronegative commercial sex workers participating in the Dakar cohort study.

Characteristic	Number	Per cent	Range	Mean (SD*)	PYrs	HIV-1 inc	HIV-2 inc	HIV inc†
Nationality:								
Senegalese	1,417	73.2			7,911	1.26	0.63	1.90
Ghanaian	256	13.2			727	2.34	1.51	3.85
Other	262	13.5			703	1.56	1.00	2.56
Age at cohort entry			19–56	30.45 (6.49)				
15–24	502	26.1			1,628	1.54	0.55	2.09
25–34	1,027	53.3			5,474	1.06	0.60	1.66
35–44	345	17.9			1,939	2.01	0.98	2.99
≥ 45	52	2.7			260	1.92	2.69	4.62
Date of cohort entry‡			2/85–7/99	4/90 (4.16)				
1985–1987	686	35.2			4,669	0.96	0.92	1.88
1988–1990	560	28.7			2,885	1.63	0.62	2.25
1991–1993	247	12.7			983	1.63	0.31	1.93
1994–1996	277	14.2			668	2.69	0.60	3.29
1997–1999	178	9.1			173	1.16	0.00	1.16
Years of follow-up			0.00–14.73	4.73 (4.02)				
0–2.99	893	45.8			1,284	5.61	3.35	8.96
3–5.99	413	21.2			1,821	1.70	0.99	2.69
6–8.99	285	14.6			2,124	1.13	0.19	1.32
≥ 9	357	18.3			4,150	0.02	0.07	0.10
Years of registered prostitution			0–26	2.70 (4.20)				
0	1,232	63.3			3,531	1.59	0.57	2.15
1–4	314	16.1			2,717	1.29	0.99	2.28
5–9	247	12.7			1,842	0.87	0.87	1.74
10–14	124	6.4			1,032	1.55	0.29	1.84
≥ 15	28	1.4			254	1.97	0.79	2.75
Average number of sexual partners per week§			0–56	6.77 (6.14)				
0–2.99	266	26.0			1,647	0.49	0.18	0.67
3–5.99	368	35.9			2,590	0.58	0.19	0.77
6–9.99	216	21.1			1,435	0.70	0.21	0.91
10–19.99	125	12.2			752	1.33	0.27	1.59
≥ 20	49	4.8			170	1.18	1.18	2.36

\* SD, standard deviation.

† Overall HIV incidence, irrespective of type.

‡ The covariate was calculated as the years since 7 February 1985, the date the first participant enrolled in the cohort study.

§ Averaged over repeat measurements. The reduced sample size is due to missing data.

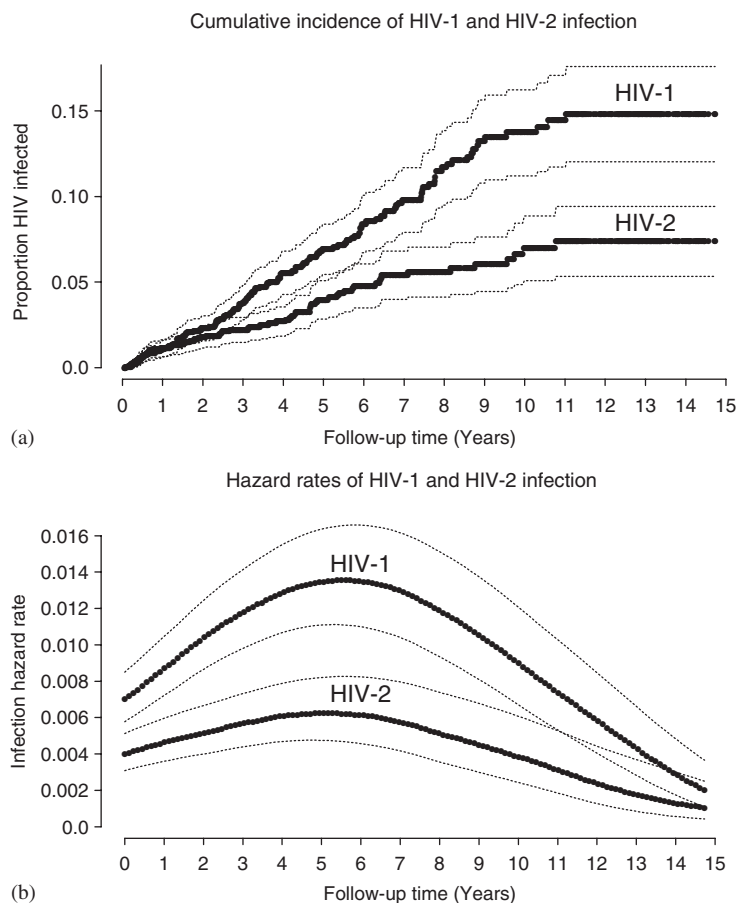


Figure 3. For the initially HIV seronegative Dakar sex worker cohort ( $n = 1948$ ) followed between 1985 and 1999: (a) estimated HIV-1 and HIV-2 cumulative incidence curves with 95 per cent confidence limits; (b) non-parametric Epanechnikov kernel estimates of the HIV-1 and HIV-2 hazard rates  $\lambda_1(t)$  and  $\lambda_2(t)$  from smoothed estimates of the cumulative type specific hazard functions [26, 27]; normal approximation 95 per cent confidence limits calculated by transforming the symmetric confidence limits [27].

exceeded one throughout the follow-up period, we infer that HIV-1 infectivity was significantly greater than HIV-2 infectivity. Because the infectivity ratio appeared constant over time, as supported by the statistical test described in reference [28] ( $P > 0.50$ ), we estimated the common value  $r_1/r_2$  over time. The inverse-variance weighted estimate was 3.55. At the selected time points by which 25, 50 and 75 per cent of the cumulative 196 infection events occurred, 1.37 years, 3.20 years and 5.74 years, respectively, the 95 per cent adjusted confidence intervals about the infectivity ratio were (2.37, 4.47), (2.34, 4.80) and (2.32, 5.65), respectively.

