Viral burden in genital secretions determines male-to-female sexual transmission of HIV-1: a probabilistic empiric model

Hrishikesh Chakraborty, Pranab K. Sen\textsuperscript{a}, Ronald W. Helms\textsuperscript{a}, Pietro L. Vernazza\textsuperscript{b}, Susan A. Fiscus\textsuperscript{c}, Joseph J. Eron\textsuperscript{d}, Bruce K. Patterson\textsuperscript{e}, Robert W. Coombs\textsuperscript{f}, John N. Krieger\textsuperscript{g} and Myron S. Cohen\textsuperscript{d}

Objective: To develop a model to predict transmission of HIV-1 from men to women.

Design: HIV-1 in seminal plasma, and endocervical CCR5 receptors were correlated with epidemiological studies of HIV-1 transmission to develop a probabilistic model.

Settings: Semen samples were collected from patient subjects in Seattle Washington, Chapel Hill, North Carolina, and St. Gallen, Switzerland. Endocervical biopsy specimens were obtained from women in Chicago, Illinois.

Participants: Eighty-six men (not receiving antiretroviral therapy) in whom CD4 cell count and semen volume were available, and 24 women in whom the number of endocervical CCR5 receptors were determined.

Main outcome measures: Prediction of transmission of HIV-1 from men to women per episode of vaginal intercourse based on the absolute burden of HIV (volume $\times$ HIV RNA copies/ml seminal plasma).

Results: The model suggests efficient heterosexual transmission of HIV-1 when semen viral burden is high. When semen contains 100,000 copies of non-syncytium-inducing (NSI) HIV RNA the probability of HIV-1 transmission is 1 per 100 episodes of intercourse; conversely, with 1000 copies NSI HIV RNA in semen, transmission probability is 3 per 10,000 episodes of intercourse.

Conclusions: This model links biological and epidemiological data related to heterosexual HIV-1 transmission. The model can be used to estimate transmission of HIV from men with high semen viral burden from inflammation, or reduced burden after antiretroviral therapy. The results offer a biological explanation for the magnitude of the HIV epidemic in places where earlier studies have shown men have high semen viral burden, such as in sub-Saharan Africa. The model can be used to develop and test HIV-1 prevention strategies.

© 2001 Lippincott Williams & Wilkins

Keywords: HIV-1, transmission, semen, heterosexual

AIDS 2001, 15:621–627

From the Department of Biostatistics, Rollin School of Public Health, Emory University, Atlanta, Georgia, the \textsuperscript{a}Department of Biostatistics, University of North Carolina at Chapel Hill, North Carolina, USA, the \textsuperscript{b}Department of Medicine and Institute for Clinical Microbiology, Kantons-spital, St. Gallen, Switzerland, the Departments of \textsuperscript{c}Microbiology and Immunology and \textsuperscript{d}Medicine, University of North Carolina at Chapel Hill, North Carolina, the \textsuperscript{e}Department of Pediatrics/Infectious Diseases, Children’s Memorial Hospital, Northwestern University Medical School, Chicago, Illinois, and the Departments of \textsuperscript{f}Laboratory Medicine and \textsuperscript{g}Urology, School of Medicine, Seattle, Washington, USA.

Requests for reprints to: Myron S. Cohen, Department of Medicine, 3003 Old Clinic Bldg, CB7005, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7005, USA.

Received: 4 August 2000; revised: 20 December 2000; accepted: 10 January 2001.

ISSN 0269-9370 © 2001 Lippincott Williams & Wilkins 621
Introduction

HIV-1 can be transmitted through contaminated blood and blood products, from mother to child, or through sexual contact [1]. The predominant mode of transmission of HIV worldwide is heterosexual intercourse [2–4].

