PDA Journal
of Pharmaceutical Science and Technology

In-Process Microbial Testing: Statistical Properties of a Rapid Alternative to Compendial Enumeration Methods

Emil M. Friedman, Mark Warner, Sam C. Shum, et al.

PDA J Pharm Sci and Tech **2015**, *69* 264-269
Access the most recent version at doi:10.5731/pdajpst.2015.01019

RESEARCH

In-Process Microbial Testing: Statistical Properties of a Rapid Alternative to Compendial Enumeration Methods

EMIL M. FRIEDMAN*, MARK WARNER, SAM C. SHUM, and FRED ADAIR

MannKind Corporation, One Casper Street, Danbury, CT ©PDA, Inc. 2015

ABSTRACT: In-process tests are used between manufacturing steps to avoid the cost of further processing material that is apt to fail its final tests. Rapid microbiological methods that return simple negative or positive results are attractive in this context because they are faster than the compendial methods used at product release. However, using a single such test will not reliably detect barely unacceptable material (sensitivity) without generating an undesirable number of false rejections (poor specificity). We quantify how to achieve a balance between the risks of false acceptance and false rejection by performing multiple rapid microbiological methods and applying an acceptance rule. We show how the end user can use a simple (and novel) graph to choose a sample size, the number of samples, and an acceptance rule that yield a good balance between the two risks while taking cost (number of tests) into account.

KEYWORDS: Alternative microbial method, Binomial distribution, Colony-forming units (CFU), False alarm rate, Most probable number (MPN), Operating Characteristic (OC) curve, Poisson distribution, Rapid microbiological method (RMM), Sensitivity, Specificity.

LAY ABSTRACT: In-process tests are used between manufacturing steps to avoid the cost of further processing material that is apt to fail its final tests. Rapid microbiological methods that return simple negative or positive results are attractive in this context because they are faster than the compendial methods used at product release. However, using a single such test will not reliably detect barely unacceptable material (sensitivity) without generating an undesirable number of false rejections (poor specificity). We quantify how to achieve a balance between the risks of false acceptance and false rejection by performing multiple rapid microbiological methods and applying an acceptance rule. We show how the end user can use a simple (and novel) graph to choose a sample size, the number of samples, and an acceptance rule that yield a good balance between the two risks while taking cost (number of tests) into account.

Introduction

A method for estimating the most probable number (MPN) of microorganisms (colony-forming units, or CFUs) in a batch of material has been in use for almost a century (*1–4*). It involves using several different size samples (or *dilutions*) of the batch. One then determines how many samples of each of the different sizes are positive for the presence of microorganism(s), but, unlike compendial plate count methods—for example, USP General Chapter <61> (*5*)—one *cannot* determine how many CFUs had originally been present in any particular sample.

When performing an in-process test, one does not necessarily need a numerical estimate of the CFU concentration in the batch. Instead, one may merely need assurance at some specified level of confidence that the concentration does not exceed a particular acceptance limit. This is because the manufacturer simply wants to decide whether to expend resources on further processing steps and rely on a more time-consuming compendial method—for example, USP General Chapter (*5*)—when testing the final product prior to release.

In such a case, a streamlined and more economical test is possible. Miller (*6*) proposed such a rapid microbiological method (RMM) using a technique

*Corresponding Author: Emil M. Friedman, Telephone: 203-790-2507; e-mail: efriedman@mannkindcorp.com
doi: 10.5731/pdajpst.2015.01019

called *dilute to specification* wherein a sample is diluted so that an average sample will have one CFU if the batch is at its acceptance limit. However, as we will show, that method only has a 63% (1 – 1/*e*) chance of rejecting a batch that just barely exceeds its acceptance limit.

This contribution refines Miller's (*6*) proposal using the same sort of probabilistic logic that Cochran (*3*) used. In particular, we propose a series of RMM tests, but unlike the MPN method, all the tests in the series use a common level of dilution. Acceptance or rejection of moving the batch to the next step in the process hinges on whether a sufficient number of tests in the series fail to detect CFUs (the *acceptance rule*). A *test scheme* is defined by its level of dilution, the number of tests performed, and its acceptance rule. We theoretically derive the *sensitivity* (ability to detect batches that exceed their acceptance limit) and false alarm rate (*specificity,* a common statistical term that simply refers to the probability that a batch that is truly below the acceptance limit will correctly be accepted) of any such test scheme, and point out several practical and impractical test schemes. By comparing the sensitivity, false alarm rate, and cost of the different test schemes, a scheme appropriate for the in-process situation can be chosen.

