Review

Validation, verification and comparison: Adopting new methods in water microbiology[#]

David P Sartory

Sartory Water Microbiology Consulting, 7 Sunnybank, Little Ness, Shrewsbury, SY4 2LQ, UK

Abstract

Until recently there has been little formal guidance on procedures for adopting new methods in water microbiology. However, the *European Union Drinking Water Directive* of 1998 specified methods that were to be used for the microbiological parameters, most being ISO methods, but allowed the use of alternative methods that were "at least as reliable". At that time, there were no published procedures for demonstrating equivalency of performance between methods. Work commissioned by the UK Drinking Water Inspectorate (DWI) developed suitable analytical and statistical protocols for comparing microbiological methods. The statistical aspects have been refined and recently published as *ISO 17994*. ISO has also recently published guidance on the validation of methods for water microbiology (*ISO/TR 13843*), which gives guidance for developers of new media on what performance information is required. These developments provide a framework for the enhancement of validation and verification procedures within a laboratory's quality system for evaluating new methods prior to their adoption. This paper overviews these developments in light of the author's experience in their use and discusses issues relating to the analytical procedures and the statistical rationale employed (including the concept of "equivalency" of performance between methods).

Keywords: water microbiology, methods, validation, verification, comparison, AQC

Introduction

Many methods used widely in water microbiology have not undergone full validation of performance, having been widely used and accepted historically, and there has been little formal guidance on demonstrating equivalent or superior performance of a new method prior to its adoption by a laboratory. Most laboratories have accepted performance claims in published scientific papers or from manufacturers. It should, however, never be assumed that a method will perform as claimed in every laboratory and for every type of matrix for which it may be used. Therefore, verification of the claimed performance, and where appropriate, demonstration of equivalent or inferior/superior performance compared to the method in current use are required. While there has been some guidance on assessment of performance characteristics for water microbiology methods (Havelaar et al., 1993; Lightfoot and Maier, 1998; ISO, 1999a) that on comparing the performance of two methods has been very limited. Typically, this has been in the nature of analysing a limited number of samples, usually 20 to 30, and using simple analysis of variance statistics (Lightfoot and Maier, 1998). This is really only sufficient to detect any gross differences between two methods. A procedure for comparing two presence/absence (P/A) methods was developed by the USEPA (1995) which required multiple analyses (20) of ten natural sources of coliform bacteria, which were then chlorine-challenged ensuring that the numbers present in the analysed test portions covered the range 1 to 10 per 100 mℓ. Similarly, Covert et al. (1992) used ten sub-samples from 22 chlorine-challenged samples containing low numbers of *E. coli* when evaluating

defined-substrate P/A tests. These approaches have been successfully employed for P/A methods where the results are simply positive or negative and only a limited set of data is required for statistical analysis (typically non-parametric) to demonstrate superior or inferior performance of one method against another.

When it comes to comparing two quantitative methods, however, there are some aspects of variability in micro-organisms that have to be taken into account. Firstly, it must be borne in mind that micro-organisms are not solutes like ions, which for chemical analyses can be assumed to be homogeneously distributed. When introduced into water, micro-organisms do not form a perfect solution but a suspension, which imparts a significant degree of inherent heterogeneity (Tillett and Lightfoot, 1995; BSi, 2003). This may be exacerbated by any reactions between the target micro-organisms and any others or particles present in the sample. Additionally, there is a variability imparted by the microbial cells in a sample (even at species level) because they will be at differing stages of growth, states of stress response and metabolic status at the time of analysis (BSi, 2003; Tillett and Sartory, 2004). This will impact on the response of target bacteria in quantitative methods requiring selective growth. All these factors result in a significant natural variability in recovery of micro-organisms from water which must be taken into account when devising analytical and statistical protocols for comparing quantitative methods, particularly when the numbers that are encountered in routine samples tend to be very low, as in the case of drinking water monitoring.

The *European Union Drinking Water Directive* of 1998 (European Union, 1998) specified for the first time the methods that were to be used for the microbiological parameters for regulatory monitoring. For *E. coli*, coliform bacteria, Enterococci and heterotrophic plate counts the methods cited are the current respective ISO standards (*ISO 9308-1, 7899-2* and *6222*) (ISO, 1999b; 2000a; 2000b), whilst for *Clostridium perfringens* mem-

brane filtration using mCP agar (Armon and Payment, 1988) was stipulated. However, some of these methods are not widely used in Europe (particularly those for *E. coli* and coliforms, and *Cl. perfringens*). The Directive, however, also stated that alternative methods "may be used, provided that the results obtained are at least as reliable as those produced by the methods specified" (European Union, 1998). At that time, there were no published procedures for demonstrating equivalency between various methods.

