

Supplementary Note 4:

Virus exposure (module IV)

In this note we derive a statistical model coupling donor viral loads to viral transmission, following different modes-of-transmission (e.g. homosexual- & heterosexual intercourse) and estimate the effect of PrEP after realistic viral exposures, after integration with module III (Supplementary Note 3).

SN4 Statistical Model of Viral Exposure

Sexual transmission is the main mode of HIV-1 infection. A number of studies have shown that the main factor that affects the probability of infection per coitus is the viral load in donor^{1,2}. In absence of a complete cure and an effective vaccine against HIV-1, novel prevention strategies such as treatment as prevention (TasP) and pre-exposure prophylaxis (PrEP) have been devised to contain the ongoing spread of HIV-1. Treatment for prevention targets HIV-1 infected individual (potential donors), whereas PrEP aims at the uninfected individuals with a considerable risk of acquiring HIV-1, e.g. sero-discordant couples, MSM (men who have sex with men), sex workers, etc. In this note, we present a mathematical model linking the viral load in the donor with the number of virus being transmitted to the recipient. The presented model can be used to assess and analyze the effects of different strategies (e.g. TasP or PrEP).

SN4.1 Relation between number of transmitted viruses and donor viral load

We assume that the number of viruses transmitted and reaching a target cell environment n is related to the virus load in the donor k through a binomial model with success rate r (probability of successfully transmitting a donor virus to the recipient).

$$n \sim \mathcal{B}(f(k), r), \quad (\text{SN4.1})$$

where the term $f(k)$ is a power function of the viral load k in the donor, i.e.

$$f(k) = \|k^m\|, \quad (\text{SN4.2})$$

where m is an exponent of the viral load k and $\|\cdot\|$ is the next integer function.

The probability of transmitting n viruses to the recipient when the viral load in the donor is k is then:

$$P(V_0 = n | \text{VL} = k) = \binom{\|k^m\|}{n} \cdot r^n \cdot (1-r)^{(\|k^m\| - n)} \quad (\text{SN4.3})$$

Function $f(k)$ in eq. (SN4.2) is motivated by observations from Hughes et al.² and Wilson et al.³. The authors observed a linear relationship between the log (viral load) in the transmitter/donor and the log (infection probability) in the $P(\text{inf}) \ll 1$ regime. A detailed derivation is given in subsection SN4.4.2.

SN4.2 Distribution of viral load in potential donors/transmitters

A number of studies have reported that the viral load in untreated HIV-1 infected individuals is log-normal distributed^{1,4}, i.e.:

$$k \sim \log \mathcal{N}(\mu, \sigma), \quad (\text{SN4.4})$$

enabling us to devise a function, which computes the probability of a viral load k in a potential transmitter,

$$P(\text{VL} = k) = g(k, \mu, \sigma). \quad (\text{SN4.5})$$

30 SN4.3 Mean infection probabilities

31 The mean probability of infection *per coitus* is the weighted average infection probability over all potential trans-
32 mitters and their respective viral loads at the time of contact:

$$\bar{P}(\text{inf}) = \int_{k=0}^{\infty} P(\text{VL} = k) \cdot P(\text{inf} | \text{VL} = k) \quad (\text{SN4.6})$$

33 where the term $P(\text{VL} = k)$ is the probability that a donor has viral load k and the term $P(\text{inf} | \text{VL} = k)$ is the
34 probability of infection per coitus conditional that a donor has viral load k .

35 We assume that the number of viruses transmitted to the recipient is drawn from a binomial distribution,
36 depending on the viral load in the donor (see above). The term $P(\text{inf} | \text{VL} = k)$ can be decomposed into

$$P(\text{inf} | \text{VL} = k) = P(V_0 = n | \text{VL} = k) \cdot P(\text{inf} | V_0 = n), \quad (\text{SN4.7})$$

37 where the term $P(V_0 = n | \text{VL} = k)$ represents the probability that n transmitted viruses enter a target cell com-
38 partment in the recipient conditioned that the donor/transmitter had viral load k . The term $P(\text{inf} | V_0 = n)$ denotes
39 the conditional probability that infection is established in the recipient after n viruses had been transmitted (see
40 **Supplementary Note 3**). Taken together, we get

$$\bar{P}(\text{inf}) = \int_{k=0}^{\infty} P(\text{VL} = k) \cdot \left(\sum_{n=0}^{\infty} P(V_0 = n | \text{VL} = k) \cdot P(\text{inf} | V_0 = n) \right) \quad (\text{SN4.8})$$

