Maximum Likelihood Estimators For Colony Forming Units

K. Michael Martini, [a,c*] **Satya Spandana Boddu,**[a] **Ilya Nemenman,**[a,b,c] **Nic Vega,** [a,b,c*]

Department of Physics, Emory University, Atlanta, GA 30322[a]
Department of Biology, Emory University, Atlanta, GA 30322[b]
Initiative in Theory and Modeling of Living Systems, Emory University, Atlanta, GA 30322[c]

Compiled May 18, 2023
This is a draft manuscript, pre-submission

Address correspondence to Nic Vega, nvega@emory.edu, K. Michael Martini, karl.michael.martini@emory.edu.

ABSTRACT There is a recognized need to measure the abundance of microbes in hospital environments, in the sanitation industry, and in food preparation. Doctors, microbiologists, and food safety experts have been addressing this need by using serial dilution methods to grow bacterial colonies in small enough numbers to count and, from these counts, to infer bacterial concentrations measured in Colony Forming Units (CFUs). There are two primary types of such methods: plating bacteria on a growth medium and counting their resulting colonies or counting the number of tubes at a given dilution that have growth. Traditionally, these types of data have been analyzed separately using different analytic methods. Here we build a direct correspondence between these approaches, which allows one to extend the use of the Most Probable Number (MPN) method from the liquid tubes experiments, for which it was developed, to the growth plates. We also discuss how to combine measurements taken at different dilutions, and we review several ways of analyzing colony counts, including the Poison and truncated Poison methods. For all methods, we discuss their relevant error bounds, assumptions, strengths, and weaknesses. We provide an online calculator for these estimators.

IMPORTANCE Estimation of the number of microbes in a sample is an important problem with a long history of confusion. We provide a comparison of methods for estimating abundance of microbes and detail a mapping between different methods, which allows to extend their range of applicability. This mapping enables higher precision estimates of Colony Forming Units (CFUs) using the same data already collected for traditional CFU estimation methods. Furthermore, we provide recommendations for how to combine measurements of colony counts taken across dilutions, correcting several misconceptions in the literature.

INTRODUCTION

Extrapolation of viable microbial counts from suspensions of live cells is a longstanding—and surprisingly complicated—problem. The fundamental problem is simple: there exists a volume V_0 with some unknown concentration of live microorganisms, which an experimentalist wants to measure. That initial volume will be serially diluted (usually in a ten-fold series), and fixed-volume aliquots (sub-samples) of the resulting suspensions will be cultured. If these aliquots are spread or dropped onto agar plates, the resulting data will be in the form of colony counts. Alternately, multiple aliquots may be taken from a single dilution and used to seed a number of wells or tubes of liquid culture, or a number of plates. Then the number of volumes showing growth when seeded from a particular dilution, as a fraction of the total number of volumes

inoculated, can be used to calculate the Most Probable Number (MPN) of live agents in the initial volume (1, 2).

Best practice for this apparently simple and ubiquitous scenario has been the subject of debate for over a century (1, 3, 4, 5, 6, 7, 8, 9, 10, 11). There are technical considerations to this problem. For example, all plates or tubes used for growth should have the same ability to support growth of the organism(s) being studied, and the sample must be sufficiently homogenized to ensure that microbes are free in solution and not adhered to one another or to a substrate (4). However, such considerations are case-specific and beyond the scope of the present work.

These counts are subject to counting errors as well. At one extreme, when the sample is too concentrated, the number of resulting colonies will be too numerous to count (TNTC; sometimes "too many to count", TMTC). At these high concentrations, colonies merge, breaking the assumption that each microbe corresponds to one colony (6, 12). At the other extreme, when the sample is very diluted, the number of colony initiating bacteria in the sample is subject to small-number statistical (sampling) fluctuations, resulting in high relative error (ratio of the standard deviation to the mean) (13, 14). Finally, experimental errors, such as inaccuracies in pipetting, can emerge and compound over the steps of a serial dilution. However, the latter source of error is expected to be negligible for equipment calibrated to usual standards, and technical replication further reduces effects of this variation (15).

Thus the problem at hand is: How can CFU density best be estimated from plate counts, given the error produced by sampling fluctuations, colony crowding, and (to a lesser degree) pipetting? These errors will contribute differently to different experimental designs. For a *single sample represented by one count of colonies* n_k at one dilution d_k (because only one dilution was measured, or because only one spot or one plate in a series was countable), statistical error of counts (presumably Poisson) is inevitable, and pipetting error will contribute but may not be significant. For a *single sample represented by more than one count of colonies* (representing counts at different dilutions within a single dilution series, and/or technical replicates where one sample was measured multiple times), the same errors apply, but pipetting bias may not be constant across measurements (for example, one failing O-ring on a multichannel pipette can lead to bias in a single column of a 96-well plate).

For *multiple samples of the same type measured in parallel* (biological replicates), we can no longer expect variation across samples to reflect a Poisson-distributed sampling error. Indeed, individual measurements will be subject to sampling variation, but variation across samples will be biological (or otherwise inherent), and demographic (accumulating over time) in addition to sampling. This was the basis for the Nobel-prize winning experiments of Luria and Delbruck, who used the distribution of fluctuations to distinguish Darwininan vs. Lamarckian evolution (16). This is also frequently the case in environmental samples, where different samples from the same source (e. g., water samples from different parts of the same lake) will produce measurements that have super-Poisson variation (aka, over-dispersed). In such cases, some of the variation is "real" due to inhomogeneities in the source, and it cannot be modeled as mere sampling error. Biological variation is problem dependent and often carriers in it the imprint of the underlying fundamental biology (16); it will not be dealt with here. Instead, we will focus on estimation of CFU density within an individual sample, which may be represented by a single set of measurements or by technical replicates, in which one sample is measured multiple times.

The main objective of this paper is to propose methods for accurately estimating colony forming units (CFUs), while taking into account the effects of crowding and sam-

pling fluctuations, without losing valuable data from counts. Drawing from previous research (13, 12, 17, 14), we present simple analytical formulas that can be used to combine counts from different dilutions and to obtain precise CFU estimates along with accurate error bars. First, we examine existing methods in the literature, assessing their strengths and weaknesses. Next, we introduce the "Poisson with a cutoff" method, which clarifies the impact of crowding on CFU density estimation and demonstrates how to minimize the effects of sampling error by combining measurements of "uncrowded" counts. Finally, we use a crowding-explicit model to demonstrate the relationship between canonical plate-based counts and the Most Probable Number method for presence/absence of growth in liquid media. We conclude by providing practical recommendations for experimentalists on how to select appropriate dilution and replication schemes and how to combine data from multiple observations. We also have provided a calculator for these estimators available on Hugging Face spaces, named CFUestimator (18).

