ENVIRONMENT AGENCY

The Microbiology of Drinking Water (2002) - Part 3 - Practices and procedures for laboratories

Methods for the Examination of Waters and Associated Materials

The Microbiology of Drinking Water (2002) - Part 3 - Practices and procedures for laboratories

Methods for the Examination of Waters and Associated Materials

This booklet contains details of practices and procedures that should be adopted within laboratories.

Within this series there are separate booklets dealing with different topics concerning the microbiology of drinking water. Other booklets include

Part 1 - Water quality and public health
Part 2 - Practices and procedures for sampling
Part 4 - Methods for the isolation and enumeration of coliform bacteria and *Escherichia coli* (including *E. coli* O157:H7)
Part 5 - A method for the isolation and enumeration of enterococci by membrane filtration
Part 6 - Methods for the isolation and enumeration of sulphite-reducing clostridia and *Clostridium perfringens* by membrane filtration
Part 7 - Methods for the enumeration of heterotrophic bacteria by pour and spread plate techniques
Part 8 - Methods for the isolation and enumeration of *Aeromonas* and *Pseudomonas aeruginosa* by membrane filtration
Part 9 - Methods for the isolation and enumeration of *Salmonella* and *Shigella* by selective enrichment, membrane filtration and most probable number techniques
Part 10 - Methods for the isolation of *Yersinia*, *Vibrio* and *Campylobacter* by selective enrichment

Whilst specific commercial products may be referred to in this document this does not constitute an endorsement of these particular materials. Other similar materials may be suitable and all should be confirmed as such by validation of the method.

Contents

Glossary

CI	confidence interval
COSHH	Control of Substances Hazardous to Health
E. coli	*Escherichia coli*
MPN	most probable number
MPR	most probable range

About this series

Introduction

This booklet is part of a series intended to provide authoritative guidance on recommended methods of sampling and analysis for determining the quality of drinking water, ground water, river water and sea water, waste water and effluents as well as sewage sludges, sediments and biota. In addition, short reviews of the most important analytical techniques of interest to the water and sewage industries are included.

Performance of methods

Ideally, all methods should be fully evaluated with results from performance tests. These methods should be capable of establishing, within specified or pre-determined and acceptable limits of deviation and detection, whether or not any sample contains concentrations of parameters above those of interest.

For a method to be considered fully evaluated, individual results from at least three laboratories should be reported. The specifications of performance generally relate to maximum tolerable values for total error (random and systematic errors) systematic error (bias) total standard deviation and limit of detection. Often, full evaluation is not possible and only limited performance data may be available.

In addition, good laboratory practice and analytical quality control are essential if satisfactory results are to be achieved.

Standing Committee of Analysts

The preparation of booklets within the series "Methods for the Examination of Waters and Associated Materials"

and their continuing revision is the responsibility of the Standing Committee of Analysts. This committee was established in 1972 by the Department of the Environment and is now managed by the Environment Agency. At present, there are nine working groups, each responsible for one section or aspect of water quality analysis. They are

1 General principles of sampling and accuracy of results
2 Microbiological methods
3 Empirical and physical methods
4 Metals and metalloids
5 General non-metallic substances
6 Organic impurities
7 Biological methods
8 Biodegradability and inhibition methods
9 Radiochemical methods

The actual methods and reviews are produced by smaller panels of experts in the appropriate field, in co-operation with the working group and main committee. The names of those members principally associated with this booklet are listed at the back of the booklet.

Publication of new or revised methods will be notified to the technical press. An index of methods is available from the Secretary.

Every effort is made to avoid errors appearing in the published text. If, however, any are found, please notify the Secretary.

Dr D Westwood
Secretary

January 2002

Warning to users

The analytical procedures described in this booklet should only be carried out under the proper supervision of competent, trained analysts in properly equipped laboratories.

All possible safety precautions should be followed and appropriate regulatory requirements complied with. This should include compliance with the Health and Safety at Work etc Act 1974 and all regulations made under the Act, and the Control of Substances Hazardous to Health Regulations 1999 (SI 1999/437). Where particular or exceptional hazards exist in carrying out the procedures described in this booklet, then specific attention is noted.

Numerous publications are available giving practical details on first aid and laboratory safety. These should be consulted and be readily accessible to all analysts. Amongst such publications are; "Safe Practices in Chemical Laboratories" and "Hazards in the Chemical Laboratory", 1992, produced by the Royal Society of Chemistry; "Guidelines for Microbiological Safety", 1986, Portland Press, Colchester, produced by Member Societies of the Microbiological Consultative Committee; and "Safety Precautions, Notes for Guidance" produced by the Public Health Laboratory Service. Another useful publication is "Good Laboratory Practice" produced by the Department of Health.

Practices and procedures for laboratories

1 Introduction and scope

The microbiological analysis of water and associated materials involves the use of selective procedures and media. In addition, the nature of the organisms being isolated and enumerated can present challenges to analysts. It is, therefore, important that the media used by a laboratory are prepared, and the procedures conducted, in such a way that the results reflect the quality of the water being tested and that the data generated are reliable. It is essential that a laboratory is able to demonstrate that results produced are fit for the purpose for which they are to be used. This can be achieved by implementing an appropriate programme of quality assurance. Statutory regulations define the performance criteria of methods for compliance purposes for the monitoring of drinking water supplies. Methods should be capable of establishing, within acceptable limits of deviation and detection, whether the sample contains numbers of selected groups of micro-organisms which contravene prescribed values. Depending on the test being used, it is necessary to be able to demonstrate the presence (or absence) of particular micro-organisms or class of micro-organisms in a given sample volume, and to estimate their numbers. The detection of small numbers of organisms is particularly important. Any effective quality assurance programme should, therefore, cover the whole process from sample collection to reporting and interpretation of results. The programme should also include a system of internal quality control, and participation in an appropriate proficiency testing scheme.

Any laboratory where the analysis of water and associated materials is undertaken should operate a quality system. The main functions of such a system are to define the processes that have been put into place to ensure that results are reliable and have been performed to recognised procedures by properly trained staff using suitable equipment. A good quality system enables analytical data to be audited and provides documentary evidence that data generated are accurate and precise. A quality system also provides the basis for documenting structures for the laboratory and staff, equipment and associated service and calibration, and methods that the laboratory uses. The quality system also acts as a reference system for any documentation relating to the laboratory and its operation.

The quality system will depend on the content of a number of documents, each of which is inter-dependent on other documents for its correct function. This booklet describes the basic requirements of a quality system, coupled with criteria for equipment and materials, which enable the reliable analysis of water and associated materials to be undertaken. Guidance on basic analytical procedures, and statistical considerations concerning results, is also given, together with protocols for comparing methods prior to adoption of a new method within a laboratory.

2 The quality manual

Most quality systems begin with a quality manual that defines the laboratory and its policy towards quality. The manual should be broad in its approach, covering all aspects that are considered important. It should also be simple in that it is easy to read and be understood by all members of staff, and easy to maintain in the ever-changing circumstances of the laboratory. In broad outlines, the manual should contain a quality statement, details of the laboratory in terms of location and staff structure, and how it will set, maintain and check quality standards. In addition, the manual is a useful document for incorporating health and safety policies, safe working procedures and environmental policies.

The quality manual should define records that the laboratory should keep and maintain, the nature of measurement calibrations, the format of analytical procedures, internal and external quality assurance, and resulting remedial action which should follow in the event of a failure in documented procedures.

Every laboratory should have an organisation chart showing staff posts and associated role profiles in day to day tasks, and importantly, the chain of accountability and reporting. The post responsible for the quality assurance programme should be clearly defined and each member of staff should have a well-defined job description outlining responsibilities. A record of staff training should be maintained and regularly updated. In addition, there should also be a defined plan for provision of cover for work when staff members are absent.

The reporting system should enable results which require immediate remedial action to be communicated without delay to appropriate persons. Records of laboratory results should be kept for as long as is necessary to comply with requirements for archive and audit trails, and should include, for example the date, place and time of sampling, and the members of staff undertaking the sampling and analysis.

The quality assurance programme can only be successful if all constituent parts are well documented. The manual should provide valuable source material covering all activities of the laboratory and should only require periodic review.

3 Laboratory records

The laboratory should have a system for the purpose of recording data relating to equipment, measuring devices, quality control and environmental monitoring. The following are considered to be important components of a laboratory quality system.

3.1 Staff training records

Staff training records should show appropriate training for each documented method where training has been given. In addition, the record should show training in the use of major items of equipment. There should also be documented evidence that training has been both adequate and successful. This may involve staff analysing external quality assurance samples where their data can be compared with data from other analysts or laboratories. Alternatively, spiked or raw water samples, containing low numbers of target organisms, may be used provided that replicate samples are also analysed by a fully trained member of staff. In order to demonstrate confidence, an appropriate number of replicates should be tested using the full analytical procedure. Details of the comparisons of the test results could also be documented in the training record. Training records can also be used to store documentary evidence of additional training, for example, courses, conferences, workshops etc that the analyst has attended.

3.2 Analytical procedures

There should be appropriate documentation of all analytical procedures in current use. Such documentation should be controlled, for example with an appropriate issue number, date of issue, appropriate pagination and known documented circulation. Where methods are revised, the original method should be stored for possible future reference. A master copy of all methods should be kept in a secure place and relevant copies issued to the laboratory. Such copies need not contain all the methods and may include only those applicable to a particular section (for example, a drinking water microbiology section). It may be useful to store

validation data with the appropriate method, together with the protocol used for conducting the validation exercise.

3.3 Equipment lists and service records

All major items of equipment should be catalogued and include data on, for example, the date of purchase, the name of the supplier, the frequency of servicing and calibration, and, if appropriate, the location of instruction manuals. The type of equipment typically covered may include, for example, incubators, water baths, autoclaves, refrigerators and microscopes. Service records of such equipment may also be stored and include reports prepared by the service engineer and details of any calibration carried out on the equipment. Details of equipment faults, repairs and the upgrading of any equipment may also be stored.

3.4 Laboratory calibration records

Laboratories should have well documented procedures for the calibration of all equipment that involves, for example, recording weight, volume, temperature or time. Calibration standards for such equipment may include certified standards, for example thermometers, and should include a certificate that traces the calibration to national standards. Alternatively, uncertified standards that have been calibrated by an accredited laboratory may be used. Certified standards need not be used routinely, but could be used to calibrate uncertified working standards on a regular basis. Once calibrated, these standards should themselves be re-calibrated as appropriate to national standards.

3.5 Environmental monitoring

Because of the ubiquitous nature of most of the microbes of interest, it is essential to ensure that any organisms that are detected have originated from the original sample and have not been introduced inadvertently during sampling or subsequent analysis. Laboratories should consider appropriate monitoring of the environment, both related to the sampling procedure and the analysis within the laboratory, to ensure that any micro-organisms detected do not adversely contribute to any result.

4 Materials used in a microbiology laboratory

Chemicals used in a microbiology laboratory should be of analytical grade quality. Reagents and chemicals should be stored and used in accordance with manufacturer's instructions and discarded if the expiry date, ie the date by which the reagent should be used, has passed. Reagents and chemicals are usually supplied with hazard assessment and toxicity data. Records of these data should be kept, and any specific hazard documented in each method and COSHH assessment[1]. Chemicals and reagents should always be handled with care and any spillages that occur should be cleaned up immediately.

4.1 Media

Manufacturers can supply most microbiological media as dehydrated formulated preparations. Media that are supplied as dehydrated powders should be stored in a cool dry place, and containers labelled clearly with the date of receipt, and the dates when containers are opened. When a container is opened for the first time, the laboratory should allocate an expiry date to the formulation relative to its potential deterioration. Manufacturers may provide details of expiry dates and storage conditions. Whilst details of the preparation and sterilisation of media are usually provided by the manufacturer, these should also be documented in the

analytical method (see section 8.1). Most powdered media are hygroscopic. Hence, after the containers have been opened and media removed, the lid should be tightly secured to reduce further absorption of moisture. Over a period of time, some media absorb excessive amounts of water and solidify. This usually results in discoloration of the medium and denaturation of the nutrients or selective properties. Any such medium should, therefore, be discarded, even if the expiry date has not passed.

Media are prepared by weighing out the appropriate amount of material and adding distilled, deionised or similar grade water (see section 4.4). Many media contain selective chemicals and where these are supplied as powders, appropriate containment measures should be taken. Media, as broth, dissolve readily in cold water. Those that contain agar do not completely dissolve in water without heating. Agar is an extract of seaweed and is usually added to media to a concentration of between 1.0 - 1.5 % depending on the purity of the agar. Agar only dissolves completely in water if the solution is heated to boiling point. Agar solutions solidify or set at approximately 42 °C. Any broth medium may be solidified by the addition of the appropriate amount of agar.

4.2 Preparation of media

The appropriate weight of the dehydrated formulated preparation, or ingredients as listed in the method, is added to the appropriate volume of water. Whilst it is often not essential to have to adjust the pH of the medium, in certain circumstances it will be necessary to do so. Adjusting the pH should be carried out by the addition of small amounts of an appropriate concentration of hydrochloric acid or sodium hydroxide solution until the required pH value is achieved. This is often carried out before sterilisation takes place. For example, when membrane lauryl sulphate broth is used for the enumeration of coliform bacteria in water, it should have a pH of 7.4 ± 0.2 after sterilisation. Sterilising the solution by heating may cause breakdown of the lactose in the medium and result in a lowering of the pH. It may, therefore, be necessary to raise the initial pH of the medium by 0.2 - 0.4 units before sterilisation, in order to achieve a final pH of 7.4 ± 0.2 after sterilisation. The pH of a medium should be measured as soon as practicable after it has reached room temperature. Media should not be allowed to stand at room temperature for significant periods of time (ideally, not more than 2 hours) before testing, as the pH may change over time.

All dehydrated media should be completely dissolved before being dispensed and sterilised. Failure to do so may result in charring of some of the media ingredients during the sterilisation process. Some media, for example membrane enterococcus agar used for the enumeration of enterococci in water, do not require sterilisation because they are sufficiently selective to prevent the growth of heat resistant organisms. Heating is required only to dissolve the agar and excessive heating may adversely affect the medium. Such media should be cooled quickly after heating to minimise breakdown of heat sensitive constituents.

A small amount of prepared medium can be used to check the final pH of the medium and to determine growth and selectivity characteristics before the remaining bulk of the medium is used.

Typically, media are sterilised by autoclaving at 115 °C for 10 minutes or 121 °C for 15 minutes. It is important that media should not be overheated. Media should not be autoclaved more than once. Sterilised media should be removed from the autoclave as soon as practicable after sterilisation is complete and due care should be taken as media may be super-heated and boil rapidly. Once removed from the autoclave, the tops on the containers may be tightened and the media allowed to set. Solidified media may be stored in the dark at room temperature

(no greater than 25 °C), should be subjected to quality control tests and used only when the tests have shown that the medium gives satisfactory performance. A prepared medium should be allocated an expiry date and this should be stated in the method. Media can be melted by heating in steam at normal atmospheric pressure or by using a microwave oven on low power.

Before a medium is poured into Petri dishes or tubes, it should be cooled, for example in a water bath or incubator, to approximately 50 °C prior to the addition of any supplements. Media should not be left standing at 50 °C for long periods of time. Media at temperatures above 50 °C should not be poured into Petri dishes, as this may lead to excessive condensation in the Petri dish during the setting process and subsequent storage. When the medium is poured into a Petri dish, a lid is added and the medium is allowed to solidify. When the medium has solidified, the Petri dish should be inverted and stored at a temperature of between 2 - 8 °C in such a way as to prevent excessive drying. This may be achieved by storing the Petri dish in an air-tight container. When the medium in a Petri dish shows excessive drying, the Petri dish should be discarded. The preparation of media should be planned to ensure adequate supplies are available, yet minimise discarding unwanted media.