41 Using eqs. (SN3.14) (**Supplementary Note 3**) and eqs. (SN4.3),(SN4.5), the mean infection probability per
42 coitus can be rewritten as

$$\bar{P}(\text{inf}) = \int_{k=0}^{\infty} g(k, \mu, \sigma) \cdot \left(\sum_{n=0}^{\infty} \binom{\|k^m\|}{n} \cdot r^n \cdot (1-r)^{(\|k^m\|-n)} \cdot (1-(1-\alpha)^n) \right) \quad (\text{SN4.9})$$

43 In summary, the equation above for the mean infection probability per coitus $\bar{P}(\text{inf})$ summarized the effects of:

- 44 1. the **viral load in the donor/transmitter**,
- 45 2. the **number of transmitted viruses reaching a target cell compartment in the recipient**,
- 46 3. the **infection probability when n viruses have been transmitted**.

47 This equation can therefore be used to assess the epidemiological effectiveness of methods that reduce viral load,
48 e.g. treatment as prevention (TasP, item 1.), circumcision and condom use as means to reduce the number of viruses
49 being transmitted (item 2.) and pre-exposure prophylaxis (PrEP, item 3.).

50 SN4.4 Parametrization

51 Eq. (SN4.9) requires a number of variables:

- 52 • the infection probability given a single transmitted virus in the recipient (α) which has been derived in
53 **Supplementary Note 3**,
- 54 • the mean (μ) and standard deviation (σ) of the \log_{10} viral load in the donor/transmitter population,
- 55 • the exponent (m) and the success rate (r) for the binomial distribution (number of transmitted viruses).

56 Next, we will explain the parametrization step by step.

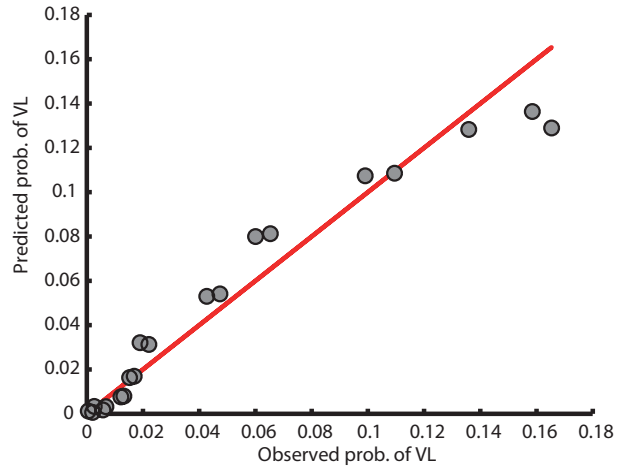


Figure SN4.1: Comparison of observed vs. predicted distribution of \log_{10} viral load (copies per ml blood) for the German Sero-converter data⁵

57 **SN4.4.1 Virus load distribution in potential transmitters**

58 In order to infer the distribution of a viral loads in treatment naive potential HIV-1 transmitters, we analyzed viral
 59 load data from recently infected treatment-naive individuals (German HIV-1 sero-converter study and acute sero-
 60 converters, $N = 1213$)^{5,6}. Since viral transmission occurs preferentially shortly after infection⁶, this particular
 61 population will provide the best possible source of relevant information regarding viral load distribution at the time
 62 of transmission. Figure 2A (main manuscript) shows the histogram of \log_{10} viral load data with a superimposed
 63 red line representing the probability of viral load assuming normal distribution with mean $\mu = 4.51$ and standard
 64 deviation $\sigma = 0.98$. Figure SN4.1 shows a QQ plot with observed probability of viral load in x axis and predicted
 65 probability of viral load in y axis. Both figures shows that the viral load is log-normal distributed. Further,
 66 the estimated mean and standard deviation from our data set (German Sero-converter study) are also in good agreement
 67 with the RAKAI study¹ (see Table SN4.1).

