Biophysical Journal
Article

Biophysical Society

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

Bhaven A. Mistry,[1] Maria R. D'Orsogna,[1,2] and Tom Chou[1,3,*]
[1]Department of Biomathematics, University of California, Los Angeles, Los Angeles, California; [2]Department of Mathematics, California State University, Northridge, California; and [3]Department of Mathematics, University of California, Los Angeles, Los Angeles, California

ABSTRACT Many biological assays are employed in virology to quantify parameters of interest. Two such classes of assays, virus quantification assays (VQAs) and infectivity assays (IAs), aim to estimate the number of viruses present in a solution and the ability of a viral strain to successfully infect a host cell, respectively. VQAs operate at extremely dilute concentrations, and results can be subject to stochastic variability in virus-cell interactions. At the other extreme, high viral-particle concentrations are used in IAs, resulting in large numbers of viruses infecting each cell, enough for measurable change in total transcription activity. Furthermore, host cells can be infected at any concentration regime by multiple particles, resulting in a statistical multiplicity of infection and yielding potentially significant variability in the assay signal and parameter estimates. We develop probabilistic models for statistical multiplicity of infection at low and high viral-particle-concentration limits and apply them to the plaque (VQA), endpoint dilution (VQA), and luciferase reporter (IA) assays. A web-based tool implementing our models and analysis is also developed and presented. We test our proposed new methods for inferring experimental parameters from data using numerical simulations and show improvement on existing procedures in all limits.

INTRODUCTION

Understanding viral dynamics is an important task in medicine, epidemiology, public health, and, in particular, for the development of antiviral therapies and vaccines. Drugs that hinder viral infection include blockers of viral entry into the host cell (1–6) and inhibitors of genetic activity and protein assembly inside the cytoplasm and nucleus (7–9). Mechanistic models of drug action have recently emerged as useful tools in helping design ad hoc experiments to study drug efficacy and in interpreting results (10–13). Mathematical models typically assume prior knowledge of given physical quantities pertaining to the virus, host cell, or the biological assay being studied. Once these parameters are assigned, viral and cell population dynamics and their statistical properties can be predicted. Among the different experimental assays, one often seeks to evaluate the number of virus particles in a stock solution or the number of viruses that have successfully infected host cells (6,14–19).

In the case of virus quantification assays (VQAs), performing repeated controlled experiments on viral dynamics or comparing results across multiple studies requires knowing how many viruses are present in the initial stock solution of each experiment (4,5). Furthermore, antigens that induce immune responses against viral infections may be engineered from viral components such as capsid proteins, viral enzymes, and genetic vectors (20) and may be used in the development of vaccines. Being able to determine the exact number of virus-derived antigens helps control the efficacy of vaccines and optimize yield (21–23).

Given the central role of VQAs, several assays have been designed to estimate viral particle counts. These include plaque (24) and endpoint dilution (23,25) assays, which will be discussed in more detail in the remainder of this work. For now, we note that these assays involve repeatedly diluting an initial solution of virus particles in the presence of a layer of plated cells until viral concentrations are low enough that the dynamics of an individual virus can be extrapolated. At these low particle counts, however, the discrete nature of the infection process cannot be neglected and can cause substantial discrepancies when replicating experiments. Average quantities are not necessarily representative, and a more in-depth approach in quantifying virus-cell interactions is necessary.

Infectivity assays (IAs), on the other hand, aim to quantify the number of viruses that have successfully infected host cells under varying antiviral drug environments (14–16). IAs may measure the total transcription activity across all cells, such as the luciferase reporter assay

Submitted November 27, 2017, and accepted for publication May 2, 2018.
*Correspondence: tomchou@ucla.edu
Editor: Ruth Baker.
https://doi.org/10.1016/j.bpj.2018.05.005

(15,26), or may count the number of host cells that were successfully infected, such as the enzyme-linked immunosorbent assay and the immunofluorescence assay with fluorescence-activated cell sorting (4,14,15,18). These assays are performed using undiluted solutions with large numbers of viral particles, reducing stochastic variability. The average number of viruses that infect a cell is estimated as the ratio of the number of viruses in solution to the number of plated cells, a quantity known as the multiplicity of infection (MOI) (19). However, each cell may be infected by different numbers of viruses distributed around the average given by the MOI. In these cases, one may be interested in the complete probability distribution for the number of virus infections in each plated cell and in the related statistical variance.

In this work, we derive a probability model for the distribution of viral infections per host cell, which we call the statistical multiplicity of infection (SMOI). The SMOI can be used as a starting point to help estimate the number of viral particles in solution in VQAs and to determine a viral strain's ability to successfully infect host cells in IAs. In Probabilistic Models of Statistical Multiplicity of Infection, we present the mathematical foundations for the SMOI in the two experimentally relevant parameter regimes of small and large viral particle counts and derive a probability model for the total number of infected cells under any dilution level. In Plaque Assay, we apply our models to the plaque assay and, to our knowledge, formulate a new method of analyzing plaque count data. In Endpoint Dilution Assay, we employ a special case of the derived probability distribution to the endpoint dilution assays and compare our results to those arising from traditional titration techniques such as the Reed and Muench (RM) (27) and Spearman-Karber (SK) methods (28). In Luciferase Reporter Assay, we use the large particle limit of our model to describe the luciferase reporter assay. Lastly, a discussion of our results, a side-by-side comparison with existing methods, and a link to web-based data analysis tools are provided in Conclusions. Mathematical appendices and further discussion of experimental attributes such as cell-size variability, coinfection, viral interference, and optimal experimental design using parameter-sensitivity analysis are presented in the Supporting Materials and Methods.

METHODS

Probabilistic models of statistical multiplicity of infection

A typical viral assay is initiated by laying a monolayer of M cells on the bottom of a microtiter well, as illustrated in Fig. 1 (17,24,25). Although variability exists among experiments, M is often set within the range of 10^4–10^5 (14,26) and is assumed to be a known experimental parameter. A supernatant containing N_0 virus particles in the range of 10^5–10^7 (24–26) is then added to the microtiter well. Although, theoretically, all N_0 particles are capable of infection, not all will successfully infect a cell. Because infection of a host cell requires a complex sequence of

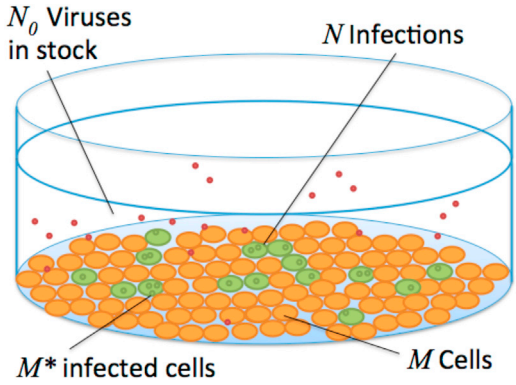


FIGURE 1 A typical assay includes a plate of M host cells inoculated with a solution of N_0 viruses. Each viral particle has some probability of infection, and the total number N of infections are distributed to the M^* infected cells. The probability of infection is roughly estimated with the reciprocal of the a priori measured particle to PFU ratio Q. To see this figure in color, go online.

biochemical processes that may include receptor binding, membrane fusion, reverse transcription, nuclear pore transport, and DNA integration (10,19), virus particles that fail at one or several of these sequential steps lead to abortive infections. To differentiate, the particles that do succeed are called infectious units (IU) or plaque-forming units (PFU). We will denote the number of IUs as $N \leq N_0$. Depending on the strain of virus, the particular experimental protocol used, and specific conditions of the assay, the random quantity N is distributed according to N_0 and the overall effective probability that an arbitrary viral particle successfully infects a host cell. A proxy that is typically used in place of this effective probability is the "particle to PFU ratio" Q, an experimentally determined parameter that quantifies, on average, the minimal number of particles required to ensure at least one infected cell (29,30). Q is often treated as an a priori measured quantity, primarily associated with the particular strain of virus being studied. Low values of Q, such as with poliovirus ($Q = 30$) (30), have a high likelihood of successful infection compared to viruses with large Q, such as human immunodeficiency virus type 1 (HIV-1) ($Q = 10^7$) (31). Thus, the reciprocal Q^{-1} can be interpreted as the probability for a single virus to infect a host cell. Assuming an initial stock of N_0 particles, the discrete probability density function of N is

$$\Pr(N = n \,|\, N_0, Q) = \binom{N_0}{n} \left(Q^{-1}\right)^n \left(1 - Q^{-1}\right)^{N_0 - n}, \quad (1)$$

which defines a binomial distribution with parameters N_0 and Q^{-1}. Although we assume Q to be a priori known, in actuality, the probability of a virus successfully infecting a host is highly dependent on the methods used to harvest the virus stock, the experimental parameters of the assay, the host receptor concentrations and binding rates, and the dynamics of the physiological processes leading to infection (29,32). A thorough investigation into these processes would be necessary to mechanistically model Q and is outside of the scope of this work. However, we will discuss in Conclusions how, with direct measurements of certain other parameters, especially N_0, our derived methods may also be used to infer Q.

