Supplemental Information

The Effects of Statistical Multiplicity of Infection on Virus Quantification and Infectivity Assays

Bhaven A. Mistry, Maria R. D'Orsogna, and Tom Chou

Supplementary Information

A Mathematical Appendices

SMOI Probability

To derive Eq. 2, we index all cells with $i \in \{1, \cdots, M\}$ and define A_i^r as the event that cell i is infected by exactly r IUs. Then, given N IUs across all M cells, the probability of A_i^r is given by

$$\Pr(A_i^r | M, N) = \binom{N}{r} \left(\frac{1}{M}\right)^r \left(1 - \frac{1}{M}\right)^{N-r}. \tag{S1}$$

Since cell sizes are assumed to be homogeneous, the probability in Eq. S1 is the same for all cells, but the events $\{A_1^r, \cdots, A_M^r\}$ are not independent as the number of IUs N shared among the M cells is finite. Thus, we use the inclusion-exclusion principle [1] to derive

$$
\begin{aligned}
\Pr(M_r = m_r | M, N) &= \sum_{j=m_r}^{M} (-1)^{j-m_r} \binom{j}{m_r} \sum_{\substack{I \subset \{1,\cdots,M\} \\ |I|=j}} \Pr\left(\bigcap_{i \in I} A_i^r\right) \\
&= \sum_{j=m_r}^{M} (-1)^{j-m_r} \binom{j}{m_r} \binom{M}{j} \Pr\left(\bigcap_{i=1}^{j} A_i^r\right) \\
&= \sum_{j=m_r}^{M} (-1)^{j-m_r} \binom{j}{m_r} \binom{M}{j} \binom{N}{r, \cdots, r, (N-rj)} \left[\prod_{i=1}^{j}\left(\frac{1}{M}\right)^r\right] \left(\frac{M-j}{M}\right)^{N-rj} \\
&= \sum_{j=m_r}^{M} \binom{j}{m_r} \binom{M}{j} \binom{N}{r, \cdots, r, (N-rj)} \frac{(-1)^{j-m_r} (M-j)^{N-rj}}{M^N}.
\end{aligned} \tag{S2}
$$

Note that the inner summation in the first identity above is over every possible collection of cells of size j, but as each cell is identical, the sum can be reduced to a single joint probability with the binomial degeneracy $\binom{M}{j}$.

Expected Value and Variance

For the generalized c-th moment $\mathrm{E}[M_r^c]$ of the number of cells M_r infected by exactly r viruses, we start with Eq. 2 to obtain

$$
\begin{aligned}
\mathrm{E}[M_r^c] &= \sum_{m_r=0}^{M} \sum_{j=m_r}^{M} m_r^c (-1)^{j-m_r} \binom{j}{m_r} \binom{M}{j} \left(\frac{N!}{(r!)^j (N-rj)!}\right) \frac{(M-j)^{N-rj}}{M^N} \\
&= \sum_{j=0}^{M} \left[\sum_{m_r=0}^{j} m_r^c (-1)^{j-m_r} \binom{j}{m_r}\right] \binom{M}{j} \left(\frac{N!}{(r!)^j (N-rj)!}\right) \frac{(M-j)^{N-rj}}{M^N}.
\end{aligned} \tag{S3}
$$

To aid our derivation, we define the function $u(j,c)$ as

$$
\begin{aligned}
u(j,c) &= \sum_{m=0}^{j} m^c (-1)^{j-m} \binom{j}{m} \\
&= j \sum_{k=0}^{j-1} (k+1)^{c-1} (-1)^{j-1-k} \binom{j-1}{k} \\
&= j \sum_{i=0}^{c-1} \binom{c-1}{i} \sum_{k=0}^{j-1} k^i (-1)^{j-1-k} \binom{j-1}{k} \\
&= j \sum_{i=0}^{c-1} \binom{c-1}{i} u(j-1, i).
\end{aligned} \tag{S4}
$$

