**Introduction.** Interventions against HIV including prophylactics, vaccines, and treatments, are often pre-clinically tested in animal models, frequently the macaque because of the similarity in disease progression between simian immunodeficiency virus and HIV (Koff et al. 2006). The conventional method for infecting these animal models has historically been a single high-dose intravenous or mucosal inoculation, ensuring that the probability of infection is near 1 (Nolen et al. 2015). The accompanying method for assessing the efficacy of the intervention, then, is usually binary: whether or not the infection takes hold (Strbo et al. 2013, Gordon et al. 2016).

Clever experimental frameworks, notably repeated low-dose challenges (RLC), have been used with some success to improve the fidelity of such statistical data (Nolen et al. 2015), but they do not remove the implicit limitations of analyzing binary variables. The basic reproduction number R0 is defined as the expected number of cells that a single infected cell itself will infect (Giorgi et al. 2010). This continuous number can be used to establish whether a particular infecting viral strain will result in systemic infection. To be precise, if a particular founding virus strain has an R0 < 1, it is reproducing below replacement and its progeny will consequently go extinct. In contrast, if R0 > 1, the founding virus’s offspring will likely amplify into systemic infection. The magnitude of R0 provides useful information: if R0 is much larger than 1, the virus is likely to thrive. If R0 is only slightly larger than 1, the infection has a correspondingly smaller chance of survival.

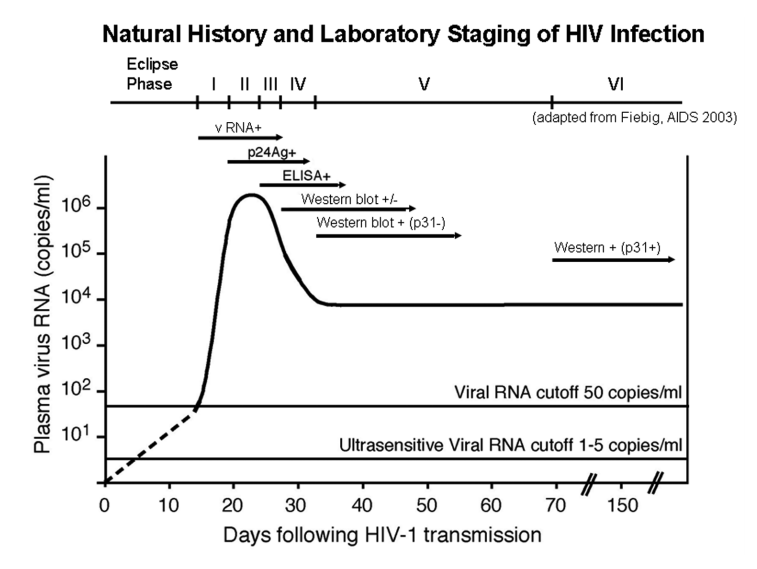
Thus, measuring changes in the initial R0 in response to specific interventions would allow a quantitative measurement of their efficacy in mitigating HIV’s progression in the body, even if infection still occurs. Moreover, the key insight of RLC experiments is that interventions may not be equally effective against high- and low-dose challenges and the initial R0 provides a natural way to quantify the strength of challenge in such methodologies. The importance of developing strong frameworks for testing HIV therapies and especially vaccines cannot be understated: prevalence of first-line anti-retroviral resistance has climbed higher than 50% in sub-Saharan Africa (TenoRes Study Group 2016).

It is important to note that R0 likely varies greatly as HIV progresses in the human body. HIV proliferates classically by so-called cell-free spread, whereby viral particles bud from an infected T cell and enter the extracellular fluid to infect another T cell encountered by chance (Zhang et al. 2015). Recent research has elucidated a second mode of transmission, with two unique mechanisms: cell-to-cell spread that occurs either by transmission of the virus by virological synapses between adjacent T cells or by in trans capture and transfer of virions between spatially close macrophages and dendritic cells (Zhang et al. 2015). Cell-to-cell spread is much more efficient than cell-free spread and is understood to be particularly important in the environment of lymphoid tissues where CD4+ T cells are much more densely packed (Zhang et al. 2015). Therefore, such an environment likely corresponds to a much higher R0. In contrast, the relative paucity of T cells outside the lymphatic and circulatory systems suggests that HIV has much more difficulty establishing itself when introduced outside these environments.