RESULTS

A Brief History of Counts Colony Forming Units (CFUs) are a proxy for the concentration of microbes within a sample. The experimental procedure for estimating CFUs consists of serially diluting homogeneous samples in a sterile aqueous buffer, then plating aliquots of these dilutions on growth-supporting agar and later counting the resulting colonies. If an appropriate dilution has been reached, each microbe will form an independent colony that is countable by eye. The simplest way of estimating CFUs is to multiply the number of colonies by the reciprocal of the dilution factor to find the concentration of colony-forming microbes in the original suspension (1, 5, 17). For example, say there is a single sample represented by one countable 10 cm plate in a dilution series, where we observe 100 distinct colonies after plating 100 μL of a 1:100 dilution (dilution 2 in a ten-fold series) from the original sample. In this case, following this simple procedure, we would obtain:

$$\frac{\text{CFU}}{\text{Volume}} = \frac{\text{counts}}{\text{FracOriginalVolume}} = \frac{100}{0.1\text{mL} \cdot 0.01} = 100 \cdot 10^3 = 10^5\,\text{CFU/mL}. \quad (1)$$

This is exactly equivalent to multiplying the number of counts by a volume correction factor (1/(size of aliquot in mL)) and multiplying by the base of the dilution series raised to the power of the number of dilution steps:

$$\frac{\text{CFU}}{\text{Volume}} = \frac{1}{0.1\text{mL}} \cdot 100\,\text{colonies} \cdot 10^2 = 100 \cdot 10^3 = 10^5\,\text{CFU/mL}. \quad (2)$$

This simple calculation follows from a more general Poisson model, explained below. This method works reasonably well under ideal conditions: all samples should be represented by a single count of colonies, and each count should be large enough to minimize small-number sampling fluctuations, and yet small enough to avoid crowding on the plate. When any of these conditions are not met, accurate estimation of CFU density becomes more complicated.

There is a broad literature of methods proposing to ensure that estimates of CFU density are "good". A *good* estimator should be accurate. Formally, this means that such estimators should have the true value of the CFU density as their expected value. In other words, they must be unbiased. *Good* estimators must also be precise, so that variance in the estimate is small and samples are repeatable. Therefore, an ideal solution to this problem should provide an estimator that is provably unbiased and with a

minimal variance. The solution to this problem is well known in statistics: if we can assume that data follows a specific probability distribution, then the *maximum likelihood estimator (MLE)* for that distribution will have these properties.(13) While this is formally true only for very large samples, MLE estimators generally perform well even for small samples. Further, an ideal method should be straightforward to use in the hands of researchers without advanced mathematics background. Unfortunately, many of the available methods fail one or the other of these requirements, being either simple to use, but statistically sub-optimal, or mathematically correct, but incomprehensible to many bench scientists.

Straight-forward to use methods focus largely on designing protocols that avoid data in error-prone extremes. For example, the FDA recommends (19) that the best dilution range for coliform bacteria results in 25 to 250 colonies per 10-cm plate, with the ideal count closest to 250. Restriction on the high end limits counting errors due to growth limitation by nutrient depletion as well as outright merging of colonies, which would bias the number of counts downward. Conversely, restriction on the lower end limits the sampling error associated with small numbers of counts. Specifically, under the assumption that counts represent random draws from a given sample and are, therefore, Poisson-distributed, the error scales as the square root of the number of counts. Thus, for small counts, the error becomes an unacceptably large fraction of the mean. Within the example above, our dilution 2 count of 100 colonies should have a standard deviation (SD) of $\sqrt{100} = 10$, giving a coefficient of variation (CV) of 10%. At dilution 3, we might obtain 10 counts, with a SD of $\sqrt{10} \approx 3.16$, and a CV of 31.6%.

From here, the simplest approach that is often used in practice is to choose only the plate or spot that has the "best" count in the acceptable range, and to estimate CFU density based on that single count. Often only the dilution at the high end of the countable range is used since it has the smallest sampling fluctuations; all other measurements are discarded (17). We call this the "pick-the-best" method for later reference. If counts in the acceptable range can be consistently achieved, this method is straightforward and reasonably accurate. However, discarding data is rarely advisable, and over- and under-crowded measurements can, in fact, be used to improve CFU estimates.

Simplest "Good" Estimator: Poisson One simple and reasonably accurate model for calculating CFUs assumes that the number of colonies are Poisson distributed, with variation due to sampling. That is, for a particular dilution, the mean colony count for that dilution is the same as the variance. This model ignores crowding effects but works well for modeling sampling fluctuations. By this model, the most likely estimator for the density of microbes is simply the ratio of the total number of colonies counted from all plates divided by the total amount of liquid used from the original sample in all plates (see *Supplementary Information*). If there is only one countable measurement for a given sample, this simplifies to "pick-the-best".

The Poisson model implicitly assumes that the original sample is well mixed and each microbe plated will result in its own separate and countable colony. It further assumes that experimental volume is spread uniformly across the agar surface, resulting in cells being randomly distributed, independent of the locations of where other cells landed. Formally, these assumptions mean that there is a uniform and well mixed population density r of microbes per unit volume in an initial volume of liquid V. The liquid is diluted by a factor $d_k = V_k/V$, where V_k is the volume of the liquid from the original sample used on the plate or the spot k. Plating will result in n_k colonies, where n_k is Poisson distributed with the parameter $\lambda = rd_kV = rV_k$. That is, the average number of colonies per experiment is rd_kV with variance rd_kV. Using these assumptions,

the MLE estimator of the density of microbes r_{mle} and its standard error are:

$$r_{\text{mle}} = \frac{\sum_k n_k}{V \sum_k d_k} = \frac{\sum_k n_k}{\sum_k V_k}, \qquad \sigma = \frac{r_{\text{mle}}}{\sqrt{\sum_k n_k}}. \tag{3}$$

In other words, the best estimator for the concentration, r_{mle}, is the total number of colonies divided by the total amount of the original volume of liquid used. However, as noted earlier, this ignores crowding and counting errors. In practice, this method should be avoided unless all measurements are from well-dispersed, uncrowded plates, as crowding effects can make a large difference in the estimator, resulting in underestimating the microbial density as colonies merge and are under-counted.

If technical replicates exist (multiple measurements of the same sample), it is straightforward to test whether the data adhere to a Poisson distribution using the following test, known as the dispersion index test. If there are j measurements of a given sample, with average number of counts $\bar{N}$ and variance of counts s_N^2, then the index of dispersion D^2 is:

$$D^2 = \frac{(j-1)s_N^2}{\bar{N}}, \tag{4}$$

which will be distributed as χ^2 with $j-1$ degrees of freedom (13). If D^2 is greater than the upper $1-\alpha$ quantile of that distribution, where α is the needed significance p-value, then we reject the null hypothesis that these replicates are drawn from the same Poisson distribution. This can indicate technical problems that are introducing an excess of variation, possibly by biasing replicates differently from one another (e. g., the failing O-ring example above), or biases due to a too-lenient cutoff for countability.