We assume each viral particle in solution acts independently of others and that host cell infection attempts are random events. At high ratios N_0/M of particles to cells, a quantity referred to as the MOI, it becomes increasingly probable for more than one IU to infect the same host cell. We define M_0 as the count of cells not infected by any IU, M_1 as the count of cells infected by exactly one IU, and so on, up to M_N, the number of cells infected by all N IUs. The SMOI is defined as the ensemble of cell counts $\{M_0, M_1, \cdots, M_N\}$. Note that two constraints must

hold: $\sum_{r=0}^{N} M_r = M$ to account for all infected and uninfected cells and $\sum_{r=0}^{N} rM_r = N$ for conservation of the total number of IUs. If we assume all M cells are of identical size and volume, they carry equal probability of being infected by a particular virus. Thus, evaluating the probability distribution that M_r takes on the value m_r reduces to the well-known occupancy problem of randomly placing balls into identical urns (33), and we derive

$$\Pr(M_r = m_r \,|\, M,N) = \sum_{j=m_r}^{M} \binom{j}{m_r} \binom{M}{j} \times \binom{N}{r,\cdots,r,(N-rj)} \frac{(-1)^{j-m_r}(M-j)^{N-rj}}{M^N}, \tag{2}$$

where the r term is repeated j times in the lower argument of the multinomial coefficient. The derivation of Eq. 2 is detailed in Appendix A in the Supporting Materials and Methods, and an investigation into the effects of inhomogeneous cell sizes is presented in Appendix B. Furthermore, in Appendix A, we derive the expected value and variance of M_r as

$$\mathrm{E}[M_r] = M\binom{N}{r}\left(\frac{1}{M}\right)^r\left(1-\frac{1}{M}\right)^{N-r} \tag{3}$$

and

$$\mathrm{Var}[M_r] = M\binom{N}{r}\left(\frac{1}{M}\right)^r\left(1-\frac{1}{M}\right)^{N-r} + \frac{M(M-1)N!(M-2)^{N-2r}}{(r!)^2(N-2r)!M^N} - \frac{M^2(N!)^2(M-1)^{2N-2r}}{(r!)^2[(N-r)!]^2M^{2N}}. \tag{4}$$

Note that the variance is equal to the expected value with two additional correction terms that cancel each other as N and M increase, indicating the probability distribution of M_r is Poisson-like for large N and M. A plot of a representative probability distribution and a test of agreement between our analytical result and numerical simulation is provided in Fig. 2.

We also derive the joint probability $\Pr(M_0 = m_0,\cdots,M_N = m_N|M,N)$ that the SMOI $\{M_0,M_1,\cdots,M_N\}$ takes on the set of values $\{m_0,m_1,\cdots,m_N\}$ as

$$\Pr(M_0 = m_0,\cdots,M_N = m_N|M,N) = \frac{1}{M^N}\binom{M}{m_0,m_1,\cdots,m_N} \times \binom{N}{0,\cdots,0,1,\cdots,1,\cdots,N,\cdots,N} = \frac{M!N!}{M^N}\prod_{r=0}^{N}\frac{1}{m_r!(r!)^{m_r}}. \tag{5}$$

The first and second multinomial expressions enumerate the degeneracy of how the M identical cells are distributed across the configuration $\{m_0,\cdots,m_N\}$ and how the N identical IUs are chosen for those cells, respectively. Although the second expression in Eq. 5 is more succinct, it must be explicitly conditioned on the constraints $\sum_{r=0}^{N} m_r = M$ and $\sum_{r=0}^{N} rm_r = N$.

The expressions in Eqs. 2 and 5 provide an exact discrete description of the stochasticity of the MOI but are computationally expensive to evaluate for large values of N and M. In a typical virology experiment, the number of viral particles N_0 and host cells M are large enough for certain asymptotic methods to be applicable. Furthermore, for intermediate values of Q and based on Eq. 1, the expected number of IUs N would be similarly large. We can thus take the mathematical limit $N,M\to\infty$ while keeping the ratio $\mu = N/M$ fixed and approximate Eq. 2 as

$$\Pr(M_r = m_r \,|\, M,N) \approx \frac{1}{m_r!}\left[\frac{M\mu^r e^{-\mu}}{r!}\right]^{m_r}\exp\left[-\frac{M\mu^r e^{-\mu}}{r!}\right]. \tag{6}$$

Equation 6 implies that M_r is Poisson-distributed, with mean and variance

$$\mathrm{E}[M_r] = \mathrm{Var}[M_r] \approx \frac{M\mu^r e^{-\mu}}{r!}. \tag{7}$$

A mathematical justification of Eq. 6 is given in Appendix A, and comparisons of Eq. 6 and the analytical result in Eq. 2 to simulations are shown in Fig. 3.

Under the same large M,N limit and using Eq. 6, we show in Appendix A

$$\Pr(M_0 = m_0,\cdots,M_N = m_N|M,N) \approx \prod_{r=0}^{N}\Pr(M_r = m_r|M,N), \tag{8}$$

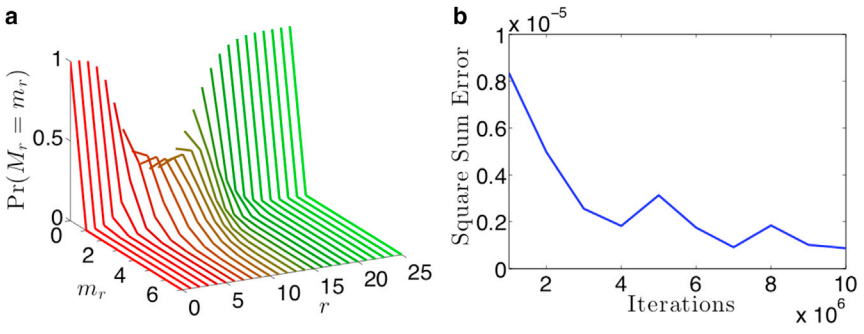


FIGURE 2 (*a*) A collection of curves of the probability of finding m_r cells that have been infected by exactly r IUs, given a total number of IUs $N = 100$ and a total number of cells $M = 10$, using Eq. 2. With $N/M = 10$, we expect very few cells to be uninfected, resulting in the probability distribution concentrated close to 0 for low values of r. Similarly, we expect few cells to be infected by a very large number of IUs, accumulating the probability distribution close to 0 for large r. Only at intermediate values of $r \approx N/M = 10$ do we observe a Poisson-like distribution. (*b*) We perform a numerical study to show empirically that our analytical result in Eq. 2 matches the statistical frequency of virus cell counts from a simulation of $N = 100$ IUs being randomly assigned to $M = 10$ cells. The square sum error between the simulated proportions and the analytical result was calculated with increasing numbers of iterations of the simulation. For iterations around 10^6, our square sum error is on the order of 10^{-6}, indicating strong agreement between our model and simulation. To see this figure in color, go online.

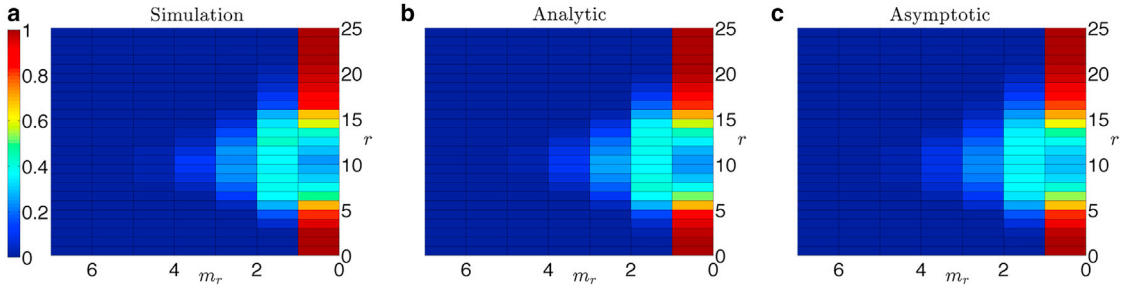


FIGURE 3 Heat maps of the probability distribution $\Pr(M_r = m_r \mid M,N)$ of finding m_r cells that have been infected by exactly r IUs given a total number of viruses $N = 100$ and $M = 10$ cells. (*a*) The statistical frequency of virus cell counts after simulating IUs randomly distributing to the M cells is shown, averaged over 1000 iterations. (*b*) The analytical result obtained from Eq. 2 is shown. (*c*) The asymptotic approximation with $M = 10$ and $\mu = (N/M) = 10$ is shown, using the expression in Eq. 6. There is close agreement between the simulated and analytical results. The relatively low values of M and N make the asymptotic formula in Eq. 6 inappropriate for this parameter regime, explaining the discrepancy between the asymptotic result and the exact analytical result. However, it is noteworthy how qualitatively small that deviation is, which will continue to vanish as M and N increase in value. To see this figure in color, go online.

which implies that as $M,N \to \infty$, the random variables $M_0,\cdots,M_N$ are independently distributed. In the next section, we will apply results of our probability model of SMOI to the case of a repeatedly diluted solution of virus particles, a procedure used in many VQAs.