Freshly prepared media should not be left on benches for any length of time, particularly if these are exposed to sunlight. The action of sunlight on media produces peroxides and other toxic substances that inhibit the growth of certain bacteria. When a medium has been prepared, it should be transferred to a dark environment as quickly as practicable. Media that show obvious signs of contamination or deterioration should be discarded.

Media that have been stored at a temperature of between 2 - 8 °C should be dried before use by leaving at room temperature (no greater than 25 °C) for 2 hours or at 37 °C for 30 minutes. The medium should be poured into a Petri dish to a depth (typically, 3 - 7 mm) sufficient to ensure the medium does not dry out during incubation. During the pouring and subsequent cooling of agar media in Petri dishes, a film of moisture is often left on the surface of the solidified agar. The incubation of an agar medium that has not been dried may result in the spread of bacteria, particularly motile bacteria, across the agar, with the result that colonies may not be isolated on the agar. This may hinder subsequent sub-culture and render them meaningless. When large numbers of agar Petri dishes are being dried, the drying time may need to be extended, or a small amount of drying agent (for example self-indicating silica gel) added to the drying chamber to absorb excess moisture.

4.3   Quality control of laboratory media

Each batch of prepared medium should be identifiable, for example marked with a batch number. For complex media requiring aseptic additions after sterilisation, each bottle of medium may need to be treated as a separate batch. Where commercial media are used, the manufacturer's batch number should be recorded. When a medium has been prepared it should be labelled with its batch number and expiry date. Details of the sterilisation should be available for all media that require autoclaving and these should be recorded together with results of any time cycle checks, for example autoclave temperature charts. The signature of appropriate staff can also be included in these records as a means of demonstrating a suitable audit trail to show correct preparation of the media.

Once preparation is complete, a small aliquot of the medium should be checked for pH. The medium should be within ± 0.2 pH units of that stated in the method. If the medium is outside the defined pH limits it should be discarded. The pH of the medium should not be adjusted after sterilisation because of the risk of introducing microbial contamination. Such effects may not be immediately obvious but may become significant during storage.

When the sterilised medium has been poured into Petri dishes, a representative number should be checked to ensure that they are satisfactory. These Petri dishes should be incubated at a temperature and time appropriate for the medium and examined for any contaminant microbial growth. Where liquid media are prepared bottles or tubes should be similarly checked. In addition, the broth may be inoculated onto an appropriate nutrient agar medium and incubated at an appropriate temperature to assess effective sterilisation. Media should also be checked to establish that they support the growth of the target organisms for which they have been prepared, and differentiate or are selective against non-target organisms. Appropriate organisms should be identified and tested with each medium. A list of typical organisms is given in Table 4.1.

Table 4.1   Examples of bacterial strains used for quality control of media for water microbiology

Bacterial strain	Reference strains	Media or reagents
Escherichia coli	NCTC 9001	membrane lauryl sulphate broth
	NCIMB 11943	membrane lactose glucuronide agar
		negative control for oxidase reagent
Enterobacter aerogenes	NCTC 10006	membrane lauryl sulphate broth
Enterococcus faecalis	NCTC 775	membrane-enterococcus agar
Clostridium perfringens	NCTC 8237	tryptose sulphite cycloserine agar
	NCIMB 6125	
Pseudomonas aeruginosa	NCTC 10662	pseudomonas agar
	NCIMB 8295	positive control for oxidase reagent
		non-target control for membrane lauryl sulphate broth
Salmonella dublin	NCTC 9676	Rappaport-Vassiliadis enrichment broth
		xylose lysine desoxycholate agar

NCIMB = National Collections of Industrial and Marine Bacteria
NCTC = National Collection of Type Cultures

4.4   Quality of distilled and deionised water

In most circumstances, tap water is unsuitable for the preparation of laboratory media. Tap water may contain unwanted organic carbon, and inorganic and metal ions. These substances may be inhibitory to bacteria, and important components of the medium may change or precipitate out of solution.

Water used for preparing media can be produced in a variety of ways. Distillation should be carried out in Pyrex glass vessels. A single distillation process may not be sufficient to remove all organic material likely to be present in tap water, unless the water is first deionised. Deionisation will remove many of the organic and inorganic compounds found in water, will also minimise scale formation in the distillation apparatus and can be used to produce good quality water. Whilst ion exchange resins will remove inorganic ions from water, carbon filtration will remove organic material. Reverse osmosis can be used where substantial amounts of water are required.

Once produced, water will readily absorb substances, for example carbon dioxide and ammonia, from the atmosphere and the container in which it is stored. Electrical conductivity

measurements are often used as a measure of the ionic content of water. The electrical conductivity of water suitable for bacteriological analysis should, ideally, measure less than 3 μS/cm. Electrical conductivities of water of less than 0.5 μS/cm are generally achievable using deionising or double distillation sytems. Single distillation systems may produce water with electrical conductivities of up to 5 μS/cm. The electrical conductivity of the water should be measured regularly, depending upon use, and the results recorded. Water stored in large containers may contain sufficient nutrients to allow microbial growth. The bacteriological quality (for example the enumeration of heterotrophic bacterial content, see Part 7 in this series) of the water should, therefore, be monitored regularly, depending upon the rate of production and its usage. Distilled or deionised water is not sterile, and may contain low numbers of organisms per millilitre. Where bacteriological counts are found to be excessive, for example over 1000 per ml in freshly produced water, or over 10000 per ml in stored water, then consideration should be given to changing the manner in which the water is produced and stored. Consideration should also be given to cleaning and disinfecting the container on a regular basis.

If water is stored for long periods, it may be appropriate to determine chlorine residual, total organic carbon and ammonia levels. Respective levels should be less than 0.02, 0.5 and 0.05 mg/l.

5   General laboratory equipment

It is important that all equipment is kept clean, is well maintained and is clearly identified so that appropriate records can be traced quickly. Any spillages should be cleaned up immediately. Cleaning, servicing and maintenance should be carried out to a standard following manufacturer's instructions. See also sections 3.3 and 3.4.

5.1   pH meters

pH meters are designed to measure the hydrogen ion concentration at ambient temperature (ie 15 - 25 °C). When not in use, pH electrodes should be stored according to manufacturer's instructions. The pH meter should be calibrated according to manufacturer's instructions and this may involve a daily operation using 2 or more freshly prepared buffer solutions in the appropriate pH range. The results of calibration data should be recorded. The response of electrodes (for example slope and millivolt output) should also be checked regularly. Flat-tip membrane electrodes or spear-tip electrodes are suitable for measuring pH values of solid media, simply by touching the surface or spearing the agar. Flat-tip membrane electrodes may require the filling solution to be replaced at regular intervals, according to the manufacturer's instructions, as electrolyte often leaches from the end of the electrode. Particular attention should be paid to rinsing the electrodes after use, as a build-up of organic material can severely inhibit electrode response.

5.2   Incubators

Incubators are available in many shapes and sizes with, or without, fan assistance. Fan-assisted incubators provide a more even temperature distribution inside the cabinet. The inside of the incubator should be made of stainless steel (for easy cleaning) and there should be a glass or Perspex door inside the main door (to help minimise temperature loss but allow viewing of the incubator contents or thermometers when the main door is opened). The temperature of the incubator should be measured at regular intervals. The minimum number of readings that should be taken include one at the beginning of the working day, before cultures have been removed, and one at the end of the working day, particularly with

temperature cycling incubators. Continuous temperature monitoring (with associated alarm systems) of the internal environment should be considered as it provides a complete appraisal of incubator performance, particularly for incubators with temperature cycling. Continuous monitoring allows realistic assessments of temperature fluctuations within incubators. Whether incubators are, or are not, fan-assisted, it is important that an even temperature distribution is established within the incubator. This can be achieved by placing thermometers, or temperature recording devices, in different parts of the incubator over a period of time, for example one week, and recording the temperatures at regular intervals. The temperature profile of the incubator should show no significant differences wherever the thermometers are placed. The temperature distribution may also depend upon the manner in which the incubator is loaded. For example, stacking Petri dishes in columns of greater than six dishes may affect the temperature distribution, and result in the temperature profile of individual dishes being different. Correct incubator temperature control is vital for the satisfactory performance of microbiological enumeration and detection. The temperature control of incubators is described in section 5.8.

5.3   Pipettes and pipettors

Many laboratories use sterile, plastic disposable 1 ml and 10 ml pipettes for routine microbiological purposes. These pipettes deliver the measured volume between the graduation and the tip of the pipette. Any pipettes that are damaged, or broken, should therefore be discarded. Volumes are usually dispensed with the aid of a pipette bulb and pipettes are plugged with non-absorbent cotton wool to prevent contamination of the contents of the pipette and the bulb when pipetting samples and cultures. A representative number of pipettes from each batch should be checked to confirm delivery of correct volumes.

Automatic pipettors and pipette tips can be used to dispense fixed or adjustable volumes of liquids. This is achieved by air displacement using a manually operated or electrically powered piston within the pipettor. There is a risk of the pipettor barrel or piston becoming contaminated and, therefore, plugged pipette tips should be used. A pipette tip of the correct size for the pipettor should be used in accordance with manufacturer's recommendations. Loosely fitting tips may leak, may not deliver the correct volume or may fall off the end of the pipettor when being used. Automatic pipettors should be used, stored when not in use, and calibrated when appropriate according to manufacturer's instructions.

Pipette tips can be purchased sterile, packaged either as individually wrapped, or in small convenient numbers. Pipette tips can also be placed into suitable containers and sterilised by autoclaving at 121 °C for 15 minutes. If containers are wet on removal from the autoclave they should be dried, by placing them in an incubator, before being used. Should the outside of the pipettor become contaminated during use it should be disinfected, by wiping with 70 % ethanol or 2-propanol, prior to continued use.

New pipettors should be calibrated before use, and at suitable intervals, according to manufacturer's instructions. The volumes chosen should represent the range of volumes for which the pipettor is likely to be used. For each volume chosen, the data are recorded and used to calculate mean volume dispensed, standard deviation and coefficient of variation. Ideally, the coefficient of variation should be less than 1 % and the bias should be less than 2 % of the volume chosen. Details, for example dates and staff undertaking calibrations, should be recorded and stored for each pipettor.

5.4   Refrigerators and freezers

The temperature of refrigerators used for the storage of media, reagents and samples should be between 2 - 8 °C. Each refrigerator should contain a calibrated thermometer which is used to record the temperature on a regular basis. Where possible, media and reagents should be stored in separate refrigerators or compartments, and should not be stored in such a manner that the temperature of the compartment is adversely affected. Samples should not be stored in the same refrigerator as media. Modern refrigerators and freezers are usually available as frost-free items, but older equipment may require regular defrosting. The use of a sodium bicarbonate solution, to wash down the inside of refrigerators and freezers, helps reduce the growth of fungi.

Freezers usually operate at a temperature of - 20 °C, but cabinets that operate at a temperature of - 70 °C are available. As with refrigerators, the temperature of freezers should be checked regularly.

5.5   Anaerobic jars

Anaerobic jars are used to encourage the growth of anaerobic and micro-aerobic bacteria. They usually comprise a polycarbonate jar with a close fitting lid held in place by a clamp. Before incubation, oxygen in the jar is removed by generating hydrogen and reacting the two gases (to form water) with the aid of a platinum catalyst. Hydrogen can be generated following the addition of water to a sachet of sodium borohydride. Great care should be taken to ensure the correct size of sachet for the appropriate volume of jar is used. The efficiency of the platinum catalyst is adversely affected by moisture, and therefore the catalyst should be dried after use by placing it in an oven at a temperature of between 80 - 100 °C for 30 minutes. Alternatively, the catalyst can be placed in a sealed container with a suitable desiccant, such as silica gel, and stored in an incubator at a temperature of between 37 - 44 °C between periods of use. As an alternative technique, oxygen can be removed from the jar by absorbing the gas using a selection of chemicals. Anaerobic jars should be cleaned after use, or disinfected if necessary.

Cultures should be stacked loosely in the jar. The caps of screw-topped containers should be loose. After loading the jar, the appropriate conditions are established, together with a means of establishing whether the conditions have been attained. This can be achieved using anaerobic indicator strips, or the inclusion of two bacterial cultures, one, which is aerobic, and another, which is micro-aerobic or anaerobic. The correct incubation of materials is only achieved if the indicator strips or bacterial cultures show that suitable internal atmospheric conditions have been achieved.

Anaerobic and micro-aerobic systems based upon small plastic bags and gas generating systems are commercially available. These are suitable for the incubation of small numbers of cultures. Before use, new batches should be performance tested with appropriate anaerobic or micro-aerobic organisms.

5.6   Autoclaves

The principle of sterilisation by autoclaving is based on moist heat transfer. Autoclaving is used to sterilise media, bottles and other equipment used in the bacteriological examination of waters. Heat is applied in the form of steam, under pressure in the absence of air. Steam may be generated in a boiler that is separate to the sterilisation chamber. Alternatively, steam can be generated by the direct heating of water in the bottom of the chamber. Where steam is

generated in a boiler separate to the sterilisation chamber, air is displaced more quickly than it is when steam is generated in the bottom of a chamber. Hence, the medium heats up and cools down faster. Sterilisation is timed from the moment when materials in the autoclave attain the appropriate sterilisation temperature. In order for correct sterilisation to take place, it is essential that steam penetrates the load, and that the heating time is not adversely affected by overloading the autoclave, both in terms of large numbers or volumes of objects in the autoclave.

Autoclaves vary in complexity, and range from simple pressure cooker systems to complex microprocessor-controlled machines, capable of a variety of sterilisation cycles. The autoclave cycle comprises an initial heating period, a period of free steaming (where air is purged from the chamber), a further heating period (where the contents are raised to the sterilisation temperature), a holding period at the sterilisation temperature, and finally a cooling period. Guidance on use, performance and construction of laboratory autoclaves is given elsewhere[2, 3].

It is important that the correct time and temperature are achieved during each sterilisation cycle and that these are monitored and recorded. The internal temperature of the autoclave should be established and verified during a sterilisation cycle. This can be achieved, for example using a multi-point thermocouple calibration procedure conducted on an appropriate load within the autoclave. Calibrations should be carried out following installation, repair or modification to the autoclave and all data recorded and stored. Thermocouples should be calibrated to national standards before calibration procedures are undertaken.

Details of each cycle of the autoclave should be recorded, together with the contents of the autoclave. A temperature cycle or sterilisation time-temperature record provides an audit trail to show correct sterilisation has taken place. Individual loads can, in addition, be marked with heat-indicating tape to demonstrate that they have been subject to a heat process.

Autoclaves should be serviced at regular intervals according to manufacturer's instructions. For autoclaves where steam is generated by the direct heating of water in the bottom of the chamber, the water should be replaced frequently. Failure to do may result in corrosion within the chamber, particularly the heating elements.

Contaminated materials and media should not be autoclaved together. Furthermore, bottles of media should not be filled completely, and caps or stoppers should be loose fitting. Failure to loosen the cap or stopper may result in the bottle exploding. It should not be possible to open autoclaves until the sterilisation cycle is complete and the temperature has cooled down to a designated safety level. Although the temperature inside the autoclave may register, for example 80 °C, the temperature of the contents may remain above this. It is important, therefore, that when the sterilisation cycle is complete, the autoclave is opened carefully and that appropriate safety equipment is used when the contents are removed.

5.7   Water baths

Water baths may be used to incubate certain cultures or keep agar media in a molten state until ready for use. They usually comprise a stirrer or circulating pump with a heating element and thermostat. When water baths are used to incubate cultures, the water should be distilled or deionised, always be stirred or circulated within the bath and switched on only when the water has reached the recommended level. When not in use, water baths should be stored according to manufacturer's instructions and should be drained at regular intervals and wiped with a suitable disinfectant (for example 70 % ethanol or 2-propanol) before being refilled.