Epidemiological and mathematical models have been developed to estimate the likelihood of HIV-1 transmission during a single episode of sexual intercourse. Such models are confounded by difficulty in collecting appropriate empirical data from discordant couples (when HIV-1 positive and negative people engage in sex) and from limitations in different kinds of estimation. For example, most published estimates of the probability of sexual transmission of HIV-1 have assumed constant infectivity between couples, ignoring the possibility that acquired immunity might reduce the efficiency of transmission [5].

The probability of per-partner sexual transmission of HIV-1 has been examined in 11 different studies, [6] whereas the per-sex-act probability of transmission has been reported in 13 studies [7–19]. The probability of transmission of HIV-1 from male to female during an episode of intercourse has been examined in seven of these studies [7,14–19]. Analysis of data from North American and European studies of heterosexual couples provide estimates of per-sex-act HIV-1 transmission of approximately 1 in 1000 (0.001, ranging from 0.0008 to 0.002) [6], although the magnitude of the HIV-1 epidemic would argue that these estimates might be unreasonably low.

The transmission of HIV-1 is ultimately a biological event, which depends upon the infectiousness of the HIV-1-infected index case [5] and the susceptibility of the uninfected partner [20]. Infectiousness is likely determined by the concentration of virus in the genital secretions and by the viral phenotype [5]. We [21,22] and others [23,24] have developed assays to measure the concentration of HIV-1 in male genital secretions, the genotype of HIV-1 in male genital secretions, [25] and the number of receptors for HIV-1 in the endocervix [26]. We have used these data to develop a model of transmission of HIV-1 from the male to female which correlates biological and epidemiological data. The results can be used to understand better the distribution of HIV-1 transmission probabilities, and to develop better HIV-1 prevention strategies. The results demonstrate that the per-contact transmission probability for transmission of HIV-1 from men to women may be considerably greater in many parts of the world than estimated in epidemiological studies.

Methods

A probabilistic model was developed to estimate the male-to-female per-contact HIV-1 transmission probability for a known transmitter and receptor cell counts by using the conditional and unconditional probability theory. This type of model has the advantage that empirical data from different, independent sources can be applied.

We assumed that the best predictor of infectiousness of the male partner is the cell-free virus measured in seminal plasma. It is not known whether HIV-1 is transmitted from cell free virus in the seminal plasma, or from cellular HIV-1 [5]. However, in the absence of genital tract inflammation, cell free HIV-1 in seminal plasma reflects the number of HIV-1 infected cells in semen [27,28]. We also assumed that the risk of HIV-1 transmission remains the same for each episode of intercourse with a partnership, although some have argued that exposure leads to some degree of immunity [29]. We also assumed that total non-synctium-inducing (NSI) HIV-1 RNA concentration \( x_1 \) and CD4 + CCR5 receptor cells \( x_2 \) were represented by a Pearson’s type-1 distribution that could be transformed into a Beta distribution (subtracting the minimum value and divided by the range) [30]. Data with highly varied configurations can be modeled with a Beta distribution. It should be noted, however, that isolates other than NSI can be transmitted sexually [31] and cells expressing other receptors for HIV-1 may prove receptive [32].

The natural choices to model a discrete response (infected or not-infected) is to use a logistic probability model. When the likelihood of an event is small we can describe the logistic probability as [33]:

\[
P_{t/x_1,x_2} = \frac{e^{g(x_1,x_2)}}{1 + e^{g(x_1,x_2)}}
\]

Where \( P_{t/x_1,x_2} \) is the conditional probability of HIV-1 transmission given the values of \( x_1 \) and \( x_2 \) (see above).