As with Cochran's (3) explanation of the MPN method, we assume that (i) the organisms are distributed randomly throughout the batch; (ii) if a sample contains one or more microorganisms, the RMM is certain to detect it; (iii) if a sample does not contain microorganisms, the RMM will not return a false positive; and (iv) each test is performed independently. Violation of assumption (ii) will lower the sensitivity. Violation of assumption (iii) will raise the false alarm rate. No method short of 100% testing is likely to detect a huge number of CFUs concentrated in a miniscule portion of a batch (a gross violation of assumption (i).

Ideally, independence means that each test is done separately *and* that each sample is taken separately. For example, taking a single sample from a batch, diluting it, and then subdividing the dilutions would be a gross violation of independence. Using a multichannel pipette to withdraw a dozen samples from a liquid batch would violate independence because all the samples would be drawn from a relatively small portion of the batch and they might tend to all have lower or higher CFU concentrations than the batch as a whole. In addition, the errors in their volumes might tend to be correlated. In practice, judgment calls are often made. For example, requiring each test to be made on a different day by a different technician using a different piece of equipment would be impractical.

This paper derives a statistical model that calculates the sensitivity and false alarm rate of any chosen test scheme and then applies it to a variety of practical schemes.

Statistical Model

Suppose that the batch contains, on average, B CFUs per gram and one takes a random sample containing Q grams. In practice, a 0.01 g sample might be obtained by diluting a 1 g sample to a total volume of 100 mL and then testing a 1 mL sample of that dilution.

The number of CFUs in such a sample will be Poisson-distributed (*7, 8*), so the expected number of CFUs per sample is BQ and the probability that a particular sample contains *no* CFUs is $p = \exp(-BQ)$ where "exp" denotes the natural exponential function.

Now suppose we *independently* take n such random samples. The probability that k or fewer of them contain at least one CFU (and that at least $n - k$ contain no colony forming units) can be calculated using the binomial probability distribution (*3, 8*):

The probability that no more than k units contain CFU's is

$$\sum_{i=0}^{k} \frac{n!}{i!(n-i)!}(1-p)^i p^{n-i} \qquad (1)$$

Or, equivalently, the probability that more than k contain CFU's is

$$\sum_{i=k+1}^{n} \frac{n!}{i!(n-i)!}(1-p)^i p^{n-i} \qquad (2)$$

where p, the probability that a single sample contains *no* CFUs, is $\exp(-BQ)$ as above.

If one *assumes* that the test method will always give a positive result if the sample being tested contains one or more CFUs and will always give a negative

result if it contains no CFUs, eq 1 or eq 2 lets one calculate the probability of permitting a batch to be moved to the next processing step as a function of the (unknown) CFU/gram using the acceptance rule: The batch will be processed further if no more than k samples test positive (i.e., at least $n - k$ samples test negative).

Applying the Model

Let's call the acceptance limit L CFU/g. Suppose a batch is at the acceptance limit ($B = L$). If we choose a sample size of Q = 3/L grams, the expected number of CFUs in the sample will be $B \times 3/L = 3$, and the probability that a *single* sample will contain at least 1 CFU will therefore be $1 - \exp(-3) = 0.95$. Let's call the *expected number of CFUs per sample when the batch is at its acceptance limit* the ECAL. Note that the ECAL does not need to be an integer.

Unfortunately, if we use an ECAL of 3.0 as above, a single sample from a batch having a CFU concentration of 24% of the acceptance limit will have slightly better than a 50/50 chance of being rejected because the expected number of CFUs in a sample of size $3/L$ is $0.24L \times 3/L = 0.72$ and $1 - \exp(-0.72) = 0.51$. The 1/L "dilute to specification" sample size (ECAL = 1.0) recommended by Miller (*6*) has a lower false alarm rate (better specificity) but has only a 63% chance of rejecting a batch that just barely exceeds (i.e., has a CFU concentration more than 100% of) its acceptance limit.