Serial dilution

Low viral-particle concentrations in assays are typically obtained via serial dilution processes to increase the sensitivity to individual viral infections (4,24,25). The initial viral stock containing N_0 particles is diluted by a fixed factor of D, and the process is repeated $d_{\max}$ times. At each dilution number d, an assay can be performed to determine if the concentration of virus particles in the diluted solution is sufficient to generate a qualitative signal of infection, known as a "cytopathic effect" (CE). For example, the diluted stock can be administered in vivo to a model organism such as a mouse. The mouse's death would indicate that at least one lethal unit of the virus was present at that dilution level. Alternatively, an in vitro assay can be carried out to measure a signal that, for example, quantifies the exact number of plated cells that were successfully infected. To model these assays, we first define M^* as the number of host cells infected by at least one IU and that are capable of producing new viruses. In Appendix A we derive the discrete probability density function for finding $M^* = m$ infected cells at a given dilution number d and find

$$\Pr(M^* = m) = \binom{M}{m}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]^m \times \exp\left(-\frac{N_0}{QMD^d}\right)^{M-m}. \tag{9}$$

Equation 9 shows that the number of infected cells M^* is binomially distributed with expected value

$$\mathrm{E}[M^*] = M\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right] \tag{10}$$

and variance

$$\mathrm{Var}[M^*] = M\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]\exp\left(-\frac{N_0}{QMD^d}\right). \tag{11}$$

We can define the probability of observing a CE at dilution number d as the probability of finding one or more infected cells:

$$\Pr(\text{“Cytopathic effect”}) \equiv \sum_{m=1}^{M} \Pr(M^* = m) = 1 - \exp\left(-\frac{N_0}{QD^d}\right). \tag{12}$$

The definition we use in Eq. 12 assumes an in vitro assay that can exhibit a cytopathic signal after a single cell infection or more. For in vivo assays, the probability that m infected cells are sufficient for a CE will depend on many complex physiological factors such as immune pressure, in-host viral evolution, and virion burst size (34). A plot of how the initial particle count N_0 and dilution factor D effect the characteristic functional form of Eq. 12 is shown in Fig. 4. Although both Eqs. 9 and 12 assume each IU contains all viral genes required for in-host replication, an extended probability model that factors in genetic mutation and degradation is provided in Appendix C. Furthermore, for the case of retroviruses, infectious processes inside the host cytoplasm may be suppressed by previous infections, known as viral interference, and is explored in Appendix D. In Plaque Assay, we will use Eq. 9 to analyze the plaque assay. Equation 12 will be used for "binary" assays that are only concerned with the presence or absence of a CE, such as the endpoint dilution assay, which we will explore in Endpoint Dilution Assay.

RESULTS AND DISCUSSION

Plaque assay

The plaque assay is an example of a VQA in which the objective is to infer the total number of viruses N_0 present in a solution, assuming the PFU to particle ratio Q has been independently measured and estimated (24,25,35). After d serial dilutions, the viral stock is added to a monolayer of M cells, and a layer of agar gel is added to the well to inhibit the diffusion of virus particles in the plate. If a virus successfully infects a host cell, the agar will limit the range of new infections to the most adjacent cells. Viral infection thus spreads out radially from the initial nucleation infection and forms a visible discoloration in the plate called a "plaque." For high particle concentrations, the number of plaques formed

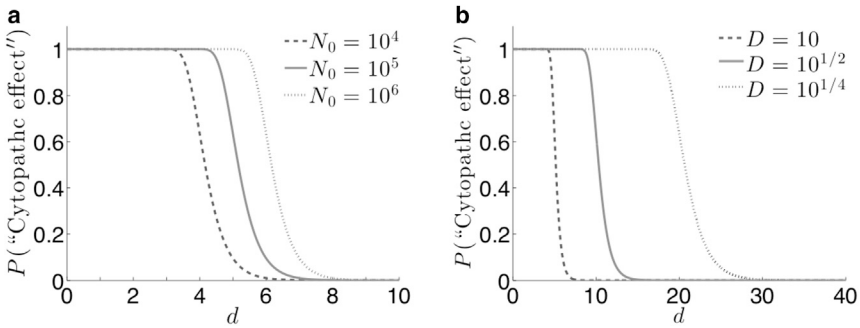


FIGURE 4 The probability of observing a cytopathic effect (CE) given in Eq. 12 as a function of the dilution number d and with $Q = 1$. (*a*) For $D = 10$, as the initial particle count N_0 increases, the critical dilution moves toward higher d. (*b*) Common dilution factors include logarithmic dilution ($D = 10$), half-logarithmic dilution ($D = 10^{1/2}$), and quarter-logarithmic dilution ($D = 10^{1/4}$). Logarithmic dilution requires a lower number of dilutions to cause the characteristic decrease in probability, requiring less individual assays to perform. Quarter-logarithmic dilution, though requiring more dilutions, has a slower transition from high to low probability across d, making the assay less sensitive to experimental error or noise. The plot above can be used to quantify the tradeoffs between the choices of D.

may be large enough to cover the entire plate surface. After a sufficient critical dilution number d_c, however, the number of plaques formed are low enough to be visibly distinct and countable. For each dilution number d, the assay can be performed for T number of trials. The "signal" data arising from the plaque assay $P_{d,t}$ is defined as the number of visible plaques counted, where $t = 1, \cdots, T$ is the trial number. The standard method of obtaining an estimate $\widehat{N}_0$ of the true particle count N_0 is to apply the sample mean of the data $P_{d_c,t}$ at the critical dilution level d_c to the formula

$$\widehat{N}_0 = D^{d_c}\left(\frac{1}{T}\sum_{t=1}^{T} P_{d_c,t}\right), \tag{13}$$

which posits that the average number of plaques is directly proportional to the particle count N_0. Equation 13 assumes that each infected cell corresponds to one IU, which is not necessarily true in the context of SMOI. Furthermore, although data corresponding to dilution numbers $d < d_c$ are unusable, data for $d > d_c$ corresponding to countable plaques are not used at all in Eq. 13.

To improve on Eq. 13 by using the entire set of plaque counts $P_{d,t}$ for our estimate of N_0, we propose a maximal likelihood estimation (MLE) scheme. Using the mathematical models derived above, we can construct an expression $\mathcal{L}(P_{d,t} \,|\, N_0)$ of the probability that the observed data $P_{d,t}$ can be generated assuming a particular value for N_0, known as a likelihood function. A value for N_0 that maximizes $\mathcal{L}(P_{d,t} \,|\, N_0)$ corresponds to the most probable estimate $\widehat{N}_0$ that could have generated the data. As each nucleation of a plaque corresponds to a distinct infected cell (and assuming that overlapping lesions of necrotic cells are still discernible as distinct plaques), we can equate $P_{d,t}$ to the total number of successfully infected cells M^*. We will ignore the dynamics of coinfection and viral interference. Using Eq. 9, we propose the following likelihood function of the data given N_0:

$$\mathcal{L}(P_{d,t} \,|\, N_0) = \prod_{d=d_c}^{d_{\max}} \prod_{t=1}^{T} \binom{M}{P_{d,t}} \left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]^{P_{d,t}} \times \exp\left(-\frac{N_0}{QMD^d}\right)^{M-P_{d,t}}. \tag{14}$$