Examining the genetic diversity of the virus at early stages of infection also provides information about the fecundity of the infecting strain. A sharp reduction in viral sequence diversity has been studied extensively in the context of sexual or mother-to-child transmission of HIV (Wolinsky et al. 1992, Delwart et al. 2001, Derdeyn et al. 2004). Correspondingly, it has been estimated that about 80% of all HIV infections arise from a single founding viral sequence (Keele et al. 2008, Haaland et al. 2009, Love et al. 2016).

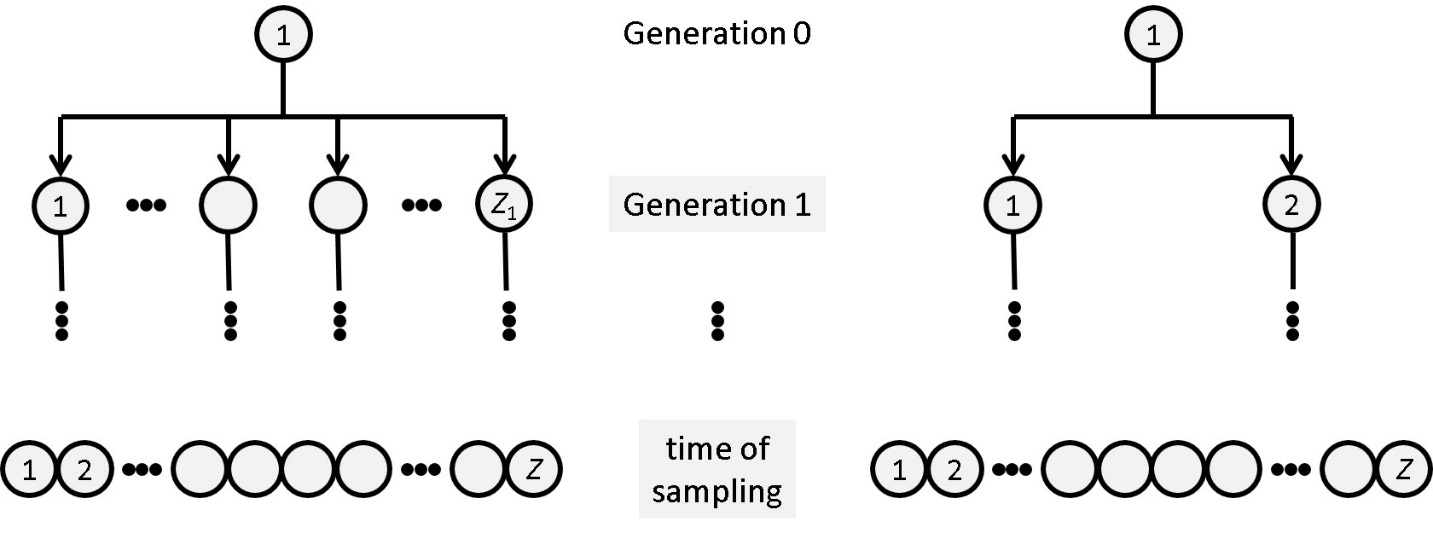
This is further corroborated by the large discrepancy in the number of per-act transmissions per 10,000 exposures between routes of infection. Sexual exposure ranges from less than 4 to 138, needle-sharing produces similar numbers (63). Meanwhile, vertical transmissions between mother and child have a risk 2260 per 10,000 exposures, while blood transfusions have the highest risk at 9250 (Patel et al. 2014).

Together, these facts suggest that the reproductive potential of HIV is starkly limited until it escapes its initial environment into systemic circulation and that the reproductive fate of the earliest invaders determines whether or not systemic infection occurs. Unfortunately, the initial R0 is typically inaccessible due to practical limitations in detecting HIV before systemic infection takes place. After the genetic transmission bottleneck described above, infection is established in local tissues at the exposure site rapidly, but systemic infection is, on average, only detectable in the blood (via viral RNA) a full 10 days after exposure (Kahn and Walker 1998, Fiebig 2003, Cohen et al. 2010). Once the virus reaches the lymphoid tissues, HIV replication reaches a crescendo, with a doubling time of 20 hours (Fiebig 2003). This calls into question the validity of using the current estimates of R0 ≈ 6 (Stafford et al. 2000) or R0 ≈ 8 (Ribeiro et al. 2010) for the entirety of infection, and particularly early in the infection. These estimates were calculated when there were at least 20 viruses/ml (the lower limit of detectability at the time of the studies) or about 105 viruses in the total blood volume of 5L (Kosaka et al. 2017). Clearly, by the time HIV is detectable in blood samples it has long since established itself in the body.