The testing procedure [29] confirmed the higher infectivity of HIV-1; the unadjusted test gave  $P < 0.001$ , and 84 of the 100 re-sampled test statistics gave  $P < 0.01$ .

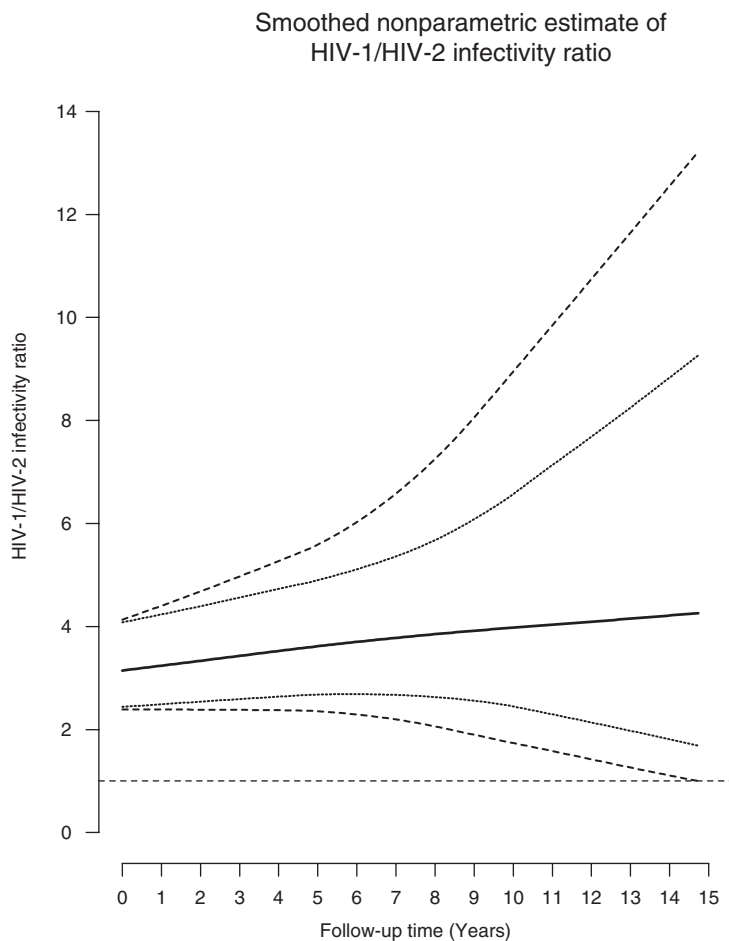


Figure 4. Smooth non-parametric estimates of  $r_1(t)/r_2(t)$ , with unadjusted 95 per cent confidence limits (dotted lines) and adjusted confidence limits (dashed lines). The confidence limits for  $r_1(t)/r_2(t)$  were calculated using the asymptotic variances of  $\hat{\lambda}_{1\text{adj}}(t)$  and  $\hat{\lambda}_{2\text{adj}}(t)$  and the delta method. The method of computing the adjusted confidence limits is given in the Appendix.

#### 4.3. Covariate-adjusted analyses

To account for covariate effects, we first evaluated each variable listed in Table I as a potential risk factor for HIV-1 or HIV-2 infection, and as a potential predictor of differential risk of HIV-1 versus HIV-2 infection. We then used the re-coded Cox model (3) to make inference on the infectivity ratio with adjustment for the covariates that were found to be significantly associated with HIV-1, HIV-2 or HIV-1/HIV-2 infection. In regression analyses the nationality covariate was given three categories, Senegalese, Ghanaian, and other, with the most prevalent category (Senegalese) used as the reference category.

At the 0.05 significance level, univariable Cox models identified four risk factors for HIV-1 infection: nationality (Ghanaians at increased risk); later date of cohort entry; older age at

Table II. Univariable Cox models of significant risk factors for HIV-1 infection and for HIV-2 infection of initially HIV seronegative commercial sex workers participating in the Dakar cohort study.

HIV type	Variable	Hazard ratio*	95 per cent CI hazard ratio	<i>p</i> -value	<i>p</i> -value (PH) <sup>†</sup>
HIV-1	Nationality:			0.044	0.19
	Senegalese	1.00	—	—	—
	Ghanaian	1.90	(1.13,3.20)	0.015	0.072
	Other	1.28	(0.68,2.40)	0.44	0.98
	10 years later date of cohort entry <sup>‡</sup>	2.92	(1.71,4.97)	<0.001	0.21
	10 years older at cohort entry	1.30	(1.00,1.69)	0.054	0.028
	10 more sexual partners per week	1.60	(1.08,2.35)	0.018	0.83
HIV-2	Nationality:			0.054	0.98
	Senegalese	1.00	—	—	—
	Ghanaian	2.18	(1.13,4.23)	0.021	0.92
	Other	1.43	(0.64,3.18)	0.38	0.88
	10 years later date of cohort entry <sup>‡</sup>	0.28	(0.12,0.68)	0.0043	0.30
	10 years older at cohort entry	1.71	(1.22,2.40)	0.0017	0.17
	10 more sexual partners per week	2.02	(1.24,3.29)	0.0042	0.46

\* Estimated hazard ratio of HIV-1 infection or of HIV-2 infection from the univariable Cox model.