This is a recursive relationship from which we can evaluate any $u(j, c)$ using all $u(j-1, i)$ such that $0 \leq i < c$. We evaluate the first three cases $u(j, 0) = \delta_{0,j}$, $u(j, 1) = \delta_{1,j}$, and $u(j, 2) = \delta_{1,j} + 2\delta_{2,j}$, where $\delta_{0,j}$ is the Kronecker delta operator that returns the value 1 when the two subscript arguments are equal and 0 otherwise. We use the result for $c = 1$ and Eq. S3 to calculate the expected value of M_r as

$$
\begin{aligned}
\mathrm{E}\left[M_r\right] &= \sum_{j=0}^{M} \delta_{1,j} \binom{M}{j} \left(\frac{N!}{(r!)^j (N-rj)!}\right) \frac{(M-j)^{N-rj}}{M^N} \\
&= M \binom{N}{r} \left(\frac{1}{M}\right)^r \left(1-\frac{1}{M}\right)^{N-r}.
\end{aligned}
\tag{S5}
$$

We obtain the second moment $\mathrm{E}\left[M_r^2\right]$ using the same method in order to obtain the variance of M_r as

$$
\begin{aligned}
\mathrm{Var}\left[M_r\right] &= \mathrm{E}\left[M_r^2\right] - \mathrm{E}\left[M_r\right]^2 \\
&= M \binom{N}{r} \left(\frac{1}{M}\right)^r \left(1-\frac{1}{M}\right)^{N-r} + \frac{M(M-1)N!(M-2)^{N-2r}}{(r!)^2(N-2r)!M^N} - \frac{M^2(N!)^2(M-1)^{2N-2r}}{(r!)^2\left[(N-r)!\right]^2 M^{2N}}.
\end{aligned}
\tag{S6}
$$

Asymptotic Approximation

For the derivation of Eq. 6, we take the mathematical limit $N, M \to \infty$ while keeping the ratio $\mu = \frac{N}{M}$ fixed and approximate Eq. 2 as follows:

$$
\begin{aligned}
\Pr(M_r = m_r | M, N) &= \sum_{j=m_r}^{M} \frac{j!M!N!(-1)^{j-m_r}(M-j)^{N-rj}}{m_r!(j-m_r)!j!(M-j)!(N-rj)!(r!)^j M^{N-rj} M^{rj}} \\
&= \frac{1}{m_r!} \sum_{j=m_r}^{M} \frac{(-1)^{j-m_r}}{(j-m_r)!(r!)^j} [M \cdots (M-j+1)] \frac{[N \cdots (N-rj+1)]}{M^{rj}} \left(1-\frac{j}{M}\right)^{N-rj} \\
&\approx \frac{1}{m_r!} \sum_{j=m_r}^{M} \frac{(-1)^{j-m_r}}{(j-m_r)!(r!)^j} M^j \mu^{jr} e^{-\mu j} \\
&\approx \frac{1}{m_r!} \left[\frac{M\mu^r e^{-\mu}}{r!}\right]^{m_r} \exp\left[-\frac{M\mu^r e^{-\mu}}{r!}\right].
\end{aligned}
\tag{S7}
$$

Note that, although the first approximation requires j in the summation to be sufficiently smaller than M, any contribution from the summation for j close to M vanishes due to both the $(j-m_r)!$ term in the denominator and the $\left(1-\frac{j}{M}\right)^{N-rj}$ term approaching 0. Under the same large M, N limit, we can derive an asymptotic approximation of the joint probability distribution by taking the natural log of both sides of Eq. 5:

$$
\begin{aligned}
\ln \Pr(M_0 = m_0, \cdots, M_N = m_N) &= \ln\left(\frac{1}{M^N}\right) + \ln M! + \ln N! + \sum_{r=0}^{N} \ln\left(\frac{1}{m_r!(r!)^{m_r}}\right) \\
&\approx -N \ln M + M \ln(M) - M + N \ln(N) - N + \sum_{r=0}^{N} \ln\left(\frac{1}{m_r!(r!)^{m_r}}\right) \\
&= \ln \mu \left(\sum_{r=0}^{N} r m_r\right) + (\ln M - \mu)\left(\sum_{r=0}^{N} m_r\right) - M e^{-\mu}\left(\sum_{r=0}^{\infty} \frac{\mu^r}{r!}\right) \\
&\quad + \sum_{r=0}^{N} \ln\left(\frac{1}{m_r!(r!)^{m_r}}\right) \\
&= \sum_{r=0}^{N} \ln\left[\frac{\mu^{rm_r} M^{m_r} e^{-m_r \mu}}{m_r!(r!)^{m_r}} \exp\left(-\frac{M e^{-\mu} \mu^r}{r!}\right)\right] - \mathcal{O}\left(\frac{M\mu^N}{N!}\right) \\
&\approx \ln\left[\prod_{r=0}^{N} \frac{1}{m_r!}\left[\frac{M\mu^r e^{-\mu}}{r!}\right]^{m_r} \exp\left(-\frac{M\mu^r e^{-\mu}}{r!}\right)\right].
\end{aligned}
\tag{S8}
$$

Since the argument in the right-hand-side of the last approximation is the same as Eq. 6, we arrive at the result in Eq. 8.

Number of Infected Cells

To derive Eq. 9, we first define N_d as the number of virus particles present in the viral solution after dilution by a factor of D^d. Obtaining N_d is effectively analogous to taking a volume of the initial viral stock scaled by D^{-d} and counting the number of particles captured in the volume. Thus, we expect N_d to be Poisson-distributed with mean $N_0 D^{-d}$ and discrete probability density function given by

$$\Pr\left(N_d = n_d \mid N_0\right) = \frac{1}{n_d!}\left(\frac{N_0}{D^d}\right)^{n_d}\exp\left(-\frac{N_0}{D^d}\right). \tag{S9}$$

Once N_d is chosen from the above distribution, for a given "particle to PFU ratio" Q, the number of IUs N follows a binomial distribution with a probability function similar to Eq. 1, but with N_0 replaced with N_d. Note that, given an SMOI $\{M_0, \cdots, M_N\}$, it is immediate that $M^* = M - M_0$. Using this modified density of N and Eqs. 2 and S9, we can derive the discrete probability density function of M^* at a given dilution number d as

$$
\begin{aligned}
\Pr\left(M^* = m\right) &= \sum_{n_d=0}^{N_0}\sum_{n=0}^{n_d}\Pr(N=n|N_d=n_d)\Pr(M_0=M-m|N=n)\Pr(N_d=n_d) \\
&= \sum_{j=M-m}^{M}(-1)^{j-M+m}\binom{j}{M-m}\binom{M}{j}e^{-\frac{N_0}{D^d}}\sum_{n_d=0}^{N_0}\frac{\left(\frac{N_0}{D^d}\right)^{n_d}}{n_d!}\left[1-Q^{-1}+Q^{-1}\left(1-\frac{j}{M}\right)\right]^{n_d} \\
&\approx \sum_{j=M-m}^{M}(-1)^{j-M+m}\binom{j}{M-m}\binom{M}{j}\exp\left[\frac{N_0}{D^d}\left(1-\frac{j}{QM}\right)-\frac{N_0}{D^d}\right] \\
&= \binom{M}{m}\left[1-\exp\left(-\frac{N_0}{QMD^d}\right)\right]^m\exp\left(-\frac{N_0}{QMD^d}\right)^{M-m}.
\end{aligned}\tag{S10}
$$

Note that the approximation that closes the exponential term in the final result employs the assumption that N_0 is sufficiently large.

B Inhomogeneous Cell Size

We derived the probability distribution in Eq. 2 assuming the plated host cells are of identical size and volume. This may not necessarily be the case as each cell exists at different stages of the mitotic cycle, will attach to the plate bottom at random locations, and contains deformities in shape and size. Assuming cells cover the entire surface of the well bottom, Pineda et al. [2] showed that the cell size proportion p_i for cell i is gamma distributed with probability density

$$f(p_i) = \frac{M^\nu \nu^\nu p_i^{\nu-1}\exp(-\nu M p_i)}{\Gamma(\nu)}, \tag{S11}$$