Consequently, the UK Drinking Water Inspectorate commissioned work to develop suitable analytical and statistical protocols for comparing two microbiological methods (DWI, 2000) which have been incorporated into UK guidance (Standing Committee of Analysts, 2002), the statistical part of which was subsequently developed into a recently published ISO standard *ISO 17994* (ISO, 2004). Over the same time, ISO also published guidance on validation of microbiological methods (*ISO/TR 13843*) (ISO, 2000c).

These developments have provided a sound framework for the validation and verification of performance of new methods prior to adoption by a water microbiology laboratory. This paper overviews these developments and outlines some of the analytical procedures and the statistical rationale that underpin them (including the concept of equivalency of performance between methods).

Validation (primary validation)

Many methods used in water microbiology have not had substantial validation of performance, some having been developed 40 or more years ago (e.g. membrane-lauryl sulphate broth (mLSB) and m-ENDO agar for coliform bacteria, and m-Enterococcus agar/Slanetz and Bartley agar for Enterococci). Their continued use is a result of their widespread (national or international) employment as well as frequent incorporation in national methods. However, many of these methods were originally adopted after only a review of data in scientific publications and limited in-house evaluation. Validation of methods is a requirement of *ISO 17025* (ISO, 1999a), which gives limited information on undertaking such work. For water microbiology this has been resolved with the publication of *ISO/TR 13843* (ISO, 2000c), which defines (primary) validation as "an exploratory process with the aim of establishing the operational limits and performance characteristics of a new, modified or otherwise inadequately characterised method". The standard describes the information required for the derivation of the numerical and descriptive specifications of a method. A key component is the requirement for an unambiguous description of the target organism. This is particularly important for water quality monitoring where *E. coli* and coliform bacteria are still universally used as regulatory indicators and for which there are several quantitative methods available based upon differing detection criteria. For example, the widely used mLSB medium relies on lactose fermentation at 44 °C for the detection of presumptive *E. coli*, whilst membrane-lactose glucuronide agar (mLGA) (Sartory and Howard, 1992) and the Colilert™ QuantiTray™ system (Fricker et al., 1997) rely upon the detection of the diagnostic enzyme ß-glucuronidase. Similarly, mLSB and mLGA detect presumptive coliforms as lactose-fermenting colonies at 37°C, whilst the Colilert™ system defines coliforms upon the ability to express ß-galactosidase. This has caused issues for some water suppliers who found that on adopting the Colilert™ system that their water quality appeared to deteriorate significantly, as several members of the Enterobacteriaceae (e.g. strains of *Serratia*) are ß-galactosidase positive, but lactose negative (usually due to lack of lactose permease). It is, therefore, essential to understand the basis of new methods (definition of target or-

ganism), so that if differences are found when comparing a new method with an established one, they can be explained.

Validation of a method will provide information on specification of performance, not only with respect to the recovery and enumeration of the target organism(s), but also the analytical requirements of the method (e.g. incubation temperature and time, media preparation and storage conditions, and sample storage or pretreatment). Key information will relate to recovery efficiency, upper and lower working (detection) limits, selectivity and specificity (false-positives and false-negatives), counting uncertainty (methodological and analyst) and a general estimate of precision. Since these data will provide the initial assessment of performance of a new or modified method it is strongly recommended that analysts with considerable experience in microbiological methods conduct the work.

Although it may be unreasonable to expect validation work based upon ISO 13843 to be undertaken for methods that have been widely used for several decades, it is appropriate that the new methods that are being developed to replace them should have full validation. Generation of appropriate validation data should be the responsibility of the research team or manufacturer developing the method, and laboratories should demand such information from commercial suppliers before any consideration of verification of performance and adoption in their laboratory.