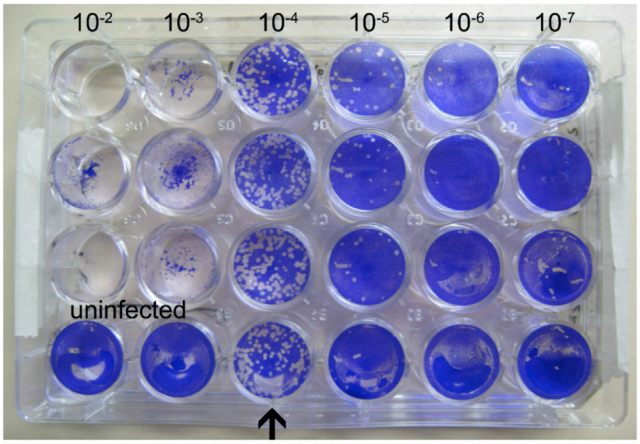
To obtain the MLE $\widehat{N}_0$, we take the derivative of the natural log of Eq. 14 with respect to N_0 and set the result to zero to obtain

$$0 = \sum_{d=d_c}^{d_{\max}} \sum_{t=1}^{T} \frac{M \exp\left(-\frac{\widehat{N}_0}{QMD^d}\right) - M + P_{d,t}}{QMD^d\left[1 - \exp\left(-\frac{\widehat{N}_0}{QMD^d}\right)\right]}. \tag{15}$$

We can solve Eq. 15 for $\widehat{N}_0$ using numerical methods such as Newton-Raphson (36), an iterative scheme that approaches the solution of an equation asymptotically starting from an initial guess $\widehat{N}_0^{\text{init}}$. To increase the stability of convergence to the solution, we choose $\widehat{N}_0^{\text{init}}$ by equating the sample average of plaque counts $(1/T)\sum P_{d_c,t}$ with the expected number of infected cells $\mathrm{E}[M^*]$ in Eq. 10 at the critical dilution d_c to derive

$$\widehat{N}_0^{\text{init}} = -QMD^{d_c} \ln\left[1 - \frac{1}{M}\left(\frac{1}{T}\sum_{t=1}^{T} P_{d_c,t}\right)\right]. \tag{16}$$

An example of raw plaque-count data and the resulting estimates for N_0 are given in Fig. 5. To quantify the relative improvement of the MLE of N_0 over the standard method in Eq. 13, we simulate plaque assay data assuming a fixed, known N_0 value. In our simulation, we use the models established in Probabilistic Models of Statistical Multiplicity of Infection to sample the N_0 particles according to Eq. S9 in Appendix A to account for serial dilution and sample again the resulting particles according to Eq. 1 to obtain the number of IUs N. The IUs are distributed randomly to the M cells with equal probability, and the resulting number of



Plaque Count Data from Sloutskin et al. (35)						
Trial	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
1	–	–	111	24	1	0
2	–	–	126	24	1	0
3	–	–	121	13	0	0

FIGURE 5 An example of raw plaque count data taken from Sloutskin et al. (35). A viral solution was assayed in a plate of $M = 3 \times 10^5$ cells at dilution numbers $d = 2$, 3, 4, 5, 6, and 7 at a dilution factor of $D = 10$. The particle to PFU ratio is assumed to be $Q = 1$. For $T = 3$ separate trials, the number of plaques were counted at each dilution level. The bottom row of plates used as a control is ignored. For dilution numbers $d = 2$ and 3, the entire plate of cells shows cytotoxicity so that the numbers of plaques were undiscernible and, thus, the countable data starts at $d_c = 4$. For the old method featured in Eq. 13, the estimate for N_0 is $\widehat{N}_0 = 1.19 \times 10^6$, and for the MLE derived from Eq. 15, $\widehat{N}_0 = 1.26 \times 10^6$. This results in a relative difference of 5.5%. Furthermore, when applying these parameters and the $\widehat{N}_0$ estimate to Eq. 17, we observe a 10.7% decrease in the estimate variation using the MLE technique. To see this figure in color, go online.

infected cells M^* is recorded. Because plates of cells with too many infections render the number of plaques uncountable, a "countable plaque threshold" renders the data unusable when the number of infected cells exceed the threshold. Thus, the resulting plaque data $P_{d,t}$ for a given dilution d and trial t is assigned the number of simulated infected cells if the latter is less than the given threshold. A scatter plot of the data $P_{d,t}$ of one such simulation is shown in Fig. 6 *a*, and the corresponding likelihood function from Eq. 14 is plotted in Fig. 6 *b*. Because the MLE method utilizes a full probabilistic model of the plaque count distribution instead of relying only on the expected value at the single critical dilution d_c, it produces an estimate consistently closer to the original N_0 that generated the data. To better quantify this property, in Appendix E, we derive an asymptotic approximation of the variance of $\widehat{N}_0$ as

$$\mathrm{Var}\left[\widehat{N}_0\right] \approx \left[\sum_{d=d_c}^{d_{\max}} \frac{T \exp\left(-\frac{N_0}{QMD^d}\right)}{Q^2 M D^{2d}\left[1 - \exp\left(-\frac{N_0}{QMD^d}\right)\right]}\right]^{-1}. \tag{17}$$

The variance is an explicit function of Q, which is assumed to be a priori known. If there is uncertainty in the value of Q, Eq. 17 can quantify how sensitive the distribution of $\widehat{N}_0$ is to variation in Q, as shown in Fig. 7 *a*. We can see that for small assumed Q, such as in poliovirus (30), error in this measurement can cause a large relative change in the accuracy of $\widehat{N}_0$. This type of sensitivity analysis on estimation variance can be done with any experimental parameter included in the likelihood function in Eq. 14. Furthermore, for directly controllable parameters, such as the serial dilution factor D, Eq. 17 can provide insight into optimizing the assay protocol, as shown in Fig. 7 *b*. Although it is evident that small D would increase the accuracy of the $\widehat{N}_0$ estimate, doing so requires more serial dilutions, which increases the time and expense of the assay. Thus, our sensitivity analysis provides a quantitative method for making experimental design choices between minimizing uncertainty versus the cost of an assay protocol. Lastly, if we compute the variance of the standard method in Eq. 13 due to the known variance in the data $P_{d,t}$ and compare with Eq. 17, we find, when using realistic parameter values from Fig. 5, the standard method results in a 10.7% higher variance than that of our method. Although the significance of the relative increase in precision of estimating N_0 found using our method is highly dependent on the context of the experimental study for which the assay

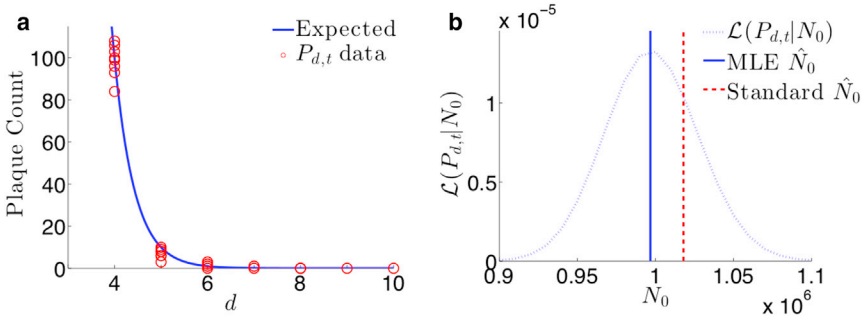


FIGURE 6 Results of plaque assay simulation for parameters $N_0 = 10^6$, $M = 10^5$, $Q = 1$, $D = 10$, $d_{\max} = 10$, and $T = 10$. (*a*) The scatter plot of simulated data $P_{d,t}$ (*circles*) and the expected value of plaque counts as given by Eq. 10 show close agreement. (*b*) The likelihood function $\mathcal{L}(P_{d,t}\,|\,N_0)$ with respect to N_0 using the same simulated data. The MLE obtained by iteratively solving Eq. 15 is $\widehat{N}_0 = 9.97 \times 10^5$ and is relatively closer to the true value of N_0 than the estimate calculated from the standard method in Eq. 13, $\widehat{N}_0 = 1.02 \times 10^6$. To see this figure in color, go online.

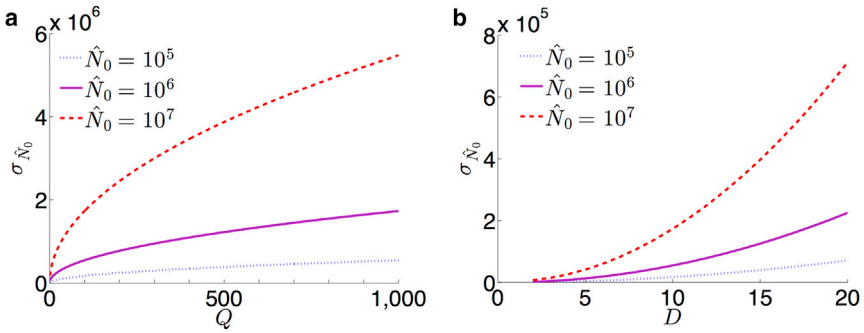


FIGURE 7 Approximations of the standard deviation $\sigma_{\widehat{N}_0} = \mathrm{Var}[\widehat{N}_0]^{1/2}$ of MLEs for the plaque assay using Eq. 17 and parameters $\widehat{N}_0 = 10^5$, 10^6, and 10^7, $M = 3 \times 10^5$, $d_c = 4$, $d_{\max} = 7$, and $T = 3$, corresponding to the assay displayed in Fig. 5. (*a*) For $D = 10$, the standard deviation increases proportional to the square root of Q. (*b*) For $Q = 1$, we can see a low dilution factor D will increase the accuracy of the estimate $\widehat{N}_0$. To see this figure in color, go online.

was performed, similar sensitivity analysis can be used to determine such tolerances.