where ν is a parameter that can be estimated, for example, by fitting imaging data of cells. Under a specific realization of cell size distributions $\{p_1, \cdots, p_M\}$, we define A_i^r as the event that cell i is infected by exactly r viruses with probability

$$\Pr(A_i^r) = \binom{N}{r}p_i^r(1-p_i)^{N-r}. \tag{S12}$$

Using the inclusion-exclusion principle as above, we derive the conditional probability distribution of the number of cells M_r that were infected by exactly r viruses as

$$
\begin{aligned}
\Pr(M_r = m_r | p_1, \cdots, p_M) &= \sum_{j=m_r}^{M} (-1)^{j-m_r} \binom{j}{m_r} \sum_{|\{i_w\}|=j} \Pr\left(\bigcap_{w=1}^{j} A_{i_w}^r \right) \\
&= \sum_{j=m_r}^{M} (-1)^{j-m_r} \binom{j}{m_r} \sum_{|\{i_w\}|=j} \binom{N}{r, \cdots, r, (N-rj)} p_{i_1}^r \cdots p_{i_j}^r \left(1 - \sum_{w=1}^{j} p_{i_w} \right)^{N-rj} \\
&= \sum_{j=m_r}^{M} (-1)^{j-m_r} \binom{j}{m_r} \sum_{|\{i_w\}|=j} \frac{N!}{(r!)^j (N-rj)!} \left(\prod_{w=1}^{j} p_{i_w} \right)^r \left(1 - \sum_{w=1}^{j} p_{i_w} \right)^{N-rj}. \quad \text{(S13)}
\end{aligned}
$$

In order to obtain the full probability, we first take note that each cell size proportion p_i is dependent on each other as they are constrained by $\sum_i^M p_i = 1$. We avoid this dependency by noticing the expression in Eq. S11 approaches zero very rapidly as p_i moves away from the expected value $1/M$. If we define a sufficiently large proportion $\hat{p}$ such that the interval $[0, \hat{p}]$ contains the majority of the area under the probability density in Eq. S11, we can make the approximation

$$
\begin{aligned}
\Pr(M_r = m_r) &= \int_0^1 \cdots \int_0^1 \Pr(M_r = m_r | p_1, \cdots, p_M) f(p_1, \cdots, p_M) \mathrm{d}p_1 \cdots \mathrm{d}p_M \\
&\approx \left[\frac{M^\nu \nu^\nu e^{-\nu}}{\Gamma(\nu)} \right]^M \int_0^{\hat{p}} \cdots \int_0^{\hat{p}} \Pr(M_r = m_r | p_1, \cdots, p_M) \left(\prod_{w=1}^{M} p_w \right)^{\nu-1} \mathrm{d}p_1 \cdots \mathrm{d}p_M. \quad \text{(S14)}
\end{aligned}
$$

It is clear that introducing cell size inhomogeneity dramatically increases the complexity of our probabilistic SMOI model. For relatively small numbers of cells M, image processing can be used to determine an estimation of a particular realization of cell size distribution $\{p_1, \cdots, p_M\}$ for a given experiment and factored into Eq. S13. Note that once the probability distribution of cell counts $\{M_0, \cdots, M_N\}$ is determined for a given realization of cell sizes $\{p_1, \cdots, p_M\}$, all subsequent analysis and derivations follow the same way as in the homogeneous cell size assumption.

C   Coinfection

As a vector for infection, the primary function of a single virus particle is to deliver its genetic contents into the host cell cytoplasm or nucleus [3–5]. The typical model for viral infection assumes each virus contains all the genetic material required to replicate within a host cell [6, 7]. Certain plant and fungi viruses, however, require two or more particles to successfully replicate within a host cell since each particle contains only part of the complete genome [8]. Similarly, RNA viruses that target animal cells undergo error prone replication, resulting in partially complete genome sequences. These damaged viral genes may encode proteins needed for the host cell to successfully replicate new viruses. In this case, regardless of a successful viral infection, new viruses capable of infecting further host cells will not be produced. Additional viral infections that contain the missing sequence fragments, though, can "rescue" the cell's ability to replicate the virus, a phenomenon known as coinfection. In the context of our definition of SMOI, we now make the distinction between M_r, the number of cells that have been infected by viral genomes from exactly r distinct virus particles, and M_r^*, the number of cells that are fully capable of replicating new functioning viruses upon undergoing r distinct viral infections. It is clear that each $M_r^* \leq M_r$ and their sum $M^* \equiv \sum_{r=1}^{N} M_r^* \leq M - M_0$, so the results in Eqs. 9 and 12 are not sufficient to quantify the total number of virus-producing cells.

