STATISTICS IN MEDICINE
*Statist. Med.* 2003; **22**:573–593 (DOI: 10.1002/sim.1342)

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

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

[1]*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.*
[2]*Department of Statistics, Florida State University, Tallahassee, FL 32306, U.S.A.*
[3]*Harvard AIDS Institute and Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, U.S.A.*
[4]*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@scharp.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.



*Received July 2001*
*Accepted April 2002*

## 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,

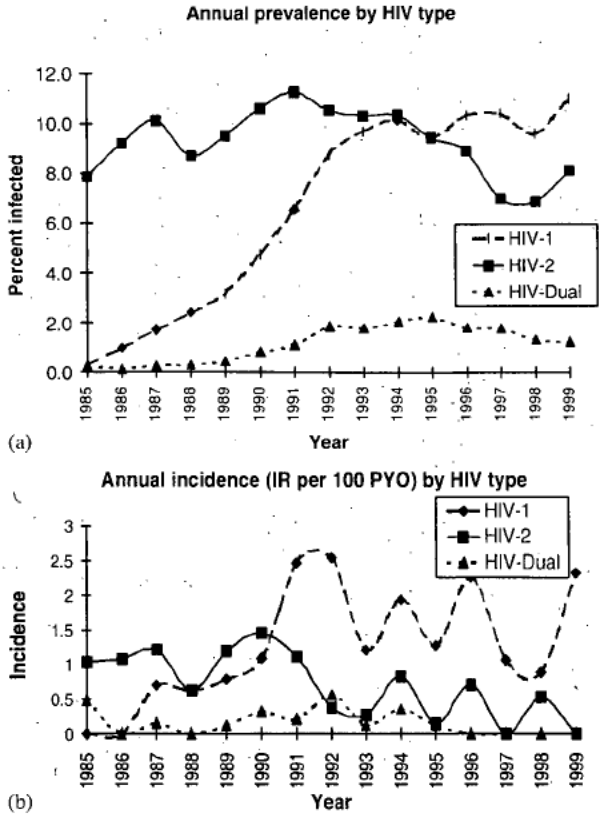


Figure 1. (*a*) Annual prevalence and (*b*) annual incidence of HIV-1 and HIV-2 infection among sex workers participating in the Dakar, Senegal, prospective cohort study. Prevalence rates were calculated from all registered sex workers visiting the clinic in the index year regardless of HIV serostatus 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

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**
Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**
Sync your system to PACER to automate legal marketing.