When in use, the temperature of the water should be measured regularly using a calibrated thermometer.

5.8   Thermometers (or temperature recording devices) and temperature control

Thermometers may be of the mercury-in-glass or alcohol-in-glass type and are available in a wide variety of temperature ranges and sizes. They may be available calibrated to national standards or be un-calibrated. Certified and calibrated thermometers require re-calibration and certification every 5 years. Certified thermometers can be used to calibrate working reference thermometers that may subsequently be used to calibrate working thermometers used to measure temperatures within the laboratory.

When thermometers, or temperature recording devices, are used, they should be capable of measuring a given temperature within a specified tolerance. For a temperature of between 20 - 40 °C, for example in incubators, the maximum fluctuation around the given temperature should be ± 1 °C. In these circumstances, thermometers capable of measuring within a tolerance of ± 0.5 °C can be used, but those measuring to within ± 0.2 °C would be better. For regulatory drinking water compliance purposes, it may be more appropriate to use thermometers that measure to within ± 0.1 °C. For incubators set at 40 °C or above, the maximum fluctuation around the given temperature should be ± 0.5 °C and thermometers should measure to within ± 0.1 °C. In practical terms, if dual temperature incubation is used (for example, incubation at 30 °C and 44 °C) then two thermometers may be required.

With such sensitive equipment and in order to prevent heat loss when the temperature is measured, thermometers can be placed in suitable plastic or glass containers filled with an inert liquid. Suitable liquids comprise glycerol, liquid paraffin or polypropylene glycol. These liquids stabilise temperature measurements and prevent a sudden drop in the temperature when the thermometer and container are removed from the incubator for reading. Bottles containing thermometers should be placed in the incubator in locations reflective of the incubated samples or materials. Mercury-in-glass thermometers are fragile and may, if broken, present a health hazard. Consequently, they should be placed inside protective cases that do not interfere with the temperature measurements. Thermometers should not be used if the mercury or alcohol column is broken.

Working thermometers should be calibrated at regular intervals and any errors compared with the reference thermometer should be no greater than the tolerance of the reference thermometer. Electronic temperature recording devices (digital thermometers and thermocouples) are now available. These can be used to monitor temperatures continuously and data can be stored, retrieved and inspected. It is essential that such devices be regularly calibrated against certified thermometers or temperature recording devices.

5.9   Balances

Balances should possess a sensitivity that is appropriate for the substance being weighed. They should be clean and serviced regularly, and calibrated to national standards at pre-determined frequencies. The preparation of media and reagent requires careful and accurate use of balances and it is essential that balances are located in suitable positions away from sources of excessive vibration, temperature variation and air movements.

Balances used for general purposes, for example top pan balances, should be accurate to ± 0.01 g. Where greater accuracy is required, for example analytical balances used for weighing amounts of less than 1.0 g, balances should be accurate to at least ± 0.0001 g. A

range of calibrated weights, appropriate for the balance in use, should be available for checking the accuracy of the balance. This should be undertaken on a regular basis, for example weekly or daily, depending on use. Balances not within specified tolerances should not be used until re-calibrated.

5.10   Microscopes

Microscopes can be used to examine material and facilitate the classification of micro-organisms. The microscope performs efficiently only if it is serviced regularly depending to its usage, is aligned correctly[4], is protected from environment contamination and is used according to manufacturer's instructions. Details of service records should be recorded and stored. When not in use, the microscope should be protected with a dust cover to prevent optical surfaces from becoming covered in dust, which might affect its use. In addition, the optics and stage should be cleaned with lens tissue after use. When microscopes with ultra-violet illumination are used, the period of use should be recorded and bulbs replaced at appropriate frequencies. When ultra-violet bulbs are replaced, safety gloves and eye protection should be worn as these types of bulbs can explode during replacement. Direct human contact with bulbs should be avoided so as not to etch the glass.

Microscopes can be calibrated using an eyepiece graticule and stage micrometer. The graticule is fitted inside the eyepiece and, generally, is marked in a scale of 0 - 100 units. The slide micrometer is usually marked with a 1 mm scale with 0.01 mm divisions. Calibration is normally performed by assessing the number of graticule units that equate to 1 mm on the slide micrometer. When a microscope is used for taking measurements of biological material, calibration should be undertaken on a regular basis.

5.11   Microwave ovens

Microwave ovens can be used to heat liquids, and melt agar quickly and easily before it is dispensed. However, certain precautions need to be taken when microwave ovens are used. When bottles of liquid are heated in a microwave oven, the liquid sometimes becomes super-heated and tends to boil, especially if the bottles are shaken when they are removed from the oven. Using low power for longer periods of time will minimise the risk of liquids becoming super-heated. Also, when bottles are removed from the oven they should not be shaken. Sealed containers can explode within the oven. Bottle caps or stoppers should therefore be loose before the bottles are placed into the oven, and bottles should not be removed from the microwave oven as soon as the heating process is complete, but should be left to cool down. If bottles of liquid are heated in a microwave oven, they should only be half full.

Microwave ovens should always be clean and any spillages that occur should be cleaned up immediately. Microwave ovens should be checked regularly for radiation leakage and to ensure that doors are well sealed.

5.12   Glassware

All items of glassware, such as pipettes, flasks and Petri dishes etc, used in the preparation of media or handling of samples should be of suitable quality. They should also be free from inhibitory substances and be adequately cleaned and sterilised before use. In many instances, pre-sterilised plastic items can be used. The accuracy of volumetric equipment should be traceable to national standards.

6   General laboratory environment

The laboratory environment where microbiological examinations of water are undertaken should comply with guidelines[5] for category 2 containment. Guidelines include provision of sealed non-absorbent floor surfaces, work surfaces that are impervious and resistant to chemicals, and separate hand-washing facilities that are close to the exit of the laboratory. In addition, laboratory cupboards should be labelled with their contents and lighting for all purposes should be adequate. The laboratory environment should be cleaned frequently and in particular, work surfaces should be cleaned and disinfected often. If micro-organisms in a category 3 containment level are intentionally sought and isolated (for example *Salmonella typhi*) then separate and additional requirements[5] need to be complied with.

Whilst laboratory-acquired infection is rare, staff should be adequately trained in the prevention of infection, not only to themselves but also to other colleagues. Training should include the understanding of risks from micro-organisms associated with ingestion, inhalation and skin absorption. Further guidance is given elsewhere [5, 6, 7].

7   Laboratory staff

All laboratory staff should have training records that should show relevant education and experience that have been acquired, see also section 3.1. Failure by an analyst to perform satisfactory should lead to a thorough investigation including both the adequacy of, and response to, the training. Wherever possible, staff should be encouraged to broaden their understanding and to make contact with people from similar organisations. It is important that staff should understand the principles of the tests carried out and the reasons why they are carried out.

In addition to receiving training in analytical techniques, laboratory staff should be made aware of the required hygiene practices that are so important in minimising the risks of infection when handling samples or cultures. These include basic requirements for wearing laboratory coats, the need for hand washing and personal hygiene and the disinfection of laboratory work surfaces and cleaning up of spillages. New employees should be made aware of these practices which should be observed and maintained at all times.

8   Analytical techniques

All analytical techniques should be fully documented, see also section 3.2. This includes giving descriptions of the micro-organisms, the preparation of media, the test procedures and any confirmatory tests that may be required. Each method used in the laboratory should be held in a reference file and be available to all staff. Details of procedures for preparing suspensions of test organisms (for validation purposes) and the practical details of how validation trials are conducted should also be recorded. Methods should also include reference to organisms used as positive and negative controls that are used in the isolation procedures and confirmatory tests, as well as quality control tests for assessing media.

As new methods are developed, test procedures will be replaced. Thus, it is important that new test procedures are properly validated and their performance assessed against the old test procedures. New test procedures should only be adopted after they have been shown that they are equivalent to, or better than, the old test procedures. Details of, and data generated for, the new test procedures should be fully documented. Details how to validate new test procedures are given in section 9.

8.1   Standard operating procedures

Details of laboratory methods can be described in standard operating procedures that should include particulars of the scope of the method, the equipment required, the preparation of media and reagents, full analytical procedures and the calculation and reporting of results. A suitable format for a standard operating procedure is given in section 8.1.1. See also Parts 4 - 10 in this series as indicated on page 2.

8.1.1   Format of a standard operating procedure

1. Introduction - This section gives a brief discussion of the organism for which the method is designed. Details of its significance to drinking water in terms of water quality, indicator value, pathogenicity and occurrence are also, generally, given.

2. Scope - Details of the type of waters that can be analysed are given.

3. Definitions - The organism is defined with specific regard to the method being described.

4. Principle - Brief details of the method are given.

5. Limitations - Brief details of those circumstances where the method is not suitable are given.

6. Health and safety - References to relevant COSHH information and special hazards associated with the method are noted.

7. Equipment - Details of equipment and special apparatus specific to the method. Reference to standard equipment requirements (and performance criteria) in general guidance to laboratories section.

8. Media and reagents - Details are given of all reagents and media that are employed in the method, together with instructions for their preparation and storage and, where appropriate, whether commercial formulations and kits are available.

9. Analytical procedure - This section gives details of the procedures that need to be carried out. The section is often sub-sectioned as follows:-

9.1 *Sample preparation* - guidance is given on volumes, special storage conditions prior to analysis, and pre-treatment or dilution preparations.

9.2 *Sample processing* - Details are given on the technique (ie, membrane filtration, MPN inoculation, pre-enrichment etc) including incubation conditions.

9.3 *Reading of results* - Details of how results are read and recorded are noted (including colony counting, biochemical tests etc.).

9.4 *Confirmation tests* - Details of any biochemical, serological or other tests used in confirmation tests are referred to.

10. Calculations - Details of the procedures required for the calculation of results are presented.

11. Expression of results - Information is given on the terms and units used for the reporting of results.

12. Quality assurance - Information on media, reagents and specification of reference organisms is described.

13. References - Technical and allied references relevant to the method are given.

8.2   Methods for the isolation and enumeration of indicator and other organisms

Two procedures for isolating and enumerating organisms are commonly used in water microbiology. These involve membrane filtration and the multiple tube or most probable number (MPN) technique. The media and incubation conditions differ with both methods according to the organism being sought. In practice, for most conventional testing of clean waters, the membrane filtration procedure is the most widely used, as it is simple to conduct and can be applied to a wide range of organisms.

8.2.1   Preparation of samples

The volume of sample submitted to the laboratory should be sufficient to ensure that all routine examinations can be carried out. Any excess sample that is not required may be stored in a refrigerator until the initial examination has been completed. This sample can then be discarded or, if required, used for additional or repeat tests in the event of unexpected high counts or possible mishaps. The enumerated counts obtained for these additional or repeat tests should either be used with caution and additional comment, or not reported, as they may not be comparable to those of the original sample examination obtained before storage.

To facilitate mixing, an air space should be present in the sample bottle. The sample bottle is inverted rapidly several times to ensure adequate mixing. If ten-fold dilutions of the sample are required, they can be prepared at this stage. Sterile solutions of quarter-strength Ringer's solution or maximum recovery diluent are suitable for preparing dilutions. Known volumes of sterile diluent solution are measured out (for example 90 ml or 9 ml) into sterile dilution bottles or tubes. Alternatively, volumes of diluent, pre-sterilised in screw-capped bottles can be used. In these cases however, it should be recognised that some bottles may suffer a loss of diluent on sterilisation or storage. Volumes should, therefore, be checked and any bottles showing obvious signs of incorrect quantities must be discarded. Whilst the bottle is held in one hand, the stopper or cap is removed with, and retained in, the other hand. A dilution of the original sample is then made, by transferring one volume of sample to nine volumes of diluent. Using a fresh, sterile pipette each time, the process is repeated as often as is necessary to ensure the correct dilution range has been prepared. Each prepared dilution is carefully and thoroughly mixed before additional dilutions are prepared. A sufficient quantity of each dilution should be prepared to enable all tests to be carried out.

8.2.2   The multiple tube (or MPN) technique

In the multiple tube technique, measured volumes of sample, or diluted sample, are added to a series of tubes containing a liquid differential medium. It is assumed that on incubation, each tube with one (or more) target organism will exhibit growth in the medium, and produce characteristic changes in the medium. Provided that some of the tubes exhibit no characteristic growth (ie the results are negative) and some of the tubes exhibit characteristic growth (ie the results are positive) then the MPN of organisms in the sample can be estimated from probability tables. Counts are typically expressed as the MPN of organisms per 100 ml

of sample. Confirmation, that positive results are due to the growth of the targeted organism sought, can be obtained by subculture to confirmation media.

The multiple tube method is particularly suitable for the examination of sludges and waters containing sediment.

8.2.3   The membrane filtration method

In the membrane filtration method, a measured volume of the sample, or diluted sample, is filtered through a membrane filter, typically composed of cellulose-based, or similar, fibres. The pore size of the membrane filter is such that the targeted organisms to be enumerated are retained on or near the surface of the membrane filter, which is then placed, normally face upward, on a differential medium, selective for the targeted organisms sought. The selective medium may be either an agar medium or an absorbent pad saturated with broth. After a specified incubation period, it is assumed that the targeted organisms retained by the membrane filter will form colonies of characteristic morphology and colour. The growth of non-target organisms is inhibited, or if present, can be distinguished by their colonial appearance. The colonies of the target organism sought are counted and the result, taking into account any dilutions made, is typically expressed as the presumptive number of organisms per 100 ml of sample. The presumptive count may then be confirmed, by sub-culturing all, or a representative number, of colonies formed.

The membrane filtration apparatus consists of a base supporting a porous disc. The filter funnel, which may be graduated, is secured to the base, for example by means of clamps, screw-threads or magnets. The filtration apparatus is connected to a vacuum source. For the examination of large numbers of samples, multiple filtration units may be used. The filtration apparatus should be sterilised on a regular basis between batches of analyses. Spare funnels as required can be disinfected by immersion in boiling distilled water for at least one minute between samples. After disinfection, each funnel should be placed in a stand and allowed to cool before use. Alternatively, a fresh pre-sterilised funnel may be used for each sample. Known polluted and non-polluted samples should be filtered using separate filtration equipment. Alternatively, polluted samples should be processed after non-polluted samples.

Membrane filters, typically 47 mm in diameter, with a nominal pore-size of 0.45 μm retain most of the bacteria commonly enumerated in water. A pore size of 0.2 μm is, however, necessary for the isolation of species of *Campylobacter* and some environmental bacteria (for example *Vibrio* species). The use of membrane filters with grid-marks facilitates counting of colonies. Where there is a need to filter large volumes of sample (for example 500 ml of river water) which may block standard 47 mm diameter membrane filters, then a large volume filtration apparatus may be useful. Membrane filters of 90 mm or 142 mm diameter and appropriate porosity can be housed in stainless steel filtration units and the samples passed through the filter using a suitable pump.

Periodically, it is necessary to check that membrane filters are suitable for the targeted organisms being sought. Quality assurance is important and membranes should be free from toxic substances that inhibit bacterial growth. When membrane filters with grid-marks are used, bacterial growth should not be inhibited or stimulated along the grid-marks. Membrane filters should be pre-sterilised before use and should not be re-used. Membrane filters should not be used beyond their shelf life date.

Absorbent pads, for use with broth media, should be at least the same diameter as the membrane filters and approximately 1 mm in thickness. The pads should be made of high

quality paper fibres, and be uniformly absorbent and free from any toxic substances that inhibit bacterial growth. Absorbent pads need not be sterile if they are of the appropriate quality. This should be verified for each batch of pads prior to use. If necessary, pre-sterilised absorbent pads are available, or pads can be sterilised by autoclaving at 121 °C for 20 minutes, either in containers or wrapped in waterproof paper or metal foil.