Distribution parameters	RAKAI study	German sero-converter data
Mean (μ)	4.02	4.51
Sigma (σ)	0.76	0.98

Table SN4.1: Comparison of \log_{10} viral load distribution between German Seroconverter data and RAKAI study (Uganda)¹

68 SN4.4.2 Derivation of power function $f(k)$ and parametrization

69 We will briefly explain the rationale behind the choice of $f(k)$ as a power function of the viral load in the transmitter/donor. The function f relates the viral load in the transmitter/donor to the number of viruses that reach a target cell compartment in the recipient. Since it is impossible to assess the number of transmitted viruses (that reach a target cell compartment), instead we will use function f together with eq. (SN3.14) (**Supplementary Note 3**) to determine the relation between the viral load in the donor and the infection probability (see eq. SN4.7).

74 Previously, a number of studies have shown a linear relationship between the \log_{10} of probability of infection for a viral load and \log_{10} of viral load in donor^{2,3}, which is however only valid in the $P(\text{inf}) \ll 1$ regime. In the aforementioned regime we will see that this approximation is in line with our previous observations (eq. (SN4.3)).

77 First, we will simplify eq. (SN4.7):

$$\begin{aligned}
 P(\text{inf}|\text{VL} = k) &= \sum_{n=0}^{\infty} P(V_0 = n|\text{VL} = k) \cdot P(\text{inf}|V_0 = n) \\
 &= \sum_{n=0}^{\infty} P(V_0 = n|\text{VL} = k) \cdot (1 - (1 - \alpha)^n) \quad \text{using eqs (SN3.14),(SN3.27)} \\
 &= \underbrace{P(V_0 = 0|\text{VL} = k) \cdot (1 - (1 - \alpha)^0)}_{=0} + \sum_{n=1}^{\infty} P(V_0 = n|\text{VL} = k) \cdot (1 - (1 - \alpha)^n) \quad \text{(SN4.10)} \\
 &= \sum_{n=1}^{\infty} P(V_0 = n|\text{VL} = k) \cdot (1 - (1 - \alpha)^n). \quad \text{(SN4.11)}
 \end{aligned}$$

78 As $n \rightarrow \infty$, we have $(1 - (1 - \alpha)^n) \rightarrow 1$. In the following, we will subsume all terms where at least one virus is transmitted. Therefore, we can replace the sum

$$\begin{aligned}
 P(\text{inf}|\text{VL} = k) &= \sum_{n=1}^{\infty} P(V_0 = n|\text{VL} = k) \cdot (1 - (1 - \alpha)^n) \\
 &= (1 - P(V_0 = 0|\text{VL} = k)) \cdot \phi \quad \text{(SN4.12)}
 \end{aligned}$$

$$= 1 - P(V_0 = 0|\text{VL} = k) - \chi \quad \text{(SN4.13)}$$

$$\approx 1 - P(V_0 = 0|\text{VL} = k) \quad \text{(SN4.14)}$$

80 with $0 < \phi \leq 1$ and $0 \leq \chi < P(V_0 = 0|\text{VL} = k)$. The equation states that in the $P(\text{inf}) \ll 1$ regime the infection probability is roughly given by the probability that *at least* one virus is being transmitted to the recipient. We then get

$$\begin{aligned}
 P(\text{inf}|\text{VL} = k) \approx 1 - P(V_0 = 0|\text{VL} = k) &= 1 - \binom{f(k)}{0} \cdot r^0 \cdot (1 - r)^{f(k)-0} \quad \text{using eq. (SN4.3)} \\
 &= 1 - \frac{f(k)!}{0! \cdot f(k)!} \cdot \underbrace{r^0}_{=1} \cdot (1 - r)^{f(k)} \\
 &= 1 - (1 - r)^{f(k)}. \quad \text{(SN4.15)}
 \end{aligned}$$

83 For $r \cdot f(k) \ll 1$ (the expected number of transmitted viruses is small/on average less than 1), the Taylor approximation $1 - (1 - r)^{f(k)} \approx r \cdot f(k)$ can be applied

$$P(\text{inf}|\text{VL} = k) \approx r \cdot f(k) \quad \text{(SN4.16)}$$

85 Taking the logarithm on both sides

$$\begin{aligned}
 \log_{10}(P(\text{inf}|\text{VL} = k)) &\approx \log_{10}(r \cdot f(k)) \\
 &= \log_{10}(r) + \log_{10}(f(k)) \\
 &= \log_{10}(r) + \log_{10}((k)^m) \quad \text{using eq. (SN4.2)} \\
 &= \log_{10}(r) + m \cdot \log_{10}(k). \quad \text{(SN4.17)}
 \end{aligned}$$