Endpoint dilution assay

Another widely used assay for quantifying the initial viral particle count N_0 is the endpoint dilution or endpoint titration assay (23,25,37). It is often used in place of the plaque assay, as it can be more rapidly performed and is useful for viral strains that are unable to form plaques. Here, serial dilutions at a factor of D are employed, and at every dilution number d, an assay is performed T times to test for a successful CE. The number E_d of observed CEs among the T trials at a given dilution number d is recorded as the signal. For low dilution, we expect many cells to be infected, and the probability of observing a CE, as shown in Eq. 12, is close to 1. If every trial of the assay is likely to display a CE, then E_d is expected to be close to T. However, at high dilution, the probability in Eq. 12 rapidly decreases to 0, as shown in Fig. 4, and E_d will be similarly small. For a large initial stock of viral particles N_0, a larger dilution number d is needed to ensure the dramatic change in probability in Eq. 12. Thus, the critical dilution at which E_d most rapidly decreases from T can be used to estimate the particle count N_0. This occurs at the point of inflection when $d = \log_D(N_0 Q^{-1})$ and corresponds to when the expected number of successful trials $\mathrm{E}[E_d] = T\,(1 - e^{-1})$, as shown in Fig. 8.

One commonly used way to estimate N_0 is the RM method, which utilizes the two dilution numbers that capture the greatest change in the data E_d (27). We first define a critical dilution number $d_{50\%}$ to be the largest dilution such that at least 50% of the trials exhibit a CE. The estimate $\widehat{N}_0$ for the particle count N_0 is given by

$$\log_{10}\left(\widehat{N}_0\right) = d_{50\%} + \frac{E_{d_{50\%}} - 0.5T}{E_{d_{50\%}} - E_{d_{50\%}+1}}. \tag{18}$$

The RM method effectively attempts to approximate the steepest descent of the CE probability given in Eq. 12 with a line connecting the assay data at dilutions $d_{50\%}$ and $d_{50\%+1}$, as displayed in Fig. 8 *a*. Unfortunately, this line always rests above the actual expectation curve of E_d, so any estimate $\widehat{N}_0$ obtained from this method will overestimate the true N_0. Another commonly used estimation scheme is the SK method, which uses the critical dilution number $d_{100\%}$, the largest dilution such that 100% of trials exhibit a CE (28,37). The SK estimate $\widehat{N}_0$ is given by

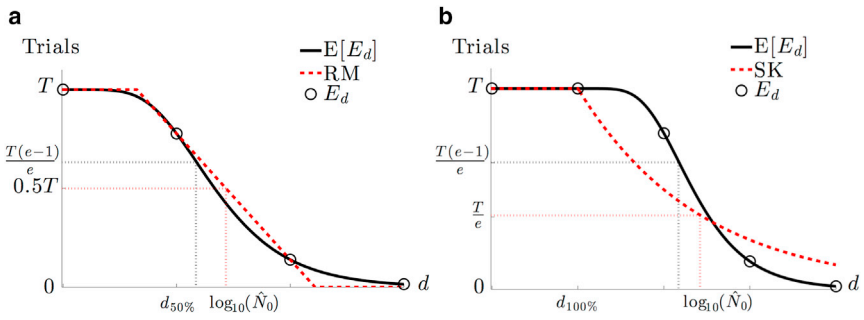


FIGURE 8 An illustration of the consistent overestimation of the Reed and Muench (RM) and Spearman-Karber (SK) methods using the expected curve $\mathrm{E}[E_d]$ of CEs given T trials as a function of the dilution number d derived from Eq. 12. (*a*) The RM method approximates the steepest decent of the expectation curve with a line connecting the two data points $E_{d_{50\%}} \leq 0.5T < E_{d_{50\%}+1}$. Because of the relative convexity of the expected curve, the linear approximation consistently rests above the curve and results in an overestimate of $\log_{10}(\widehat{N}_0)$. (*b*) From the last dilution $d_{100\%}$ such that all trials exhibit a CE, the SK method assumes an exponential decay of the expectation. Obtaining the characteristic decay rate of the exponential involves calculating the area under the curve, which is done numerically using the data E_d. However, according to our model, many of the expected values of E_d exist above the exponential, causing the numerical integration to overestimate the area and, thus, decay too slowly. This gradual decrease in the exponential curve results in a larger estimate of $\log_{10}(\widehat{N}_0)$. To see this figure in color, go online.

$$\log_{10}\left(\widehat{N}_0\right) = d_{100\%} - \frac{1}{2}\log_{10}(D) + \log_{10}(D)\sum_{d=d_{100\%}}^{d_{\max}}\frac{E_d}{T}. \tag{19}$$

In this method, the downward slope for the expectation of E_d is assumed to follow a decaying exponential starting at dilution $d_{100\%}$, as shown in Fig. 8 *b*. The intention is to find the dilution at which Te^{-1} CEs are expected by calculating the area under the exponential curve, given by the summation term in Eq. 19. However, the actual values of E_d will follow the expected curve from our model, leading to an overestimate of the area and, by extension, a larger value for $\widehat{N}_0$. Both standard methods were derived from the heuristic observation that E_d exhibits sigmoidal behavior as a function of the dilution number d, but an underlying probabilistic model was missing, resulting in consistent overestimation of the true N_0. Furthermore, neither method uses the "particle to PFU ratio" Q, accounts for the stochasticity of serial diluting viral samples, considers the dynamics of SMOI, or employs the entire set of data E_d.

We present an alternative way to infer N_0, using Eq. 12 to establish an MLE scheme. We restrict ourselves to in vitro assays in which a single infected cell is sufficient to display a CE. Then each cytopathic count is binomially distributed with parameters T and the probability given in Eq. 12. Thus, for a set of data $\{E_1, E_2, \cdots, E_{d_{\max}}\}$, we propose the likelihood function

$$\mathcal{L}(E_d \,|\, N_0) = \prod_{d=1}^{d_{\max}}\binom{T}{E_d}\left[1-\exp\left(-\frac{N_0}{QD^d}\right)\right]^{E_d} \times \exp\left(-\frac{N_0}{QD^d}\right)^{T-E_d}. \tag{20}$$

Equation 20 is an expression of the probability of the data $\{E_1, \cdots, E_{d_{\max}}\}$ given the current assumed value of N_0. To obtain the best estimate $\widehat{N}_0$ of N_0, we maximize the likelihood function by taking the log and derivative of $\mathcal{L}(E_d \,|\, N_0)$ with respect to N_0 and set it equal to zero to obtain

$$0 = \sum_{d=1}^{d_{\max}}\frac{E_d - T + T\exp\left(-\frac{\widehat{N}_0}{QD^d}\right)}{QD^d\left[1-\exp\left(-\frac{\widehat{N}_0}{QD^d}\right)\right]}. \tag{21}$$

As with Eq. 15, solving Eq. 21 for $\widehat{N}_0$ requires a numerical method such as Newton-Raphson. As an appropriate initial estimate for $\widehat{N}_0$, the formula

$$\widehat{N}_0^{\text{init}} = -0.5QD^{d_c}\left[\ln\left(1-\frac{E_{d_c}}{T}\right) + D\ln\left(1-\frac{E_{d_c+1}}{T}\right)\right] \tag{22}$$

can be used, where d_c is the largest dilution number such that at least half of the trials exhibit a CE. Equation 22 is the average of the N_0 estimates at dilutions d_c and d_{c+1} when setting the CE probability in Eq. 12 to 1/2. For a comparison of our MLE method with the RM and SK methods, we simulate data similar to that described in Plaque Assay. Here, we take the number of trials such that the simulated count of infected cells is greater than zero as the values of E_d for a given dilution number d. We plot the likelihood from Eq. 20 and compare the MLE of N_0 with those derived by the RM and SK methods in Fig. 9 *a*. Although both RM and SK estimate very similar values of $\widehat{N}_0$, they both consistently overestimate the a priori set N_0 relative to the MLE method. This demonstrates the advantage of a probabilistic model for parameter inference over heuristically determined formulas.

