Enclosure A – Measurement Error

Methodology

As mentioned in the executive summary, the AOAC questions posed are broad and the SAWG therefore narrowed its scope to identify important features that could lead to further development, as needed. Certain assumptions were made. One primary underlying assumption is that a "representative" sample can be obtained. Thus SAWG did not address outright "errors" due to mislabeling of samples, cross-contamination, incorrect readings from a machine, etc.

To address measurement error, the SAWG examined a publication that thoroughly attempts to quantify the measurement variation of basic microbiological methods (Niemela 2002). This document focuses on the measurement error associated with counting colonies and/or other discrete entities. The SAWG report identifies important factors that contribute to laboratory measurement variability and recommends that these factors be addressed with all methods for the purpose of controlling them and quantifying them, if possible.

The SAWG focused on examples of method protocols to examine where the measurement error variation can occur. Examination of these examples led to the identification of major sources of variation that the SAWG will consider:

A. Dilution
B. Recovery
C. Counting
D. Organism confirmation
E. Organism variability
F. Overall statistical considerations

A short description of the each of the sources of variation is given below, followed by section F that provides an example of how the operating characteristics for a plan could be constructed taking into consideration these sources of variation.

Both the enumeration of microbiological counts and the identification of microbes' genus and species involve a number of steps. Some of the steps include: sample collection, sample preparation including dilution(s), and (in some cases) maceration or mixing (or both). In determination of genus and species of pathogenic organisms, there is often an incubation step in a selective media prior to implementing one of the methods used for detection. All of these steps may have some error or variation associated with them.

A. Dilution error

Dilution errors are those errors associated with the sample preparation from the time a sample is gathered until the time the organisms are either counted or identified. Without proper training, the opportunity for error could be substantial, and the impact may vary from small to great. The first dilution error is initial sample size. If a sample of a particular size is to be placed into a fixed amount of diluent, the size of the initial sample (either larger or smaller than nominal) would cause under-dilution or over-dilution error, respectively. Similarly, if a surface area constituted the sample, then sampling an area larger than the specified area would result in an under-dilution, and an area smaller than the specified area (or incomplete swabbing of the specified area) will cause over-dilution.

With regard to setting up blanks for dilutions**,** some additional errors may occur. First, volumetric errors related to the use of graduated cylinders and pipettes could occur. Second, dilution blanks may be prepared volumetrically correct, but then autoclaved causing volume reduction. Errors may also be associated with incorrectly reading the meniscus in graduated cylinders and pipettes. Pipettes may differ in the way volumes are correctly measured. Plastic and glass pipettes' and cylinders' menisci are not read the same way, and may not have the same reliability.

Pipetting errors, of course, can occur for all dilutions after the original sample is placed into the original blank and the subsequent dilutions are made. Also, when micro-volumes of samples are pipetted into test containers, errors can occur due to pipetting technique. Additional errors may occur due to debris restricting the filling or emptying of the pipette or pipette tip, thus causing a non-representative sample to be placed in the testing container.

These errors do not include the obvious errors associated with spillage or leakage, but if unobserved or uncorrected, these factors could contribute greatly to the error associated with dilution.

B. Recovery error

The recovery rate for a microbiological method is the proportion of target cells (or spores) in the test sample that is presented to the detection method. With rare exception, recovery entails multiplication of the target cells to the high numbers required by the detection portion of the test. If multiplication does not occur or is impeded, the microbial count is underestimated. Factors confounding multiplication are as follows:

1. Incubation conditions: Test methods specify the time, temperature, and atmosphere of incubation. Ranges are generally provided for time and temperature. In many cases these ranges may be very broad and may compound in

methods that have multiple steps, e.g. liquid culture incubation for time ±2 hr, temp ± 1C, then subculture incubation for time ±2 hr, temp ± 1C, and finally agar plate culture for time ±2 hr, temp ± 1C. Validation studies rarely validate the extremes, as doing so is costly.

2. Media: Biological media is generally not calibrated from lot-to-lot or supplier-to-supplier, with perhaps the exception being Standard Methods Agar which is calibrated lot-to-lot per supplier. A great deal of variation can and does occur related to the components of basic culture media. Selective media adds another level of variation, since selective agents may vary in toxicity depending upon lot, preparation, and storage.

3. Product Matrix: Organisms may be in or on a product, in clumps or as single cells. Accurate enumeration requires full release from the matrix for repeatable enumeration. Some foods may be inhibitory to growth—spices are common examples. In addition, foods may contain competitive flora which may inhibit growth or outgrow the target population.

4. Target Flora: Recovery rates may vary among genera, species, or even subspecies and strains. The target flora may be injured and thus variation in recovery may increase.