In order to model coinfection, we need to consider the genome of the virus species of interest. Specifically, we assume the genome is made up of G distinct genes. For example, many variants of HIV-1 carry a gene sequence containing $G = 9$ genes [3]. In our model, we assume each gene encodes a protein that is essential for replication. Though individual nucleotide changes due to random mutations may result in an amino acid chain that is no longer functioning, some genes may be robust to these changes due to codon degeneracy or the gene's shear length [9]. Thus, we assume each gene $g = 1, \cdots, G$ contained within a viral particle has a probability q_g of losing function. If a cell is infected by exactly r viral genomes, we define B_g^r as the event that gene g is still no longer functional, so that $\Pr(B_g^r) = q_g^r$. To quantify the probability that k genes are no longer

functional in a host cell that has been infected by exactly r viral genomes, we use the inclusion-exclusion principle [1] to derive

$$
\Pr\left(\text{``}k\text{ failed genes given }r\text{ infections''}\right) = \sum_{j=k}^{G}(-1)^{j-k}\binom{j}{k}\sum_{\substack{I\subset\{1,\cdots,G\}\\|I|=j}}\Pr\left(\bigcap_{g\in I}B_g^r\right)
$$

$$
= \sum_{j=k}^{G}(-1)^{j-k}\binom{j}{k}\sum_{\sigma_1=0}^{1}\cdots\sum_{\sigma_G=0}^{1}\mathbb{1}_{\sum_{g=1}^{G}\sigma_g=j}\prod_{g=1}^{G}q_g^{\sigma_g r}, \tag{S15}
$$

where $\mathbb{1}_{\sum_{g=1}^{G}\sigma_g=j}$ is an indicator function that returns zero when the number of nonzero σ_g is not exactly j. The infected cell is only capable of producing viable viruses if none of the genes have failed and is equivalent to setting $k=0$ in Eq. S15. Then we define the probability H_r that a cell infected by exactly r viral genomes will successfully produce new viruses as

$$
H_r = \sum_{j=0}^{G}(-1)^j\sum_{\sigma_1=0}^{1}\cdots\sum_{\sigma_G=0}^{1}\mathbb{1}_{\sum_{g=1}^{G}\sigma_g=j}\prod_{g=1}^{G}q_g^{\sigma_g r}. \tag{S16}
$$

Note that the probability that a cell not infected by any viral genome will produce viruses is $H_0=0$. Then, given an SMOI $\{M_0,\cdots,M_N\}$, the number of cells M_r^* capable of virus replication after being infected by exactly r viral genomes is binomially distributed with parameters M_r and H_r. The probability of M^* cells producing viruses is given by

$$
\Pr\left(M^*=m|M_0,\cdots,M_N,M,N\right) = \sum_{M_1^*,\cdots,M_N^*}\binom{m}{M_1^*,\cdots,M_N^*}\prod_{r=1}^{N}\binom{M_r}{M_r^*}H_r^{M_r^*}(1-H_r)^{M_r-M_r^*}. \tag{S17}
$$