86 The equation above justifies (under the assumptions made) the linear relationship between the \log_{10} probability
 87 of infection and the \log_{10} of viral load in the donor/transmitter that has previously been stated^{2,1}, i.e. each \log_{10}
 88 increase in the donor viral load increases the probability of infection in the $P(\text{inf}) \ll 1$ regime. In the following,
 89 we fixed the slope m to the value $\log_{10}(2.45)$ reported by Wilson et al.³.

90 Up to now, we have explained the parametrization of all variables in the eq. (SN4.9) except for the success rate
 91 r in eq. SN4.3 (proportion of donor/transmitter viruses that enter a target cell environment in the recipient).
 92 The mean infection probabilities per coitus $\bar{P}(\text{inf})$ for different modes of transmission are known from various stud-
 93 ies, see e.g.⁷ and can be used together with the derived formulas to estimate this parameter. Thus, in eq. (SN4.9)
 94 we fixed all the parameters except r and estimated r for different modes of transmission using least squares mini-
 95 mization. Table SN4.2 shows the estimates of r for different modes of transmission.

Mode of intercourse	Mean probability of infection per coitus ($\bar{P}(\text{inf})$)	estimated success rate r
homosexual	0.03	3.7136×10^{-3}
heterosexual	0.003	3.6266×10^{-4}

Table SN4.2: **Estimation of r for different modes of transmission.** Mean probability of infection per coitus $\bar{P}(\text{inf})$ were taken from Royce et al.^{7,8,9}

96 The following Table lists all parameter values used in the eq. (SN4.9) to estimate the mean infection probability
 97 per coitus.

Parameter	Value	Ref.
infect. prob. for single virus (α)	0.0996	10
mean \log_{10} viral load in donor (μ)	4.51	5
standard deviation of \log_{10} viral load in donor (σ)	0.98	5
exponent (m)	0.3892	1
success rate for <i>homosexual</i> intercourse (r)	3.7136×10^{-3}	-
success rate for <i>heterosexual</i> intercourse (r)	3.6266×10^{-4}	-

Table SN4.3: **Parameters values for eq. (SN4.9).**

98 **SN4.5 Relation between donor virus load and infection probability (no treatment)**

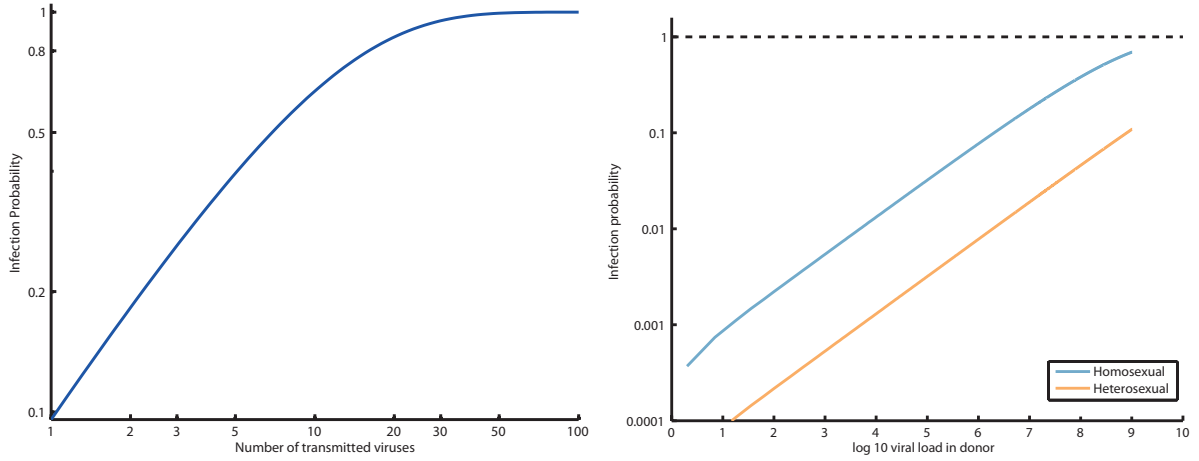


Figure SN4.2: **Relation between of infection probability number of transmitted viruses n and virus load in donor for different modes of transmission.** **Left:** Infection probability depending on the number of transmitted viruses in the absence of PrEP \emptyset . Both axes are in logarithmic scale. **Right:** Dependence of the infection probability *per coitus* on the viral load in a donor for different modes of transmission. The green and blue lines represent unprotected homosexual- and heterosexual intercourse respectively. The horizontal black dashed line marks the infection probability of 1.