Suppose, however, that we take eight samples of size $3/L$ (ECAL = 3.0) and apply the following acceptance rule: Accept processing the batch further if two or more samples test negative and reject if fewer than two test negative (i.e., more than six text positive)." Using eq 2, the probability of *correctly* rejecting a batch with $B = L$ CFU/g is still nearly 95% because with $k = 6$

$$\sum_{i=k+1}^{n} \frac{n!}{i!(n-i)!}(1-p)^i p^{n-i} = 8(1-e^{-3})^7 e^{-3} + (1-e^{-3})^8 = 0.94 \quad (3)$$

However, a batch at 24% of the acceptance limit now has only a 4% chance of being incorrectly rejected and it takes 54% of the acceptance limit to exceed a 50/50 chance of being incorrectly rejected.

Figure 1 shows, for a variety of test schemes, the probability of correctly rejecting a batch when it is at its acceptance limit, L CFU/g, *versus* the CFU concentration corresponding to a 50/50 false alarm rate as calculated from either eq 1 or eq 2. It includes all possible combinations of 1, 4, or 8 tests with ECALs of 1.0, 2.0, or 3.0 whose sensitivity and false alarm rates are within the limits of the vertical and horizontal axes shown. (The 17 schemes outside of the axes shown are unlikely to be useful.)

Three of the testing schemes identified by boxed labels were described earlier in this section. The scheme identified by the label (1, 1, 3.0) is the high sensitivity (towards the top of the graph)/high false alarm rate (towards the left on the graph) scheme. The first number in the label is the number of tests. The second is the minimum number of negatives required. The third is the ECAL. The (1, 1, 1.0) testing scheme is Miller's (*6*) "dilute to specification" (lower sensitivity, lower false alarm rate) scheme. The (8, 2, 3.0) testing scheme has a lower false alarm rate and about the same sensitivity as the (1, 1, 3.0) testing scheme.

Schemes with short labels are identified as follows: Symbols indicates the expected number of CFUs when the batch is at its acceptance limit (the ECAL). Short numeric labels denote the number of negatives required to pass the batch. The solid line connects schemes involving eight tests. The short dashed and long dashed lines connect schemes involving one and four tests, respectively. No ability to interpolate should be inferred. Other schemes with long labels are discussed in the Discussion section.

Discussion

Figure 1 ought to be more useful to a scientist, engineer, or microbiologist than a set of operating characteristic (OC) curves. Figure 1 lets one compare the advantages and disadvantages of many potential test schemes. An OC curve (*9*) could easily be calculated for each of the schemes in Figure 1, but a graph showing more than a few of them would be unreadable. The choice of 50/50 for accept/reject on the x-axis corresponds to the ISO Standard 3534-2 "Point of Indifference" (*10*). In other contexts a different probability of rejection could be