Combining Data: Common Bad Estimators The primary reason for the "pick-the-best" approach is that it eliminates confusion over how to combine multiple measurements for a given sample, particularly when counts belong to more than one dilution. First notice that combining measurements from technical replicates that are taken at the same dilution is straightforward. For example, let's assume an original 200 μL volume V contains $r = 3 \cdot 10^8$ CFU. We can create simulated serial dilutions from this original volume by assuming that each pipetting step (ten-fold dilutions and plating onto agar) is a binomial sampling event (14) that comes with experimental noise. In one such simulation, triplicate plating 100 μL aliquots results in counts n_6 =(162, 141, 148), all from the sixth ten-fold dilution. The fraction of the original volume plated in each case is $V_6 = 0.5 \cdot 10^{-6} = 5 \cdot 10^{-7}$. These numbers can be combined via the Poisson method shown in the previous section to estimate CFU density in V:

$$\text{CFU} = \frac{162+141+148}{5 \cdot 10^{-7} + 5 \cdot 10^{-7} + 5 \cdot 10^{-7}} = \frac{162+141+148}{3 \cdot 5 \cdot 10^{-7}} = 3.007 \cdot 10^8. \tag{5}$$

Alternately, counts taken from the same dilution can be averaged across technical replicates, then adjusted by the volume plated and the dilution read (20):

$$\text{CFU} = 2 \cdot \frac{162+141+148}{3} \cdot 10^6 = 2 \cdot \frac{162+141+148}{3 \cdot 10^{-6}} = \frac{162+141+148}{3 \cdot 0.5 \cdot 10^{-6}} = 3.007 \cdot 10^8. \tag{6}$$

Clearly, these two most common approaches are algebraically identical.

In contrast, combining counts across different dilutions is less straightforward. In fact, some commonly-used methods for combining measurements are statistically inadmissible. For example, if there are multiple measurements in the countable range, the USDA recommends (20) that researchers calculate the estimated CFU for each dilution separately using the average colony count across technical replicates at a given dilution and then average the results of the separate dilutions. If the two estimates are more than a factor of 2 apart, the researcher is instructed to instead only use the

counts from the higher-density plates. This commonly used method, incorrectly combines the data using a simple average, thus increasing the variance of the estimated CFU density. Indeed, continuing the example above, let's suppose that, on the plates corresponding to the seventh ten-fold dilution from these three technical replicates, we observe (13, 17, 20) colonies. The Poisson estimator gives us:

$$\mathrm{CFU} = \frac{162 + 141 + 148 + 13 + 17 + 20}{3(5 \cdot 10^{-7}) + 3(5 \cdot 10^{-8})} = 3.036 \cdot 10^{8}. \tag{7}$$

The USDA averaging method gives:

$$\begin{aligned} \mathrm{CFU} &= \frac{1}{2}\left(2\frac{162 + 141 + 148}{3} \cdot 10^{6} + 2\frac{13 + 17 + 20}{3} \cdot 10^{7}\right) \\ &= \frac{1}{2}(3.06 \cdot 10^{8} + 3.33 \cdot 10^{8}) = 3.2 \cdot 10^{8}. \end{aligned} \tag{8}$$

On these data, averaging was substantially less precise, with an error of 7% as compared with the Poisson method's error of 1% (recall that the true density in this simulated example is $3.0 \cdot 10^{8}$ CFU per 200 μL). The USDA method improperly averages across dilutions, amplifying fluctuations associated with small colony number counts, whereas the simple Poisson model properly combines measurements across dilutions by effectively re-weighting small counts by the small volumes with which they are associated. In a later section, we demonstrate that averaging across dilutions will, as a rule, increase the variance of CFU estimates.

Too Few and Too Many Further, there is the problem of what to do with zero counts. These data are inevitably limited by some threshold of detection (TOD), representing the smallest CFU density at which counts can be detected. This "left-censoring" is a well-known issue (21, 22, 23) with many proposed work-arounds, including but not limited to: substituting zeros with a small value (which may be the average of the undetectable range, a maximum-likelihood inferred value, or some other small number), reporting zeros as "below TOD" or "<1" rather than as a value, and pretending they didn't happen (not generally recommended; although if zeros are rare, it won't make much difference) (21, 23). Sometimes, a threshold of quantification (TOQ) representing the lowest "acceptable" (sufficiently precise) count is used along with or instead of TOD (17), with values below this threshold omitted from analysis.

The "best" approach to zero-contaminated count data depends on what else is in the data and what the data will be used to do. If a sample is represented by zero and non-zero measurements, the Poisson model explicitly allows zero counts to be incorporated as outcomes of the random sampling process. For example, if a hypothetical $V = 200\ \mu$L sample contains $5 \cdot 10^{7}$ CFU, one simulation of serial dilution and plating in triplicate with 100 μL per plate produces dilution-6 counts of (31, 26, 20) and dilution-7 counts of (4, 0, 0). Using just the dilution-6 counts, we estimate

$$\mathrm{CFU} = \frac{31 + 26 + 20}{3(5 \cdot 10^{-7})} = 5.13 \cdot 10^{7} \pm 0.59 \cdot 10^{7}. \tag{9}$$

If we use the lower dilution as well, we obtain

$$\mathrm{CFU} = \frac{31 + 26 + 20 + 4 + 0 + 0}{3(5 \cdot 10^{-7}) + 3(5 \cdot 10^{-8})} = 4.91 \cdot 10^{7} \pm 0.55 \cdot 10^{7}. \tag{10}$$

In this case, incorporating data from zeros (in the form of the additional volume that was plated but contained no counts) improved precision. Alternately, when zeros are common because the density in the original sample is close to the TOD, non-zero counts are useful for making a distinction between samples where no organisms are

detectable (and density might be zero) and those where the density of organisms cannot be zero. Although the actual density cannot be estimated accurately or precisely from very low counts, the distinction between "<TOD" and ">1" for a given sample is important (23).

At the other end of the range, researchers must deal with crowding and set thresholds for "too many to count". Defining an optimal range for "countable" data is not always straightforward, and this determination is very important to ensure that CFU estimates are accurate. Since the sampling-based standard error of counts scales as $\frac{n_k}{\sqrt{n_k}}$, the number of colonies counted n_k should be as large as reasonably possible.

However, there are consequences for pushing this too far. As cell density in the aliquot increases, counts will be biased downwards due to merging of colonies and colony stunting or failure to grow. These data are then "right-censored", with an upper limit past which the number of counts observed does not increase in proportion to an increase in the density in the original sample. Densities above this point result in "crowded" samples, with counts that are lower than the true number of colony forming units. Further, as the number of colonies per plate or spot increases, data collection becomes more time-consuming; it is common for researchers to minimize effort on plates near the top of the "acceptable" range by dividing plates into sections, counting colonies in one section, and multiplying this count by the number of sections to get an estimated final count for the whole plate (19). While this approach is sufficient for a rough estimate of CFU density, it introduces additional sampling variation due to both reduction in counts and imperfect division of plates, and it does not remove bias due to crowding. We will demonstrate the consequences later in this paper.