If we let $m=0$ and sum over the density in Eq. 5 for all possible SMOI, given an IU count N, we can derive the probability of observing a cytopathic effect as

$$
\begin{aligned}
\Pr(\text{``Cytopathic effect''}|N) &= 1-\sum_{M_0,\cdots,M_N}\frac{1}{M^N}\binom{M}{M_0,\cdots,M_N}\binom{N}{0,\cdots,0,1,\cdots,1,\cdots,N,\cdots,N}\prod_{r=1}^{N}(1-H_r)^{M_r}\\
&= 1-\frac{M!N!}{M^N}\prod_{r=0}^{N}\sum_{M_r=0}^{M}\frac{(1-H_r)^{M_r}}{M_r!(r!)^{M_r}}\\
&\approx 1-\frac{M!N!}{M^N}\exp\left[\sum_{r=0}^{N}\frac{1-H_r}{r!}\right],
\end{aligned} \tag{S18}
$$

where the approximation is due to the assumption that the number of cells M is large. For intermediate values of N, computing the summation in the exponential is numerically viable, assuming the probabilities of gene failure $q_1,\cdots,q_G$ are known. Though this expression may be used in place of Eq. 12 to analyze some virus quantification assays, for large values of N, numerically evaluating H_r becomes computationally expensive.

D Viral Interference

To infect healthy cells, all species of viruses must undergo a series of events including cell attachment, entry via membrane fusion or endocytosis, and intracellular transport. Retroviruses, such as HIV-1, must also undergo reverse transcription, nuclear pore transport, and DNA integration in order to use the host cell's transcription machinery to produce viral protein. In the models developed in this paper, the probabilities of success for each of these processes was assumed to be subsumed into the *a priori* estimated particle to PFU ratio Q. However, for certain retroviruses, it has been observed that after an initial infection, subsequent infections from the same virus species become less likely [10, 11]. This phenomenon, known as viral interference, is often due to the host producing new viral proteins after a refractory period that can inhibit one or more of the intracellular processes leading to integration of subsequent viral infections. To include this dynamic into our models, we first decouple the probabilities of integration from Q and define N as the number of viruses that have successfully completed viral entry into the host cytoplasm, but before all intracellular processes that lead to integration. Note that all of our results concerning the statistical multiplicity of infection (SMOI) still hold and we make the distinction between the number M_r of cells infected by r of the

N infectious units and the number M_s^* of cells with exactly s integrations. Furthermore, some species of virus can contain multiple copies of their genome, such as HIV-1 which contains two copies per particle [3]. Let a be the number of genomes contained in a single virus particle to be integrated into the host cell. Then the maximum number of possible integrations for a cell from M_r is ra. Let p_s be the probability of a viral genome integrating into the host DNA given that $s-1$ integrations have already occurred. Define $H_{r,s}$ as the probability a cell contains s successful integrations given that it was infected by exactly r distinct virus particles and is given by

$$H_{r,s} = \begin{cases} p_1 p_2 \cdots p_s (1-p_{s+1})^{ra-s} & 0 \leq s \leq ra \\ 0 & s > ra. \end{cases} \tag{S19}$$

If we define $M_{r,s}^*$ as the number of cells with s integrations after infection by exactly r virus particles, then given an SMOI $\{M_0, \cdots, M_N\}$ and N, we can derive the probability function

$$\Pr(M_{r,s}^* = m | M_0, \cdots, M_N, N) = \binom{M_r}{m} H_{r,s}^m (1 - H_{r,s})^{M_r - m}. \tag{S20}$$

Noting that $M_s^* = \sum_{r=0}^N M_{r,s}^*$ is the number of cells with exactly s integrations, we can use Eqs. 6 and S20 to derive the expected value as

$$\begin{aligned} \mathrm{E}\left[M_s^*|N\right] &= \sum_{r=0}^N \mathrm{E}\left[M_{r,s}^*|N\right] \\ &= \sum_{r=0}^N H_{r,s} \mathrm{E}\left[M_r|N\right] \\ &= M e^{-\mu} \sum_{r=0}^N \frac{H_{r,s} \mu^r}{r!}, \end{aligned} \tag{S21}$$

where $\mu = \frac{N}{M}$. Note that if we are concerned with the total number $M^* = M - M_0^*$ of cells with at least one integration, as is the case for the probability distributions derived for assays employing serial dilution, the issue of viral interference is negligible, allowing us to subsume the probability of the first integration into the particle to PFU ratio Q as before and leave all subsequent virus quantification analysis unchanged from the results in **Plaque Assay** and **Endpoint Dilution Assay**. However, for assays that attempts to quantify the total number of integrations, such as the luciferase reporter assay, the expectation in Eq. S21 can be used, assuming the probabilities $p_1, \cdots, p_N$ have *a priori* been estimated.