<sup>†</sup> *P*-value for test of proportional hazards (PH), based on the method of Grambsch and Therneau [28].

<sup>‡</sup> The covariate was calculated as the years since 7 February 1985, the date the first participant enrolled in the cohort study.

entry, and more sexual partners per week (Table II). The same variables were significant univariable risk factors for HIV-2 infection, with hazard ratios of comparable magnitude to those for HIV-1, with exception that an earlier date of cohort entry predicted an increased HIV-2 infection risk. For HIV-1 and HIV-2, Table III shows the best-fitting multivariable Cox models for predicting infection risk. Since nearly half the sample missed the sexual partners covariate, this covariate was made ineligible for entering the model. The model for HIV-1 contained the variables nationality and date of cohort entry, with hazard ratio estimates and confidence intervals similar to those in the univariable models. The model for HIV-2 contained these two variables plus the age at entry, also with estimates and confidence intervals similar to those for the univariable models.

Using model (3), the only covariate that predicted a significantly differential HIV-1 versus HIV-2 infection risk was the date of cohort entry. The HIV-1/HIV-2 ratio of hazards  $\lambda_1(t)/\lambda_2(t)$  was increased an estimated 10.38-fold (95 per cent CI (3.68,29.27),  $P < 0.001$ ) for each ten years later date of entry into the cohort.

To study the infectivity ratio, we next fit model (3) (with the estimated partner relative prevalence entered as a time-dependent offset term), using the risk factor covariates nationality, date of cohort entry, and age at entry. The date of cohort entry and the age at entry

Table III. Multivariable Cox models of significant risk factors for HIV-1 infection and for HIV-2 infection of initially HIV seronegative commercial sex workers participating in the Dakar cohort study.

HIV type	Variable	Hazard ratio*	95 per cent CI hazard ratio	<i>p</i> -value	<i>p</i> -value (PH) <sup>†</sup>
HIV-1	Nationality:			0.027	0.12
	Senegalese	1.00	—	—	—
	Ghanaian	1.97	(1.17,3.32)	0.011	0.056
	Other	1.00	(0.53,1.90)	0.99	0.65
	10 years later date of cohort entry <sup>‡</sup>	3.08	(1.79,5.32)	<0.001	0.13
HIV-2	Nationality:			0.072	0.55
	Senegalese	1.00	—	—	—
	Ghanaian	2.07	(1.07,4.01)	0.032	0.95
	Other	1.83	(0.81,4.12)	0.14	0.84
	10 years later date of cohort entry <sup>‡</sup>	0.27	(0.11,0.67)	0.0046	0.27
	10 years older at cohort entry	1.65	(1.18,2.32)	0.0036	0.16

\* Estimated hazard ratio of HIV-1 infection or of HIV-2 infection from the multivariable Cox model, controlling for the other covariates.

<sup>†</sup> *P*-value for test of proportional hazards (PH), based on the method of Grambsch and Therneau [28].

<sup>‡</sup> The covariate was calculated as the years since 7 February 1985, the date the first participant enrolled in the cohort study.

were centred by their mean values. With covariate value specified at these mean values and at Senegalese nationality, the adjusted estimate of  $r_1/r_2$  was 3.40, 95 per cent CI (2.35,4.93),  $P < 0.001$ . It was valid to estimate the common parameter  $r_1/r_2$  because the test for time-varying  $r_1(t)/r_2(t)$  was non-significant ( $P > 0.50$ ). We next re-fit the Cox model including the additional risk factor covariate average number of sexual partners per week, centred by its mean. The adjusted estimate of  $r_1/r_2$  was 3.65, 95 per cent CI (1.47,9.09),  $P = 0.0053$  and 3.86, 95 per cent CI (2.35,6.32),  $P < 0.001$  for the models using variants (i) and (ii) of the sexual contacts variable described in Section 3.3, respectively. These estimates were meaningful because the tests for time-varying  $r_1(t)/r_2(t)$  were non-significant. Note that for each method used for handling missing data in the sexual partners variable, the covariate-adjusted estimates of the infectivity ratio were similar to one another and to the unadjusted estimate. Table IV summarizes the models and resulting estimates of  $r_1/r_2$ , indicating consistent results, with estimates ranging between 3.40 and 3.86.

The testing procedure [29] for evaluating differential infectivity was carried out three times, controlling for the same three sets of covariates. For the first analysis that excluded the sexual partners covariate, the unadjusted test statistic gave  $P < 0.001$ , and all 100 of the re-sampled test statistics gave  $P < 0.01$ . The same result was obtained for the analysis that also adjusted for the sexual partners covariate (ii). The complete-case analysis that included the sexual partners covariate (i) gave a less significant result, with 54 of 100 of the re-sampled test statistics giving  $P < 0.01$ . The reduced significance level can be explained by the smaller sample size.

Table IV. Summary of estimates of the HIV-1/HIV-2 infectivity ratio  $r_1/r_2$ .