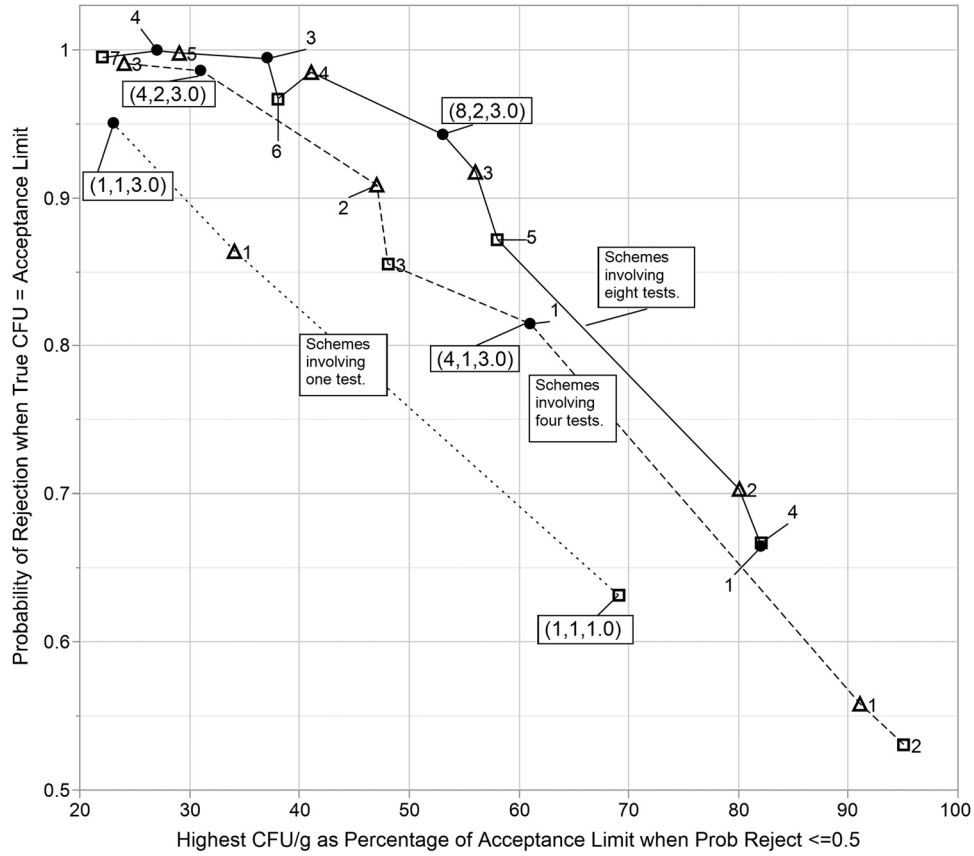

Figure 1

Expected performance of various testing schemes. Symbols indicates the expected number of CFUs when the batch is at its acceptance limit (the ECAL). Square = 1.0. Triangle = 2.0. Filled circle = 3.0. Short numeric labels denote the number of negatives required to pass the batch. Schemes with long numerical labels are discussed in the text; they denote the number of tests, number of negatives required, and the ECAL. The solid line connects schemes involving eight tests. The short dashed and long dashed lines connect schemes involving one and four tests, respectively. No ability to interpolate should be inferred. Points towards the top of the plot correspond to high sensitivity to detect batches that exceed the acceptance limit. Points towards the right have lower false alarm rates (better specificity).

chosen. A .jmp file and an Excel spreadsheet, either of which can help an interested reader evaluate other candidate testing schemes, are available from the corresponding author. It would also be straightforward to reproduce the logic in a programming language. If one's software permits one to point to individual data points and ask for more information, one could create a denser plot than Figure 1 and explore it for attractive test schemes.

Ultimately, the best testing scheme is a somewhat subjective compromise that depends on the sensitivity and specificity needed at the process stage of interest, the stage one is at in the production process, whether earlier or later in-process tests will provide additional safety margins, and the testing time and costs involved. This is because in-process tests are used so that one can discard material before incurring additional processing costs if the material is "too likely" to fail later (quantifying "too likely" is beyond the scope of this paper). When determining the cost of a scheme it is essential to recognize that the each test in the scheme must be performed on independently obtained and prepared samples (see Introduction). Historical data regarding typical CFU levels may also be useful in choosing a scheme. In addition, if different schemes are used at different points in the process, one may want to keep some level of consistency between them to reduce the likelihood of mistakes. (Different schemes might be used at different points in the pro-

cess depending on the relative costs of different processing steps.)

For example, if CFU levels are expected to be far below the acceptance limit, one might prefer the more sensitive (4, 2, 3.0) or even the (1, 1, 3.0) to the more labor-intensive (8, 2, 3.0) scheme. The latter two schemes have essentially the same sensitivity (ability to detect lots that exceed the acceptance limit), but the false alarm rate rises as one reduces the number of tests performed. Schemes towards the lower right [such as (8, 3, 1.5) scheme] might be used if false alarms (horizontal axis) are less financially tolerable, and if false negatives (vertical axis) will be detected in some later stage before prohibitive additional processing costs are incurred. When the number of negatives are typically well above the minimum required, one can compare the actual number to alert or action levels. For example, someone using the (4, 2, 3.0) scheme for acceptance might select an alert level of (4, 1, 3.0).