Previous works (12) have modeled crowding using shifted Poisson distributions. In these models, the variance of estimates from crowded data would be the same as if there was no crowding and the mean would be shifted down due to colonies merging together. However, this is *a priori* unlikely to be true. As we will show below, if colonies are crowded, both the mean and the variance will be shifted relative to the pure Poisson (uncrowded) distribution. The reason for this is that the variance of the large colony counts is shifted downward due to a "ceiling" effect—there is an upper bound to the total number of colonies, which limits upwards fluctuations. In other words, the use of a shifted Poisson distribution is a reasonable approximation, but the variance must also be modified.

Better Estimators: Poisson With Cutoff, aka What's Countable, Exactly? The main problem with the naive Poisson model is that it does not account for counting errors due to crowding. The simplest way to take account of the crowding is to assume that there is a threshold of colonies, M, below which crowding is negligible, which in practice will often be smaller than the largest number of counts we are willing to attempt. We can then segment our data into two parts: plates with counts above the threshold where crowding is important, and plates with counts below the threshold where crowding is not important. If we have identified our cutoffs well, the naive Poisson estimator above is correct for all measurements where the number of colonies counted $n_k \leq M$. The calculation is, therefore, exactly the same as for the naive Poisson estimator given above, with the caveat that only measurements $n_k \leq M$ are used. Here, the indicator function $I(n_k < M)$ is 1 when $n_k < M$, and 0 otherwise. Similarly, $I(n_k > M)$ is 1 when $n_k > M$, and 0 otherwise. Due to its balance between simplicity and accuracy, this method is the easiest to use in practice.

$$r_{\text{mle}} = \frac{\sum_k I(n_k \leq M) n_k}{\sum_k I(n_k \leq M) V_k}, \qquad \sigma = \frac{r_{\text{mle}}}{\sqrt{\sum_k I(n_k \leq M) n_k}}. \tag{11}$$

If we want to incorporate data from measurements above this threshold M, the calculation becomes slightly more complicated. Using "crowded" measurements as if they were uncrowded will bias the naive Poisson estimator downward, resulting in underestimation of CFU density (Fig. 2). In a *sophisticated* version of this model, we can use the number of plates/spots that were above the crowding threshold M, along with the colony counts from plates/spots below this threshold at the same dilution, to estimate CFUs. This will be applicable when plate counts at a given dilution are toward the high end of the countable range, such that some technical replicates fall below this threshold and others above it by chance. To estimate the CFU density in the original sample r, the following equation should be solved numerically (see SI for the derivation):

$$\sum_k I(n_k \leq N)\left(\frac{n_k}{r} - d_k V\right) + \sum_k I(n_k > N)\frac{d_k V (d_k r V)^N e^{-d_k r V}}{\int_0^{d_k r V} t^N e^{-t} dt} = 0. \tag{12}$$

The first term is equivalent to the simple Poisson model and uses the counts from uncrowded samples directly, whereas the second term reflects the probability of counts being above the threshold M. Inference of r can be done in Excel using SOLVER or using numerical solvers in R, Python, MATLAB, etc. An equivalent model is shown in (13).

This model properly accounts for two error sources: the sampling fluctuations and the crowding effect. The simple Poisson, using only counts from uncrowded plates, gives a good estimate for the CFU counts and properly combines multiple measurements at different dilution factors. The more sophisticated form of the model has greater precision, but the greater computational effort may or may not be worth it to an investigator depending on the effect size and the structure of the experiment. In the next section, we present an alternate estimator based on the Most Probable Number approach, which we argue provides a better trade-off between effort and estimator performance when incorporating data from crowded samples.

Crowding and the Most Probable Number For the final model we consider the effects of crowding in space. To account for crowding, we will divide each plate into $N \approx \frac{A_{\text{plate}}}{A_{\text{colony}}}$ regions, each approximately the size of a full colony. We make the assumption that if more than one microbe lands in one of these regions, the colonies that form from these cells will grow together and be counted as one colony. For each region, the number of cells landing in that region will be Poisson distributed with parameter $\lambda = \frac{d_k r V}{N}$.

These assumptions are equivalent to that of quantal-based methods for microbial quantification, such as the commonly used Most Probable Number (MPN) method. In the MPN assay, a known quantity (volume of original sample) is introduced into each of a series of N replicate tubes, and the dilution of the original sample is adjusted to find a region where some of the tubes contain viable growth and some do not. The results of this assay are therefore, for each dilution volume V_k from the original sample, out of the N_k tubes inoculated, a number n_k that is positive for growth.

A direct mapping to tube-based assays is possible if space on a plate (or within a spot) is considered as a set of colony-sized bins. Each of the N colony-sized regions on a plate or within a spot corresponds to one tube. The presence of colonies in a particular region corresponds to when a tube has growth. Hence a plate that is divided into N regions can be thought of as N tubes being tested in parallel, cf. Figure 1.

Therefore, the probability of n_k successes in N colony-size regions on the agar surface can be described using a crowding-explicit model based on the binomial distribution. Assuming that the cells in the original sample are well-mixed, the probability