E   Sensitivity Analysis

The probability models derived in **Probabilistic Models of Statistical Multiplicity of Infection** allow us to construct the likelihood functions for the plaque, endpoint dilution, and luciferase reporter assays in Eqs. 14, 20, and 27 for the primary purpose of inferring unknown parameters such as N_0 and μ. The utility of these functions can be extended to performing sensitivity analysis on these maximum likelihood estimates (MLE) and optimizing experimental design. This requires constructing a Fisher Information Matrix (FIM), a quantitative measure of the information one can extract from a likelihood function with an arbitrary set of data [12, 13]. The FIM, which we will denote as J, is constructed by computing the gradient of the log of the likelihood function with respect to the parameters being inferred. For example, for the plaque assay and potentially inferred parameters N_0, Q, and M, J is given by

$$J = \mathrm{E}\left[(\nabla \ln \mathcal{L})(\nabla \ln \mathcal{L})^{\mathrm{T}} \right] = \begin{bmatrix} J_{N_0, N_0} & J_{N_0, Q} & J_{N_0, M} \\ J_{Q, N_0} & J_{Q, Q} & J_{Q, M} \\ J_{M, N_0} & J_{M, Q} & J_{M, M} \end{bmatrix}, \tag{S22}$$

where we derive

$$J_{N_0,N_0} = \mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial N_0}\right)^2\right] = \sum_{d=d_c}^{d_{\max}} \frac{T \exp\left(-\frac{N_0}{QMD^d}\right)}{Q^2 M D^{2d}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}, \tag{S23}$$

$$J_{Q,Q} = \mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial Q}\right)^2\right] = \sum_{d=d_c}^{d_{\max}} \frac{T N_0^2 \exp\left(-\frac{N_0}{QMD^d}\right)}{Q^4 M D^{2d}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}, \tag{S24}$$

$$J_{M,M} = \mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial M}\right)^2\right] = \sum_{d=d_c}^{d_{\max}} \frac{T N_0 \exp\left(-\frac{N_0}{QMD^d}\right)}{Q M^2 D^d\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}, \tag{S25}$$

$$J_{N_0,Q} = J_{Q,N_0} = \mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial N_0}\right)\left(\frac{\partial \ln \mathcal{L}}{\partial Q}\right)\right] = -\sum_{d=d_c}^{d_{\max}} \frac{T N_0 \exp\left(-\frac{N_0}{QMD^d}\right)}{Q^3 M D^{2d}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}, \tag{S26}$$

$$J_{N_0,M} = J_{M,N_0} = \mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial N_0}\right)\left(\frac{\partial \ln \mathcal{L}}{\partial M}\right)\right] = -\sum_{d=d_c}^{d_{\max}} \frac{T N_0 \exp\left(-\frac{N_0}{QMD^d}\right)}{Q^2 M^2 D^{2d}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}, \tag{S27}$$

$$J_{Q,M} = J_{M,Q} = \mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial Q}\right)\left(\frac{\partial \ln \mathcal{L}}{\partial M}\right)\right] = \sum_{d=d_c}^{d_{\max}} \frac{T N_0^2 \exp\left(-\frac{N_0}{QMD^d}\right)}{Q^3 M^3 D^{2d}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}. \tag{S28}$$

In particular, the elements of the main diagonal of J, known as Fisher Information Numbers, are interpreted as the "precision" of each MLE and can inform an experimentalist of the potential variation in their inferred parameter with respect to data defined by the likelihood function. Comparing the main diagonal elements can offer insight into experimental design. To illustrate, in the example above, it is immediately apparent that the ratio of $J_{Q,Q}$ to J_{N_0,N_0} is N_0^2/Q^2, where it is understood that N_0 is typically several orders of magnitude higher than Q. This implies that the likelihood function of Eq. 14, and, by extension, the plaque assay itself contains far more information about the parameter Q than N_0. This provides an analytical way to decide which parameter estimation should be the focus of a particular assay.