C. Counting error

It is often believed (assumed) that the distribution of the results of a count follows the Poisson distribution. This is based on assuming that the cells are distributed uniformly so that, per unit of product, there is a single expected level, r. From this assumption, it can be shown (Jaynes, 2003: Probability Theory: The logic of Science Cambridge University Press) that the distribution of the number of cells in any volume, v, of product is a Poisson distribution with expected value rv, and standard deviation $(rv)^{1/2}$. However, cells, larvae or other types of microbiological contaminates are usually not distributed in nature as a Poisson, but rather are distributed in clumps, or colonies, either because there are factors that would attract microbes to cluster or because of the cell division process creating a tendency for colonies to form (e g, Campylobacter). Because the assumption of uniformity cannot be assumed, in order for the Poisson distribution to hold, it is necessary to homogenize the sample. Variation beyond that expected from the Poisson thus can be introduced when the sample is not homogeneous.

Additional variation is also introduced due to the non–exactness of the counting of colonies of a specified species. The counting may differ appreciably between persons for a given sample on a given medium. Familiarization thus with the counting procedures is an important requirement for analysts.

D. Organism confirmation error - selection and testing of colonies

During many analytical procedures colonies need to be selected for further testing. Methods usually specify that a certain number of colonies, or a certain proportion of the colonies meeting the description of the target species are selected for further testing. The sampling errors involved in this procedure depend upon the differential power of the primary isolation medium and upon the ratio of target species to non-target species that meet the description of colonies to be further selected.

D.1. Differential power of primary isolation medium. The differential power of the medium is the ability of the medium to cause target species to appear sufficiently different from non-target species, to facilitate the efficient selection of the target species for confirmatory testing (if required).

The differential power of the medium may be graded as follows:

Absolute: in which every colony on the plate is counted and no further testing is required. Examples: Standard Plate Count, Aerobic Plate Count.

Highly differential- in which we can be almost certain that colonies that meet the description belong to the target species, other colonies are clearly differentiated. Example: B. cereus on MYP or PREYPA or PEMBA.

Moderately differential- in which colonies meeting the description of typical strains may belong to the target species. Some non-target species may fail to be differentiated from the typical colonies of the target species and/or some atypical colonies may belong to the target group. Examples include: Salmonella sp. on BSA, Coagulase positive staphylococci on Baird Parker Agar, Listeria species on MOX.

Poorly differential- in which colonies of the target species and some (related) non-target species are not differentiated. Example: L. monocytogenes on Oxford Agar or PALCAM.

The magnitude of error that may be associated with this factor increases as the differential ability of the medium decreases.

D.2. Ratio of target to non-target colonies on the primary isolation medium. The ratio of target to non-target colonies affects the ability of the operator to select the most likely colonies for further testing. This factor operates in two different ways that may interact: the selectivity of the medium and/or the differential power of the medium.

The selectivity of the medium refers to the ability of the medium to suppress non-target species. If a medium is highly selective, it is more likely that a colony on the

agar will belong to the target species. In extreme cases, poorly selective media may allow non-target species to overgrow target species to the extent that the target species cannot be detected (for example, Citrobacter overgrowing *Salmonella* on XLD agar). In many cases, this reason is not a large problem if the medium is sufficiently differential.

The differential ability of the medium is discussed above. If poorly differential agars are used, then selecting colonies of the target species will depend entirely on the ratio of target and similarly-appearing non-target species. For example, both L. monocytogenes and L. innocua will have identical colonies on Oxford Agar, but only one of these species is the target. In qualitative tests: if the method requires a number of colonies to be selected, and only one (1) colony needs to be confirmed as positive for the target species to be reported as detected, then it will be possible to calculate the likelihood of selecting a colony of the target species depending upon the number of colonies of each of the target species and the similarly appearing non-target species. In quantitative tests, a number of colonies are selected and the proportion of these colonies found to be confirmed positive is used as a factor by which the presumptive positive count is multiplied to determine the confirmed positive count.

D.3. Confirmatory Testing. Approaches to confirmatory testing vary depending upon the target organism. In particular, the number of tests to be performed varies from target organism to target organism. Sometimes, only one test result is required (e.g., coagulase), whereas other times a range of confirmatory tests are required (e.g., BAM method for B. cereus). Each of these tests has its own characteristics (rates of false positive, false negative etc.).