We have to choose the function \( g(x_1, x_2) \) in such a way that if there is no NSI HIV RNA then there will be no transmission, and if there are no receptor cells then there will be no transmission. We evaluated different functions for \( g(x_1, x_2) \) and the following function that satisfies our conditions:

\[
P_{t/x_1,x_2} = \exp\{b_1 \log x_1 + b_2 \log x_2\} \tag{1}
\]

To estimate \( b_1 \) and \( b_2 \) we can write the unconditional HIV-1 transmission probability \( p_t \) as:
A model of sexual transmission of HIV Chakraborty et al. 623

\[ p_t = \int_0^1 \int_0^1 f(x_{i1}) f(x_{i2}) (p_t / x_{i1}, x_{i2}) dx_{i1} dx_{i2} \]

where \( t = 1, 2; i = 1, 2 \ldots n_i \) and \( j = 1, 2 \ldots n_j \)

\[ p_t = \frac{\Gamma(\alpha_{1t} + \beta_{1t})}{\Gamma(\alpha_{1t})} \Gamma(\beta_{1t}) \int_0^1 e^{(b_1 \log x_{i1}) (x_{i1})^{\alpha_{1t}-1} (1 - x_{i1})^{\beta_{1t}-1}} dx_{i1} \]

\[ \times \frac{\Gamma(\alpha_{2t} + \beta_{2t})}{\Gamma(\alpha_{2t})} \Gamma(\beta_{2t}) \int_0^1 e^{(b_2 \log x_{i2}) (x_{i2})^{\alpha_{2t}-1} (1 - x_{i2})^{\beta_{2t}-1}} dx_{i2} \]

Where \( \alpha_t \) and \( \beta_t \) are Beta distribution parameters. After some algebraic manipulation the final equations can be written as follows [33]:

for \( t = 1 \) we have

\[ p_1 = \frac{\Gamma(\alpha_{11} + \beta_{11}) \Gamma(\alpha_{2} + \beta_{2})}{\Gamma(\alpha_{11}) \Gamma(\alpha_{2} + b_{1}) \Gamma(\alpha_{11} + \beta_{11} + b_{1})} \Gamma(\alpha_{2} + b_{2}) \] (2)

for \( t = 2 \) we have

\[ p_2 = \frac{\Gamma(\alpha_{12} + \beta_{12}) \Gamma(\alpha_{2} + \beta_{2})}{\Gamma(\alpha_{12}) \Gamma(\alpha_{2} + b_{1}) \Gamma(\alpha_{12} + \beta_{12} + b_{1})} \Gamma(\alpha_{2} + b_{2}) \] (3)

The above two equations do not have algebraic solutions for \( b_1 \) and \( b_2 \). Therefore, we used the successive approximation method to get an estimate of \( b_1 \) and \( b_2 \). Substituting the values of \( b_1 \) and \( b_2 \) in (1) we estimate the male-to-female penile-vaginal per-sexual-act HIV-1 transmission probability with a known infectiousness measure for the male partner and a known susceptibility measure for the female partner.

The model uses extensive data from four different study sites (see Results). Semen specimens were collected at the University of North Carolina, University of Washington, and St. Gallen. Endocervical CCR5 receptors were studied at Northwestern University. The methods used for measurement of HIV-1 in seminal plasma [22,24] and CCR5 receptor density [26] have been reported previously.

**Results**

Nine studies (eight from the USA and Europe and one from Africa) have reported the concentration of HIV-1 in seminal plasma [5]. The three largest studies were conducted in Chapel Hill \((n = 88)\), [22,34–36], Seattle \((n = 165)\), [24] and St. Gallen \((n = 100)\) [22]. The data used for the current analysis included additional subjects who were not available for study at the time the cited papers were written. We evaluated the data from these three centers from the inception of the research up to and including July 1999. We considered only samples collected from visits at which the patients were receiving no antiviral therapy (as antiviral therapy may be expected to reduce HIV-1 in semen [5,37]), and for which the seminal HIV-1 RNA count/ml, semen volume, and CD4 cell count were available.