**Figure 1. The stages of acute human immunodeficiency virus infection characterized by detection of viral particles and antibody responses.**

Fortunately, the relatively high mutation rate of retroviruses means that the population history of a particular infection can be reconstructed, at least qualitatively. To illustrate, consider two hypothetical founder viruses, where the first produces some large number of daughters Z1, corresponding to a large initial R0, and the second produces a small number, say 2, of daughters Z2, corresponding to an R0 of order 1. If the phylogenies are allowed to propagate until the total population size reaches some large number N and a sample of M is taken such that Z1 >> M(M-1)/2, the M sampled sequences are likely descended from different daughters of the first virus and are unlikely to share any mutations away from the founder sequence. If a mutation occurs in either of the 2 children of the second virus, however, it is expected that about half of the sample will share this mutation. This, in principle, demonstrates the utility of examining the distribution of mutations in a given sample of some large population to extract information about the early R0 of the virus.



**Figure 1: Two viral ancestries, illustrating the effect of the initial R0 on samples.**

Typical studies sampling the HIV gp120 gene sequence from human patients have so far taken samples ranging from 16 to 30 per patient (Lee et al. 2009). The advent of modern sequencing techniques, particularly deep sequencing, has allowed for the practical measurement of sequence diversity in HIV-infected individuals (Salazar-Gonzalez et al. 2008). Sequencing speed and capacity are continuing to grow exponentially and the cost to sequence millions of base-pairs is dropping rapidly (McElroy et al. 2014). This presents a unique opportunity to use sequence diversity data to reconstruct the initial R0.

It has been demonstrated that the so-called site frequency spectrum (SFS) – a vector η where each component η\_m counts the number of sites where a mutation appeared in exactly m of the M sampled sequences – can be generated as a function of R0 (Spouge 2019). The SFS itself is practically inaccessible because the founder sequence is typically unknown; however, the consensus sequence of sampled viruses can be used as a proxy. By counting the number of minority letters (i.e. letters differing from the consensus), a folded site frequency spectrum (η’m = η\_m + η\_M-m) can be generated which counts the number of alignment columns where the number of minority letters is m.

This folded site-frequency spectrum can be used to calculate a likelihood function L(R0; **η**), which can then be maximized and inverted for the initial R0. Alternatively, population moments (i.e. the expected values of powers of the site frequency spectrum) can be expressed as functions of the initial R0. By setting these functions equal to the sample moments, the initial R0 can be recovered. This paper seeks to show that these so-called method of moments and maximum likelihood estimations of the initial R0 as functions of the folded site frequency spectrum are possible and effective in regimes where the initial R0 is small and there is only one founding virus strain, both reasonable assumptions given the analysis above.

**Methods.** Before a direct accounting of the algorithms responsible for generating a viral ancestry, it is useful to elucidate the biologically relevant parameter ranges. The gene itself is about 2550 nt long and each replication averages

ε ≈ 2.16 x 10-5 point mutations/base/replication (Giorgi et al. 2010). Simple multiplication yields µ ≈ 0.0551 mutations per replication in gp120. Estimates of HIV replication time range from 1.76 to 4.2 days (Love et al. 2016). (Talk about parameters of Poisson and gamma distributions here?)

Generate an array of R0 values to test. For each value of R0, perform 1000 realizations. Each realization consists of generating a random seed, letting a population reproduce from a single founder to 6000 live (un-burst) viruses, and calculating method of moments and maximum likelihood estimations of R0 for each individual realization and the mean estimated error of the R0 estimate. These estimates are pushed onto an array. Once 1000 realizations of a particular R0 have been calculated, the sample standard deviation in the estimates of R0 is calculated.

The population reproduces as follows: first, a single founder virus is created and added to a new list which will contain unburst viruses. While the size of this list is below the 6000 threshold, the list is first sorted by age and a random number, according a Poisson distribution, of daughters is added to the oldest virus in the list, with each daughter assigned a burstday randomly according to a Gamma distribution. These daughters are then added to the list and the parent virus is removed. If, at any point, the list becomes empty – i.e. the founder virus’s lineage goes extinct – a new random seed is generated and the process begins anew.