Model	Covariates adjusted for	Estimated $r_1/r_2$	95 per cent CI	$p$ -value
Non-parametric*	None	3.55	(2.34,4.80) <sup>†</sup>	<0.001
Re-coded Cox model <sup>‡</sup>	Nationality Date of cohort entry Age at entry	3.40	(2.35,4.93)	<0.001
Re-coded Cox model <sup>‡</sup>	Nationality Date of cohort entry Age at entry Number of sexual partners/week (i) <sup>§</sup>	3.65	(1.47,9.09)	<0.001
Re-coded Cox model <sup>‡</sup>	Nationality Date of cohort entry Age at entry Number of sexual partners/week (ii) <sup>§</sup>	3.86	(2.35,6.32)	<0.001
Donnelly <i>et al.</i> [12]	Number of sexual partners/week (i) <sup>§</sup>	4.24	(1.97,7.52)	<0.001
Donnelly <i>et al.</i> [12]	Number of sexual partners/week (ii) <sup>§</sup>	4.75	(3.35,6.15)	<0.001

\* Inverse variance weighted average of the estimated  $r_1(t)/r_2(t)$  over time, using non-parametric kernel smoothing, as described in the Appendix.

<sup>†</sup> Calculated at the time point by which 50 per cent of the cumulative total 196 infections occurred, with bootstrap adjustment for uncertainty in the partner relative prevalence estimate.

<sup>‡</sup> Re-coded Cox proportional hazards model (3), as described in reference [30].

<sup>§</sup> (i) complete-case analysis; (ii) analysis of all sex workers imputing the missing sexual contacts values to be equal to the mean.

#### 4.4. Sensitivity analysis

When  $K=2$  times the HIV-1/HIV-2 relative prevalence ratio  $\pi_1(\cdot)/\pi_2(\cdot)$  in sex workers was used to approximate the partner HIV-1/HIV-2 relative prevalence ratio, the common estimate of the infectivity ratio  $r_1/r_2$  over time was 2.38, and the adjusted lower 95 per cent confidence limit exceeded one through 10 years of follow-up. Without adjustment for covariates or for uncertainty in the partner relative prevalence, the test for higher HIV-1 infectivity gave  $P<0.01$  for all  $K\leq 2.50$ . The test repeated with adjustment for the three risk factors nationality, date of cohort entry and age at entry gave  $P<0.001$  for all  $K\leq 2.94$ . Thus, the sensitivity analysis confirmed that the inference of greater HIV-1 infectivity persisted when the key assumption was substantially violated.

#### 4.5. Comparison with an analysis by the Donnelly *et al.* [12] method

To check if the method of Donnelly *et al.* [12] gave a different result than the new methods, we applied it to the data set. We took the known, constant HIV-1 prevalence in partners,  $p_1$ , to be the inverse variance weighted average of the 15 annual HIV-1 prevalence estimates between 1985 and 1999 in sex workers. The HIV-2 partner prevalence  $p_2$  and the dual HIV-1/2 partner prevalence  $p_{12}$  were calculated similarly. This gave  $p_1=0.0437$ ,  $p_2=0.105$  and

$p_{12} = 0.00658$ . We applied the method twice: (i) first in a complete case analysis in the 1029 sex workers with data on the average number of sexual partners per week; (ii) second in an analysis in all 1948 sex workers that imputed missing sexual contacts values to be equal to the mean. The first analysis gave  $\hat{r}_1/\hat{r}_2 = 0.000311/0.0000733 = 4.24$ , 95 per cent CI (1.97,7.52), Wald test  $P < 0.001$ , and the second analysis gave  $\hat{r}_1/\hat{r}_2 = 0.000563/0.000119 = 4.75$ , 95 per cent CI (3.35,6.15), Wald test  $P < 0.001$ . The confidence intervals were computed using the delta method and the observed information matrix for  $(\hat{r}_1, \hat{r}_2)'$ , with matrix entries given in [12]. The results are comparable to those produced using the new methods; note that both confidence intervals about  $r_1/r_2$  include all of the estimates 3.40, 3.55, 3.65 and 3.80 listed in Table IV.

## 5. DISCUSSION

The analysis suggests that heterosexual HIV-1 infectivity is greater than heterosexual HIV-2 infectivity. Adjusting for risk factors in the analysis gave similar inferences about higher HIV-1 infectivity as the unadjusted analysis, and a sensitivity analysis showed robustness of the inferences. The infectivity ratio estimate was approximately constant over the follow-up period, with unadjusted common estimate 3.55 and covariate-adjusted common estimates ranging between 3.40 and 3.86. The conclusion of higher HIV-1 infectivity was consistent with other studies that found HIV-1 to have higher viral load [3,4] and a higher perinatal transmission rate [9–11].