The expressions we derived in Eqs. 14 and 20, applied to simulated data, can also help quantify tradeoffs in experimental design. As discussed above, there exist viruses that cannot form plaques, restricting the options of VQAs to endpoint dilution. However, for many cases, the choice between using one assay over the other can be one of convenience. More specifically, endpoint dilution assays can often be performed more rapidly than plaque assays. Using the same simulated data for both assays, we plot Eqs. 14 and 20 together in Fig. 9 *b*. The plots clearly show the superiority of the plaque assay for estimating the viral stock number N_0 in respect to both how close the MLE infers the true N_0 value and the amount of variance in that estimate. Although the amount of variability and error that is tolerable for an experiment may be context dependent, the plots in Fig. 9 *b* provide a quantitative way to differentiate between the two methods.

Luciferase reporter assay

The luciferase reporter assay is commonly used to measure the infectivity of a viral strain. Here, the ratio $\mu = N/M$ of total infections over the number of plated cells is estimated by measuring the transcription activity of viral proteins (14–16). The reporter employs an oxidative enzyme luciferase that facilitates a reaction when introduced to the substrate luciferin, resulting in bioluminescence. The protocol begins with attaching the luciferase gene to the viral genome. The altered viral strain is cloned to a total particle count N_0, which, in this case, is assumed to be fixed and known. The solution of viruses is added to a plated monolayer of M host cells. An incubation time is allowed for transcription of viral proteins and, incidentally, the luciferase enzyme. Subsequently, all cells are lysed to release all cytoplasmic contents into the solution, upon which luciferin is added. The oxidation of luciferin is facilitated by the luciferase enzyme, and the resulting bioluminescence yields a measurable signal (38). The light intensity is thus a measure of total transcription activity of the viral genome in all infected cells and can be used as a proxy for the total number of viruses N that successfully infected host cells.

Although there is stochasticity in transcription-factor binding and, in the case of retroviruses, the number of integration

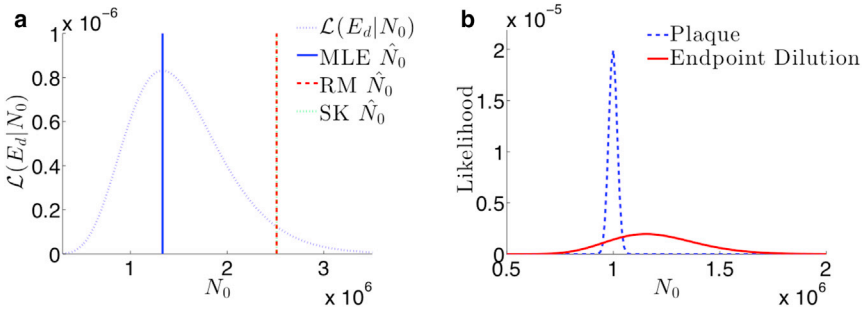


FIGURE 9 (*a*) The likelihood function $\mathcal{L}(E_d \mid N_0)$ in Eq. 20 for the endpoint dilution assay and the corresponding maximal likelihood, RM, and SK estimates given simulated data generated with $N_0 = 10^6$, $Q = 1$, $D = 10$, and $d_{\max} = 10$. The estimates for maximal likelihood ($\widehat{N}_0 = 1.33 \times 10^6$), RM ($\widehat{N}_0 = 2.51 \times 10^6$), and SK ($\widehat{N}_0 = 2.51 \times 10^6$) all overestimate N_0, but the smaller relative error of the MLE is an improvement on the errors of the existing two methods. (*b*) The likelihood functions $\mathcal{L}(P_{d,t} \mid N_0)$ and $\mathcal{L}(E_d \mid N_0)$ for the plaque and endpoint dilution assays, respectively, are shown, given simulated data. The data were generated with the parameters $N_0 = 10^6$, $M = 10^5$, $Q = 1$, $D = 10^{1/4}$, $d_{\max} = 30$, and a "countable plaque threshold" of 150. The plaque assay likelihood is concentrated close to the true N_0 value, whereas the endpoint dilution likelihood is far more spread out and overestimates N_0. This direct quantitative comparison can inform an experimentalist when choosing between the two methods. To see this figure in color, go online.

sites on the host DNA, we will assume that each successful virus infection contributes one viral genome to be transcribed and each transcription occurs at a constant rate proportional to the total number of integrated viral genomes. Note that the limited number of transcription factors, ribosomes, and other cell machinery necessary to produce viral proteins and the luciferase reporter causes the production rate to saturate as the number of infecting viruses r per cell increases. Thus, transcription activity saturates with increasing number of infections r. We can model this effect by defining a monotonically increasing function $f(r)$ representing the number of transcribed viral proteins when a cell is infected by r viruses over the course of the assay. Thus, for a given SMOI $\{M_0, \cdots, M_N\}$, we will model the intensity signal L of the total luciferase reporter luminescence with

$$L = \sum_{r=0}^{N} L_0 f(r) M_r, \tag{23}$$

where L_0 is the fluorescence intensity arising from a single luciferase reporter present in the solution. Although $f(r)$ may take on many functional forms, a commonly used model for transcription-factor kinetics is the Hill function (39) given by

$$f(r) = \frac{f_{\max} r^h}{K + r^h}, \tag{24}$$

where $f_{\max}$ is the maximal transcription activity of luciferase, h is the Hill coefficient that effectively describes cooperative binding of multiple transcription factors at a promoter region, and K is an effective dissociation constant relating the binding and unbinding rates of transcription factor. The functional form of Eq. 24 accounts for the limited transcription machinery available for the multiple copies of viral genome present in the cell. In Fig. 10 *a*, we calculate the discrete probability distribution $\Pr(L = \ell)$ by considering the cumulative weight of every allowable configuration of N viruses infecting M cells through Eq. 23.

Because luciferase reporter assays typically involve large values of initial virus count N_0 and cell count M, we can use the asymptotic approximations in Eqs. 6 and 7 along with the central limit theorem (40) to assume L is normally distributed with expected value

$$\mathrm{E}[L] = L_0 f_{\max} M e^{-\mu} \sum_{r=0}^{N} \frac{r^h \mu^r}{(K + r^h) r!} \tag{25}$$

and variance

$$\mathrm{Var}[L] = L_0^2 f_{\max}^2 M e^{-\mu} \sum_{r=0}^{N} \frac{r^{2h} \mu^r}{(K + r^h)^2 r!}. \tag{26}$$

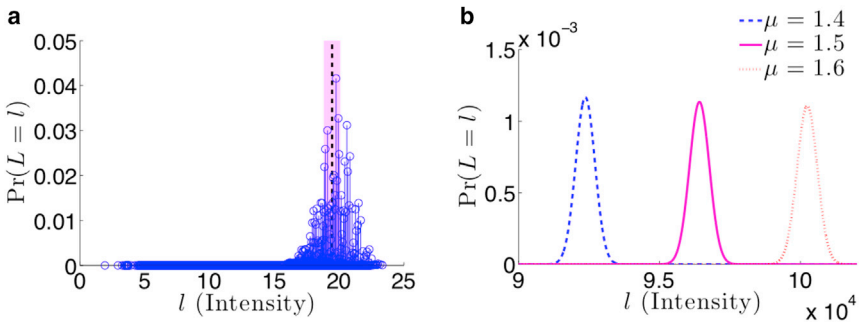


FIGURE 10 Probability distributions of the luciferase assay fluorescence intensity L from Eq. 23. (*a*) A toy example of a discrete probability distribution of allowable fluorescence intensities for $N = 30$ viruses infecting $M = 20$ cells is shown. Due to the M^N finite number of allowable configurations of the SMOI, there are a corresponding finite number of intensities with specific probabilities determined by Eq. 5 and represented by a unique circle. The parameters used for the reporter kinetics are $f_{\max} = 2$, $h = 1$, $K = 1$, and $L_0 = 1$. The mean intensity of the fluorescence signal is $\mathrm{E}[L] = 19.5$, represented by the vertical dotted line, and variance is $\mathrm{Var}[L] = 1.49$, represented by the shaded region. (*b*) The normally distributed approximation of fluorescence intensity using $M = 1 \times 10^5$, $f_{\max} = 2$, $h = 1$, $K = 1$, and $L_0 = 1$ is shown. The distributions are plotted for $\mu = 1.4$, 1.5, and 1.6 by computing the expected values $\mathrm{E}[L] = 9.23 \times 10^4$, 9.64×10^4, and 1×10^5 and the variances $\mathrm{Var}[L] = 1.7 \times 10^5$, 1.24×10^5, and 1.3×10^5 respectively. To see this figure in color, go online.