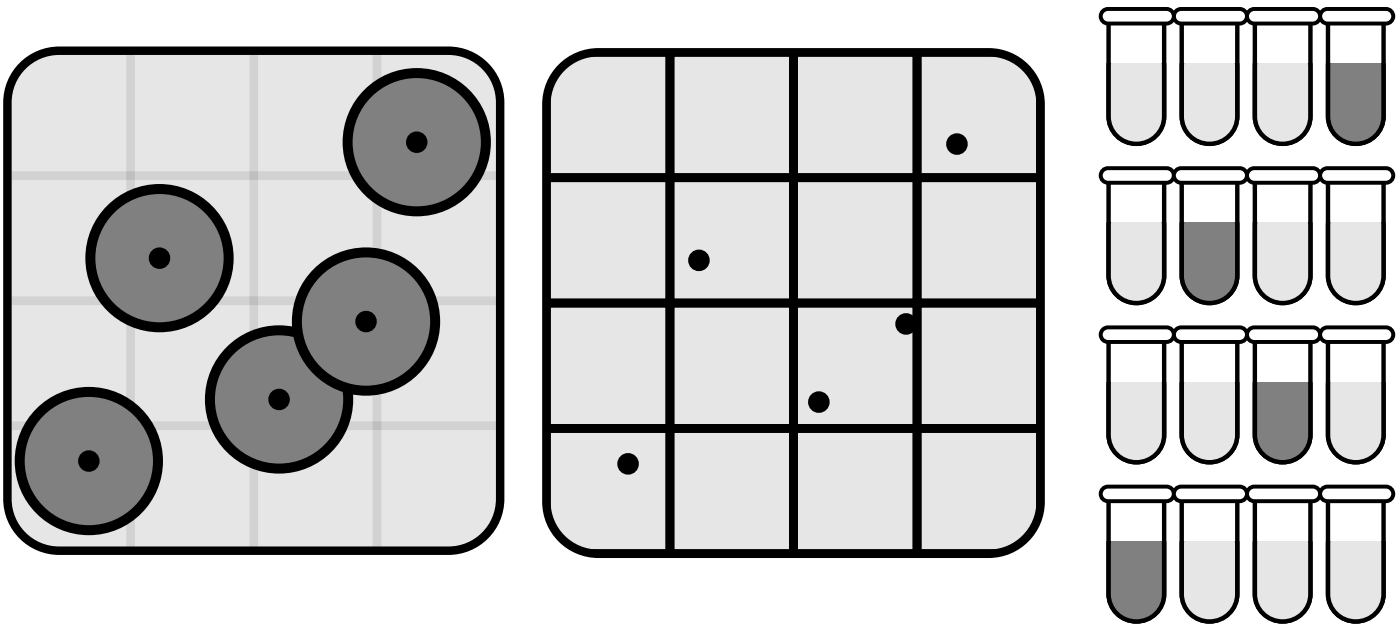

FIG 1 Visual equivalence between plate and tube based assays. The left panel is a cartoon of a typical plate containing colonies, where the growing colonies are shown as dark disks. In the middle panel, the plate is divided into N (here 16) approximately colony-sized regions. If a region contains one or more colony centers (black dots), this region can be mapped to a positive (dark) tube as shown in the right panel. Similarly regions containing no colony centers are mapped to negative (light) tubes. This demonstrates that plating is equivalent to a massive parallel version of a tube based assay with $N \approx \frac{A_{\text{plate}}}{A_{\text{colony}}}$ tubes. Furthermore it demonstrates that the MPN method can be used for plate data.

of zero cells landing in a particular region is (from the Poisson) $p_0 = e^{\frac{-d_k rV}{N}}$ and the probability that at least one cell lands in that region is therefore $p_> = 1 - e^{\frac{-d_k rV}{N}}$. Assuming that the original sample is well mixed, each region is independent of all other regions in our crowding model, so that

$$p(n_k) = \binom{N}{n_k} p_>^{n_k} p_0^{N-n_k} = \binom{N}{n_k} (1 - e^{\frac{-d_k rV}{N}})^{n_k} e^{-d_k rV(N-n_k)}. \tag{13}$$

We can maximize this probability to find the MLE CFU density, r_{mle} (see the SI fo the full derivation). We can accomplish this by numerically solving the following equation for r:

$$\sum_k \frac{d_k n_k}{N(1 - e^{-r d_k V/N})} = \sum_k d_k. \tag{14}$$

This expression for r is the same as that of the MPN estimator(19, 24). In the SI we show that, in the limit where concentrations and colony counts are low, this model simplifies to the Poisson model. Outside the "uncrowded" regime, the mean and the variance of data from the crowding model are not the same as in the Poisson. Therefore, the two approaches are not equal to each other, though both are depressed due to the "ceiling" effect described earlier. In the SI, we also find that the error associated with the maximum likelihood estimator r_{mle} of the MPN method can be minimized at an optimal dilution factor, which falls into the crowded regime.

The MPN procedure can generate biased estimates of the original sample density, and the precision and accuracy of results depend strongly on the number of tubes used (13). The bias on the maximum likelihood estimator results in an over-estimate of 20-25% with 5 tubes, which is reduced to a few percent with 50 tubes (see SI). By back of the envelope calculation, an average 10 cm plate (inside diameter 86 mm,

surface area 58 cm^2) can fit a maximum of approximately 5000 medium-sized (1 mm outside diameter) "tubes", whereas a single grid square on a 10 x 10 cm square plate (typically gridded 6 x 6) can fit 200 of these colony-sized spaces. All of these are well above the threshold where the bias in this estimator (25) makes much difference in the value. (Note that this refers to the number of *colony-sized spaces* available and is independent of the number of colonies observed.) This also means that the standard error of the estimator will, in theory, be minimized at a plating density that is much higher than the threshold for "uncrowded" plates and, in fact, is well into a range of densities where a minority of colonies will be distinct. Fortunately, the standard error is still well behaved over a broad space in fraction of regions occupied (SI), meaning that plate counts into the "uncrowded" range will still produce good estimates with this method. In fact, this produces a result equivalent to that of the Poisson method in the fully uncrowded regime. However, the MPN method is most useful as plating densities encroach into the crowded regime, allowing precise and accurate estimation of CFU density from plates that would provide severely biased estimates using a naive Poisson model.

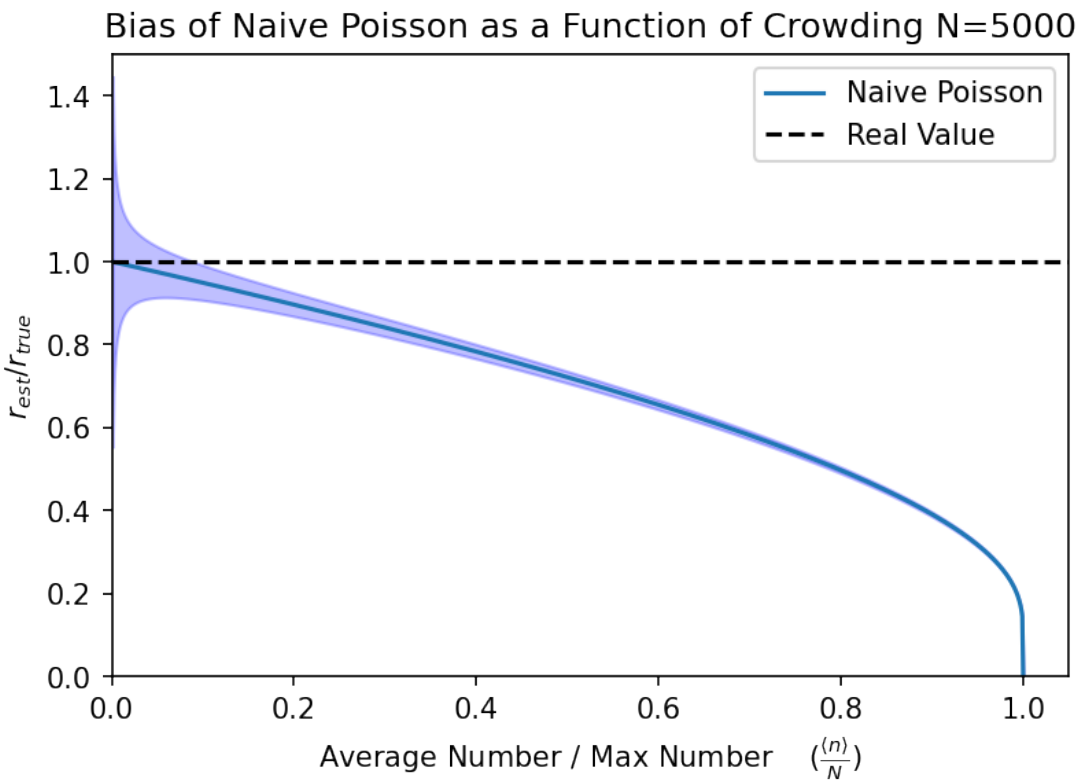

FIG 2 The naive Poisson estimator underestimates the true concentration and becomes more biased as a function of crowding. We illustrate this by plotting the ratio of the estimated concentration (with the error bands denoting $\pm$ one s. e. m. at $N = 5000$) to the true concentration. Here crowding is measured by the ratio of the average number of colonies to the maximum number of colonies that can fit within a plate $f = \frac{\langle n \rangle}{N}$. At low crowding values, the naive estimator has low bias, but large uncertainty. At a crowding value of 0.2 the naive-Poisson estimator underestimates the true concentration by about 10%, and many-fold underestimation is possible as crowding approaches 1.