We discuss limitations of the methods. First, the methods did not explicitly model the interval-censored nature of the follow-up times. This is not expected to appreciably affect the results, because HIV infection was rare. Second, bias could occur if there were dependencies among cohort participants in the risk of HIV-1 infection or of HIV-2 infection. Such dependencies could arise if clusters of sex workers shared client populations. Third, since direct measurements of the HIV-1 versus HIV-2 male partner relative prevalence over time were not available, it was necessary to approximate the ratio with the relative prevalence over time of a surrogate population (female sex workers). Inferences about differential infectivity could be biased to the extent that the relative prevalence in sex workers differed from that of their partners. If the main source of HIV infection of sex workers and of their partners was sexual contacts with each other, then under the null hypothesis of equal HIV-1 and HIV-2 infectivity, the relative prevalence in sex workers would be expected to represent the relative prevalence in partners. If HIV-1 infectivity exceeded HIV-2 infectivity, or if the male to female infectivity was higher than the female to male infectivity (as some evidence supports [31,32], though a recent study suggested comparable infectivities [18]), then the HIV-1/HIV-2 relative prevalence in sex workers would be expected to overestimate the HIV-1/HIV-2 relative prevalence in partners. A mathematical argument supporting this supposition is provided in the Appendix. If overestimation occurred, then the estimate of the HIV-1/HIV-2 infectivity ratio obtained here would be a lower bound, lending further robustness to the finding of higher HIV-1 infectivity. The claim that the conclusion is conservative is qualified by the observation that influxes of new men into the client pool could alter the partner relative prevalence in unpredictable ways, conceivably causing it to lag ahead or behind the relative prevalence in sex workers. Fourth, since the HIV-1 and HIV-2 infection statuses of partners were unknown, it was necessary to make the homogeneity assumption that at each calendar time, the partner



populations of all at-risk sex workers had the same relative prevalence. Fifth, the unavailability of risk factor information from partners precluded adjusting the analysis for potential partner confounders such as the duration of infection. Sixth, covariate information quantifying risk for HIV-1 and HIV-2 infection, both in intrinsic susceptibility and in exposure, was incomplete and measured with error. In particular, data on the frequency of sexual contacts with partners were missing on many sex workers, and many of the reported frequencies may have been inaccurate.

Some of the above limitations could be addressed by expanding the methods to account for unmeasured heterogeneity in susceptibility and exposure [33], and for the error in covariate measurement. Other limitations, however, could not be addressed in the present study, due to the lack of available information. Improvements in the quality and completeness of the behavioural data collected from sex workers, and collection of data on the HIV-1/HIV-2 relative prevalence ratio and on risk factor covariates directly from clients, would provide the information needed for addressing these limitations. Inferences made using the methodology presented here would be most reliable when these kinds of data are available.

In light of these limitations, we believe the study still provides strong evidence that HIV-1 infectivity exceeded HIV-2 infectivity, because the magnitude of the result was large and comparable for unadjusted and covariate-adjusted analyses, the result was consistent with knowledge gained from other studies, and bias introduced by using the sex worker population to approximate the relative HIV-1 versus HIV-2 prevalence in their partners would most plausibly tend to attenuate the result towards the null hypothesis.

This study demonstrates the utility of prospective studies of HIV at-risk populations, where the kinetics of HIV infection *in vivo* can be readily evaluated using novel time to event modelling methodologies. These multidisciplinary studies not only contribute to our understanding of the complexities of HIV epidemiology but also allow us to use quantitative methods in describing the dynamics of these epidemics at a population level, thus assessing their impact on the future of the HIV epidemic and on the design of HIV vaccine trials. Furthermore, the methodology described here may be useful generally for evaluating relative type-specific infectivity of any heterogeneous pathogen that exposes a cohort followed prospectively, especially when data are available on type-specific prevalence over time in the exposing population.

## APPENDIX

The Nelson–Aalen-type estimates of  $\Lambda_{1\text{adj}}(t)$  and  $\Lambda_{2\text{adj}}(t)$  are given by

$$\hat{\Lambda}_{1\text{adj}}(t) = \int_0^t \frac{\sum_{j=1}^n \hat{\pi}_1^{-1}(t_{c_j} + u) d[N_{1j}(u)]}{\sum_{j=1}^n Y_j(u)}, \quad \hat{\Lambda}_{2\text{adj}}(t) = \int_0^t \frac{\sum_{j=1}^n \hat{\pi}_2^{-1}(t_{c_j} + u) d[N_{2j}(u)]}{\sum_{j=1}^n Y_j(u)}$$

respectively, where  $n = 1948$  is the number of cohort participants,  $N_{ij}(t)$  is a counting process recording infection events of type  $i$  for subject  $j$ ,  $Y_j(t)$  is the at-risk process for HIV (either type) for subject  $j$ , and  $\pi_i(t_{c_j} + u)$  is the relative HIV- $i$  versus HIV partner prevalence at calendar time  $t_{c_j} + u$  that corresponded to follow-up time  $u$  for subject  $j$ . Note that, if  $\hat{\pi}_i^{-1}(t_{c_j} + u)$  in  $\hat{\Lambda}_{i\text{adj}}(t)$  is replaced by one, then  $\hat{\Lambda}_{i\text{adj}}(t) = \hat{\Lambda}_i(t)$  is the ordinary Nelson–Aalen estimator of  $\Lambda_i(t)$ .