Verification (secondary validation)

Verification (termed secondary validation in ISO/TR 13843) is a simplified validation process. Its purpose is to answer the basic question "Does this new method perform to its specification in my laboratory?" There is limited guidance on verification in *ISO/TR 13843*, simply that a number of natural samples should be used, analysed as split samples or replicate dilution series with duplicate counting to verify expected counting performance. However, how many samples and of what nature? This author suggests that a limited number of samples using an appropriate quantitative reference material (e.g. Lenticules™ as described by Lightfoot et al., 2001) to confirm target and non-target colony morphology and colouration or reaction colour. This also allows the analysts to become acquainted with the new method without any issues of interferences associated with natural samples. Once the analysts are proficient, then natural samples appropriate to the laboratory are analysed. It should be remembered that these samples will typically contain target and non-target micro-organisms in some state of stress and probably reduced metabolic status. This may result in differing appearance or reactions compared to those using pure culture reference organisms. In addition, the species or strains in these samples are likely to be different from those encountered by the laboratory or manufacturer that undertook the original validation work. There is, therefore, the possibility of encountering atypical growth or reactions that may be specific to the laboratory. There is no recommendation on the number of natural samples that should be analysed for verification of performance, but about 30, covering the range of water types or matrices typically analysed by the laboratory, would appear to be reasonable. The laboratory should analyse several samples of each water type or matrix, as a single result from a sample source may not be truly reflective of how the method performed on that water type. Additionally, if the types of bacteria normally encountered by the laboratory with their current method are subject to seasonal variation, it may be appropriate to conduct the verification exercise over a period of time that would take that source of variability into account.

It is essential that the identity of the target bacteria isolated

by the method is confirmed and *ISO/TR 13483* recommends that 100 presumptive positives should be isolated and verified (using appropriate biochemical or serological protocols). This author suggests that the identity of a number (up to 50) of non-target presumptive isolates is also subject to identification to check the false-negative rate.

Successful performance of a new method after the verification exercise can result in a laboratory adopting the method. If, however, the new method were to replace one already being used by a laboratory it would be appropriate to assess the new method against the current method, and to generate verification of performance data at the same time. One of the key benefits of this would be the generation of data that can be used to explain to customers why the method has been changed (e.g. greater recovery or specificity/selectivity, etc.), any additional benefits (e.g. more rapid analyses) and any potential impact it may have on the results from their future samples.

Comparison of methods

Probably the most useful recent developments that can be used by laboratories as part of their procedures for adopting new methods are the protocols for comparing methods developed for the UK Drinking Water Inspectorate (DWI) (DWI, 2000; Standing Committee of Analysts, 2002), the statistical aspects of which were further developed in *ISO 17994* (ISO, 2004). Although these protocols were originally developed in response to the specific issue of EU Member States wishing to use alternative national methods to the ISO methods for the microbiological parameters under the EU Drinking Water Directive, they do provide a robust and statistically sound method for a laboratory to evaluate new alternative methods prior to adoption.

The main issues that had to be addressed in the work commissioned by the DWI were:

- What is meant by "at least as reliable", and what statistical approach would be appropriate?
- How many samples would be needed to generate sufficient data to demonstrate "at least as reliable", taking into account the inherent variability of dispersion of micro-organisms in water samples?
- Are spiked samples appropriate for generating chlorine-stressed target micro-organisms, and how can sufficient samples be generated?

The statistical approach developed to address the question of "at least as reliable" was the testing of the hypothesis that there is no difference between counts, on average, and the "confidence levels" of the estimated average difference between counts. The data are generated from paired samples (with counts in the range 20 to 50 target organisms per test portion) and the 95% confidence interval for the average difference count between the two methods is calculated (from either untransformed or $\log_{10}$ transformed data using appropriate parametric or non-parametric statistics, depending upon the outcome of testing for normality). If the new method gives significantly lower (or higher) counts, then the 95% confidence interval will lie entirely below (or above) zero-average-difference, and a clear conclusion follows. If there is no significant difference, then the conclusion of "at least as reliable" is only accepted if the 95% confidence interval lies entirely above the value that would indicate that the new method could find 10% fewer organisms than the reference method. In other words, if the lower confidence interval point was not less than 90% of the mean count of the reference method, the new method was con-

sidered acceptable or, in UK terms, "equivalent". It is suggested that 150 samples would normally be sufficient to generate the required data. It should be noted, however, that the inherent variability in numbers of micro-organisms in the samples may result in the 95% confidence interval being too wide, resulting in an inconclusive comparison, which would require further samples to be analysed. The procedure also suggested that at least four sources, and up to ten, of water type (e.g. treated drinking water sourced from upland rivers or lowland rivers, from springs or groundwaters, etc.) or spiking material reflecting the types of water normally analysed, should be used, with at least 15 samples producing paired data from each source. In studies undertaken at Severn Trent Water, we have found more consistent results when using river water, and generally try to use raw water from the source from which the drinking water under test was derived as this would contain micro-organisms appropriate for potential occurrence in that drinking water.