The calculations proceed as follows: first, M samples are taken from the live list and Am (m = 1,2,...,M) is calculated by counting the non-founder ancestors that have m descendants in the sample. ηm (m = 1,2,...,M) is calculated by sampling Poisson distributions around mu = 0.0551 multiplied by each Am. η’m (m = 1,2,...,M\_tilde) is then calculated by folding ηm. From η’m. R0 is estimated by maximum likelihood estimation, with and without an Euler-Maclaurin approximation, and the method of moments. A grid search initially bounds the search space and a golden-section search refines the estimate of R0. Standard deviations are calculated from these estimated R0’s and pushed onto the same output list.

Once the entire simulation is complete, these lists are outputted directly into a “raw\_data” CSV file. Means of this list are then taken and sample standard deviations are calculated. These results are then pushed onto another list and outputted into a “means” CSV file.

**Results.**

**Discussion.**

**References.**

Cohen, Myron S., Cynthia L. Gay, Michael P. Busch, and Frederick M. Hecht. “The Detection of Acute HIV Infection.” *The Journal of Infectious Diseases* 202 Suppl 2 (October 15, 2010): S270-277. <https://doi.org/10.1086/655651>.

Delwart, Eric, Magdalena Magierowska, Maya Royz, Brian Foley, Lorraine Peddada, Richard Smith, Charles Heldebrant, Andrew Conrad, and Michael Busch. “Homogeneous Quasispecies in 16 out of 17 Individuals during Very Early HIV-1 Primary Infection.” *AIDS* 16, no. 2 (January 25, 2002): 189–95.

Derdeyn, Cynthia A., Julie M. Decker, Frederic Bibollet-Ruche, John L. Mokili, Mark Muldoon, Scott A. Denham, Marintha L. Heil, et al. “Envelope-Constrained Neutralization-Sensitive HIV-1 After Heterosexual Transmission.” *Science* 303, no. 5666 (March 26, 2004): 2019–22. <https://doi.org/10.1126/science.1093137>.

Fiebig, Eberhard W., David J. Wright, Bhupat D. Rawal, Patricia E. Garrett, Richard T. Schumacher, Lorraine Peddada, Charles Heldebrant, et al. “Dynamics of HIV Viremia and Antibody Seroconversion in Plasma Donors: Implications for Diagnosis and Staging of Primary HIV Infection.” *AIDS (London, England)* 17, no. 13 (September 5, 2003): 1871–79. <https://doi.org/10.1097/01.aids.0000076308.76477.b8>.

Haaland, Richard E., Paulina A. Hawkins, Jesus Salazar-Gonzalez, Amber Johnson, Amanda Tichacek, Etienne Karita, Olivier Manigart, et al. “Inflammatory Genital Infections Mitigate a Severe Genetic Bottleneck in Heterosexual Transmission of Subtype A and C HIV-1.” *PLoS Pathogens* 5, no. 1 (January 23, 2009). <https://doi.org/10.1371/journal.ppat.1000274>.

Keele, Brandon F., Elena E. Giorgi, Jesus F. Salazar-Gonzalez, Julie M. Decker, Kimmy T. Pham, Maria G. Salazar, Chuanxi Sun, et al. “Identification and Characterization of Transmitted and Early Founder Virus Envelopes in Primary HIV-1 Infection.” *PNAS* 105, no. 21 (May 27, 2008): 7552–57. <https://doi.org/10.1073/pnas.0802203105>.

Koff, Wayne C., Philip R. Johnson, David I. Watkins, Dennis R. Burton, Jeffrey D. Lifson, Kim J. Hasenkrug, Adrian B. McDermott, et al. “HIV Vaccine Design: Insights from Live Attenuated SIV Vaccines.” *Nature Immunology* 7, no. 1 (January 2006): 19–23. <https://doi.org/10.1038/ni1296>.

Kosaka, Priscila M., Valerio Pini, Montserrat Calleja, and Javier Tamayo. “Ultrasensitive Detection of HIV-1 P24 Antigen by a Hybrid Nanomechanical-Optoplasmonic Platform with Potential for Detecting HIV-1 at First Week after Infection.” *PLoS ONE* 12, no. 2 (February 15, 2017). <https://doi.org/10.1371/journal.pone.0171899>.

Lee, Ha Youn, Elena E. Giorgi, Brandon F. Keele, Brian Gaschen, Gayathri S. Athreya, Jesus F. Salazar-Gonzalez, Kimmy T. Pham, et al. “Modeling Sequence Evolution in Acute HIV-1 Infection.” *Journal of Theoretical Biology* 261, no. 2 (November 21, 2009): 341–60. <https://doi.org/10.1016/j.jtbi.2009.07.038>.