The infectivity ratio  $r_1(t)/r_2(t)$  was estimated by

$$\left(\frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}\right) = \frac{\widehat{\lambda}_{1\text{adj}}(t)}{\widehat{\lambda}_{2\text{adj}}(t)}$$

where  $\widehat{\lambda}_{1\text{adj}}(t)$  and  $\widehat{\lambda}_{2\text{adj}}(t)$  are Epanechnikov kernel estimates that smooth  $\widehat{\Lambda}_{1\text{adj}}(t)$  and  $\widehat{\Lambda}_{2\text{adj}}(t)$ , respectively [26, 27]. Specifically

$$\widehat{\lambda}_{i\text{adj}}(t) = \frac{1}{h} \sum_{k=1}^{n_i} K\left(\frac{t - T_{ik}}{h}\right) I(|T_{ik}| \leq t + h) \frac{\sum_{j=1}^n \widehat{\pi}_i^{-1}(t_{cj} + T_{ik})}{\sum_{j=1}^n Y_j(T_{ik})}$$

where the Epanechnikov kernel function  $K(x) = 0.75(1 - x^2)I(|x| \leq 1)$ ,  $h$  is a bandwidth, and  $T_{i1} < \dots < T_{in_i}$  denote the successive HIV- $i$  jump (infection) times of the counting process  $N_i(t) = \sum_{j=1}^n N_{ij}(t)$ . Treating  $\widehat{\pi}_1(\cdot)$  and  $\widehat{\pi}_2(\cdot)$  as known, we computed a 95 per cent CI for  $r_1(t)/r_2(t)$  by transforming symmetric confidence limits for  $\log[r_1(t)/r_2(t)]$  as

$$\left( \left(\frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}\right) \exp \left\{ -1.96 \sqrt{\left[ \widehat{\text{var}}\left(\frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}\right)\right] / \frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}} \right\}, \right. \\ \left. \left(\frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}\right) \exp \left\{ 1.96 \sqrt{\left[ \widehat{\text{var}}\left(\frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}\right)\right] / \frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}} \right\} \right)$$

The variance of  $\left(\frac{\widehat{r_1(t)}}{\widehat{r_2(t)}}\right)$  was estimated using the delta method and the formula (see reference [27], p. 232) with

$$\widehat{\text{var}}(\widehat{\lambda}_{i\text{adj}}(t)) = \frac{1}{h^2} \sum_{k=1}^{n_i} K\left(\frac{t - T_{ik}}{h}\right)^2 I(|T_{ik}| \leq t + h) \\ \times \left( \left[ \sum_{j=1}^n Y_j(T_{ik}) \right] - \Delta N_i(T_{ik}) \right) \frac{\sum_{j=1}^n \widehat{\pi}_i^{-1}(t_{cj} + T_{ik})}{(\sum_{j=1}^n Y_j(T_{ik}))^3}$$

where  $\Delta N_i(T_{ik}) = N_i(T_{ik}) - N_i(T_{ik-1})$ .

To calculate adjusted confidence limits for  $r_1(t)/r_2(t)$  that account for the uncertainty in  $\widehat{\pi}_1(\cdot)$ , the vector of annual log HIV-1/HIV-relative prevalences for calendar years 1985 through 1999 were sampled from a 15-variate normal distribution with mean  $\widehat{\mu}$  and variance-covariance  $(\widehat{\sigma}_{ij})$ ,  $i, j = 1, \dots, 15$ . The mean vector  $\widehat{\mu}$  was calculated from ratios of observed HIV-1 and HIV annual prevalences. The variances  $\widehat{\sigma}_{ii} = \widehat{\text{var}}(\widehat{\pi}_1(i))$  were calculated using the delta method and the estimated binomial variances for the HIV-1 and HIV-2 prevalences in year  $i$ . Estimates of covariance terms  $\widehat{\sigma}_{ij} = \widehat{\text{cov}}(\widehat{\pi}_1(i), \widehat{\pi}_1(j))$ ,  $i \neq j$ , were calculated using a bootstrap procedure which took into account the correlation between the relative prevalence estimates induced by including some sex workers in both calculations. A similar procedure was used to account for the uncertainty in  $\widehat{\pi}_1(\cdot)/\widehat{\pi}_2(\cdot)$  when applying the test of reference [29].

Table V. Ratio ( $\rho$ ) of HIV-1/HIV-2 prevalence ratios for sex workers versus clients.

$r_{1,2} = \frac{r_1}{r_2}$	$r_{m,f} = \frac{r_{\text{male to female}}}{r_{\text{female to male}}}$	$\rho$
1	1/4	1
1	1/2	1
1	1	1
1	2	1
1	4	1
2	1/4	1.01
2	1/2	1.03
2	1	1.06
2	2	1.06
2	4	1.06
4	1/4	1.02
4	1/2	1.04
4	1	1.09
4	2	1.09
4	4	1.10

We now explore mathematically the claim made in the Discussion that the HIV-1/HIV-2 relative prevalence in sex workers may plausibly overestimate the HIV-1/HIV-2 relative prevalence in clients, and as a result the obtained estimates of the infectivity ratio are likely lower bounds. This issue is complicated, depending on many factors including the male to female type-specific infectivity probabilities, the contact rates of sex workers with clients and vice versa, rates of import and export of sex workers and clients into the population, and starting conditions for the prevalence of the two virus types in sex workers and in clients. Here we consider a simple simulation. Assume a closed system, with no entry or exit of individuals, in which all sexual contacts are between sex workers and their clients. Also assume common infectivities  $r_i^{mf}$  and numbers of sexual contacts with clients per year  $c^{sw}$  among sex workers, and common infectivities  $r_i^{fm}$  and annualized numbers of contacts with sex workers per year  $c^{cl}$  among clients. Here the ‘mf’ and ‘fm’ superscripts refer to male to female and female to male infectivity probabilities, respectively. At some initial time  $t$ , suppose there are  $n^{sw}$  sex workers and  $n^{cl}$  clients, and let  $f_i$  be the prevalence of HIV- $i$ ,  $i = 1, 2$ , assumed to be the same for sex workers and for clients, and put  $f = f_1 + f_2$ . With this set-up, the HIV-1/HIV-2 prevalence ratio at time  $t$  is equal in sex workers and in their clients. Given various values for the infectivities and the contact rates, we consider how the prevalence ratios compare one year after time  $t$ . To study this, first note that during the year there are an expected  $(1 - f)n^{sw}r_i^{mf}c^{sw}f_i$  sex workers newly infected with HIV- $i$ , and an expected  $(1 - f)n^{cl}r_i^{fm}c^{cl}f_i$  clients newly infected with HIV- $i$ . It follows that, 1 year after time  $t$ , the expected HIV-1/HIV-2 prevalence ratio in sex workers divided by the expected HIV-1/HIV-2 prevalence ratio in clients equals