With these limitations, 41 subjects seen in Chapel Hill between July 1994 and February 1996 provided 64 samples \((1–5/subject)\). The total seminal HIV-1 RNA count in one ejaculate ranged from 2000 to 2790 000 with a mean of 143 455 and a median of 8100. Seventeen subjects from Seattle, with 40 separate visits, were included. The number of samples collected from subjects ranged from one to three between April 1994 and July 1997. The total HIV-1 RNA in semen in one ejaculate ranged from 30 to 39 795 copies with a mean of 2623 copies and a median of 480 copies. Twenty-eight subjects from the Swiss cohort were included: each subject provided only one sample between October 1994 and December 1997. The total HIV-1 RNA in semen in one ejaculate ranged from 200 to 13 935 418 copies with a mean of 971 510 and a median of 2488 copies.

The total number of samples was divided into two groups: in one group were visits at which subjects had CD4 cell counts \(< 200 \times 10^6/l\) and in another group were visits at which subjects had CD4 cell counts \(> 200 \times 10^6/l\). A CD4 cell count of 200 was chosen as a cutoff because of a comparable epidemiological study [18]. In the first group 40 samples from 33 different patients were used and in the second group 92 samples from 53 patients subjects were considered (Table 1). In the first group CD4 cell count was in the range 5–189 \(\times 10^6/l\) (median, 105 \(\times 10^6/l\)) and in the second group CD4 cell count was in the range 202–1240 \(\times 10^6/l\) (median, 395 \(\times 10^6/l\)).

Semen volume per ejaculate ranged from 0.10 ml to 7.30 ml with a mean of 2.56 ml and a median of 2.30 ml and the distribution was similar in two groups. The mean (median) HIV-1 RNA count/ml was 275, 202 (1302). Total seminal HIV-1 RNA count in one ejaculate was calculated by multiplying the HIV-1 RNA count/ml by the total semen volume. The HIV-1 RNA/ejaculate distribution was different in two groups, as expected based on several studies demonstrating increasing HIV-1 in seminal plasma as CD4 cell counts fall [22,34]. The degree of variation in HIV-1 RNA in semen was greater in the CD4 cell count
The transmission of HIV-1 requires that the infectious strain utilize very specific receptors [20,26]. Recent data suggest that HIV-1 variants which use CCR5 receptors (NSI isolates) are preferentially sexually transmitted [38]. As the precise number of NSI isolates in a swarm of HIV-1 in semen is unknown we assumed that it is similar to a swarm of HIV-1 in blood. In Centers for Disease Control (CDC) stage IV CII patients studied by Schuitemaker et al. [39] 70% of the swarm was NSI whereas in CDC stage II 100% of the swarm was NSI, and this distribution was used for our calculations.

Observations were correlated because the visits of an individual patient are not independent. The bootstrap resampling process was used to calculate the Beta distribution parameter estimates for two groups. First, one observation was selected randomly from each subject and the minimum value and the range for the set were calculated. Second, all of the selected observations were transformed by subtracting the minimum value and dividing it by the range. From the transformed variable, the Beta distribution parameter estimates of \( \alpha \) and \( \beta \) were calculated by using the maximum likelihood method. Third, this process was repeated 1000 times to obtain 1000 Beta distribution parameter estimates of \( \alpha \) and \( \beta \). Finally, the mean of those 1000 estimates of \( \alpha \) and \( \beta \) was calculated. The bootstrap resampling for the two groups was carried out independently. The Beta distribution parameter estimates for the CD4 cell count \( \leq 200 \times 10^6/l \) group were \( \alpha_{12} = 0.242, \beta_{12} = 1.428 \).

The number of receptors for HIV-1 will also determine the efficiency of transmission. The receptor cell distribution parameter was estimated from studies in which the CD4 + CCR5 cell count/mm\(^2\) in the endocervix was actually measured [26]. The mean (median) receptor cell count was 176.0/mm\(^2\) (184.8/mm\(^2\)) with a minimum of 12.6/mm\(^2\) and a maximum of 449.4/mm\(^2\). The receptor cell values were transformed by subtracting the minimum value of 12.6/mm\(^2\) and dividing by the range of 436.8/mm\(^2\). By using the scaled data Beta distribution parameter estimates of \( \alpha_2 = 0.769 \) and \( \beta_2 = 1.143 \) were calculated by using maximum likelihood method.