99 Fig.SN4.2(left) shows the relation between the infection probability and the number of transmitted viruses
 100 reaching a target-cell compartment to the recipient. The infection probability given a single virus α is computed in
 101 **Supplementary Note 3** and equal to 0.0996 i.e roughly 10 %. Using the Eqn (SN3.14), the infection probabilities
 102 for $n > 1$ number of transmitted virus can be computed. The figure shows a linear relation between the logarithm of
 103 the infection probability and the logarithm of the number of transmitted viruses until ≈ 20 viruses are transmitted.
 104 Thereafter the infection probability saturates and approaches 1.

105 Fig.SN4.2 (right) shows the dependence of the infection probability *per coitus* on the viral load in a donor for
 106 different modes of sexual transmission (blue = homosexual and orange = heterosexual). For both transmission
 107 modes, the logarithm of the infection probability is linearly related to the logarithm of the viral load in donors,
 108 however for the homosexual transmission mode the infection probability approaches the probability of 1 after a
 109 donor viral load of roughly 10^8 copies/ml. Moreover, the infection probability is higher for the homosexual- than
 110 for the heterosexual transmission mode.

References

- 111
- 112 [1] T. C. Quinn, M. J. Wawer, N. Sewankambo, et al. Viral load and heterosexual transmission of human im-
113 munodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med*, 342(13):921–929, Mar 2000.
- 114 [2] James P. Hughes, Jared M. Baeten, Jairam R. Lingappa, et al. Determinants of per-coital-act HIV-1 infectivity
115 among african HIV-1-serodiscordant couples. *J Infect Dis*, 205(3):358–365, Feb 2012.
- 116 [3] David P. Wilson, Matthew G. Law, Andrew E. Grulich, David A. Cooper, and John M. Kaldor. Relation
117 between HIV viral load and infectiousness: a model-based analysis. *Lancet*, 372(9635):314–320, Jul 2008.
- 118 [4] J. W. Mellors, CR Rinaldo, Jr, P. Gupta, et al. Prognosis in HIV-1 infection predicted by the quantity of virus
119 in plasma. *Science*, 272(5265):1167–1170, May 1996.
- 120 [5] S. Duwe, M. Brunn, D. Altmann, et al. Frequency of genotypic and phenotypic drug-resistant HIV-1 among
121 therapy-naive patients of the german seroconverter study. *J Acquir Immune Defic Syndr*, 26(3):266–273, Mar
122 2001.
- 123 [6] K. P. Yousef, K. Meixenberger, M. R. Smith, et al. Inferring HIV-1 transmission dynamics in germany from
124 recently transmitted viruses. *JAIDS, submitted*, 2016.
- 125 [7] R. A. Royce, A. Seña, W Cates, Jr, and M. S. Cohen. Sexual transmission of HIV. *N Engl J Med*,
126 336(15):1072–1078, Apr 1997.
- 127 [8] Kimberly A. Powers, Charles Poole, Audrey E. Pettifor, and Myron S. Cohen. Rethinking the heterosexual
128 infectivity of HIV-1: a systematic review and meta-analysis. *Lancet Infect Dis*, 8(9):553–563, Sep 2008.
- 129 [9] Marie-Claude Boily, Rebecca F Baggaley, Lei Wang, et al. Heterosexual risk of HIV-1 infection per sexual
130 act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis*, 9(2):118–129, Feb
131 2009.
- 132 [10] Max von Kleist, Stephan Menz, and Wilhelm Huisinga. Drug-class specific impact of antivirals on the
133 reproductive capacity of HIV. *PLoS computational biology*, 6(3):e1000720, 2010.