A visualization of the normal approximation of the probability distribution of L is shown in Fig. 10 *b*. Furthermore, with Eqs. 25 and 26, we can derive the likelihood function $\mathcal{L}(L_t^{\text{data}} \mid \mu)$ of the data L_t^{data}, given μ:

$$\mathcal{L}\left(L_t^{\text{data}} \mid \mu\right) = \prod_{t=0}^{T} \frac{1}{\sqrt{2\pi \text{Var}[L]}} \exp\left[-\frac{\left(L_t^{\text{data}} - \text{E}[L]\right)^2}{2\text{Var}[L]}\right], \tag{27}$$

where $1 \leq t \leq T$ is the trial number. Because of the complicated functional form of the mean and variance of L, creating a maximal likelihood scheme to estimate μ from experimental data is intractable, so we use Eq. 25 by replacing the expected value with the experimental average of measurements L_t^{data}. If we assume no cooperative transcription binding ($h = 1$), we solve for the estimate $\widehat{\mu}$ by applying the Newton-Raphson iterative method to the equation

$$0 = \frac{1}{T}\sum_{t=0}^{T} L_t^{\text{data}} - L_0 f_{\max} M e^{-\widehat{\mu}} \sum_{r=0}^{N_0} \frac{r\widehat{\mu}^r}{(K+r)r!}. \tag{28}$$

The typical method, under the assumption that luminescent intensity is proportional to the number of IUs N, is to use the sample mean via the formula $\widehat{\mu}^{\text{init}} = (1/L_0 MT)\sum_{t=0}^{T} L_t^{\text{data}}$. This standard approach fails to account for the effects of SMOI, but can be used to generate an initial guess for solving Eq. 28 iteratively. To compare the two estimates, we simulate data similar the descriptions in the previous two sections. Here, we do not dilute the initial particle count and, after distributing the N IUs to the M cells with equal probability, we compile the SMOI configuration and calculate L_t^{data} using Eq. 23. The results are shown in Fig. 11. The iterative method produces an estimate $\widehat{\mu}$ far closer to the true value of μ than the former method. A similar approach can be used to compare methods for alternative functional forms of the viral protein transcription dynamics described in Eq. 24.

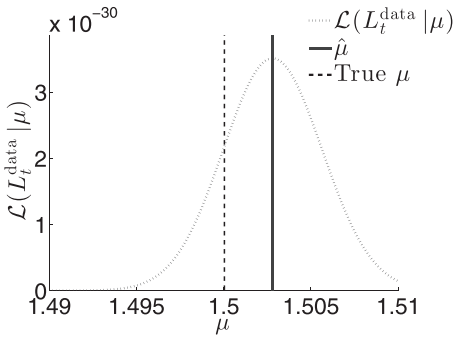


FIGURE 11 The likelihood function $\mathcal{L}(L_t^{\text{data}} \mid \mu)$ using Eq. 27 and simulated data. We set $\mu = 1.5$ and assign other parameters with $M = 1 \times 10^5$, $f_{\max} = 2$, $h = 1$, $K = 1$, and $L_0 = 1$. The estimate derived from solving Eq. 28 is $\widehat{\mu} = 1.502$, whereas the standard method based on the sample mean yields $\widehat{\mu} = 0.97$, far lower than what is displayed in the plot.

CONCLUSIONS

In this work, we derived probability models that quantify the viral infectivity of host cells in an in vitro environment. By factoring in the stochastic nature of virus-host engagement, defective and/or abortive events, and the possibility of multiple infections of a single host, we defined the SMOI and determined related probabilistic models. We analyzed two limiting regimes: small numbers of infecting viruses N and large N. For the low N regime, Eqs. 2 and 5 model how the limited number of IUs are distributed among the M host cells. Alternatively, for large N, we showed that the cell counts of the SMOI become statistically independent, as displayed in Eq. 8, and that they display a Poisson distribution (Eq. 6). Lastly, we explored the effects of serial dilution on the total number of infected cells and the probability of observing an infectious signal in Eq. 9.

Using our probability models along with reasonable assumptions of applied combinatorics and nonlinear inference, we analytically derived expressions for several virus assays to improve on existing methods of experimental data analysis. For virus quantification assays, serial dilution results in low numbers of viral particles. Using the appropriate probability model, we created to our knowledge new methods of estimating the particle count N_0 in the initial viral stock for the plaque assay and the endpoint dilution assay. For measuring infectivity of a viral strain, the objective is to determine the effective multiplicity of infection $\mu = N/M$ as the ratio of successfully infecting viruses N and the total number of cells M included in the assay. As these assays operate under no dilution, we employed the large N limit probability model to analytically derive expressions for the luciferase reporter assay to estimate μ. A summary of each estimation method along with the most commonly used counterpart is displayed in Table 1.

VQAs are primarily concerned with inferring N_0 and assume a priori knowledge of M and the particle to PFU ratio Q. In actuality, there can be variability in the number of cells present in the microtiter well, and, as discussed in Probabilistic Models of Statistical Multiplicity of Infection, the true value of Q is dependent on the particular protocol and particular conditions under which an assay was performed. If an alternative assay (RNA tagging, spectroscopy, super-resolution imaging, etc.) not using cell infection can accurately measure N_0, then, in theory, a subsequent infection assay can be used to infer a more reliable measure of Q. In fact, in our analysis of the plaque assay presented in Appendix E of the Supporting Materials and Methods, we show that one can determine a significantly higher amount of information about Q with the same assay protocol if N_0 is a priori known rather than the reverse case. Thus, one may argue that assays that employ serial dilution, such as plaque and endpoint dilution

TABLE 1 A Summary of the Analytically Derived Expressions Used to Analyze Experimental Results

Comparison of Virological Assay Analyses

Assay (parameter)	Standard Method	New Method
Plaque (N_0)	$\widehat{N}_0 = D^{d_c}\left(\frac{1}{T}\sum_{t=1}^{T} P_{d_c,t}\right)$	$0 = \sum_{d=d_c}^{d_{\max}}\sum_{t=1}^{T}\frac{M\exp\left(-\frac{\widehat{N}_0}{QMD^d}\right) - M + P_{d,t}}{QMD^d\left[1-\exp\left(-\frac{\widehat{N}_0}{QMD^d}\right)\right]}$ Initial guess: $\widehat{N}_0^{\text{init}} = -QMD^d\ln\left(1-\frac{1}{MT}\sum_{t=1}^{T}P_{d_c,t}\right)$
Endpoint dilution (N_0)	RM: $\log_{10}(\widehat{N}_0) = d_{50\%} + \frac{E_{d_{50\%}} - 0.5T}{E_{d_{50\%}} - E_{d_{50\%}+1}}$	$0 = \sum_{d=1}^{d_{\max}}\frac{E_d - T + T\exp\left(-\frac{\widehat{N}_0}{QD^d}\right)}{QD^d\left[1-\exp\left(-\frac{\widehat{N}_0}{QD^d}\right)\right]}$
	SK: $\log_{10}(\widehat{N}_0) = d_{100\%} - \left[\frac{1}{2} - \sum_{d=d_{100\%}}^{d_{\max}}\frac{E_d}{T}\right]\log_{10} D$	Initial guess: $\widehat{N}_0^{\text{init}} = \frac{-QD^{d_c}}{2}\ln\left[\left(1-\frac{E_{d_c}}{T}\right)\left(1-\frac{E_{d_c+1}}{T}\right)^D\right]$
Luciferase reporter $\left(\mu = \frac{N}{M}\right)$	$\widehat{\mu} = \frac{1}{L_0MT}\sum_{t=0}^{T}L_t^{\text{data}}$	$0 = \frac{1}{T}\sum_{t=0}^{T}L_t^{\text{data}} - L_0 f_{\max} M e^{-\widehat{\mu}}\sum_{r=0}^{N_0}\frac{r\widehat{\mu}^r}{(K+r)r!}$ Initial guess: $\widehat{\mu}^{\text{init}} = \frac{1}{L_0MT}\sum_{t=0}^{T}L_t^{\text{data}}$

For virus quantification assays, such as the plaque and endpoint dilution assays, one typically wishes to estimate the number of initial viral particles N_0. For luciferase reporter infectivity assay, the ratio $\mu = N/M$ is desired. Our improved parameter estimation methods are listed next to standard methods currently used.

assays, may be better utilized to infer Q. Because the underlying likelihood of the data in all assays would be the same, the same derivation techniques would follow with respect to Q to formulate its MLE. This analysis shows the robust utility of a full probabilistic model and data likelihood function.

Although the derived assay models provide explicit equations for inference, many of the expressions are analytically unsolvable and require numerical solutions. To improve the accessibility of some of our results, we have created a web-based tool (available at https://bamistry.github.io/SMOI/) that can accept data from plaque, endpoint dilution, or luciferase reporter assays and automatically estimate the parameter of interest. Ultimately, these tools can be used for analysis of future virological studies but may also be useful when revisiting the results of studies that stress quantifying viral infectivity (15,41). For studies that use serial dilution assays, our approach stresses the advantages of using information in the data associated with all dilution numbers rather than just that of the critical dilution.