Utility of the Models Here we demonstrate the relative utility of each model for estimation of CFU density from simulated data. First, we can use the crowding-explicit binomial sampling model described in the previous section, to estimate bias due to

crowding, and to demonstrate the importance of choosing an appropriate cutoff M, below which plates are considered to be uncrowded and countable. To do so, we solve the crowded binomial model in Eq. 13 for dV with respect to the average number of colonies $\langle n \rangle$ and the number of colony-sized regions on a plate N. Doing so we find $dV = \frac{-N}{r} \log(1 - \frac{\langle n \rangle}{N})$. We can substitute this into the Poisson estimator and find:

$$r_p = \frac{\langle n \rangle}{dV} = \frac{\langle n \rangle}{\frac{-N}{r} \ln\left(1 - \frac{\langle n \rangle}{N}\right)} = -r \frac{\frac{\langle n \rangle}{N}}{\ln\left(1 - \frac{\langle n \rangle}{N}\right)}. \tag{15}$$

Let us define the ratio of expected colony number to the number of colony-sized regions as $f = \frac{\langle n \rangle}{N}$. This ratio represents the amount of crowding, where a value of 1 is the maximum crowding and a value close to zero is in the uncrowded regime. Expressing the previous expression in terms of the crowding we see

$$\frac{r_p}{r} = -\frac{f}{\ln(1 - f)}. \tag{16}$$

This ratio indicates how close the estimated CFU concentration is to the true concentration. A ratio of 1 tells us that we have an unbiased estimator, whereas a ratio of less than 1 tells us we are underestimating the CFU density. We plot this expression in Fig. 2 to show how the simple Poisson estimator underestimates the actual concentration as a function of crowding, f. After a crowding value of $f = 0.2$ the Poisson estimator starts to be significantly biased, undershooting the true value by about 10%. This has implications for the value used in the Poisson model with a cutoff. The cutoff should be chosen such that the bias is not greater than the experimenters targeted precision. For example, if a bias must be less than 10%, then a cutoff of about 20% of the total plate capacity should be used. In the case of a 10 cm plate with an estimated 5000 1mm diameter colony-sized regions, this corresponds to a cutoff of $M = 1000$, whereas the more typical cutoff of $M = 300$ provides an essentially unbiased estimate (bias 3%), but this results in a large statistical fluctuation of 5.8%. In the case of 6mm grid grid on a 10cm by 10cm plate, there are roughly 200 grid regions in a plate. Thus an $M = 40$ would be appropriate to achieve the bias less than 10%, and a threshold of $M = 12$ colonies is required to reduce bias to 3% for colonies of this size. At these thresholds, the statistical error would be 15.8%.

To compare the performance of the different estimators discussed here, we simulated 1000 experiments and applied each of our estimators to the resulting data. Data for each experiment was modeled using the binomial crowding model with $r = 10^5$, $V = 0.2$, $N = 5000$, and dilution values (0.1, 0.1, 0.01, 0.01, 0.001, 0.001). This corresponds to two replicates for each dilution in a tenfold dilution experiment. An example set of colony counts corresponding to these dilutions is (1705, 1629, 196, 181, 21, 21). The first two dilutions are in the over-crowded regime and the last two dilutions are in the dilute uncrowded regime. The traditional methods ("pick-the-best", averaging, segment averaging) and Poisson with a cutoff will discard the first two counts as too many to count, while the other methods will use their numeric values. The resulting distributions are plotted in Fig. 3.

The results show that the MPN (most probable number) method is unbiased and has the highest degree of accuracy. The Poisson with a cutoff (which always discards counts from the least-diluted samples in these outputs) is nearly unbiased, whereas the naive Poisson is biased down due to inclusion of "crowded" data. The naive Poisson has a similar variance as that of the MPN because both are using all the data points. However, the measure around which the naive Poisson estimator varies is incorrect due to this bias. With the Poisson estimator, increasing accuracy comes at a

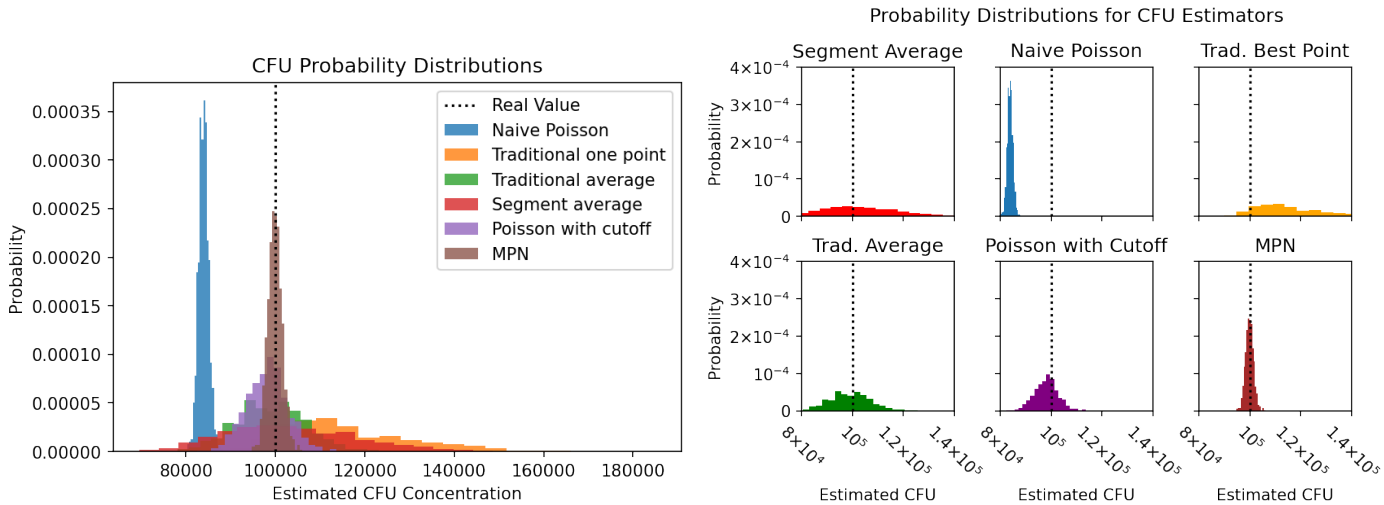

FIG 3 The probability distributions of estimated CFU concentrations from different estimators generated from 1000 independent numerical experiments with dilutions 0.1, 0.1, 0.01, 0.01, 0.001, 0.001, $r = 100000$**,** $V = 0.2$**,** $N = 5000$**. Here the Segmented average, naive Poisson, "pick the best", traditional average, Poisson with cuttoff, and MPN methods are compared. The MPN method demonstrates the best combination of high precision and accuracy.**