Protocols for the generation of chlorine-stressed bacteria using microcosms spiking with either river water or sewage effluent were developed and later successfully used in further work for the DWI, where the three methods commonly used in the UK for *E. coli* and coliform bacteria were compared to that outlined in *ISO 9308-1*. However, the variability of natural inocula, especially that associated with sewage effluent, means that it is advisable to conduct several test runs to become familiar with the procedures and understand the levels of target bacteria in the spiking material and numbers surviving after chlorine stress. Alternative sources of samples are part-treated samples from water treatment works and contaminated groundwaters that can be subjected to minimal chlorination. For methods used for environmental samples, a range of samples of known contamination is appropriate. Details of these protocols and the statistical approach are given in *The Microbiology of Drinking Water – Part 3* (Standing Committee of Analysts, 2002).

The statistical approach developed for the DWI has been further refined in *ISO 17994*, which again, although aimed at comparing methods in an inter-laboratory equivalency exercise, is readily adapted for a single laboratory, simply by replacing the number of laboratories involved with the number of water types tested. *ISO 17994* also uses paired count data (but transformed into natural logs) from which relative difference (RD) percentages are calculated as in Eq. (1), with *a* and *b* being the paired counts from the two methods.

$$RD = [\ln(a) - \ln(b)] \times 100\% \tag{1}$$

From the data generated, a mean relative difference percentage is calculated together with an expanded uncertainty (*U*, calculated using the standard deviation, *s*, and which approximately corresponds with the 95% confidence interval), Eq. (2).

$$U = 2s / \sqrt{n} \tag{2}$$

Two methods are considered quantitatively equivalent if the mean relative difference does not differ significantly from zero and the expanded uncertainty does not extend beyond a maximum acceptable deviation from zero (*D*, for which the DWI protocol suggestion of 10% is cited). Significant difference of the mean relative difference from zero indicates significantly different performance by the new method (better or worse). Where the expanded uncertainty covers both zero mean relative difference and either the +10% or –10% acceptable deviation, the comparison is inconclusive and more samples are required to reduce the width of the expanded uncertainty (for which *ISO 17994* provides a method

for calculating how many extra samples would be needed).

The main difference between the statistical approaches of the DWI protocol and *ISO 17994* is that the DWI protocol takes into account the occurrence of non-Normal data (which has frequently been encountered in UK studies), whereas the ISO method forces a log-Normal assumption. However, both approaches should generally reach the same conclusions.

One of the key recommendations of *ISO 17994* is that all presumptive target colonies/positive wells or tubes from both methods are subjected to confirmation test, and not a selected percentage that may be the normal practice for routine analyses. This will reduce any potential variability that could be introduced by the selection of isolates for confirmation. As with verification, it would be advisable to formally identify a selection of confirmed target organisms, particularly when only simple confirmation procedures are part of the method. For example, the confirmation procedure in *ISO 9308-1* for *E. coli* is simply testing for oxidase and production of indole at 44°C, a procedure that would allow a number of other coliforms (most notably strains of *Klebsiella oxytoca* and *Kluyvera* spp.) to be confirmed as *E. coli*.

ISO 17994 also gives some information on conducting preliminary statistical evaluation of the data from each participating laboratory and each water type used, suggesting simple analysis of variance or its non-parametric equivalents, but gives little other advice. The protocol developed for the DWI and published in *The Microbiology of Drinking Water – Part 3* (Standing Committee of Analysts, 2002) does describe how to undertake such preliminary analyses through testing for normality and using appropriate parametric or non-parametric methods on untransformed or $\log_{10}$-transformed data.

Adopting new methods - Conclusions

The recent work by the DWI and ISO discussed in this paper provides laboratories, for the first time, with the tools to evaluate new methods in water microbiology reliably, prior to adoption in the laboratory. The *ISO/TR 13843* gives clear guidance on what information a laboratory may expect to receive from a manufacturer/supplier of new media or methods. Additionally, the analytical protocols developed for the DWI, coupled with *ISO 17994*, provide a robust and verifiable process for demonstrating equivalent or superior performance of any new method, whether against the laboratory's own current method or a nationally stipulated one.