Increasing the number of negatives required while keeping the number of tests and the ECAL constant clearly increases a scheme's sensitivity to detect lots that exceed the acceptance limit (moving upward in the figure) at the expense of a higher false alarm rate (poor specificity, moving rightward in the figure). Increasing the ECAL has the same directional effect. Increasing the number of tests permits one to improve both sensitivity and specificity at the expense of increasing the cost of testing, but this soon reaches a point of diminishing returns. For example, one can achieve 95.4% sensitivity to lots that exceed their acceptance limit while retaining a 50/50 chance of not rejecting a lot at 81% of its acceptance limit, but that scheme would involve 100 tests, an acceptance criterion of at least 20 negatives, and an ECAL of 2.0 CFU/sample.

As mentioned by Halvorson and Ziegler (*2*), the results of any series of tests can also be used to compute a maximum likelihood estimate for the actual CFU concentration in the batch and a one- or two-sided confidence limit around it (the MPN procedure). For example, if one independently tests eight samples and three results are negative, one can estimate that the binomial proportion of *negatives* is 0.375, and Table A23a of Hahn and Meeker (*11*) tells one that a single-sided 95% *lower* confidence limit on that proportion is 0.111. If the sample size is Q grams, $p = \exp(-BQ)$ tells us that a 95% *upper* confidence limit on the CFU/g in the batch is $-\ln(0.111)/Q$. Thus, had the chosen sample size been Q = 0.03 g, the result would translate to a 95% upper confidence limit of 73.3 CFU/g.

Conclusion

When used in-process, the sensitivity and specificity of economical RMMs can be greatly enhanced by performing multiple tests on independent samples and adopting an appropriate rule such as "X out of N samples must test negative". The equations and figures shown herein can help one choose an appropriate acceptance rule.

Acknowledgements

Reviewers 1 and 2 are thanked for suggesting how to simplify an earlier version of Figure 1. All the reviewers are thanked for various suggestions.

Conflict of Interest Declaration

The authors declare that they have no competing interests.

References

1. McCrady, M. H. The numerical interpretation of fermentation-tube results. *J. Infect. Dis.* **1915**, *17* (1), 183–212.

2. Halvorson, H. O.; Ziegler, N. R. Applications of statistics to problems in bacteriology. I. A means of determining bacterial population by the dilution method. *J. Bacteriol.* **1933**, *25* (2), 101–121.

3. Cochran, W. Estimation of bacterial densities by means of the "most probable number". *Biometrics* **1950**, *6* (2), pp 105–116.

4. Blodgett, B. Appendix 2 of the FDA's Bacteriological Analytical Manual, 2010, (http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm109656.htm, accessed March 25, 2014).

5. USP General Chapter <61> Microbiological Examination of Nonsterile Products: Microbial Enumeration Tests. *United States Pharmacopeia,* 37th Rev.; U.S. Pharmacopeial Convention, Inc.: Rockville, MD, 2014.

6. Miller, M. J Case study of a new growth-based rapid microbiological method (RMM) that detects the presence of specific organisms and provides an estimation of viable cell count. *Am. Pharm. Rev*. **2012**, *15* (2), 18–25.

7. Ross, S. *A First Course in Probability,* 2nd ed.; Macmillan: New York, 1984; p 126.

8. George, E. P. Box; William, G. Hunter; Stuart Hunter, J. (1978), *Statistics for Experimenters*, Wiley, NY, 1978, chapter 5.6, 5.5.

9. Ott, E. R. *Process Quality Control;* McGraw-Hill: New York, 1975; p 174.

10. International Organization for Standards. Statistics—Vocabulary and Symbols—Part 2: Applied Statistics, *ISO 3534-2, Section 4.6.6,* 2006. (https://www.iso.org/obp/ui/#iso:std:iso:3534:-2:ed-2:v1:en, accessed 8/19/14).

11. Hahn, G. J.; Meeker, W. Q. *Statistical Intervals;* Wiley: New York, 1991; p 342.

PDA Journal
of Pharmaceutical Science and Technology