cost in precision; the Poisson with cutoff has roughly twice the standard error of the MPN method due to the fact that it does not use all the data and throws out the first two counts of each experiment. Next, the the traditional averaging method (20) has roughly five times the standard error of the MPN method, due to the fact that it gives lower-precision measurements the same weight as higher-precision large counts in the uncrowded regime. However, it is unbiased. If there are technical replicates, pick-the-best (choosing the largest number of counts in the countable range, over multiple technical replicates at each dilution) is a biased estimator (overestimating CFUs) and has a standard error roughly ten times that of the MPN method. (Pick-the-best where the best count from *each* technical replicate is used is equivalent to Poisson with a cutoff, with some loss of precision due to discarding of small counts.) Segment averaging (here, counting one-quarter of the plate, and assuming perfect segmenting such that exactly one-quarter of the colonies are counted) resulted in an unbiased estimator with the largest standard error, roughly 13 times the standard error of the MPN method.

These simulations show that the MPN method produces the most precise results and is unbiased. However, the Poisson with a cutoff is a close second, also with high accuracy and precision and with the advantage of being practical to calculate by hand. The bias of the naive Poisson (using all data) serves as a warning: if counts are not in the uncrowded regime, the Poisson assumptions do not apply, and an estimator using only number of colonies counted at each dilution will under-estimate the CFU density in the original sample. Other standard estimators (averaging, segment averaging) using the same data required for the Poisson estimator show universally poorer precision than Poisson with a cutoff and cannot be recommended.

CONCLUSION

We have presented several methods for estimating CFUs and we have provided a calculator for these estimators available on Hugging Face spaces, named CFUestimator (18). In practice, the choice of method will depend on the precision required for the

estimate of CFU density. For experiments with reasonably large expected effect size, the simplest mathematically admissible method - the Poisson estimator with a cutoff - is perfectly valid, as long as the dilutions are chosen appropriately to ensure all measurements are in the countable range. Broadly speaking, addition of unbiased data will improve the precision of an estimator. Historically, technical replicates have been used for this purpose - even technical duplication is sufficient to markedly reduce variance of the estimated CFU density, although triplicate plating is preferred to safeguard against accidents and outliers (26) (also see SI). The Poisson model allows data from technical replicates to be combined into a single mathematically interpretable estimator with definable properties - specifically, a maximum likelihood estimator, which should be an unbiased and minimally variable estimator for the true value. This is as opposed to averaging (20), which produces an estimate whose properties are not well defined. The Poisson method also allows the investigator to incorporate data from dilutions with too few counts, *in addition to* (not in place of) data from countable wells in the same dilution series - by effectively re-weighting the contribution of these wells by the total volume of original suspension that they contain, these data can be used to improve the accuracy of the estimator even though their sampling variance is high.

The correspondence shown here between using tubes and gridding a plate into subsections based on colony area allows the usage of estimator techniques typically used for quantal-based measurements of CFU density, specifically the MPN, where positive growth events (e. g., colonies) are explicitly considered to represent *one or more* originating cells. These techniques have a long history in environmental surveillance microbiology, and statistically well-founded techniques are readily available for analysis of such data (2, 27, 28). If an experimentalist wants tighter bounds for an estimated CFU count, the MPN provides a very low-variance, unbiased estimator at the cost of some extra steps. This estimator allows the experimentalist to incorporate data from normally uncountable (TMTC) plates as well as counts from uncrowded plates, maximizing the amount of information that can be gleaned from a dilution series.

The MPN model requires an estimate of the maximum number of colonies that can be packed into the growth area for each sample; we show (SI) that it is better to over-estimate this maximum than to under-estimate it. If the patch size on a plate is correctly chosen to be around the size of a typical colony, even a spot-plating assay on a 10 by 10 cm plate is equivalent to running hundreds of tubes in parallel. Further, it is necessary to estimate the number of occupied regions in the growth area. In or near the uncrowded regime, this will be equivalent to the number of counts. However, this method does not require that all colonies are individually countable - instead, image analysis(29, 30, 31) can be used to estimate both the size of an individual colony and the fraction of total area occupied by colony growth. The MPN estimator can therefore potentially provide accurate, precise estimates of CFU density for plates where exact counts cannot be obtained. However, colony size varies across different microorganisms as well as across culture conditions (media type, agar percentage, pad thickness, plate drying time and conditions, growth temperature and atmosphere, etc.) and incubation time on plates, meaning that the size range of colonies may be different even across plates within a single experiment (32, 33). This added complication of properly choosing a grid size or determining the typical size of a colony means that application of the MPN will most likely require parameters estimated for the specific experiment being analyzed. Further, the fact that colony size can decrease under crowding means that heavily-crowded plates or plate regions, where few or no distinct colonies are visible, may have very different “average” colony sizes than the same microbes in a less-crowded area. While theory suggests that the MPN estimator will be most pre-

cise when the majority of colony-sized locations are occupied ((34), also see SI), this practical limitation suggests that use of the MPN on plate count data will become less accurate with extremes of crowding, and that the best use of the MPN is likely to be in the liminal region between the technically uncrowded and the physically uncountable, where most to all growth is in the form of distinct, countable colonies but crowding produces a measurable bias in these counts.

ACKNOWLEDGEMENTS

This work was funded, in part, by NSF Grant No. 2014173. Ilya Nemenman was additionally supported by the Simons Investigator program.