Love, Tanzy M. T., Sung Yong Park, Elena E. Giorgi, Wendy J. Mack, Alan S. Perelson, and Ha Youn Lee. “SPMM: Estimating Infection Duration of Multivariant HIV-1 Infections.” *Bioinformatics* 32, no. 9 (May 1, 2016): 1308–15. <https://doi.org/10.1093/bioinformatics/btv749>.

McDermott, Adrian B., Jacque Mitchen, Shari Piaskowski, Ivna De Souza, Levi J. Yant, Jason Stephany, Jessica Furlott, and David I. Watkins. “Repeated Low-Dose Mucosal Simian Immunodeficiency Virus SIVmac239 Challenge Results in the Same Viral and Immunological Kinetics as High-Dose Challenge: A Model for the Evaluation of Vaccine Efficacy in Nonhuman Primates.” *Journal of Virology* 78, no. 6 (March 15, 2004): 3140–44. <https://doi.org/10.1128/JVI.78.6.3140-3144.2004>.

McElroy, Kerensa, Torsten Thomas, and Fabio Luciani. “Deep Sequencing of Evolving Pathogen Populations: Applications, Errors, and Bioinformatic Solutions.” *Microbial Informatics and Experimentation* 4 (January 15, 2014): 1. <https://doi.org/10.1186/2042-5783-4-1>.

Ngina, Purity, Rachel Waema Mbogo, and Livingstone S. Luboobi. “The In Vivo Dynamics of HIV Infection with the Influence of Cytotoxic T Lymphocyte Cells.” *International Scholarly Research Notices* 2017 (2017): 2124789. <https://doi.org/10.1155/2017/2124789>.

Nolen, Tracy L., Michael G. Hudgens, Pranab K. Senb, and Gary G. Koch. “Analysis of Repeated Low-Dose Challenge Studies.” *Statistics in Medicine* 34, no. 12 (May 30, 2015): 1981–92. <https://doi.org/10.1002/sim.6462>.

Patel, Pragna, Craig B. Borkowf, John T. Brooks, Arielle Lasry, Amy Lansky, and Jonathan Mermin. “Estimating Per-Act HIV Transmission Risk: A Systematic Review.” *AIDS* 28, no. 10 (June 19, 2014): 1509–19. <https://doi.org/10.1097/QAD.0000000000000298>.

Ribeiro, Ruy M., Li Qin, Leslie L. Chavez, Dongfeng Li, Steven G. Self, and Alan S. Perelson. “Estimation of the Initial Viral Growth Rate and Basic Reproductive Number during Acute HIV-1 Infection.” *Journal of Virology* 84, no. 12 (June 2010): 6096–6102. <https://doi.org/10.1128/JVI.00127-10>.

Salazar-Gonzalez, Jesus F., Elizabeth Bailes, Kimmy T. Pham, Maria G. Salazar, M. Brad Guffey, Brandon F. Keele, Cynthia A. Derdeyn, et al. “Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing.” *Journal of Virology* 82, no. 8 (April 2008): 3952–70. <https://doi.org/10.1128/JVI.02660-07>.

Stafford, M. A., L. Corey, Y. Cao, E. S. Daar, D. D. Ho, and A. S. Perelson. “Modeling Plasma Virus Concentration during Primary HIV Infection.” *Journal of Theoretical Biology* 203, no. 3 (April 7, 2000): 285–301. <https://doi.org/10.1006/jtbi.2000.1076>.

TenoRes Study Group. “Global Epidemiology of Drug Resistance after Failure of WHO Recommended First-Line Regimens for Adult HIV-1 Infection: A Multicentre Retrospective Cohort Study.” *The Lancet. Infectious Diseases* 16, no. 5 (2016): 565–75. <https://doi.org/10.1016/S1473-3099(15)00536-8>.

Wolinsky, S. M., C. M. Wike, B. T. Korber, C. Hutto, W. P. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. L. Munoz. “Selective Transmission of Human Immunodeficiency Virus Type-1 Variants from Mothers to Infants.” *Science* 255, no. 5048 (February 28, 1992): 1134–37. <https://doi.org/10.1126/science.1546316>.

Zhang, Changwang, Shi Zhou, Elisabetta Groppelli, Pierre Pellegrino, Ian Williams, Persephone Borrow, Benjamin M. Chain, and Clare Jolly. “Hybrid Spreading Mechanisms and T Cell Activation Shape the Dynamics of HIV-1 Infection.” *PLoS Computational Biology* 11, no. 4 (April 2, 2015). <https://doi.org/10.1371/journal.pcbi.1004179>.