$$\rho = \frac{1 + (1 - f)c^{sw}r_1^{mf}}{1 + (1 - f)c^{sw}r_2^{mf}} \bigg/ \frac{1 + (1 - f)c^{cl}r_1^{fm}}{1 + (1 - f)c^{cl}r_2^{fm}}$$

To study  $\rho$ , we first put  $c^{sw} = 6.77 \times 52 = 352.0$ , the mean value in the data set. It is expected that clients have fewer contacts per week than sex workers; accordingly we set  $c^{cl} = 0.10 \times 6.77 \times 52 = 35.2$ . We then vary two ratios, the relative HIV-1/HIV-2 infectivity  $r_{1,2} = r_1^{mf}/r_2^{mf} = r_1^{fm}/r_2^{fm}$  and the relative male to female versus female to male infectivity  $r_{m,f} = r_1^{mf}/r_1^{fm} = r_2^{mf}/r_2^{fm}$ . Table V shows  $\rho$  as a function of  $r_{1,2}$  at values 1, 2, 4 and  $r_{m,f}$  at values 1/4, 1/2, 1, 2, 4. In each scenario the maximum of  $r_1^{mf}, r_2^{mf}, r_1^{fm}$  and  $r_2^{fm}$  is taken to be 0.000437, the average of the two estimates of the HIV-1 infectivity obtained by Donnelly *et al*'s method in Section 4.5.

Note that if the null hypothesis  $r_{1,2} = 1$  is true, then  $\rho = 1$ , so that the relative prevalence in sex workers correctly represents that in clients. If HIV-1 infectivity exceeds HIV-2 infectivity ( $r_{1,2} > 1$ ), then  $\rho$  always exceeds 1 (by 1 per cent to 10 per cent), indicating that the relative prevalence in sex workers overestimates that in clients. Note that the overestimation is slight when female to male infectivity exceeds male to female infectivity, and is larger when male to female infectivity exceeds female to male infectivity. If the contact rate of clients  $c^{cl}$  is increased, then the pattern is the same except the extent of overestimation is less.

#### ACKNOWLEDGEMENTS

We thank Professors Chung-Cheng Hsieh and Ira Longini and the anonymous referees for valuable comments which improved the article. This work was supported by grants AI24643 and AI46703-01 from the National Institutes of Health, grant DMS-9971784 from the National Science Foundation, and U.S. Army Medical Research grants 17-C-7072, 17-C-0138, 17-95-5005 and PO1 A1 30795.

#### REFERENCES

1. Simon F, Matheron S, Tamalet C, Loussert AI, Bartczak S, Pepin J, Dhiver C, Gamba E, Elbim C, Gastaut JA. Cellular and plasma viral load in patients infected with HIV-2. *AIDS* 1993; **7**:1411–1417.
2. Dieng-Sarr A, Thior I, Kokkotou E, Sankalé JL, Marlink R, Coll-Seck EM, Essex M, Siby T, NDoye I, Mboup S, Kanki P. HIV-1 and HIV-2 dual infection: lack of HIV-2 provirus correlates with low CD4+ lymphocyte counts. *AIDS* 1998; **12**:131–137.
3. Popper SJ, Dieng-Sarr A, Travers KU, Guye-NDiaye A, MBoup S, Essex M. Lower HIV-2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. *Journal of Infectious Diseases* 1999; **180**:1116–1121.
4. Albert J, Naucier A, Bottiger B, Broliden P, Albino P, Ouattara S, Bjorkegren C, Valentin A, Biberfeld G, Fenyo E. Replicative capacity of HIV-2, like HIV-1, correlates with severity of immunodeficiency. *AIDS* 1990; **4**:291–295.
5. Hoxie JA, Brass LF, Pletcher CH, Haggarty BS, Hahn BH. Cytopathic variants of an attenuated isolate of human immunodeficiency virus type 2 exhibit increased affinity for CD4. *Journal of Virology* 1991; **65**:5096–5101.
6. Marlink RG, Ricard D, M'Boup S, Kanki PJ, Romet-Lemonne J-L, N'Doye I, Diop K, Simpson MA, Greco F, Chou MJ, DeGruttola V, Hsieh C-C, Boye C, Barin F, Denis F, McLane MF, Essex M. Clinical, hematologic, and immunologic cross-sectional evaluation of individuals exposed to human immunodeficiency virus type 2 (HIV-2). *AIDS Research and Human Retroviruses* 1988; **4**:137–148.
7. Romieu I, Marlink R, Kanki P, M'Boup S, Essex M. HIV-2 link to AIDS in West Africa. *Journal of AIDS* 1990; **3**:220–230.
8. Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, Traore I, Hsieh CC, Dia MC, Gueye EH, Hellinger J, Sankalé JL, Ndoye I, Mboup S, Essex M. Reduced rate of disease development with HIV-2 compared to HIV-1. *Science* 1994; **265**:1587–1590.
9. Poulsen AG, Kvinesdal BB, Aaby P, Lisse IM, Gottschau AK, Dias F, Lauritzen E. Lack of evidence of vertical transmission of human immunodeficiency virus type 2 in a sample of the general population in Bissau. *Journal of AIDS* 1992; **5**:25–30.
10. Abbott RC, Ndour-Sarr A, Diouf A, Ndoye T, Ba D, Tall N, Mbaye N, Kb F, Diadiou F, Essex M, Marlink R, Mboup S, Kanki P. Risk determinants for HIV infection and adverse obstetrical outcomes in pregnant women in Dakar, Senegal. *Journal of AIDS* 1994; **7**:711–717.