Our probabilistic models of viral infection can be further generalized to include, for example, the effects of cell size inhomogeneity, coinfection, and viral interference. In the Supporting Materials and Methods, we provide a framework that would allow one to explore how these confounding factors can further alter the signal of a virus assay. Future refinement of these extensions can help to ultimately derive a mechanistic model for the probability of a single virus successfully infecting a host cell, which we defined as Q^{-1}. Understanding this probability of infection can help aid further experimental design and allow better quantification and resolution of the infection dynamics of particular viral strains.

SUPPORTING MATERIAL

Supporting Materials and Methods are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)30575-7.

AUTHOR CONTRIBUTIONS

B.A.M. derived mathematical formulae, developed statistical inference framework, performed simulations, generated plots, and wrote the initial draft. B.A.M. (bamistry@ucla.edu) also developed the web-based analysis tool at https://bamistry.github.io/SMOI/. M.R.D. and T.C. verified the mathematical results, contributed to their analyses, and edited the manuscript. T.C. conceptualized, designed, and supervised the research.

ACKNOWLEDGMENTS

We are especially grateful to Dr. Nicholas Webb, Prof. Benhur Lee, and Prof. Jerome Zack for insightful discussions.

This work was supported in part by grants from the National Science Foundation (DMS-1516675) and the Army Research Office (W911NF-14-1-0472).

REFERENCES

1. Pegu, A., Z. Y. Yang, …, G. J. Nabel. 2014. Neutralizing antibodies to HIV-1 envelope protect more effectively in vivo than those to the CD4 receptor. *Sci. Transl. Med.* 6:243ra88.

2. Osbourn, J. K., J. C. Earnshaw, …, J. McCafferty. 1998. Directed selection of MIP-1 α neutralizing CCR5 antibodies from a phage display human antibody library. *Nat. Biotechnol.* 16:778–781.
3. Qiu, S., H. Yi, …, W. Li. 2012. The binding mode of fusion inhibitor T20 onto HIV-1 gp41 and relevant T20-resistant mechanisms explored by computational study. *Curr. HIV Res.* 10:182–194.
4. Platt, E. J., M. M. Gomes, and D. Kabat. 2014. Reversible and efficient activation of HIV-1 cell entry by a tyrosine-sulfated peptide dissects endocytic entry and inhibitor mechanisms. *J. Virol.* 88:4304–4318.
5. Platt, E. J., J. P. Durnin, and D. Kabat. 2005. Kinetic factors control efficiencies of cell entry, efficacies of entry inhibitors, and mechanisms of adaptation of human immunodeficiency virus. *J. Virol.* 79:4347–4356.
6. Fätkenheuer, G., A. L. Pozniak, …, E. van der Ryst. 2005. Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. *Nat. Med.* 11:1170–1172.
7. Jonckheere, H., J. Anné, and E. De Clercq. 2000. The HIV-1 reverse transcription (RT) process as target for RT inhibitors. *Med. Res. Rev.* 20:129–154.
8. Thierry, S., S. Munir, …, O. Delelis. 2015. Integrase inhibitor reversal dynamics indicate unintegrated HIV-1 dna initiate de novo integration. *Retrovirology.* 12:24.
9. Paterson, D. L., S. Swindells, …, N. Singh. 2000. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Ann. Intern. Med.* 133:21–30.
10. Wilen, C. B., J. C. Tilton, and R. W. Doms. 2012. HIV: cell binding and entry. *Cold Spring Harb. Perspect. Med.* 2 (8):a006866.
11. 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.
12. Boulant, S., M. Stanifer, and P. Y. Lozach. 2015. Dynamics of virus-receptor interactions in virus binding, signaling, and endocytosis. *Viruses.* 7:2794–2815.
13. Chou, T. 2007. Stochastic entry of enveloped viruses: fusion versus endocytosis. *Biophys. J.* 93:1116–1123.
14. Chikere, K., T. Chou, …, B. Lee. 2013. Affinofile profiling: how efficiency of CD4/CCR5 usage impacts the biological and pathogenic phenotype of HIV. *Virology.* 435:81–91.
15. Johnston, S. H., M. A. Lobritz, …, 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.
16. Webb, N. E., and B. Lee. 2016. Quantifying CD4/CCR5 usage efficiency of HIV-1 Env using the affinofile system. *Methods Mol. Biol.* 1354:3–20.
17. Killian, M. L. 2008. Hemagglutination assay for the avian influenza virus. *Methods Mol. Biol.* 436:47–52.
18. Mascola, J. R., M. K. Louder, …, M. Roederer. 2002. Human immunodeficiency virus type 1 neutralization measured by flow cytometric quantitation of single-round infection of primary human T cells. *J. Virol.* 76:4810–4821.
19. Brown, C. M., and K. D. Bidle. 2014. Attenuation of virus production at high multiplicities of infection in Aureococcus anophagefferens. *Virology.* 466-467:71–81.
20. Sette, A., and J. Fikes. 2003. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. *Curr. Opin. Immunol.* 15:461–470.
21. Gerdil, C. 2003. The annual production cycle for influenza vaccine. *Vaccine.* 21:1776–1779.
22. Tree, J. A., C. Richardson, …, D. Looby. 2001. Comparison of large-scale mammalian cell culture systems with egg culture for the production of influenza virus A vaccine strains. *Vaccine.* 19:3444–3450.
23. Neumann, G., K. Fujii, …, Y. Kawaoka. 2005. An improved reverse genetics system for influenza A virus generation and its implications for vaccine production. *Proc. Natl. Acad. Sci. USA.* 102:16825–16829.
24. Kropinski, A. M., A. Mazzocco, …, R. P. Johnson. 2009. Enumeration of bacteriophages by double agar overlay plaque assay. *Methods Mol. Biol.* 501:69–76.
25. Johnson, V. A., R. E. Byington, and P. L. Nara. 1990. Quantitative Assays for Virus Infectivity. Palgrave Macmillan, Basingstoke, UK, pp. 71–86.
26. Agrawal-Gamse, C., F. H. Lee, …, M. M. Laakso. 2009. Adaptive mutations in a human immunodeficiency virus type 1 envelope protein with a truncated V3 loop restore function by improving interactions with CD4. *J. Virol.* 83:11005–11015.
27. Reed, L. J., and H. Muench. 1938. A simple method of estimating fifty percent endpoints. *Am. J. Hyg.* 27:493–497.
28. Hamilton, M. A., R. C. Russo, and R. V. Thurston. 1977. Trimmed Spearman-Karber method for estimating median lethal concentrations in toxicity bioassays. *Environ. Sci. Technol.* 11:714–719.
29. Schwerdt, C. E., and J. Fogh. 1957. The ratio of physical particles per infectious unit observed for poliomyelitis viruses. *Virology.* 4:41–52.
30. Klasse, P. J. 2015. Molecular determinants of the ratio of inert to infectious virus particles. *Prog. Mol. Biol. Transl. Sci.* 129:285–326.
31. Layne, S. P., M. J. Merges, …, P. L. Nara. 1992. Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus. *Virology.* 189:695–714.
32. Turner, T. E., S. Schnell, and K. Burrage. 2004. Stochastic approaches for modelling in vivo reactions. *Comput. Biol. Chem.* 28:165–178.
33. Roberts, F. S., and B. Tesman. 2005. Applied Combinatorics. Pearson Prentice Hall, Upper Saddle River, NJ.
34. Gilchrist, M. A., D. Coombs, and A. S. Perelson. 2004. Optimizing within-host viral fitness: infected cell lifespan and virion production rate. *J. Theor. Biol.* 229:281–288.
35. Sloutskin, A., and R. S. Goldsten. 2014. Infectious focus assays and multiplicity of infection (MOI) calculations for alpha-herpesviruses. *Bio Protoc.* 4:e1295.
36. Lange, K. 2013. Optimization. Springer Science+Business Media, New York, NY.
37. Ramakrishnan, M. A. 2016. Determination of 50% endpoint titer using a simple formula. *World J. Virol.* 5:85–86.
38. Montefiori, D. C. 2009. Measuring HIV Neutralization in a luciferase reporter gene assay. *Methods Mol. Biol.* 485:395–405.
39. Weiss, J. N. 1997. The Hill equation revisited: uses and misuses. *FASEB J.* 11:835–841.
40. Lange, K. 2003. Applied Probability. Springer Science+Business Media, New York, NY.
41. Mistry, B., M. R. D'Orsogna, …, T. Chou. 2016. Quantifying the sensitivity of HIV-1 viral entry to receptor and coreceptor expression. *J. Phys. Chem. B.* 120:6189–6199.