E. Organism variability

Microbiological analytical tests exploit one or more microbial characteristic to differentiate between those microorganisms included within the group and those excluded. The breadth of the designated group can be large (Gram negatives, *Enterobacteriaceae*) or small (*Salmonella enterica* serotype Enteritidis phage type 4, *E. coli* O157:H7). An ideal test would detect every microorganism that is intended to be within the group (sensitivity) and ignore every microorganism intended to be excluded (selectivity). Failure to recognize a microorganism that should be in the group is termed a false negative; conversely, accepting a microorganism that should not be in the group is a false positive.

The variety of properties used to group microorganisms range from physical structure (rods, spore formers), to metabolic characteristics (ability to metabolize a particular sugar, production of hydrogen sulphide), to the ability to survive toxins (brilliant green agar, antibiotics), to production of antigenic proteins (ELISA tests), to the presence of plasmids, specific DNA sequences (PCR tests), and to the ability to produce a toxin (*C. botulinum*, Enterohemorrhagic *E. coli*).

Unfortunately for the consistent grouping of microbes, bacteria typically do not have a consistent set of characteristics within the group. Nearly all enterohemorrhagic *E. coli* O157:H7, for example, cannot ferment the sugar sorbitol (unlike nearly all other *E. coli* which can ferment it) and this characteristic is used in identification tests. However, there is a sorbitol-fermenting serotype of *E. coli* that can also produce the shiga toxins. Furthermore, microbial characteristics are not static as DNA exchange occurs between bacteria at a much higher level taxonomic level than species. Because of these inconsistencies in the presence of microbial characteristics, there is a trend to identify microorganisms of public health concern by the presence of the DNA that codes for the particular virulence factors. This attempts to include all microorganisms that can cause a particular illness, regardless of their conventional strain, species or even genus designations. But even this strategy is not without difficulties as pathogenicity is frequently the result of a cluster of virulence factors and not all pathogens that cause that illness may have an identical or complete set of the virulence factors. In addition, some strains may posses the DNA for the virulence factors but the genes are never expressed, making those strains non-pathogenic. The pre-test environment, whether in a food or an enrichment medium, can sometimes affect the expression of an identifying characteristic.

In the development of a method to detect specific organisms or groups of organisms, care must be taken to select a characteristic that is shared by all the organisms to be included and absent in those to be excluded. Validation of a test protocol by testing against a wide range of microorganisms of both groups is necessary. Quantifying the rate of false positives or false negatives, however, is difficult and rarely done. In actuality, virtually all microbial tests are not as sensitive or selective as desired. Microbiologists rely on subjective knowledge and experience of the appropriateness of most tests for the situation at hand. This is demonstrated by the classic "fecal coliform" test widely used to as an indicator of the presence of sewage contamination in shellfish and water. However, this test is inappropriate (not sufficiently selective) for detecting sewage contamination on vegetables as there frequently are non-pathogenic soil bacteria that would be declared positive by the test.

F. Overall statistical/distribution considerations

The statistical considerations for characterizing method performance can be described by taking an example method and considering the probability distributions that would be encountered at each of the steps for the method. The example that follows goes through this process.

Assumptions: It can be assumed that there is a probability distribution of levels, -cfu/ml – (or cfu/g), throughout the product being sampled. The concepts that are needed for designating distributions need to be discussed at greater length. But, for the moment, assume that any pathogen or interfering organism is uniformly distributed throughout the 100 ml of the sampled material.

In this example, it is assumed that there is one type of organism of concern - the target organism - distributed uniformly with level, r_t, and there is a another type of organism – an interfering organism - distributed uniformly with level, r_i.

Steps 1-2 Prepare a 1:5 dilution; spread 1 ml of material on three plates of one type of agar (for the moment consider just one type of agar).

It is assumed that the 0.2 ml of the 100 ml sample is randomly selected so that the number of cells, n_x, of the target or the interfering organisms is distributed as a Poisson distribution with parameter $0.2r_x$, where the subscript 'x' is either 't' or 'i'.

Step 3 Grow colonies.

The n_x cells are assumed to develop into m_x colonies. If f_x is the probability that a cell will develop into a colony, and we assume that the events of these occurrences among the n_x cells are independent, then the distribution of the number of colonies m_x is a binomial with parameters, n_x and f_x, so that the expected value of m_x, conditional on n_x is $n_x f_x$. It turns out then that the number of colonies, m_x, is distributed as a Poisson distribution with parameter $\lambda_x = 0.2r_x f_x$, so that, unconditionally, the expected value of m_x is λ_x.

Step 4 Select 5 colonies from the $m = m_t + m_i$ colonies for confirmation. If one or more is positive for the target organism then the sample is classified as positive for the target organism; otherwise it is not, and thus is classified as negative. For the moment, assume that any selected colony will be identified properly.