Also used were the unconditional probability estimates from a published paper [18] in which the male-to-female per-sex-act penile-vaginal HIV-1 transmission probability was estimated to be 0.0006 for the CD4 cell count \( \leq 200 \times 10^6/l \) group and 0.0007 for the CD4 cell count \( > 200 \times 10^6/l \) group. All of the values of \( P_{1t}, P_{2t}, \alpha_{11}, \beta_{11}, \alpha_{12}, \beta_{12}, \alpha_2, \) and \( \beta_2 \) were placed in equations (2) and (3) and used the successive approximation method with a precision of two decimal places to estimate \( b_1 \) and \( b_2 \) (model parameters). Our estimates were \( b_1 = 0.778 \) and \( b_2 = 0.604 \). Thus, the final model equation could be written as:

\[
P_{n/x_1, x_2} = \frac{X_1^{0.778}}{X_2^{0.604}}
\]

The transmission probabilities for different values of seminal viral load in one ejaculate for three different
endocervical receptor cell number densities (25%, 50% and 75%) are presented in Fig. 1, with the assumption that 100% of the HIV-1 variants in the semen express the NSI phenotype. This model predicts that the per-contact HIV-1 transmission probability ranges from 0.0001 to 0.0003 when the seminal viral load is 1000 (3.0 log10) copies per ejaculate and the endocervical receptor cells count ranges between the 25th and the 75th percentile. The model demonstrates a sharp increase in transmission probability as seminal viral load and/or receptor cells count increases. For 100 000 copies (5.0 log10) HIV-1 in an ejaculate sample, the transmission probabilities range from 0.0039 to 0.0096. It seems unlikely that only NSI variants will be detected in the semen. Accordingly, the effects of varying the concentrations of HIV-1 syncitium inducing (SI)/NSI phenotype are shown in Fig. 2.

Discussion

Estimates of the efficiency of transmission of HIV-1 have been derived from epidemiological studies and mathematical models. Epidemiological studies [7,14–19] which have included estimates of male to female sexual transmission of HIV-1 are summarized in Table 2. However, the transmission probabilities presented are so low that it becomes difficult to understand the magnitude of the HIV-1 pandemic, especially in developing countries. An alternative approach to explain the epidemic is the development of mathematical models. For example, Jacquez and coworkers have argued that the majority of sexual transmission of HIV-1 occurs during the narrow window of primary infection [41].

Greatly improved understanding of the biology of sexual transmission of HIV-1 [5] and collection of large numbers of relevant samples provides a unique opportunity to link epidemiological and biological data. We believe that HIV-1 transmission must depend on the concentration of the appropriate HIV-1 variants in the genital secretions, [5] and availability of permissive cells [20]. Based on the understanding of the biology of sexual transmission and using data collected in several different studies, we have developed a probabilistic model (equation 2). This model predicts very limited

---

Table 2. Published estimates of male-to-female per-sex-act penile-vaginal HIV-1 transmission probabilities.

<table>
<thead>
<tr>
<th>Study</th>
<th>Male-to-female per-sex-act penile–vaginal HIV-1 transmission probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Padian et al. 1987 [14]</td>
<td>0.0008–0.001</td>
</tr>
<tr>
<td>Peterman et al. 1988 [7]</td>
<td>0.0005–0.0023</td>
</tr>
<tr>
<td>Wiley et al. 1989 [15]</td>
<td>0.0008–0.001</td>
</tr>
<tr>
<td>Duerr et al. 1994 [16]</td>
<td>0.0006–0.0026</td>
</tr>
<tr>
<td>Downs et al. 1996 [17]</td>
<td>0.0005–0.0012</td>
</tr>
<tr>
<td>Leonaert et al. 1998 [18]</td>
<td>0.0006–0.0008</td>
</tr>
<tr>
<td>Shiboski et al. 1998 [19]</td>
<td>0.0006–0.0009</td>
</tr>
</tbody>
</table>

*Combined male-to-female and female-to-male transmission probability.
transmission of HIV-1 when the concentration of HIV-1 in semen is low (as is commonly the case in developed countries [36], and in subjects receiving antiretroviral therapy [37]). Markedly increased efficiency of HIV-1 transmission is expected to occur when the concentration of HIV-1 in semen becomes greater (Figs 1 and 2).