8.2.4   Advantages and limitations of the membrane filtration method

An advantage of the membrane filtration technique is the speed with which results can be obtained. For example, presumptive coliform and *Escherichia coli* (*E. coli*) counts and individual colonies for confirmatory testing can be available after 18 hours incubation. In addition, there is considerable saving in labour and in the amount of media and glassware required when compared to traditional MPN techniques. Furthermore, false-positive reactions that may occur with some media in the multiple-tube technique are less likely to occur with membrane filtration.

The membrane filtration technique is unsuitable for use with waters of high turbidity. In these circumstances, the membrane filter may become blocked before sufficient water has been filtered. Also, the accumulated deposit on the membrane filter may inhibit the growth of the organisms being recovered or enumerated. A membrane filtration technique may be unsuitable for use when waters are examined that contain small numbers of targeted organisms in the presence of large numbers of non-targeted bacteria that are also capable of growth on the medium used.

8.3   Statistical considerations

Whilst bacterial counts from a sample can yield valuable information, it is important to assess the reliability of this information. The expression "bacterial count" is used to imply the count of organisms being enumerated, for example coliform count; the same principles also apply to the counting of viruses or other pathogens. In terms of the assessment, there are two main factors to be considered. Firstly, what can be inferred from the sample regarding the level of bacterial contamination in the water source from which the sample was taken; ie the variability or fluctuation in bacterial content of the water? Secondly, how accurate and precise was the actual count for the sample using the method chosen by the laboratory? These two sources of variability are often confused, leading to misunderstandings in the interpretation of sample counts. Generally, in water microbiology, the first sample variability, the bacterial content at the water source, is the more important[8]. This type of variability should be clear to all staff with an interest in the submission of samples for analysis, the interpretation of results and any subsequent actions. The second source of variability, the accuracy and precision of the count, is generally of more concern to laboratory staff.

8.3.1   Estimating bacterial counts within the water source

The examination of a single sample gives an indication of the bacterial count of the sample at a particular location in the catchment area, or supply, at a particular time. The location where a sample is to be collected should be carefully chosen, and thus, a sample should be typical of the sampling area. The actual volume of water sampled may not however possess identical characteristics, with respect to microbiological quality, as those present in adjacent volumes of water. Indeed, only a very small volume of water is examined in the laboratory compared with the volume of source water in question. The confidence interval (CI) for the bacterial density in a body of water cannot, generally, be estimated from a single sample. A mathematical description of the distribution of bacteria is required before a range, such as a

95 % confidence interval, can be estimated. The only situation where a single sample can give such an estimate is when the bacteria are distributed randomly, and in this case, the appropriate mathematical description is the Poisson distribution which has a single statistical parameter, ie the mean, μ, has the same value as the variance. However, there is at present no evidence that bacteria are ever randomly distributed in any part of a water system.

There tends to be enormous variation in the microbiological quality of untreated waters[8]. For treated waters, contamination may be intermittent and organisms may be present as aggregates, often on particulate matter, rather than evenly or randomly distributed. Thus, samples from the same sampling point, even when taken closely adjacent in time, can show large differences in bacterial counts[9]. The statistical parameters describing the distribution of bacteria may change over time and, therefore, a series of single samples collected at different times cannot be used to estimate confidence intervals for the bacterial content of the source of water at any one time. They should be used instead to indicate trends over time.

8.3.2   Estimating the accuracy of counting bacteria from a single sample

Differences between results from the analyses of a single sample can arise for several reasons. For example, from inadequate sampling procedures and from prolonged delays or unsuitable storage conditions that occur between sampling and commencement of analysis. These types of differences will not be considered here, neither will differences caused by accidental contamination. Two possible sources of error that will be considered are those caused by the effects of dilution and the method of counting the organisms. Both these sources are inherent in laboratory methods and are manifest in the second source of error outlined in section 8.3.

8.3.2.1 Potential imprecision due to dilution of the sample

Samples of treated water should not need to be diluted before examination. Samples of heavily contaminated, untreated water may, however, require dilution in order that a suitable count can be enumerated. If the multiple tube technique is used, then some tubes, but not all, should exhibit characteristic growth of the target organism. If the membrane filtration technique is used, then the membrane filter should not be overgrown with target organisms, or indeed, non-target organisms.

It is usual practice to report the bacterial count of targeted organisms, as the number of organisms per 100 ml of sample. With undiluted waters, 100 ml of sample is examined by the membrane filtration technique, and 105 ml by the MPN technique, for example in the 11-tube series of 1 x 50 ml, 5 x 10 ml and 5 x 1 ml. If the sample requires dilution (prior to any additional dilution inherent in the multiple tube method) and this dilution is, for example 10-fold, then only 10 ml (or 10.5 ml) of the original sample will be examined. The count obtained is then multiplied by the appropriate dilution factor, and the calculated count per 100 ml is now an estimate of the number of organisms contained in 100 ml of sample and a confidence interval can be calculated using random distribution theory[10]. Some examples are shown in Table 8.1.

The variability introduced by dilution is likely to be relatively small compared with the variability in bacterial density in a water source, where numbers are sufficiently high to require dilution of the sample before examination. Confidence intervals, as shown in Table 8.1, should not be stated when results are reported, as quoting such intervals may cause misunderstandings and be taken as a statement about the likely bacterial density in the water source. Table 8.1 illustrates the additional imprecision introduced into the bacterial count when dilutions are used and volumes smaller than 100 ml are examined.

Table 8.1   **Estimated count per 100 ml and 95 % confidence intervals for the number of organisms reported in 100 ml of sample, where a sub-sample is examined, following dilution**

Number of organisms found in sub-sample	10-fold dilution		100-fold dilution	
	EC	CI	EC	CI
10	100	50-180	1000	480-1830
50	500	380-650	5000	3750-6640
100	1000	820-1200	10000	8190-12200

EC = estimated count
CI = 95 % confidence interval

8.3.2.2 Other imprecision of laboratory methods

Samples of treated waters should not contain indicator organisms, and the presence of very small numbers of such organisms in samples of water are capable of being detected, and enumerated with good precision, by methods described in this series, see page 2. Untreated waters, however, may yield moderate or high bacterial counts, and in these situations, the accuracy with which the count is made should be considered.

The overall bias and precision of a method should be established by primary validation[11] and comparison of results obtained using a reference method. A protocol for undertaking such a comparison, together with examples, is described in section 9.

8.3.2.3 Multiple tube (or MPN) method

In the multiple tube method, a series of sub-samples is taken from the original sample, and processed to ascertain which of the sub-samples show the presence of the targeted organism. A mathematical formula, based on laws of probability, is then used to estimate the MPN of organisms present in the sample[12, 13, 14]. These various mathematical approaches and the principles involved in the estimation of bacterial densities by dilution methods have been reviewed[14, 15] and tables have been developed[16, 17] which give greater detail. However, in practice, the full extent of the tables are rarely used[18, 19].

A confidence interval was often published with the MPN that demonstrated the variability of the count, not the variability of the water source, although it has often been confused with the latter. For this reason, confidence intervals of this sort should not be reported.

Modern computers now enable the determination of the probability of counts associated with each dilution series to be quantified exactly[20, 21]. While the latest calculation of the MPN shows little discrepancy with previously published values, these new calculations have highlighted two issues. Firstly, the variability of previously published confidence intervals. Secondly, for moderate or high bacterial density, the multiple tube method does not give a clear MPN. There is a "most probable range" (MPR) of counts, all of which are almost equally likely to be as correct as the MPN. All calculations are based on the assumption that the organisms present in the water are evenly or randomly distributed and the importance of thorough mixing of the sample cannot be over-emphasised. Although the multiple tube method is very sensitive for the detection of small numbers of indicator organisms, the MPN is not an exact value unless very large numbers of tubes are examined, as is the case with recently developed multi-well MPN techniques. Apparent differences between results should, therefore, be interpreted with caution. It should be appreciated that variations in bacterial

numbers in the water source may be very much greater than any variability introduced by the multiple tube method.

8.3.2.4 Membrane filtration method

If a sample of water is filtered and the membrane filter incubated, and then every relevant colony on the membrane filter is counted, and every colony is tested and confirmed, then the presumptive and confirmed counts are as precise as the method allows. No further statistical imprecision need be considered. If the sample is diluted prior to filtration then the count becomes an estimate of the bacterial density in the undiluted sample, as already described. If the presumptive count is obtained from only a segment of the membrane filter then variability is introduced comparable to that introduced by diluting the sample, assuming that the segment is typical of the whole filter.

8.3.3   Confirmation of isolated organisms

Confirmatory tests of the presumptive colonies present on a membrane filter should be carried out. When multiple colonies are present, different approaches can be adopted when consideration is given to the number of colonies that should be tested for confirmation. If the aim is to estimate the count of relevant colonies, then consideration should be given to the variability that is introduced when only a fraction of the total number of colonies present is tested for confirmation. The colonies should be chosen at random and the number tested should provide a sufficient level of accuracy. This may require sub-culture of all the colonies on a membrane filter when fewer than ten presumptive colonies are present. However, this may not be practicable and may not be necessary, especially in the case of highly specific methods where all the colonies are expected to confirm as positive.

Alternatively, if the aim is to demonstrate the presence or absence of the targeted organism, then a different approach may be chosen, provided there are no microbiological contra-indications. The presence of the organism is demonstrated as soon as one colony is tested and a positive confirmation is made. Hence, a laboratory may choose to examine fewer colonies, initially, than when the aim is to estimate the count, rather than demonstrate presence or absence. However, if the colonies that are chosen and tested for confirmation, and positive confirmations are not obtained, then the sample cannot be assumed, at this stage, to be free of confirmed organisms. This is because other colonies on the filter, which have not been chosen for confirmatory testing, may, if tested, prove positive. Hence, other colonies from the membrane filter should be tested until at least one positive confirmation is obtained, or all colonies have been tested and no confirmation has been shown. This sequential testing is acceptable only when refrigerated storage of the membrane filter is not detrimental to the survival and/or recognition of the relevant organism.

If all presumptive colonies are tested to confirm their nature, then no further imprecision (other than that due to the method) is introduced when the presumptive count is converted into a confirmed count. If only some of the colonies are tested by confirmatory methods, then further imprecision is introduced into the confirmed count. For example, if a presumptive count is made by counting all the typical colonies, N, on a filter then it is common practice to make confirmatory tests on some, but not all, of these colonies, unless N is small. If n is the number of colonies tested, and x is the number of colonies that are confirmed as the target organism, then the confirmed colony count is estimated as xN/n. For example, if 60 colonies were observed on the filter, and 20 colonies were selected at random for testing, and 15 of these colonies were confirmed, then the estimated confirmed count would be $15 \times 60 / 20 = 45$.

It is assumed that the "n" colonies are selected at random, or by some other procedure which ensures they represent a typical sub-sample of the "N" colonies. It is further assumed that all the "N" colonies are equally likely to be from the relevant organism group. The conditional probability that y is the true count, given that x colonies have confirmed can be calculated from:

$$P(x \mid y) = {}^{y}C_x \cdot {}^{N-y}C_{n-x} / {}^{N}C_n$$

The 95 % CI for the confirmed count can be found by observation of the probabilities for all possible values of y, using the observed value of x. The CI will exclude "end of range" high and low values of y, such that their cumulative conditional probabilities sum to less than or equal to 0.05[10].

As an example, if 10 colonies are observed (N) and only two colonies are tested (ie n = 2) then Table 8.2 shows the complete range of probabilities.

Table 8.2   Range of probabilities, if 10 colonies are counted and only two colonies are tested for confirmation

Given that y =	the probability that		
	x = 0 is	x = 1 is	x = 2 is
0	1.000 }		
1	0.800 }	0.200 }	
2	0.622 }	0.356 }	0.022
3	0.467 }	0.467 }	0.067 }
4	0.333 }	0.533 }	0.133 }
5	0.222 }	0.556 }	0.222 }
6	0.133 }	0.533 }	0.333 }
7	0.067 }	0.467 }	0.467 }
8	0.022	0.356 }	0.622 }
9		0.200 }	0.800 }
10			1.000 }

The brackets (}) show the range of y values which should be included in the 95 % CI for the true count. It should be noted that if one of the two colonies tested confirms (ie x = 1) then all the possible values of y are within the 95 % CI because all the probabilities exceed 0.05.

In general, if only a small number of the total colonies are tested, then the CI can be very wide. They also tend to be wider if a substantial number of colonies turn out not to confirm. This may not be of concern in cases where a positive result may mean a breach of any regulations. Some examples are shown in Table 8.3.

Table 8.3   Variation in the 95 % CI with variation in the proportion of tested colonies confirmed

Colonies observed (presumptive count) ie N	Number tested, ie n	Number confirmed, ie x	Confirmed count	95 % CI
10	2	0	0	0-7
10	2	1	5	1-9
10	2	2	10	3-10
14	7	5	10	6-12
50	10	5	25	9-41

With treated waters, where the vast majority will yield zero or very few presumptive colonies, then to improve accuracy, as many colonies as possible should be tested by confirmatory methods. For untreated waters, it may be worthwhile considering the use of the presumptive count rather than introduce the additional variation which accompanies confirmation of some but not all of the colonies. It should be noted that the practice of confirming a maximum of 10 colonies can still introduce potential significant variation, especially if the presumptive count is large and some colonies fail to confirm. However, a balance should be made between the benefits of improved accuracy and the capacity of the laboratory to undertake confirmatory tests for large numbers of colonies.

8.3.4   Comparing results with prescribed limits

Typically, prescribed microbiological limits for drinking waters concentrate on the presence or absence of indicator organisms and pathogens. Therefore, the potential problem of how to compare actual counts enumerated and estimated counts calculated need not be addressed with respect to potable waters.

8.3.5   Reporting results

The report should be a clear statement of the findings. A further statement on sample error, to qualify these findings, should not be necessary for routine samples. The sampling strategy should be designed with the aim of acquiring an adequate level of information. If it is necessary that a report for a special or unusual sample warrants a statement on accuracy and precision, then a clear distinction should be made between the variability of the water source and the variability inherent to the laboratory method.

Absence of organisms or immeasurably high counts should be reported according to the following criteria:

No organisms detected. A water sample in which no relevant organisms are detected should be reported as "none found in the volume of sample examined". It should be noted that in microbiological terms there is no equivalent to the chemical concept of "limit of detection". An expression such as "less than 1 per unit volume" has no meaning.

Overgrowth of membrane filter or all multiple tubes positive. This means that the analysis has failed to estimate the true count either because of insufficient pre-dilution or the presence of high numbers of non-target organisms. With the multiple tube method it is customary to report this, in the appropriate units, as "> 180" for the 11 tube series or "> 1800" for the 15 tube series, but recognising that the count could be very much higher. With membrane filtration and other methods the report should be "count too high to be estimated at the dilution used".

8.4   Internal quality control

To ensure that a laboratory is capable of isolating, accurately identifying and enumerating micro-organisms in a sample, and to avoid contaminating samples with extraneous micro-organisms, it is necessary to undertake a system of internal quality control. This consists of submitting quality control samples to the usual isolation, enumeration, identification and confirmatory procedures normally used with real samples. Quality control samples should contain micro-organisms similar to those normally sought, and where appropriate non-target organisms, as well as samples that are sterile. If the procedures function satisfactorily, such micro-organisms will be detected, or in the case of sterile samples, no micro-organisms will

be found. The control procedures should be undertaken in parallel with each batch of samples examined. This may necessitate several positive and negative control samples, and blank samples being set up each day, with separate quantitative testing schemes to check enumeration. See also section 8.5.

Positive control samples contain target organisms that produce typical colonies or positive reactions on isolation media and in confirmation tests. Negative control samples contain non-target organisms that do not produce colonies or positive reactions, or produce atypical colonies, on isolation media and in confirmation tests. Blank control samples are usually sterile samples used to test the integrity of the analytical procedure.