The distribution of the number of selected colonies of the target organism is hypergeometric. Let k_t be the number of targeted selected colonies. The probability of a classified positive sample is the probability that $k_t > 0$. This can be written as:

$$P(k_t > 0) = 1 - \frac{m_i!(m_t + m_i - 5)!}{(m_i - 5)!((m_t + m_i)!},$$

where here it is assumed that $m_i \geq 5$. If m_i is less than 5 then there is 100% probability of selecting at least one colony of the targeted organism (provided, of course, that $m_t > 0$).

If m_i is large, and m_t is not, then the above probability could be close to zero. If both are large, but at an expected certain ratio, say, $E(m_t) = gE(m_i)$ - that is, for every colony of the interfering kind there are expected g colonies of the targeted kind - then the percentage of colonies of the targeted kind is: $m_t/(m_t+m_i) \approx \alpha_t = g/(g+1)$ and the hypergeometric distribution can be approximated as a binomial distribution with parameters 5 and α_t. The above expression - the probability of a classified positive sample – can then be approximated as:

$$P(k_t>0) \approx 1-(1+g)^{-5}.$$

Example (for large levels):

1. If g =1, (that is, there is an expected equal number of targeted and interfering organisms) then the probability of a positive sample is $1 - 2^{-5} = 1\text{-}1/32 \approx 0.97$, or 97%. Or, in other words, there would be a 3% false negative rate.

2. If g = 1/3 – that is, for every three interfering colonies there is one targeted colony, then the false negative rate would be $(1.3333)^{-5} = 0.132 = 23.7\%$.

3. If g = 0.5 – that is, for every two interfering colonies there is one targeted colony, then the false negative rate would be $(1.5)^{-5} = 0.132 = 13.2\%$.

4. If g = 2 – that is, for every interfering colonies there are 2 targeted colonies, then the false negative rate would be $3^{-5} = 0.004 = 0.4\%$.

5. If g = 2 but instead of 5 colonies only 3 colonies were tested for confirmation, then the false negative rate would be $3^{-3} = 3.7\%$.

   The above calculations indicate that the number of tested colonies can be important when there is a significant percentage of interfering colonies expected. Even when g = 2, the false negative rate is 0.4% when there are large numbers of both types of colonies, which could be considered large in some applications. The question that needs to be addressed is: <u>what values of g are possible or likely?</u>

   The types of uniformity or distributional assumptions made in these situations are paramount to the validity of the calculations. For the above scenario, it is assumed that the types of cells are distributed independently and uniformly within the sample. However, in reality it might be more realistic to assume an 'extreme' negative correlation of some sort between the types of cells, so that values of g are either close to 0 or 1.

<u>Example Calculations</u>:

a. Suppose, r_t = 10 cfu/ml, and the likelihood for growth is f_t = 75%, so that in a 0.2 ml sample, there would an expected 1.5 colonies of the targeted organisms. For the interfering organisms, assume that r_i = 20 cfu/ml, and f_i = 75% as well, so that there would an expected 3 colonies of the interfering organisms. Thus, the value of g would be 0.5. However, the expected number of cells in the 0.2 ml sample is small, so an exact calculation for determining the probability of a false negative would be needed. The probability of a (false) negative result is, Pn = 24.35% - almost 25% of the time, the results would be negative for the target organism.

b. Double r_t (= 20 cfu/ml) and r_i, keeping everything else the same, Pn = 13.86%, close to the theoretical asymptotic result of 13.2% given above in 2).
c. r_t = 10 cfu/ml and r_i = 5 cfu/ml (so that g = 2) Pn = 22.32%.

d. r_t = 20 cfu/ml and r_i = 10 cfu/ml, then Pn = 5.1%.

e. r_t = 30 cfu/ml and r_i = 15 cfu/ml, then Pn = 1.35%.

f. r_t = 40 cfu/ml and r_i = 20 cfu/ml, then Pn = 0.6%.

g. r_t = 60 cfu/ml and r_i = 30 cfu/ml, then Pn = 0.4% (close to the asymptotic result).

h. r_t = 100 cfu/ml and r_i = 50 cfu/ml, then Pn = 0.4% (just for emphasis).

i. r_t = 100 cfu/ml and r_i = 200 cfu/ml, then Pn = 13.2%.

Figure 1 presents the operating characteristic (OC) curves for the probability of a negative finding, Pn, versus the assumed level of the target organism, r_t, assuming different levels of the interfering organism, r_i; where $r_t = gr_i$. It is assumed same growth likelihood and recovery of 75% for both types of organisms, and 5 colonies are tested for confirmation.