Acknowledgements

References

ARMON R and PAYMENT P (1988) A modification of m-CP medium for enumerating *Clostridium perfringens* from water samples. *Can. J. Microbiol.* **34** 78-79.

BSI (2003) *Water Quality – Enumeration of Micro-Organisms in Water Samples – Guidance on the Estimation of Variation of Results with Particular Reference to the Contribution of Uncertainty of Measurement. DD 260:2003.* London: British Standards Institution.

COVERT TC, RICE EW, JOHNSON SA, BERMAN D, JOHNSON CH and MASON PJ (1992) Comparing defined-substrate coliform tests for the detection of *Escherichia coli* in water. *J. AWWA* **84** 98-104.

DWI (2000) *The Comparison of Methods for Drinking Water Bacteriology – Cultural Techniques.* Report of the Drinking Water Inspectorate, Department of the Environment, Transport and Regions, London, UK. (available from http://www.dwi.gov.uk/regs/ infolett/2000/pdf/il500pro.pdf) Accessed on 19 October 2003.

EUROPEAN UNION (1998) Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption. *Off. J. Eur. Comm.* 5 Dec 1998. L330/32-L330/53.

FRICKER EJ, ILLINGWORTH KS and FRICKER CR (1997) Use of two formulations of Colilert and QuantiTray™ for assessment of the bacteriological quality of water. *Water Res.* **31** 2495-2499.

HAVELAAR AH, HEISTERKAMP SM, HOEKSTRA JA and MOOIJMAN KA (1993) Performance characterisation of methods for the bacteriological examination of water. Water Sci. Technol. **27** 1-13.

ISO (1999a) General Requirements for the Competence of Testing and Calibration Laboratories. ISO/IEC 17025:1999. Geneva: International Organisation for Standardization.

ISO (1999b) Water Quality – Enumeration of Culturable Micro-Organisms – Colony Count by Inoculation in a Nutrient Agar Culture Medium. ISO 6222:1999. Geneva: International Organisation for Standardization.

ISO (2000a) Water Quality – Detection and Enumeration of Escherichia coli and Coliform Bacteria – Part 1: Membrane filtration method. ISO 9308-1:2000, Geneva: International Organisation for Standardization.

ISO (2000b) Water Quality – Detection and Enumeration of Intestinal Enterococci – Part 2: Membrane Filtration Method. ISO 7899-2:2000. Geneva: International Organisation for Standardization.

ISO (2000c) Water Quality – Guidance on Validation of Microbiological Methods. ISO/TR 13843:2000. Geneva: International Organisation for Standardization.

ISO (2004) Water Quality – Criteria for the Establishment of Equivalency between Microbiological Methods. ISO 17994:2004. Geneva: International Organisation for Standardization.

LIGHTFOOT NF and MAIER EA (eds.) (1998) *Microbiological Analysis of Food and Water: Guidelines for Quality Assurance.* Amsterdam: Elsevier. 266 pp.

LIGHTFOOT NF, RICHARDSON IR and HARFORD JP (2001) The use of Lenticules™ for the process control of enumeration techniques in food and environmental microbiology. *J. Appl. Microbiol.* **91** 660-667.

SARTORY DP and HOWARD L (1992) A medium detecting ß-glucuronidase for the simultaneous membrane filtration enumeration of *Escherichia coli* and coliforms from drinking water. *Lett. Appl. Microbiol.* **15** 273-276.

STANDING COMMITTEE OF ANALYSTS (2002) *The Microbiology of Drinking Water 2002 – Part 3 – Practices and Procedures for Laboratories. Methods for the Examination of Waters and Associated Materials.* Nottingham: Environment Agency. (available from http://www. envirnment-agency.gov.uk/ science/399393/401849) Accessed on 19 October 2003.

TILLETT HE and LIGHTFOOT NF (1995) Quality control in environmental microbiology compared with chemistry: what is homogeneous and what is random? *Water Sci. Technol.* **31** 471-477.

TILLETT HE and SARTORY DP (2004) Measurement uncertainty in water microbiology may not estimate the laboratory's precision. *Accredit. Qual. Assur.* **9** 629-632.

USEPA (1995) *Protocol for Alternative Test Procedures for Coliform Bacteria in Compliance with Drinking Water Regulations – Presence/Absence Membrane Filter Methods for Finished Waters.* Cincinnati: US Environmental Protection Agency.