Control organisms should, wherever possible, be derived from freeze-dried, wild-type reference strains, obtainable from reference sources, and reconstituted and diluted with quarter strength Ringer's solution, or similar appropriate diluent, to contain suitable numbers of organisms. However, care should be taken in the selection of these organisms as some have been shown to give atypical results when compared to genuine wild-type organisms. It is good practice to minimise the number of sequential sub-culture operations as the biochemical characteristics of some strains may change on repeated culturing. The use of natural waters, known to contain relevant organisms, may also be suitable as analytical control samples. All confirmatory tests should include positive, negative and blank control samples.

8.5 Quantitative internal quality control

In addition to qualitative checks with positive, negative and blank control samples there should be checks on the enumeration procedures[22]. Two approaches can be considered and involve the use of appropriate reference materials and the use of split sub-samples from a source known to contain the target organism.

If automatic counting instruments are used these should be tested and calibrated against known values.

Quality control or Shewhart charts are used extensively in the water industry for demonstrating statistical control of laboratory chemical procedures. This practice can be extended for demonstrating microbiological control. However, the natural random variation in the number of organisms present in sub-samples of the same sample means that there can be a wide scatter of results between sub-sample analyses, which is to be expected. Many more samples are required for microbiological examination compared with chemical analyses in order to detect real "out of control" situations. Even then, these situations would better be described or classified as probably out of control rather than definitely out of control. Hence, for microbiological purposes, "guidance charts" are used with appropriate response lines, rather than action or warning lines that trigger remedial action or further investigation.

8.5.1 Reference materials

Guidance charts can be plotted using regular counts enumerated on samples taken from a batch of suitable reference material that may be commercially obtained or internally prepared. The usual practice is to plot the results sequentially over a period of time.

If the reference material does not possess a statement of certified mean and variance values, then these values should be estimated from a suitable number of replicated analyses. These analyses should be carried out under conditions that ensure the values are "in control" or assumed to be "in control". The chart is plotted using the values determined or constructed

from these data using transformed data, for example, square root or log count data. Response lines are then drawn on the chart at appropriate intervals. Appropriate response lines may be located at ± 2 standard deviations of the mean (equivalent to upper and lower "warning" limits) and at ± 3 standard deviations of the mean (equivalent to upper and lower "action" limits). Regular samples of the reference material are then processed with routine samples and the counts plotted sequentially. Documented action should follow if values are recorded that fall outside the range of the response lines. The following guidance is often used as a basis for action[22].

(i)   One count falls outside an action limit:
(ii)   Two out of three successive counts exceed the same warning limit:
(iii)   Nine consecutive counts fall on the same side of the mean:
(iv)   Six consecutive counts show a trend that continuously rises or falls.

It should be noted that the original estimates of the mean and variance values may not be totally reliable and may need to be further studied, especially if action is triggered repeatedly because response limits are exceeded and remedial action does not identify appropriate causes. In addition, the quality of the reference material may need to be questioned.

If possible, the counts for the reference material should be enumerated without prior knowledge of the mean and variance values. A guidance chart that does not exhibit some degree of variation in counts (in line with random variation) may be indicative of operator bias.

8.5.2   Split samples

Quality control checks for consistency in enumeration can also be made using split samples. Split samples comprise a sample divided into 2 sub-samples, each of which is analysed with each batch of routine samples. The use of split samples should involve samples that are known to contain target organisms. The duplicate sub-samples can be considered as two halves of a single sample, and the results can be plotted on a chart containing appropriate response limits.

Because of the manner in which micro-organisms are distributed in water, the examination of split samples can result in significant variation in the counts enumerated. For example, if the count reported for the first sub-sample is 5, then the 95 % CI for the count of the second sub-sample will be 0 - 14. The CIs for the count of the second sub-sample, given the count observed in the first sub-sample are given in Annex A. Thus, it may be expected that duplicate sub-samples will give counts outside of the 95 % CI, on 5 % of occasions, ie once in every 20 samples. Procedures should be adopted within the laboratory to deal with situations that occur too frequently (ie greater than 5 % of occasions) where sub-samples give counts outside of the 95 % CI. The count for the first sub-sample should be recorded on a control chart, together with the corresponding CI for the count of the second sub-sample (obtained from table A1 in annex A). The count of the second sub-sample is then recorded alongside these figures. If this count falls outside the range of the CI, then this fact should be recorded. If, over a period of time, the count of the second sub-sample falls outside the range of the CI on more than 5 % of occasions, then investigations should be carried out to determine the cause[23].

Alternatively, a more approximate statistical approach can be used with paired counts using the Index of Dispersion chi-squared test[15, 24]. For paired split samples, the formula for calculating the Index of Dispersion, D, is:

$$D^2 = (x_1 - x_2)^2 / (x_1 + x_2)$$

To construct a guidance chart, the median is plotted, as are values of the 99% and 95% confidence level limits, ie for p = 0.05 and p = 0.01 (ie 3.841 and 6.635 respectively, each with 1 degree of freedom). These values are approximately equivalent to 2 and 3 standard deviations, and act as appropriate "response" limits. The calculated values of D^2 obtained for split samples should be equally distributed on either side of the median line.

Laboratories using split sample internal quality control should carry out analyses regularly, and plot the results on guidance charts. Each sub-sample should be treated as separate samples and analysed in the normal, routine manner. The sub-samples should be randomly positioned in the incubator, and these positions should be changed frequently when different batches of samples are examined. If possible, counts should be enumerated in such a manner so as to ensure that the sub-samples are not recognised as being connected. If the variation between the counts of the sub-samples is significantly less than would be expected, then operator bias may be suspected.

8.6   External quality control

Laboratories should participate in an appropriate inter-laboratory external quality control scheme that involves the examination of samples distributed by an external organisation. The laboratory's results can then be compared with those obtained by other participating laboratories and provide an independent assessment on the quality of the laboratory's performance. It is essential that the samples distributed by the scheme organiser are treated and analysed in exactly the same way as routine samples, and that appropriate action is taken when results fall outside of the expected range.

It should be noted that the purpose of external quality control samples is to assist individual laboratories assess their own capabilities to undertake selected analyses and to correct any deficiencies which may be present. They should not used for the purpose of determining whether one laboratory performs better or worst than another participating laboratory.

9   Comparison of methods

9.1   Introduction

Methods for drinking water bacteriology should be capable of serving their intended purpose, ie to detect and/or quantify target organisms or groups of target organisms with adequate precision and accuracy. In addition, the validation of microbiological methods is important and a recent report[11] describes various procedures for carrying this out.

In certain countries methods are prescribed, in other countries, they are not. If alternative methods are used in place of statutory, regulatory or laboratory accredited methods, they should be of "equivalent or better" performance. Methods, capable of achieving a certain performance are published by a variety of sources, including those shown in the Table 9.1.

Table 9.1   Sources of methods

International Organisation for Standardisation (ISO)
European Committee for Standardisation (CEN)
Individual national standards organisations (for example BSi, DIN, AFNOR)

Standing Committee of Analysts (SCA)
The American Association of Official Analytical Chemists
American Public Health Association

The demonstration that new methods are at least as accurate and precise as reference methods is, however, a complex procedure. This section describes a protocol for comparing the recoveries of confirmed target organisms by two or more methods. The protocol and procedures described have been developed from those derived for the Drinking Water Inspectorate[25].

9.2   Scope of procedures

This section describes the procedures for establishing the equivalence of microbiological cultural methods used in drinking water bacteriology, for example the replacement of a reference or standard method by an alternative new method. Instructions, including the preparation of spiked samples and the recommended number of measurements, are described to evaluate whether a new method as a replacement for a reference method could be adopted for routine use in the laboratory. The new method should, before evaluation, be thoroughly validated. The comparisons involve two stages. Stage 1 compares the results of 2 methods using samples containing about 20 - 50 target organisms per test volume, usually 100 ml. Stage 2 compares the results of the same methods using samples with lower levels of target organisms, usually in the range of about 1-10 target organisms per test volume. Only paired samples with at least one positive result are considered, as paired samples with zero counts do not provide additional information on the comparative recovery of target organisms.

9.2.1   Statistical approach and acceptance criteria

The comparison of a new method with a reference method should be undertaken with an appropriate diversity of target and competing non-target organisms (obtained from a variety of sources) relevant to the test methods. The preparation of suitable samples (see section 9.6.2) is very important and the waters used should be derived from several sources. Each source is referred to as a "category of origin" and samples of water are taken over different periods of time. In stage 1, samples should be used which produce enumerated counts within the optimum ranges of both methods. These counts should yield sufficient numbers of organisms to provide a meaningful statistical comparison. For example, with a membrane filtration method, a suitable range of 20 - 50 target organisms per unit volume (typically, 100 ml) is estimated to be sufficient. If the method enumerates more than one target group of organism (for example *E. coli* and coliform bacteria) then separate tests may be necessary to ensure that each target group is enumerated in the range. In addition, guidance is given on the number of samples needed to provide sufficient statistical information to reach a positive conclusion.

These criteria, described in section 9.6, require a clear presentation of the data, a statistical comparison within each category of origin of samples, and finally, an overall statistical comparison. The new method should be rejected if it is shown that significantly lower average counts are obtained than those obtained using the reference method. The new method may be accepted, subject to the conclusions reached in stage 2 of the trail, if it is shown to be better than the reference method or the following criteria are satisfied:

(i)   it is demonstrated that there is no statistical, significant difference between the methods; and

(ii)   the 95 % confidence interval for the average difference lies above the value which would indicate that the new method was finding 10 % fewer organisms than that found by the reference method.

A new method, which is found to be acceptable after stage 1 has been completed, should then be tested against the reference method as described in stage 2. This should then demonstrate that there are no major differences between the two methods when much lower counts are compared.

9.3   Basic concepts and definitions

A laboratory considering an alternative new method to one currently in use should obtain sufficient comparative data to demonstrate the equivalence of the two methods before adopting the new method for routine use. If appropriate, other laboratories should then undertake the process of full data comparison and the data from all laboratories may then be pooled and reviewed to establish robustness.

9.3.1   Microbiological cultural methods

Methods are considered microbiological cultural methods when growth and multiplication of micro-organisms are essential features for their detection and/or quantification.

9.3.2   Definitions

Confirmed counts - The number of presumptive counts multiplied by the proportion confirmed, that conform to the definition of the target organism.

Error - The statistical variation including natural variation and imprecision of the method.

Over-dispersion - The variation in excess of that shown by the Poisson distribution.

Poisson distribution - Fully random distribution of particle numbers when sampling a perfectly mixed suspension, exhibiting no attraction or repulsion between micro-organisms.

Precision - The closeness of agreement between independent test results obtained under stipulated conditions.

Presumptive counts - The number of organisms that produce a response typical of the target organism in or on a primary detection medium.

Primary validation - Establishment of the specifications for the performance of a new method and/or experimental verification that a method meets theoretically derived quality criteria.

Repeatability - Closeness of the agreement between the results of successive measurements of the same measure carried out under repeatability conditions of measurement. For example, this can be calculated from replicate counts from sub-samples obtained from a well-mixed sample, analysed by one analyst using the same reagents and method.

Reproducibility - Closeness of the agreement between the results of measurements on the same measure carried out under reproducibility conditions of measurement. For example, this can be calculated from replicate counts from sub-samples obtained from a well-mixed sample,

analysed by more than one analyst or laboratory using different reagents, but the same method.

Robustness - The insensitivity of an analytical method to small changes in procedure.

Validation - The procedures undertaken which establish the performance of a new method.

Verification or secondary validation - Demonstration by experiment that an established method functions according to its specifications in the user's hands.

9.4   Sources of variation and error

As already described in section 8.3, there are several sources of variation that may complicate the evaluation of comparing alternative methods for drinking water bacteriology. These include sample variation, natural variation and systematic imprecision inherent in the methods.
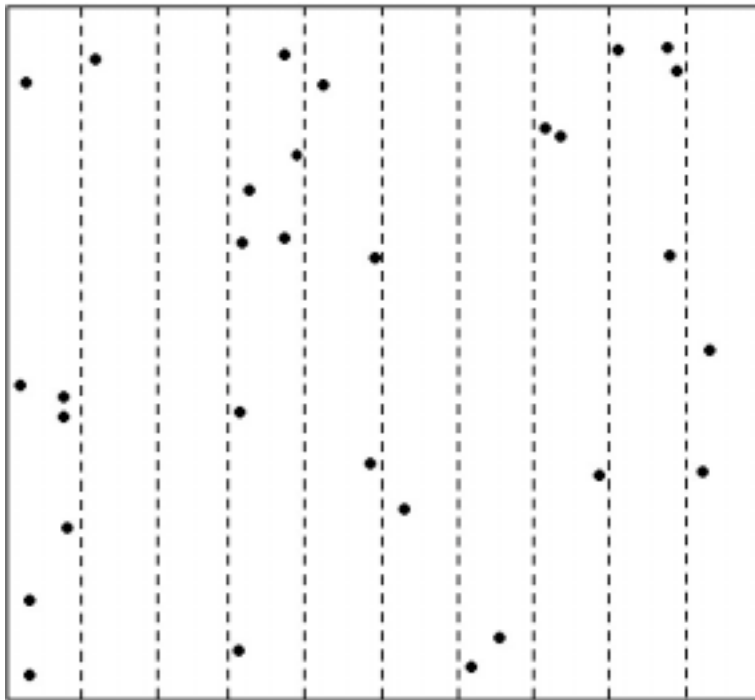
9.4.1   Sample variation

A water source, sampled for monitoring purposes, may exhibit enormous variation in its microbial content over time and between sampling sites[9]. Samples, used in comparative trials of alternative methods, should, therefore, not be collected or prepared separately. A paired or split sample approach (see also section 8.5.2) should be used. A suitable sample should be thoroughly mixed, and two aliquots of this sample taken for analysis. The analysis of each aliquot should then be carried out at the same time, the first aliquot being analysed by one method and the second aliquot being analysed by the other method. Over time and on average, the theoretically expected number of organisms in both aliquots should be the same.

9.4.2   Natural (random) variation

Figure 9.1 illustrates the volume of a thoroughly mixed sample of water containing 30 micro-organisms that are randomly distributed. For each of the ten identically marked aliquots, it is important to note that the number of organisms present in each aliquot may not be the same and that these numbers may differ purely by chance. Overall, the average number of organisms is 3 per unit aliquot. However, as depicted, the range is shown to be 0 – 7. This type of variation within a sample will always occur in water microbiology. In addition, over-dispersion may occur, as a result of the attraction or repulsion between organisms and suspended matter, laboratory equipment or other organisms.

To accommodate this natural variation, many samples need to be analysed to evaluate the systematic variations that may exist when different methods are compared. Sufficient data should be generated to average out the effects caused by the natural variation depicted in Figure 9.1. An example of this natural variation is illustrated in Figure 9.2 which shows the results of 50 paired water samples examined for the same organism using the same method[23]. As shown, the results are scattered and the correlation between the pairs of counts appears low. The correlation coefficient or product-moment statistic, r^2 is calculated as 0.39, even though it might be expected that a value of 1 should be generated under theoretical or ideal conditions. This illustrates that the use of this statistic, r^2, is not appropriate in these cases. Hence, the correlation between paired counts needs to be assessed and interpreted against this background of inevitable variability.

Figure 9.1   Random variation of organisms in aliquots from a sample containing 30 organisms

9.4.3   Other sources of variation

Other factors can affect either the number of organisms present or the numbers detected and reported. These factors include inadequate mixing of samples and inaccurate measurement of aliquot volumes. Also, errors in the number of organisms reported can be introduced by equipment, analysts or laboratory procedures, as well as by the methods used. A small amount of random variability is expected from every procedure and this can be acceptable. However, excessive random variability might indicate an imprecise method and this should become apparent during the characterisation of a new method (see section 9.5). Non-random or systematic variation, for example due to the inadequacy of the method, should be highlighted during method validation when a new method is being evaluated. Any investigation, therefore, needs to separate or distinguish the variation caused by or inherent to the methods used and that resulting from natural or random variability.