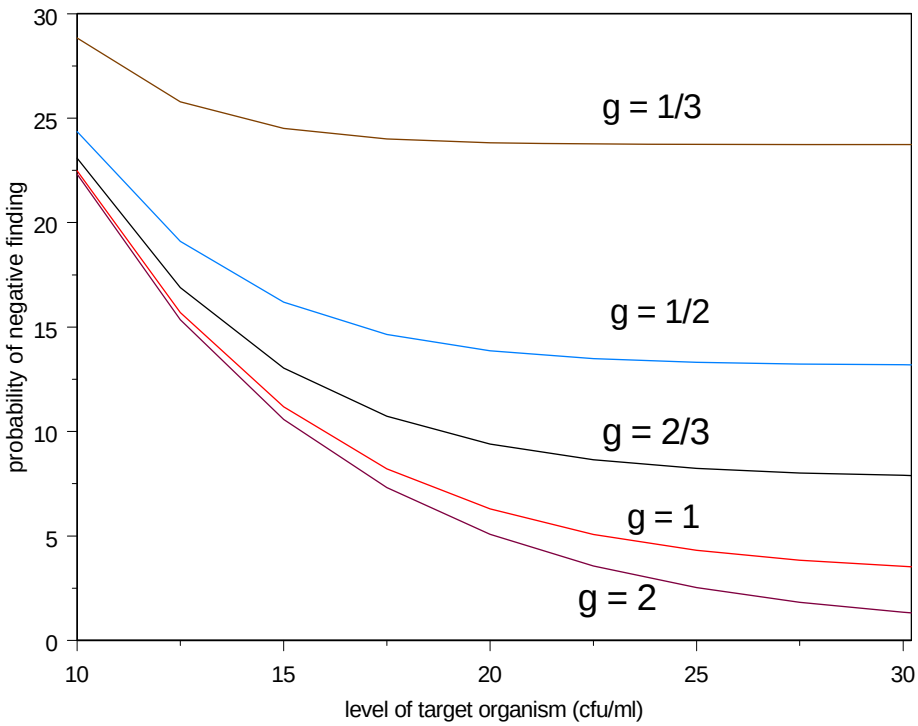

Figure 1: The OC curves for the probability of a negative finding at different levels of target and interfering organisms. For every colony of the interfering kind there are expected g colonies of the targeted, so that g = 1 means that there are expected the same number of interfering and targeted colonies; g = 2/3 means that for every 3 interfering colonies, there are an expected 2 targeted colonies

Summary and conclusion

This review is a bottom-up approach that attempts to identify and quantify all potential errors of a laboratory method. From a mathematical characterization of the results associated with each source of variation, a derivation of the expected operating characteristics of a method can be made. However, there are other unidentified sources of variation that are not likely to be captured when studying the method through a bottom-up approach that are associated with human errors, for example pertaining to equipment settings and calibrations, as well as unexpected changes in environmental conditions that cannot be captured in small, controlled, laboratory studies. The importance of these might be determined through ruggedness tests, where the parameter specifications for critical steps of the method are changed slightly from their nominal values in order to determine the effect of small changes. Ideally methods that are used are rugged in that the results are not affected greatly by small changes of the method's specifications. To the degree that a test is rugged, the bottom –up approach for determining the magnitude of variation (of results) will capture a large portion of the total variation. Thus we have recommended ruggedness testing to be part of method validation (BPMM Task Force Report Executive Summary).

While inter-laboratory studies may be needed to develop reproducibility measures that basically validate methods to be used by qualified analyst , it is still critically important to identify the sources of variability in a method and quantify their effects within laboratory. These can be used for quality control monitoring. In addition, if definitive inter-laboratory studies providing reliable measures of method performance do not exist, then performance measures determined from a series of a bottom-up studies, identifying and quantifying variability associated with the critical steps of a process should be conducted that can be used for laboratory QA. It is possible that the performance operating characteristics estimated will be accurate when using a bottom-up approach particularly so if it can be shown that the method is rugged.

References

Dahms, Susanne. 2004. Microbiological sampling plans – statistical aspects. Mitt. Lebensm. Hyg. 95, 32–44.

International Commission on Microbiological Specifications of Foods (ICMSF). 2002. Microorganisms in foods 7: Microbiological testing in food safety management. Kluwer Academic/Plenum Publishers, New York, p. 118.

Jaynes, ET. 2003. Probability Theory: The logic of Science. Cambridge University Press, p170-171.

Niemela, Seppo, I. 2002. Uncertainty of quantitative determination derived by cultivation of microorganisms. Publication J3/2002, Center for Metrology and Accreditation, Advisory Commission for Metrology, Helsinki, Finland.