REFERENCES

1. **McCrady MH**. 1915. The numerical interpretation of fermentation-tube results. The J Infect Dis p 183–212.
2. **Cochran WG**. 1950. Estimation of bacterial densities by means of the" most probable number". Biometrics 6 (2):105–116.
3. **Breed RS, Dotterrer W**. 1916. The number of colonies allowable on satisfactory agar plates. J bacteriology 1 (3):321–331.
4. **Fisher R, Thornton H, Mackenzie W**. 1922. THE ACCURACY OF THE PLATING METHOD OF ESTIMATING THE DENSITY OF BACTERIAL POPULATIONS: WITH PARTICULAR REFERENCE TO THE USE OF THORNTON'S AGAR MEDIUM WITH SOIL SAMPLES. Ann Appl Biol 9 (3-4):325–359.
5. **Halvorson H, Ziegler N**. 1933. Application of statistics to problems in bacteriology: I. A means of determining bacterial population by the dilution method. J Bacteriol 25 (2):101–121.
6. **Jennison MW, Wadsworth GP**. 1940. Evaluation of the errors involved in estimating bacterial numbers by the plating method. J Bacteriol 39 (4):389–397.
7. **Jones P, Mollison J, Quenouille M**. 1948. A Technique for the Quantitative Estimation of Soil Micro-organisms: With a Statistical Note by. Microbiology 2 (1):54–69.
8. **Skinner F, Jones P, Mollison J**. 1952. A comparison of a direct-and a plate-counting technique for the quantitative estimation of soil micro-organisms. Microbiology 6 (3-4):261–271.
9. **Johnson EA, Brown Jr BW**. 1961. The Spearman estimator for serial dilution assays. Biometrics p 79–88.
10. **Taylor J**. 1962. The estimation of numbers of bacteria by tenfold dilution series. J Appl Bacteriol 25 (1):54–61.
11. **Harris R, Sommers L**. 1968. Plate-dilution frequency technique for assay of microbial ecology. Appl microbiology 16 (2):330–334.
12. **Ben-David A, Davidson CE**. 2014. Estimation method for serial dilution experiments. J microbiological methods 107:214–221.
13. **Haas CN, Rose JB, Gerba CP**. 2014. Exposure Assessment, p 159–234. *In* Quantitative Microbial Risk Assessment. John Wiley & Sons, Ltd. doi:10.1002/9781118910030.ch6. Section: 6 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118910030.ch6.
14. **Christen JA, Parker AE**. 2020. Systematic statistical analysis of microbial data from dilution series. J Agric Biol Environ Stat 25 (3):339–364.
15. **Hedges AJ**. 2002. Estimating the precision of serial dilutions and viable bacterial counts. Int journal food microbiology 76 (3):207–214.
16. **Luria SE, Delbrück M**. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28 (6):491.
17. **Sutton S**. 2011. Accuracy of plate counts. J validation technology 17 (3):42–46.
18. **Martini KM**. Calculate Colony Forming Units (CFUs) https://huggingface.co/spaces/KMichaelMartini/CFUestimator.
19. **for Food Safety C, Nutrition A**. BAM appendix 2: Most probable number from serial dilutions. FDA https://www.fda.gov/food/laboratory-methods-food/bam-appendix-2-most-probable-number-serial-dilutions.
20. **USDA**. Quantitative Analysis of Bacteria in Foods as Sanitary Indicators. USDA https://www.fsis.usda.gov/sites/default/files/media_file/2021-03/MLG-3.pdf.
21. **Canales RA, Wilson AM, Pearce-Walker JI, Verhougstraete MP, Reynolds KA**. Oct 2018. Methods for Handling Left-Censored Data in Quantitative Microbial Risk Assessment. Appl Environ Microbiol 84 (20):e01203–18. doi:10.1128/AEM.01203-18. Publisher: American Society for Microbiology.
22. **Gijbels I**. 2010. Censored data. WIREs Comput Stat 2 (2):178–188. doi:10.1002/wics.80. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.80.
23. **EPA C**. 2014. Evaluation of Options for Interpreting Environmental Microbiology Field Data Results having Low Spore Counts. Environmental Protection Agency.
24. **Alexander M**. 1983. Most probable number method for microbial populations. Methods Soil Anal Part 2 Chem Microbiol Prop 9:815–820.
25. **Salama IA, Koch GG, Tolley DH**. Jan 1978. On the estimation of the most probable number in a serial dilution experiment. Commun Stat - Theory Methods 7 (13):1267–1281. doi:10.1080/03610927808827710. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03610927808827710.
26. **Weenk GH**. Oct 1992. Microbiological assessment of culture media: comparison and statistical evaluation of methods. Int J Food Microbiol 17 (2):159–181. doi:10.1016/0168-1605(92)90113-H.
27. **Garthright WE**. 1993. Bias in the logarithm of microbial density estimates from serial dilutions. Biom J 35 (3):299–314.
28. **Loyer MW, Hamilton MA**. Dec 1984. Interval estimation of the density of organisms using a serial-dilution experiment. Biometrics 40 (4):907–916. doi:10.2307/2531142. MAG ID: 2317336612.
29. **Chiang PJ, Tseng MJ, He ZS, Li CH**. 2015. Automated counting of bacterial colonies by image analysis. J microbiological methods 108:74–82.
30. **Brugger SD, Baumberger C, Jost M, Jenni W, Brugger U, Mühlemann K**. 2012. Automated counting of bacterial colony forming units on agar plates. PloS one 7 (3):e33695.
31. **Bewes J, Suchowerska N, McKenzie D**. 2008. Automated cell colony counting and analysis using the circular Hough image transform algorithm (CHiTA). Phys Med & Biol 53 (21):5991.
32. **Savage GM, Halvorson HO**. Mar 1941. The Effect of Culture Environment on Results Obtained with the Dilution Method of Determining Bacterial Population. J Bacteriol 41 (3):355–362. doi:

10.1128/jb.41.3.355-362.1941. MAG ID: 108945167.

33. **Chacón JM, Möbius W, Harcombe WR**. Mar 2018. The spatial and metabolic basis of colony size variation. The ISME J 12 (3):669–680. doi:10.1038/s41396-017-0038-0.
34. **Strijbosch LW, Buurman WA, Does RJ, Zinken PH, Groenewegen G**. 1987. Limiting dilution assays: experimental design and statistical analysis. J immunological methods 97 (1):133–140.
35. **Hurley MA, Roscoe M**. 1983. Automated statistical analysis of microbial enumeration by dilution series. J applied bacteriology 55 (1):159–164.
36. **Clarke K, Owens N**. 1983. A simple and versatile micro-computer program for the determination of 'most probable number'. J Microbiol Methods 1 (3):133–137.
37. **Herigstad B, Hamilton M, Heersink J**. 2001. How to optimize the drop plate method for enumerating bacteria. J microbiological methods 44 (2):121–129.
38. **Jett BD, Hatter KL, Huycke MM, Gilmore MS**. 1997. Simplified agar plate method for quantifying viable bacteria. Biotechniques 23 (4):648–650.
39. **Kirchman D, Sigda J, Kapuscinski R, Mitchell R**. 1982. Statistical analysis of the direct count method for enumerating bacteria. Appl Environ Microbiol 44 (2):376–382.
40. **Brewster JD**. 2003. A simple micro-growth assay for enumerating bacteria. J Microbiol Methods 53 (1):77–86.
41. **Tremaine SC, Mills AL**. 1987. Tests of the critical assumptions of the dilution method for estimating bacterivory by microeucaryotes. Appl environmental microbiology 53 (12):2914–2921.
42. **Gronewold AD, Wolpert RL**. 2008. Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration. Water research 42 (13):3327–3334.
43. **Horvitz DG, Thompson DJ**. 1952. A generalization of sampling without replacement from a finite universe. J Am statistical Assoc 47 (260):663–685.