9.4.4   Statistical detection of other sources of variation

Method comparison studies have been designed and analysed to detect whether other sources of variation are present, and whether they are microbiologically and/or statistically significant[26]. The sources of variation in the enumeration of the relevant organism become apparent when the components of that count are studied. For example,

$$y_i = \mu + \varepsilon_i \tag{1}$$

where: i is 1 or 2, representing the first or second aliquot in the paired sample;
              y_i is the number of organisms enumerated;
              μ is the mean value for the sample; and
              ε_i is the random error.

Figure 9.2   Pairs of replicate counts of coliform organisms
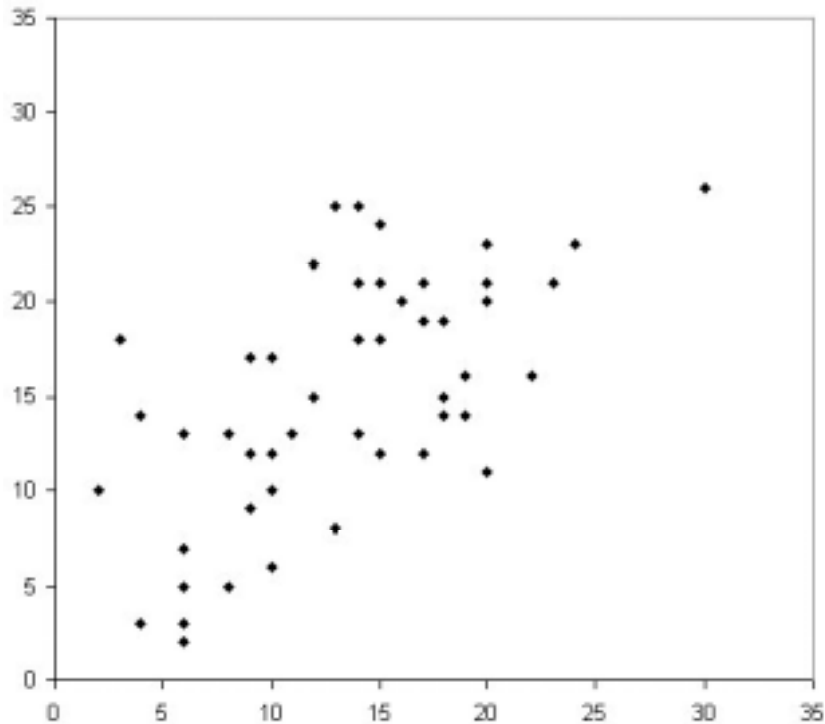
The y-axis represents counts on the first sub-sample
The x-axis represents counts on the second sub-sample

Equation (1) can be expanded to:

$$y_i = \mu_t + m_l + m_m + \varepsilon_{ti} + \varepsilon_{li}$$

where: μ_t is the true mean value of organisms present in the whole sample;
                    m_l is the laboratory effect (independent of the two methods);
                    m_m is the method effect (m_{ref} = reference method and m_{new} = new (trial) method);
                    ε_{ti} is the random or natural error between aliquots;
                    ε_{li} is the random measurement error in the laboratory.

The laboratory effect plus the method effect (ie $m_l + m_m$) is the systematic, average difference from the true mean when that method is used. It represents the bias and is inversely proportional to the "trueness" of the measurement.

Random variation reflects precision and hence, the difference between the paired counts is:

$$y_1 - y_2 = (\mu_t + m_l + m_{ref} + \varepsilon_{t1} + \varepsilon_{l1}) - (\mu_t + m_l + m_{new} + \varepsilon_{t2} + \varepsilon_{l2})$$

$$= (m_{ref} - m_{new}) + (\varepsilon_1 + \varepsilon_2)$$

If sufficient samples are examined then the random errors should average to zero. Thus, the expected value of $y_1 - y_2$ may be represented as:

$$E(y_1 - y_2) = m_{ref} - m_{new}$$

37

Any interaction between method and laboratory will be included in this expression but does not affect the conclusions about the effectiveness of the new (trial) method in a particular laboratory undertaking the trial. Because the absolute errors may be large (due to the natural random variation) the precision will be low and a large amount of data will be required for a powerful statistical estimate of ($m_{ref} - m_{new}$).

9.4.5   Limitation of errors

Errors should be minimised or eliminated by implementing a quality assurance programme that includes the use of internal quality control samples and participation in an appropriate external inter-laboratory quality assessment scheme.

9.4.6   A "confidence level" approach

When two methods are compared, it may not be statistically acceptable to analyse a pre-determined number of samples for each stage of the comparison tests. Whilst it may be possible to theorise about the number of samples that might give a 95 % probability of detecting a difference, the distributions of counts in the samples examined might be different from those predicted. A "confidence level" approach is, therefore, used to provide the statistical power necessary for the evaluation of a new validated method to be undertaken before it is accepted for routine use within a laboratory.

The approach for comparing method A (for example, a reference method) with method B (for example, a newly proposed trial method) is made on the basis of recording the difference in results obtained for paired sub-samples of a sample. This data set, when complete, is then progressively evaluated to ascertain whether the average results are comparable and the confidence intervals are acceptable.

9.5   Prerequisites for method comparisons

9.5.1   Initial characterisation of a new method

During method development of a new, modified or previously inadequately characterised method, it is paramount that certain optimum conditions should be established. Primary validation should establish the operational limits and performance characteristics. The validation should provide numerical and descriptive specifications for the performance and include a detailed and unambiguous description of the target organisms. An example of a validation study[11] is shown in Figure 9.3.

Primary validation of the method should establish the specificity, selectivity, relative recovery and other characteristics[11]. The method should be tested with samples containing low numbers of target organisms. Additionally, primary validation should establish that the repeatability and reproducibility of the method are acceptable.

Descriptions of methods for the bacteriological examination of water should provide details of performance characteristics and their specifications including the scope of the method, incubation robustness and time sensitivity, reliable working limits, and target definition and identification. Additionally, advice should be provided on, and requirements stated for, quality control of media and equipment. Protocols should provide laboratories with structured procedures to assist the application of the method and, therefore, the capability to generate valid results. For example, details should include statements on sensitivity, selectivity, counting uncertainty, replicate analysis, within-sample variation, and proportionality.

9.5.2   Common identification of target and non-target organisms

Microbiological methods should be designed to detect and/or enumerate particular types of micro-organisms, ie target organisms. All other micro-organisms, ie non-target organisms, that may be present in the sample should be not detected, or if they are, should be readily differentiated, and therefore, should not interfere with the detection or enumeration of the target organisms. Non-target organisms are often described as competitive or background flora. The definition of target organisms should reflect current microbiological understanding, and be sufficient to ensure common differentiation between target and non-target organisms when two different methods are being compared.

9.5.3   False-positive and false-negative results

If a non-target organism is mistakenly identified as a target organism, a false-positive result is obtained. Alternatively, a false-negative result is obtained when a target organism is not correctly identified. Note that a false-positive result or a false-negative result may be reported for individual colonies, as well as for the overall final result of a sample. The nature and concentration of target organisms and non-target organisms often vary considerably between samples taken from a specific location, and especially from those taken from different locations. A consequence of this is that a method that has been evaluated for a particular type of sample may not necessarily have universal applicability. Methods have been described in international standards, or prescribed as legal requirements, as a means of achieving a standardised approach to analysis. If these methods are clearly and unambiguously described then inter-method differences are eliminated. This does not mean that an international standard or prescribed method is suitable for all situations and samples. The laboratory should be responsible for evaluating the performance of a method, especially when different types of samples are analysed by the same method. The temporal variation of the performance of a method, in relation to variable characteristics of the micro-flora, should be evaluated as part of any quality assurance programme.

9.5.4   Quality assurance

It is essential that any laboratory that carries out a comparison of microbiological methods according to this guidance should adopt appropriate quality assurance systems. Accreditation, whilst beneficial to the laboratory in terms of documentation, does not guarantee a laboratory's performance. The sources of "error" or variability (as defined in 9.3.2) that affect microbiological methods include;

(i)   sample error - the variation in enumerated counts as a result of taking of sub-samples;
(ii)   natural error - the variation due to the Poisson distribution of organisms in a liquid matrix;
(iii)   systematic error - the imprecision inherent in the method used; and
(iv)   random laboratory error - the imprecision that is particular to a laboratory or analyst.

In an attempt to minimise the effects of systematic and random "errors" it is essential that laboratories make use of appropriate reference materials and take part in inter-laboratory, external quality assessment or proficiency testing schemes. In addition, attention should be paid to media, incubators and membrane filters.

Figure 9.3   Example of a validation experiment

Plate count and spread plate (repeated for different samples)

Membrane filtration (repeated for different samples)
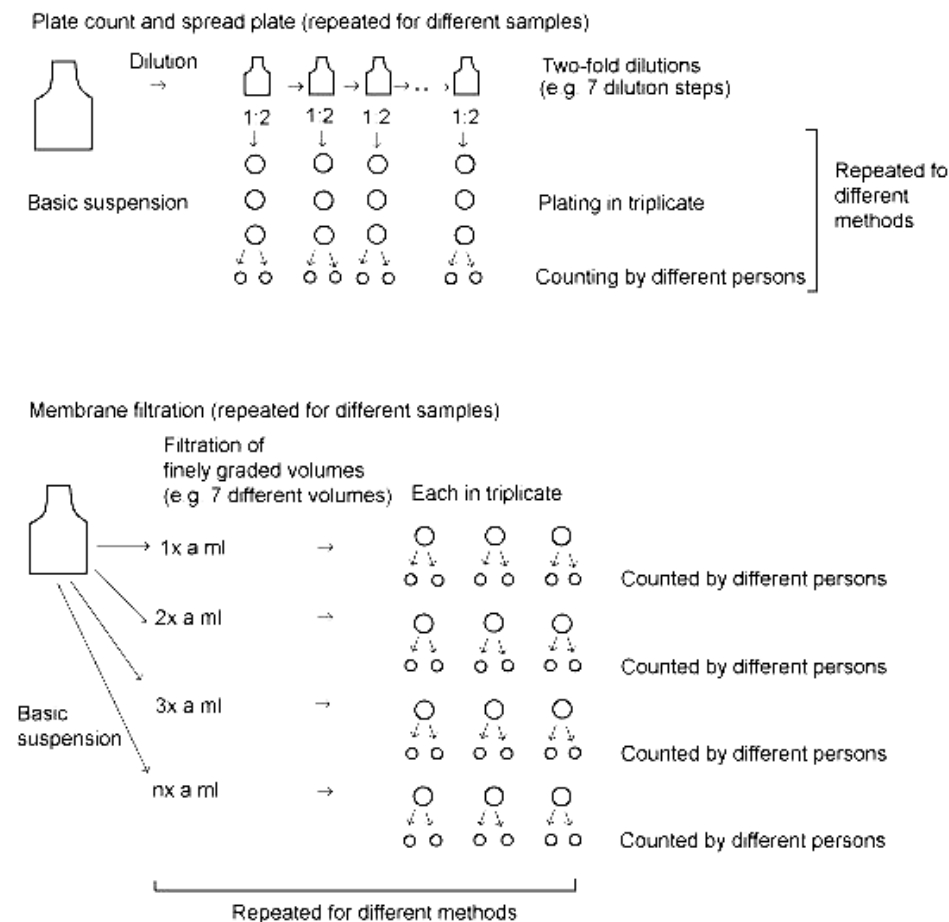
Note.
Any data on replicate counting or parallel plates available can be used for study of specific
details of validation. Sensitivity and selectivity should be proven before use of this design.

9.5.4.1 Media

It has been shown that small changes in the composition of the medium used in a particular
method can affect the method performance, in particular its ability to recover target organisms
from certified reference materials[27]. There may also be batch to batch variation in the quality
of the medium used. Ideally, the comparison of microbiological methods should use a single
batch of medium that has been performance tested using reference materials. Media should be
prepared correctly, paying attention to the conductivity of the water, the pH of the medium
(before and after sterilisation) and ensuring that the time the medium is exposed to high
temperatures is minimised.

9.5.4.2 Incubators

The temperatures of incubation used for water microbiology are, usually, critical. The
temperature in an incubator can vary between individual shelves and be affected by loading
patterns. Incubators should preferably be fan-assisted and temperatures should be controlled
within stringent limits. During incubation, temperatures should be recorded regularly,
preferably by continuous monitoring.

9.5.4.3 Membrane filters

The quality of membrane filters differs between different manufacturers and also, between different batches prepared by the same manufacturer. This can affect the recovery of target organisms. The batch number of the membrane filters used should, therefore, be recorded, and tests using reference materials should be carried out on each new batch of membrane filters to ensure they are satisfactory. Membrane filters from a single batch should be used when undertaking comparisons of microbiological methods.

9.6   Stage 1 of the comparison of method A versus method B

The following procedures apply to the comparison of enumeration methods. The comparison of two microbiological methods involves processing (at the same time) aliquots of the same sample by the two methods under study. A number of different samples from each source or category of origin and from different sources or categories of origin are similarly analysed and then a statistical analysis of all the results is carried out.

The methods to be compared should be tested with the types of samples, which it is anticipated, will be routinely analysed by the methods. These samples, generally, will comprise waters that have been subject to some form of treatment, usually including disinfection. Because of the high quality of most treated water supplies it will, generally, be necessary to prepare samples that mimic the effect of inadequate treatment. Protocols for the preparation of suitable samples, containing chlorine as disinfectant, are given in section 9.6.2. For alternative disinfectants, it will be necessary to determine by experiment those conditions appropriate for the survival of suitable numbers of target organisms.

It is estimated that a minimum of 150 samples should be included in the comparison trial, which reflect the range of source waters. The methods used should be tested with the appropriate volume of sample relevant to the target organism and the prescribed limit. This is usually 100 ml and this volume is used in this document for illustrative purposes. The samples should not be diluted and should be tested over a period of several of days, generally, testing approximately 10 - 15 samples per day.

9.6.1   Preparation of samples

There are a several ways of preparing suitable samples (based on chlorinated waters) for carrying out comparisons of microbiological methods and these are listed in order of preference.

(i)   Chlorinated tap water plus river water with the addition of extra quantities of chlorine to produce chlorine-stressed organisms, to a final concentration of chlorine of approximately 0.1 - 0.5 mg/l (see section 9.6.2.1).

(ii)   Through treatment samples (for example, following granulated activated carbon or post rapid gravity filter treatment) if necessary, with a final concentration of chlorine of approximately 0.1 mg/l.

(iii)   Chlorinated tap water plus sewage effluent with the addition of extra quantities of chlorine to produce chlorine-stressed organisms, to a final concentration of chlorine of approximately 1.2 - 2.5 mg/l (see section 9.6.2.2).

(iv)   Naturally contaminated un-chlorinated groundwater with the addition of extra quantities of chlorine to produce chlorine-stressed organisms, to a final concentration of chlorine of approximately 0.1 mg/l.

In certain situations, it may be necessary to use environmentally stressed organisms instead of chlorine-stressed organisms. In these cases, suitable samples may be prepared by prolonged storage of sewage effluent or river water samples.

9.6.2   Preparation of spiked samples

For the initial phase of stage 1, spiked samples are prepared which contain chlorine-stressed target organisms, non-target organisms and organisms closely related to target organisms. Ideally, samples should contain 20 - 50 target organisms per test aliquot (for example, 100 ml).

9.6.2.1 Generation of chlorine-stressed target organisms using river water

9.6.2.1.1   Collect approximately 10 litres of tap water from a supply that is representative of the water supplies to be tested, and cool to 2 - 8 °C (store overnight if necessary). Collect at least 1000 ml of river water. If the tap water being used is derived from surface water, then the water source from which the tap water is derived should be used.