There are several limitations to this model. First, the model was constructed with available biological data. Collection of semen specimens is difficult, and many potential subjects were excluded from consideration because they were receiving antiretroviral therapy. Second, our approach to the phenotypic requirements for HIV-1 transmission is flawed. We focused entirely on the NSI/SI phenotype whereas many other virologic properties might affect transmission [5]. Furthermore, our assumption that only NSI isolates can be transmitted is not entirely correct, as SI variants have occasionally been sexually transmitted under some conditions [31]. In addition, we assumed that the isolates in semen are similar to those in blood [39], but the SI/NSI ratio in the HIV-1 swarm in semen is unknown [40]. In addition, we would expect to detect a higher proportion of SI isolates in subjects with more advanced disease [42,43]. Third, seminal plasma HIV-1 RNA levels were measured using two different techniques [22,24]. However, a recent study suggests that the seminal and blood HIV-1 RNA measurements used by these laboratories are comparable [44].

The greatest limitation of this and other models lies in the tremendous difficulty in clinical validation. Proving the model to be correct requires examination of the concentration of HIV-1 in semen actually leading to transmission of the virus in a discordant couple. A recent study in Uganda [45] has provided an exceptional opportunity for further examination of the predictions in the model. Quinn et al. [45] measured HIV-1 in the blood of more than 15,000 study subjects, ultimately demonstrating that 415 HIV-1 infected subjects (228 infected men) were in discordant sexual partnerships. HIV-1 was not transmitted by infected subjects with less than 1500 copies of HIV-1 RNA/ml serum, whereas subjects with more than 50,000 copies HIV-1 RNA/ml serum infected sexual partners at a rate of 23 per 100 person-years over 30 months. While blood and semen clearly reside in separate and distinct biological compartments, blood viral burden can be correlated with viral burden in semen [22–24,46]. In addition, genital tract inflammation (which was commonly detected in the study in Uganda [45,47]) can increase HIV-1 in genital secretions to a level considerably greater than the level in blood [48]. The transmission frequency observed in the Ugandan study strongly suggests that the increased transmission predicted at higher concentrations of HIV-1 in semen our model must have occurred.

Prevention of transmission of HIV-1 has proven a daunting task, in part because of confusion about the benefits to be derived from different approaches. Blower and coworkers have developed an important mathematical model designed to address the effects of antiretroviral therapy on the HIV-1 epidemic [49]. This model is limited, however, by lack of empirical data. The probabilistic model presented here is actually developed around biological results. The model can be used to predict the effects of differences in semen viral burden and CCR5 receptors on HIV-1 transmission. Indeed, we and others have reported considerable variability in the concentration of HIV-1 in semen resulting from anatomical and physiological barriers between blood and the male genital tract, local genital tract replication of HIV-1 (which is greatly influenced by inflammation and sexually transmitted diseases) and the effects of antiviral therapy [5,37,48]. In addition, CCR5 receptor density is affected by a variety of factors [20,26]. Such variation may offer a biological basis for the accelerated spread of HIV-1 in some developing countries [50]. In addition, this model can be used to predict the effects of biological interventions designed to reduce viral burden, influence viral phenotype, and/or expression of receptor cells.

Sponsorship: Supported by NIH Award ROI 6K14381, the UNC CFAR, P30-037260, the Verne R. Caveness General Clinical Research Center RR00046, K/A 157810, Swiss National Science Foundation 233-048902.96.

References