A more general use for the FIM is to understand the variance of an MLE given an arbitrary set of data. Independent, but identical experiments can produce different estimates for each parameter and, according to the Cramer-Rao inequality, the matrix inverse J^{-1} will provide a theoretical lower bound on the covariance matrix of the parameter estimates [12]. Furthermore, it can be shown that the distribution of MLEs asymptotically approaches a normal distribution centered around the true experimental parameter value with covariance J^{-1} as the amount of data increases [14]. For single point estimation, the FIM reduces to the one Fisher Information Number with which the reciprocal can be used to approximate the variance of a parameter. For example, the plaque assay is typically used to infer only the parameter N_0, so using Eq. S23, we can obtain the asymptotic approximation

$$\mathrm{Var}\left[\hat{N}_0\right] \approx J_{N_0,N_0}^{-1} = \left[\sum_{d=d_c}^{d_{\max}} \frac{T \exp\left(-\frac{N_0}{QMD^d}\right)}{Q^2 M D^{2d}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}\right]^{-1}. \tag{S29}$$

This analytical expression for the variance can be used to determine confidence intervals of the MLE, perform sensitivity analysis of other parameters, and aid in optimal experimental design.

References

1. Lange, K., 2003. Applied Probability. Springer Sci. Bus. Med. Inc.

2. Pineda, E., P. Bruna, and D. Crespo, 2004. Cell Size Distribution in Random Tessellations of Space. *Phys. Rev. E* 70:066119.

3. Wilen, C. B., J. C. Tilton, and R. W. Doms, 2012. HIV: Cell Binding and Entry. *Cold Spring Harb. Persp. Med.* 2.

4. Qian, K., S. L. Morris-Natschke, and K.-H. Lee, 2009. HIV Entry Inhibitors and Their Potential in HIV Therapy. *Med. Res. Rev.* 29:369–393.

5. Boulant, S., M. Stanifer, and P.-Y. Lozach, 2015. Dynamics of Virus-Receptor Interactions in Virus Binding, Signaling, and Endocytosis. *Viruses* 7:2794–2815.

6. Chikere, K., T. Chou, P. R. Gorry, and B. Lee, 2013. Affinofile Profiling: How Efficiency of CD4/CCR5 Usage Impacts the Biological and Pathogenic Phenotype of HIV. *J. Virol.* 435:81–91.

7. Johnston, S. H., M. A. Lobritz, S. Nguyen, K. Lassen, S. Delair, F. Posta, Y. J. Bryson, E. J. Arts, T. Chou, and B. Lee, 2009. A Quantitative Affinity-Profiling System That Reveals Distinct CD4/CCR5 Usage Patterns among Human Immunodeficiency Virus Type 1 and Simian Immunodeficiency Virus Strains. *J. Virol.* 83:11016–11026.

8. Aguilera, E. R., A. K. Erickson, P. R. Jesudhasan, C. M. Robinson, and J. K. Pfeiffer, 2017. Plaques Formed by Mutagenized Viral Populations Have Elevated Coinfection Frequencies. *Am. Soc. Microbiol.* 8:1–12.

9. Stern, A., S. Bianco, M. T. Yeh, C. Wright, K. Butcher, C. Tang, R. Nielsen, and R. Andino, 2014. Costs and Benefits of Mutational Robustness in RNA Viruses. *Cell Reports* 8:1026–1036.

10. Nisole, S., and A. Saïb, 2004. Early Steps of Retrovirus Replicative Cycle. *Retrovirol.* 1:9.

11. Nethe, M., B. Berkhout, and A. C. van der Kuyl, 2005. Retroviral Superinfection Resistance. *Retrovirol.* 2:52.

12. Gunawan, R., Y. Cao, L. Petzold, and F. J. D. III, 2005. Sensitivity Analysis of Discrete Stochastic Systems. *Biophys. J.* 88:2530–2540.

13. Casella, G., and R. L. Berger, 2002. Statistical Inference. Thomson Learning Inc.

14. Sobel, M., 1982. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Socio. Meth.* 13:290–312.