9.6.2.1.2   Remove a small quantity of the cooled tap water and determine the concentration of free and total chlorine in a suitable aliquot. This determination is used to calculate the amount of chlorine that should be added to the remaining volume of tap water, to produce a free chlorine concentration of approximately 0.1 - 0.5 mg/l. The calculated amount of chlorine can be added using a solution prepared from sodium hypochlorite or chlorine-generating tablets. The chlorinated tap water should be stoppered or capped and thoroughly mixed. Store the chlorinated tap water at 2 - 8 °C.

9.6.2.1.3   Add 900 ml of the cooled chlorinated tap water to a suitable container, bottle or flask. To the container, add 100 ml of the river water, mix well, leave for 5 minutes, and then determine the free and total chlorine concentration. To a second container, add 900 ml of the cooled chlorinated tap water and 100 ml of deionised or distilled water, mix well, leave for 5 minutes and then determine the free and total chlorine concentration. These two containers are used as controls for assessing whether the chlorine demand is too high. For example, if the concentration of chlorine in the mixed tap and river water falls to non-detectable levels within the 5 minutes, then 10 litres of tap water containing a higher concentration of chlorine, ie greater than 0.1 - 0.5 mg/l, will be required. The concentration of chlorine in the tap water, required to achieve the desired concentration of free chlorine in the mixed tap and river water solution, will vary according to the pH and organic and inorganic contents of the river and tap water. It may be necessary to carry out preliminary trials to determine the optimum concentration of chlorine in the tap water. When satisfactorily resolved and 10 litres of tap water of the correct concentration of chlorine have been prepared, add 900 ml of cooled tap water containing the correct level of chlorine to each of seven suitable containers, bottles or flasks.

9.6.2.1.4   Add 100 ml of the river water to one of the containers and mix well. Allow the chlorination process to react for 1 minute ± 5 seconds, and then add 1 ml of 18 % m/v sodium thiosulphate pentahydrate solution to the container. Cap and mix well, and store at 2 - 8 °C. Repeat the procedure using each of the remaining six containers and chlorination times of 1.5, 2.0, 2,5, 3.0, 3.5, and 4.0 minutes (± 5 seconds) respectively.

9.6.2.1.5   Remove 10 ml of the mixed tap and river water from each container and analyse each of the seven samples for the target organism. A method should be used that will yield a presumptive result, ideally, within 24 hours. Store the containers at 2 - 8 °C.

9.6.2.1.6   After incubation, determine the number of organisms in each 10 ml aliquot, and identify those containers, bottles or flasks found to contain 30 - 90 target organisms in the corresponding 10 ml aliquots. This number of organisms is higher than the target range of 20 - 50, in order to allow for some decay in the population of the organisms during overnight storage.

9.6.2.1.7   For the number of identified containers possessing 30 - 90 target organisms in 10 ml aliquots, add 900 ml of fresh tap water from the original source (9.6.2.1.1) to separate clean 1000 ml containers. To each of these containers, add sufficient sodium thiosulphate pentahydrate solution to neutralise any residual chlorine and mix well. To each separate container, add 100 ml of the corresponding mixed tap and river water samples from those identified containers possessing 30 - 90 target organisms in 10 ml aliquots. Cap and mix well. Each 1000 ml of diluted mixed tap and river water sample now enables up to 10 replicate 100 ml samples to be analysed by two or more methods used in parallel by one or more analysts. Alternatively, larger volumes of diluted mixed tap and river water samples can be prepared, by increasing proportionately the volumes of fresh tap water from the original source (9.6.2.1.1) and mixed tap and river water.

9.6.2.2   Generation of chlorine-stressed target organisms using sewage effluent

9.6.2.2.1   Collect 10 litres of tap water from a supply that is representative of the water supplies to be tested, and cool to 2 - 8 °C (store overnight if necessary). Collect at least 1000 ml of sewage effluent and store for one hour at 2 - 8 °C to ensure solid material settles.

9.6.2.2.2   Prepare a solution of chlorine, containing 12 - 25 mg/l by dissolving the appropriate amount of hypochlorite solution or chlorine-generating tablets in 1 litre of distilled or deionised water. Cap and mix well.

9.6.2.2.3   Taking care not to disturb any settled solid material, transfer 500 ml of the sewage effluent into a clean 10 litre container (one fitted with a tap will make the following procedures easier to carry out) containing a magnetic stirrer bar, or other stirring mechanism. Add 8.5 litres of the tap water previously stored at 2 - 8 °C. Cap the container, mix the contents thoroughly and stand the container on a magnetic stirrer and stir vigorously.

9.6.2.2.4   Whilst maintaining the stirring action, add to the container, sufficient volume, up to 1000 ml, of the solution of chlorine to produce a free chlorine concentration in the mixed tap water-sewage effluent solution of 1.2 - 2.5 mg/l. (The exact volume of chlorine solution may have to be adjusted accordingly). Mix the contents vigorously. After 3 minutes, transfer 500 ml of the chlorinated mixed tap water-sewage effluent solution into a suitable vessel, bottle or flask containing 1 ml of 18 % m/v sodium thiosulphate pentahydrate solution. Stopper and mix well by inverting several times to ensure the chlorine is rapidly neutralised. Repeat the procedure at one-minute intervals, by transferring 500 ml of the chlorinated mixed tap water-sewage effluent solution into other, separate vessels, bottles or flasks each containing 1 ml of 18 % m/v sodium thiosulphate pentahydrate solution, until 16 samples have been taken and prepared.

9.6.2.2.5   Remove 10 ml of the mixed tap water-sewage effluent solution from each container and analyse each of the 16 solutions for the target organism. A method should be used that will yield a presumptive result, ideally, within 24 hours. Store the containers at 2 - 8 °C.

9.6.2.2.6   After incubation, determine the number of organisms in each 10 ml aliquot, and identify those containers, bottles or flasks found to contained 30 - 90 target organisms in the corresponding 10 ml aliquots. This number of organisms is higher than the target range of 20 - 50, in order to allow for some decay in the population of the organisms during overnight storage.

9.6.2.2.7   For the number of identified containers possessing 30 - 90 target organisms in 10 ml aliquots, add 900 ml of fresh tap water from the original source (9.6.2.2.1) to separate clean 1000 ml containers. To each of these containers, add sufficient sodium thiosulphate pentahydrate solution to neutralise any residual chlorine and mix well. To each separate container, add 100 ml of the corresponding mixed tap water-sewage effluent solution from those identified containers possessing 30 - 90 target organisms in 10 ml aliquots. Cap and mix well. Each 1000 ml diluted mixed tap water-sewage effluent solution now enables up to 10 replicate 100 ml samples to be analysed by two methods used in parallel by one or more analysts. Alternatively, larger volumes of diluted mixed tap water-sewage effluent solution can be prepared, by increasing proportionately the volumes of fresh tap water from the original source (9.6.2.2.1) and mixed tap water-sewage effluent solution.

9.6.3   Confirmation tests

If confirmation of presumptive target colonies is required, then this should be carried out according to the requirements of the method. Generally, for drinking water, 10 presumptive colonies should be tested for confirmation if there are more than 10 presumptive colonies present, and all colonies should be tested, if there are 10 presumptive colonies or less. Colonies should always be chosen at random, but to avoid any bias from, for example, unconscious choice of similar colonies, all the colonies in a randomly chosen segment of appropriate size should be examined. See also section 8.3.3.

9.6.4   Verification of identity of target and non-target organisms

Methods should already have undergone validation that should have included a determination of the proportion of false-positive and false-negative results. However, this determination may have been carried out on a limited range of samples or sources of organisms. It is possible that different sources or categories of origin of water may contain different spectra of organisms from those examined in the initial validation trial and this may affect the proportion of false-positive and false-negative results. It can be useful, therefore, to carry out a more extensive identification of a selection of target or presumptive target colonies and non-target or presumptive non-target colonies. This identification is distinct from any confirmation steps that may be an integral part of the methods under test.

A minimum of 100 target colonies and, where appropriate, 100 non-target colonies per method should be selected for full identification by a suitable method. For most purposes, commercial identification kits may be adequate, but other approaches to identification may be more appropriate for some organisms. If the method under test is used for more than one target or presumptive target organism then at least 100 representative colonies of each organism should be examined. For example, if a method detects *Salmonella* species and *E. coli* simultaneously, then examination of 100 presumptive colonies of *Salmonella* and

100 presumptive colonies of *E. coli* would be required. The colonies should be selected so that they are evenly distributed over the sources of water examined. When confirmation tests are conducted, the most appropriate procedure of selecting target colonies for further identification is to choose the first one identified for confirmation. The advantage of this is that it will be known whether the colony confirmed or not. Non-target colonies should be selected at random, preferably one colony per Petri dish or plate and selected so that there are similar numbers examined from each source of water.

The spectrum of target or non-target organisms detected should be compared with that expected from previous validation data. If a particular source or category of origin of water exhibits differences from other sources, examination of the identification data may facilitate an interpretation of the differences.

9.7   Interpretation of stage 1 data

Pilot work with the preparation of samples is essential. It is necessary to ensure that as many samples as possible give counts within the required range of 20 - 50 target organisms. Once the study has commenced, all enumerated counts should be recorded. If any result is higher than expected, for example, a result is too numerous to count (such as greater than 100 for membrane filtration, or in the multiple tube technique, all tubes exhibit growth in the medium) then the subsequent data analysis may be biased. If the paired results obtained by both methods are too numerous to count, then both results can be omitted from the data analysis. This is because both results contribute little or no information about whether the new (trial) method gives a higher or lower result than the reference method. However, if only one method produces a result that is too numerous to count, then both results must be included in the data analysis. The exclusion of such results could bias the conclusions. In some situations, it is possible to allocate arbitrary numbers to those results reported as being too numerous to count, for example, 181 for a result reported as greater than 180. However, if there are many such results, the use of non-parametric statistical techniques may be more appropriate because there will be no proper estimate of the difference in counts for these paired results. Clearly, the data analysis will be much easier to conduct and interpret, if the results fall within the target range.

When a zero count obtained by one method is reported but is associated with a non-zero count obtained by the other method, then both results must be recorded and included in the data analysis. If paired zero counts are reported by both methods then these results can be excluded from the data analysis because they contribute little or no information about whether the new (trial) method gives a higher or lower result than the reference method.

9.7.1   Statistical comparison of methods

When performing confirmation tests, representative colonies should be chosen at random and typically, 10 colonies per plate are chosen. Depending upon circumstances, it may be necessary to test more colonies. See also section 8.3.3.

9.7.1.1 Stage 1 - spiked samples (20 - 50 target organisms/aliquot)

The procedures described in sections 9.6.2.1 and 9.6.2.2 should enable aliquots of samples to be prepared that contain 20 - 50 target organisms. The samples may be stored and appropriate aliquots withdrawn and tested by both methods. This procedure should then be repeated on different days. However, on every occasion, the sample should be thoroughly mixed before the appropriate volumes are withdrawn for analysis by both methods. The results from both

methods must be recorded as a "paired sample" result and examples of data analysis are given in Annex B.

It is preferable that the prepared samples described in sections 9.6.2.1 and 9.6.2.2 are derived from a selection of sources or categories of origin (see section 9.2.1). Each category of origin will involve material from a particular source (for example, a specific section of river, a treatment works, etc.). Material can be collected over a period of time. For convenience, these categories of origin are referred to as "sources", although it is noted that the actual samples prepared are not taken directly from particular sources but have involved some manipulation according to the details within sections 9.6.2.1 and 9.6.2.2.

A sufficient number of samples (at least 15) from each source or category of origin (usually 5 - 10) should be analysed to give statistical information to enable the following questions to be answered satisfactory.

Question (I)   Is the relative performance of the two methods similar for all the sources or categories of origin used in the study?

Question (II)   On average, does the new (trial) method find similar numbers of target organisms, or significantly lower or significantly higher numbers of organisms?

Question (III) If the new method finds similar numbers of organisms, does the 95 % confidence interval for the estimated difference between the average counts exclude the value where the new (trial) method finds greater than, or equal to, 10 % lower numbers of organisms than the reference method?

The analysis of at least 15 samples per source or category of origin, together with a total of not less than 150 samples for all sources or categories should provide sufficient information to answer the above questions. However, the difficulty of being able to predict the numbers of target organisms in a sample makes it difficult, in turn, to predict the statistical information in a fixed size trial. The numbers of samples and sources suggested above are, therefore, to be used as a guide and the final numbers will be dependent on the outcome of the comparison. If the comparison appears inconclusive, then more samples should be analysed.

If the results of stage 1 of the comparison indicate that the new (trial) method appears acceptable, then stage 2 of the comparison should be undertaken to confirm the outcome.

9.7.1.2 Data collection and evaluation

Analyse 15, or more, samples per source or category of origin by the two methods under investigation, ie the reference method and the new (trial) method, ensuring that at least 150 samples are examined.

Plot the paired results, differentiating each source or category of origin. Also, plot the differences (actual or transformed data) on appropriate scales.

9.7.1.2.1   From the data, ascertain whether the data are suitable for parametric analysis, ie are the data distributed in an approximately Gaussian (or normal) manner? If the answer to this question is yes, perform a data analysis, for example using the *t*-test, analysis of variance (ANOVA) test etc. If the answer to the question is no, transform or convert the data to an appropriate scale, if there is one, and carry out a parametric data analysis. Alternatively,

perform a non-parametric data analysis, using, for example Wilcoxon signed rank test. Whether a parametric or non-parametric data analysis is carried out, Question (I) should then be considered. Namely, is the relative performance of the two methods similar for all the sources or categories of origin used in the study? Depending on the data analysis, further questions need to be addressed.

For a parametric data analysis, do the t-tests or ANOVA tests show significant differences between the sources or categories of origin? If the answer to this question is no, Question (II) should be considered. If the answer is yes, then possible microbiological causes should be investigated and decisions taken whether or not the differences affect part of (ie, a particular source) or the whole of the data (ie, all sources). Depending upon these actions and decisions, the whole of the data may need to be rejected and the comparison terminated, ie, the new (trial) method is rejected. See example B2.

For a non-parametric data analysis, does the tabulation of data by source or category of origin of frequencies of paired results show differences between the sources or categories of origin? If the answer to this question is no, Question (II) should be considered. If the answer is yes, then possible microbiological causes should be investigated and decisions taken whether or not the differences affect part of, or the whole of, the data. Depending upon these actions and decisions, the whole of the data may need to be rejected and the comparison terminated, ie the new (trial) method is rejected. See example B2.

9.7.1.2.2   Question (II) concerns whether or not, on average, the new (trial) method finds similar numbers of target organisms, or significantly lower or significantly higher numbers of organisms. Depending on the data analysis, further questions need to be addressed.

For a parametric data analysis, does the mean difference between enumerated counts differ significantly from zero? If the answer to this question is no, Question (III) should be considered. If the answer is yes, then a further question needs to be addressed. Namely, does the new (trial) method show significantly lower or significantly higher numbers of organisms? If the counts are significantly lower then the comparison is terminated, ie the new (trial) method is rejected. If the counts are significantly higher, then the new (trial) method can be accepted.

For a non-parametric data analysis, does the median difference between enumerated counts differ significantly from zero? If the answer to this question is no, Question (III) should be considered. If the answer is yes, then a further question needs to be addressed. Namely does the new (trial) method show significantly lower or significantly higher numbers of organisms? If the counts are significantly lower then the comparison is terminated, ie the new (trial) method is rejected. If the counts are significantly higher, then the new (trial) method can be accepted.

9.7.1.2.3   Question (III) concerns whether or not, the 95 % confidence interval for the estimated difference between the average counts excludes the value where the new (trial) method finds greater than, or equal to, 10 % lower numbers of organisms than the reference method. Depending on the data analysis, further questions need to be addressed.