11. Adjorlolo-Johnson G, DeCock K, Ekpini E, Vetter KM, Sibailly T, Brattegaard K, Yavo D, Doorly R, Whitaker JP, Kestens L. Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast. *Journal of the American Medical Association* 1994; **272**:462–466.
12. Donnelly C, Leisenring W, Kanki P, Awerbuch T, Sandberg, S. Comparison of transmission rates of HIV-1 and HIV-2 in a cohort of prostitutes in Senegal. *Bulletin of Mathematical Biology* 1993; **55**:731–743.
13. Kanki P. Biologic features of HIV-2: an update. In *AIDS Clinical Review*, Volberding P, Jacobsen MA (eds). Marcel Dekker: New York, 1991; 17–38.
14. Kanki P, M'Boup S, Marlink RJ, Travers K, Gueye A, Boye C, Sankalé J-L, Donnelly C, Leisenring W, Siby T, Thior I, Dia M, Gueye E-H, Hsieh C, N'Doye I, Essex M. Prevalence and risk determinants of human immunodeficiency virus type 2 (HIV-2) and human immunodeficiency virus type 1 (HIV-1) in West African female prostitutes. *American Journal of Epidemiology* 1992; **136**:895–907.
15. Travers K, Mboup S, Marlink R, Gueye-Nidaye A, Siby T, Thior I, Traore I, Dieng-Sarr A, Sankalé JL, Mullins C, Mboup S, Kanki PJ. Natural protection against HIV-1 infection provided by HIV-2. *Science* 1995; **268**:1612–1615.
16. Jewell NP, Shiboski SC. Statistical analysis of HIV infectivity based on partner studies. *Biometrics* 1990; **46**:1133–1150.
17. Kim MY, Lagakos SW. Estimating the infectivity of HIV from partner studies. *Annals of Epidemiology* 1990; **1**:117–128.
18. Gray RH, Wawer MJ, Brookmeyer R, Sewankambo NK, Serwadda D, Wabwire-Mangen F, Lutalo T, Li X, vanCott T, Quinn TC. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet* 2001; **357**:1149–1153.
19. Grant RM, Wiley JA, Winkelstein W. Infectivity of the human immunodeficiency virus: estimates from a prospective study of homosexual men. *Journal of Infectious Diseases* 1987; **156**:189–193.
20. DeGruttola V, Mayer KH. Assessing and modelling heterosexual spread of the human immunodeficiency virus in the United States. *Reviews in Infectious Diseases* 1988; **10**:138–150.
21. Hu DJ, Buve A, Baggs J, van der Groen G, Dondero TJ. What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. *AIDS* 1999; **13**:873–881.
22. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *Numerical Recipes in C*. University of Cambridge Press: Cambridge, 1988.
23. Hudgens MG, Longini IM, Choopanya K, Vanichseni S, Kitayaporn D, Mastro TD, Mock PA, Halloran ME, Hu DJ, Satten GA. Subtype specific transmission probabilities for human immunodeficiency virus type 1 among injecting drug users in Bangkok, Thailand. *American Journal of Epidemiology* 2002; **155**:159–168.
24. Travers K, Eisen G, Marlink R, Kanki P, M'Boup S. Protection from HIV-1 infection by HIV-2. *AIDS* 1998; **12**:224–225.
25. Chiang CL. *An Introduction to Stochastic Processes and their Applications*. Krieger: New York, 1980.
26. Rammlau-Hansen H. Smoothing counting process intensities by means of kernel functions. *Annals of Statistics* 1983; **11**:453–466.
27. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer-Verlag: New York, 1993.
28. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**:515–526.
29. McKeague IW, Gilbert PB, Kanki P. Comparison of competing risks with adjustment for covariate effects. *Biometrics* 2001; **57**:818–828.
30. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995; **51**:524–532.
31. Nicolosi A, Correa Leite ML, Musicco M, Arici C, Gavazzeni G, Lazzarin A. The efficiency of male-to-female and female-to-male sexual transmission of the human immunodeficiency virus: a study of 730 stable couples. Italian Study Group on HIV Heterosexual Transmission. *Epidemiology* 1994; **5**:570–575.
32. European Study Group on Heterosexual Transmission of HIV. Comparison of female to male and male to female transmission of HIV in 563 stable couples. *British Medical Journal* 1992; **304**:809–813.
33. Hougaard P. *Frailty Models for Survival Data*. Laegeforeningens Forlag: Copenhagen, 1995.