For a parametric data analysis, does the 95 % confidence interval of the mean difference include the value where the new (trial) method finds greater than, or equal to, 10 % lower numbers of organisms than the reference method? If the answer to this question is no, then the results of stage 1 of the comparison would indicate the data appear acceptable and stage 2 of

the comparison should be undertaken. If the answer is yes, then stage 1 of the comparison should be expanded and more samples analysed, and the whole process repeated.

For a non-parametric data analysis, does the 95 % confidence interval of the median difference include the value where the new (trial) method finds greater than, or equal to, 10 % lower numbers of organisms than the reference method? If the answer to this question is no, then the results of stage 1 of the comparison would indicate the data appear acceptable and stage 2 of the comparison should be undertaken. If the answer is yes, then stage 1 of the comparison should be expanded and more samples analysed, and the whole process repeated.

9.8   Stage 2 evaluation

The stage 2 of the comparison should be undertaken only when satisfactory results are obtained from stage 1. Stage 2 comprises the comparison of results of paired analyses of samples containing less than 20 target organisms per unit test volume. This comparison is carried out to ensure that the results remain valid at lower levels of organisms, approaching those numbers closer to statutory limits, but not so low as to be based on presence/absence criteria.

The paired results of at least 30 samples are required, where enumerated counts in the range 1 - 10 are recorded by at least one of the methods used. The samples can be prepared in the same way as described in sections 9.6.2.1 and 9.6.2.2 but with extra dilution steps. Successive two-fold dilutions of the same sample can be prepared, but samples should be derived from more than one source or category of origin. In addition, samples should contain an appropriate diversity of organisms.

If, the paired results obtained in stage 1 contain at least 30 samples giving counts in this lower range for all sources or categories of origin, then the data from these samples can be used for the stage 2 evaluation.

9.8.1   Data analysis for stage 2

As for stage 1, all the results should be plotted. With low counts it may be more of a problem to use a parametric data analysis approach, and it becomes more efficient to use a non-parametric analysis, see the worked examples in annex B. The proportion of paired results where the count by the new (trial) method exceeds the count given by the reference method should not be significantly lower than 50 %. Thirty samples should give an estimate of the proportion, with a confidence interval that is not too large, for example ± 20 %. If the confidence interval is large and there is evidence to suggest the new (trial) method is not performing well, then more samples should be analysed to establish whether or not there is any significant difference within these bounds.

9.9   Evaluation of a method for use in a laboratory

Any laboratory considering the introduction of a microbiological method should be familiar with its use within the laboratory. For example, staff should analyse samples containing pure cultures in suspensions in waters in the range 20 - 50 target and non-target organisms per unit test volume.

9.9.1   Evolution of method introduction

On a national or international scale, the adoption of a new method involves a sequential series of events. Namely,

(i)   comparison of new (trial) method with reference method in one laboratory;
(ii)   subsequent comparison of new (trial) method with reference method in four , or more laboratories;
(iii)   assessment of robustness; and
(iv)   adoption of new method.

A new method should undergo full comparative testing, using the procedure outlined in this document, in at least five laboratories before being regarded as potentially of general applicability. Where adequate comparative assessments have been undertaken in a single laboratory and these assessments indicate that the results obtained using a new (trial) method are comparable to the results obtained by a reference method, then the new method could be adopted for routine use by that laboratory. The adoption of the new method for routine use would not depend on whether other laboratories had carried out similar studies. When five, or more, laboratories have demonstrated that the performance of a new method is equivalent to, or better than the performance of a reference method, then wider adoption by other laboratories can be considered. In these cases, the comparison exercises undertaken by other laboratories may involve fewer numbers of samples. The comparison studies carried out in the initial five, or more, laboratories may require the replicate analysis of at least 180 samples (150 samples for stage 1 and 30 samples for stage 2) in each laboratory. Ideally, all procedures described in section 9.6.1 should be used, and samples should be representative of the sources of waters that the laboratory is likely to analyse by the new method. Data from the comparison studies undertaken in the different laboratories should be then be combined and reviewed following further statistical appraisal. By combining the data, it is possible to assess more confidently the robustness, repeatability and reproducibility of the new method.

Once the robustness, repeatability and reproducibility of the new method have been satisfactory established, the new method is generally acceptable for adoption for routine use. Therefore, the numbers of samples that subsequent laboratories need to analyse by the new method can be reviewed in light of the expanding database. However, a minimum of 30 samples, containing low numbers of a variety of organisms, should be analysed and results compared with those obtained using the reference method.

9.10   Comparison of an MPN method with an enumeration method

The design of the study and the same procedures described in sections 9.6 and 9.7 should be used for comparing results obtained using an MPN method and those obtained using an enumeration method. When an MPN method is the new method, the aim of the comparison exercise is to show that the MPN method does not find significantly lower numbers of organisms than found by the enumeration method, and if this is the case, the average difference in counts is accurately established. The required level of accuracy is such that the 95 % confidence interval for the average difference in counts should not include the situation where the MPN method finds 10 % lower numbers of organisms. In situations where the MPN method is the reference method, the investigation should establish that the enumeration method does not give significantly lower counts. If this is so, then the 95 % confidence interval for the average difference in counts should not include the situation where the enumeration method finds 10 % lower numbers of organisms.

Water samples should be prepared which contain organisms from a variety of sources or categories of origin, and be typical of the samples analysed in the laboratory. Whilst the same procedures as described in sections 9.6 and 9.7 are used, the only difference will be those factors influencing the choice of statistical methods. Paired results are used but, in the situation where an enumerated method is compared with an MPN method, results are based on the comparison of enumerated counts and most probable numbers.

The range of values achievable with a multiple tube method is discontinuous within the range of the method. For example, in an 11-tube series (1 x 50 ml, 5 x 10 ml, 5 x 1 ml) if 9 tubes exhibit growth in the medium (say 1, 5, 3) then from appropriate tables, the MPN is 91 per 100 ml. If 10 tubes exhibit growth in the medium (say 1, 5, 4) then the MPN is 160 per 100 ml. It is impossible to obtain a count between 91 and 160. Appropriate tables[28] show the counts (MPNs) and ranges of counts (MPRs) corresponding to 11-tube series (1 x 50 ml, 5 x 10 ml, 5 x 1 ml) and 15-tube series (5 x10 ml, 5 x 1 ml, 5 x 0.1 ml). One approach[29], especially where a tube series exhibits large gaps in MPNs is to group the results from the counting methods and compare them with the corresponding MPN. The grouping is carried out by consideration of each count and determining the tube combination that would be the most appropriate from a sample containing this number of organisms. (This should not be confused with MPRs or confidence intervals published for MPNs[28] which are obtained from different conditional probabilities). Appropriate conditional probabilities have been published[29] and resulting ranges tabulated for tube combinations. For example, with the 11-tube series given above, it has been shown that counts of between 69-110 would probably give a tube result of 1, 5, 3 and an MPN of 91. Counts of between 111-175 would probably give a tube result of 1, 5, 4 and an MPN of 160. Enumerated counts 69-110 could be interpreted as "equivalent" to an MPN of 91. Alternatively, the MPN can be regarded as the end result and compared directly with the count from the paired result. Careful plots of the results should be made and consideration given to using non-parametric analyses along similar lines to those described in the worked examples with enumeration methods in the second example of annex B.

These problems can be reduced, by using suitable samples where less than half of the tubes in the series exhibit growth. If not, the MPN will be an approximate count, and the comparison with the enumeration method might become biased. Multiple tube methods that require large numbers of tubes (at several dilutions) are more reliable than multiple tube methods with fewer dilutions and tubes.

As with a comparison of two enumeration methods, it is important that when the study is underway, all results are used. The only results that should be discarded are those where both methods recorded zero counts, or both methods failed to give an estimate of the count because the enumeration method showed overgrowth and the MPN method resulted in all tubes within the series exhibiting growth, ie were positive.

9.11   Comparison of two MPN methods

The same procedures described in sections 9.6 - 9.8 should be used when two MPN methods are compared. The points raised in section 9.10 still apply to both MPN methods and the principles of the comparison remain the same. Again, factors may influence the choice of statistical methods, which should be decided after thorough scrutiny of the data summaries and plots. It is likely that non-parametric data analyses should be used. The preparation of samples should be such that the number of tubes in the series exhibiting growth in the medium for the reference MPN method should be less than half of those tubes inoculated.

10   References

1.   The Control of Substances Hazardous to Health Regulations 1999, Statutory Instrument 1999 No. 437, The Stationery Office.

2.   BS 2646:1993 Autoclaves for sterilisation in laboratories, Part 1 - Specification for design, construction and performance; Part 3 - Guide to safe use and operation; Part 5 - Methods of test for function and performance. British Standards Institution, London.

3.   Safety at autoclaves, Guidance Note PM73, 1998, London, Health and Safety Executive.

4.   Standing Committee of Analysts, Isolation and Identification of *Cryptosporidium* oocysts and *Giardia* cysts in waters 1999, *Methods for the Examination of Waters and Associated Materials*, in this series, Environment Agency.

5.   Categorisation of pathogens according to hazard and categories of containment, 1995. Advisory Committee on Dangerous Pathogens. London. Stationery Office.

6.   The prevention of laboratory acquired infection, Collins, C. H., Hartley, E. G. & Pilsworth, R., Public Health Laboratory Service, Monograph Series No. 6, 1974, London, Stationery Office.

7.   Safety procedures: notes for guidance, 4th edition. Public Health Laboratory Service, 1993, London.

8.   Potential inaccuracy of microbiological counts from routine water samples. *Water Science and Technology*, Tillett, H. E., 1993, **27**, 15-18.

9.   Preliminary study of microbiological parameters in eight inland recreational waters, *Letters in Applied Microbiology*, Public Health Laboratory Service Water Surveillance Group, 1995, **21**, 267-271.

10.   Inaccuracy of counts of organisms in water or other samples: effects of pre-dilution. *Letters in Applied Microbiology*, Tillett, H. E. & Farrington, C. P., 1991, **13**, 168-170.

11.   Water Quality - *Guidance on Validation of Microbiological Methods*, Technical Report ISO TR 13843:2000 International Standards Organisation, Geneva.

12.   The numerical interpretation of fermentation-tube results. *Journal of Infectious Diseases*, McCrady, M. H., 1915, **17**, 183-212.

13.   Tables for rapid interpretation of fermentation-tube results. *Public Health Journal, Toronto*, McCrady, M. H., 1918, **9**, 201-220.

14.   Estimation of bacterial densities by means of "the most probable number", *Biometrics*, Cochran, W. G., 1950, **6**, 105-116.

15.   Statistical methods and control in bacteriology. *Bacteriological Reviews*, Eisenhart, C. & Wilson, P. W., 1943, **7**, 57-137.

16.   Numerical estimation of *B. coli* by dilution method. *Indian Journal of Medical Research*, Swaroop, S., 1938, **26**, 353-378.

17.   The range of variation of the most probable number of organisms estimated by the dilution method, *Indian Journal of Medical Research*, Swaroop, S., 1951, **39**, 107-134.

18.   How probable is the most probable number? *Journal of the American Waterworks Association*, Woodward, R. L., 1957, **49**, 1060-1068.

19.   The probability of most probable numbers. *European Journal of Applied Microbiology*, de Man, J. C., 1975, **1**, 67-78.

20.   Estimated numbers of bacteria in samples from non-homogeneous batches of water. *Journal of Applied Bacteriology*, Tillett, H. E. & Coleman, R. E., 1985, **59**, 381-388.

21.   Most probable numbers of organisms: revised tables for the multiple tube method. *Epidemiology and Infection*, Tillett, H. E., 1987, **99**, 471-476.

22.   Microbiological Analysis of Food and Water: Guidelines for Quality Assurance, Lightfoot, N. F. & Maier, E. A. (eds.) 1998, Elsevier Science, Amsterdam.

23.   Duplicate split samples for internal quality control in routine water microbiology. *Letters in Applied Microbiology*, Lightfoot, N. F., Tillett, H. E., Boyd, P. & Eaton, S., 1994, **19**, 321-324.

24.   Statistical concepts in microbiology, *Bacteriological Reviews*, Steerman, R. L., 1955, **19**, 160-215.

25.   Comparison of Methods for Drinking Water Bacteriology - Cultural Techniques, Drinking Water Inspectorate, Ashdown House, 123 Victoria Street, London, SW1E 6DE or www.dwi.detr.gov.uk/regs/infolett/2000/info500.htm.

26.   Quality control in environmental microbiology compared with chemistry; What is homogeneous and what is random? *Water Science and Technology*, Tillett, H. E. & Lightfoot, N. F., 1995, **31**, 471-477.

27.   Development of microbiological reference materials. Commission of the European Communities, Community Bureau of Reference, Report EUR 14375 EN, Mooijman, K. A., IN'T Veld, P. H., Koekstra, J. A., Heisterkamp, S. H., Havelaar, A. H., Notermans, S. H. W., Roberts, D., Griepink, B. & Maier E., 1992, ISSN 1018-5593.

28.   Standing Committee of Analysts, The Microbiology of Recreational and Environmental Waters (2000), *Methods for the Examination of Waters and Associated Materials*, in this series, Environment Agency.

29.   Water quality control trials: statistical tables for direct comparison between membrane filtration bacterial counts and multiple tube methods with a description of the bacteriological method. *Epidemiology & Infection*, Tillett, H. E., Wright, A. E. & Eaton, S., 1988, **101**, 361-366.

Annex A   **95 % Confidence intervals for the (unobserved) count from the second sub-sample for the observed count from the first sub-sample (see section 8.5.2)**

Observed count in first sub-sample	95 % CI for unobserved count in second sub-sample	Observed count in first sub-sample	95 % CI for unobserved count in second sub-sample
0	0 – 5	51	33 – 73
1	0 – 7	52	33 – 75
2	0 – 9	53	34 – 76
3	0 – 11	54	35 – 77
4	0 – 12	55	36 – 78
5	0 – 14	56	37 – 79
6	1 – 16	57	38 – 80
7	1 – 17	58	38 – 82
8	2 – 19	59	39 – 83
9	2 – 20	60	40 – 84
10	3 – 22	61	41 – 85
11	3 – 23	62	42 – 86
12	4 – 24	63	42 – 88
13	5 – 26	64	43 – 89
14	5 – 27	65	44 – 90
15	6 – 28	66	45 – 91
16	6 – 30	67	46 – 92
17	7 – 31	68	47 – 93
18	8 – 32	69	47 – 95
19	8 – 34	70	48 – 96
20	9 – 35	71	49 – 97
21	10 – 36	72	50 – 98
22	10 – 38	73	51 – 99
23	11 – 39	74	52 – 100
24	12 – 40	75	52 – 102
25	13 – 41	76	53 – 103
26	13 – 43	77	54 – 104
27	14 – 44	78	55 – 105
28	15 – 45	79	56 – 106
29	16 – 47	80	57 – 107
30	16 – 48	81	58 – 108
31	17 – 49	82	58 – 110
32	18 – 50	83	59 – 111
33	19 – 52	84	60 – 112
34	19 – 53	85	61 – 113
35	20 – 54	86	62 – 114
36	21 – 55	87	63 – 115
37	22 – 56	88	63 – 117
38	22 – 58	89	64 – 118
39	23 – 59	90	65 – 119
40	24 – 60	91	66 – 120
41	25 – 61	92	67 – 121
42	26 – 63	93	68 – 122
43	26 – 64	94	69 – 123
44	27 – 65	95	69 – 125
45	28 – 66	96	70 – 126
46	29 – 67	97	71 – 127
47	29 – 69	98	71 – 128
48	30 – 70	99	73 – 129
49	31 – 71	100	74 – 130
50	32 – 72		

Annex B   Worked examples of comparison of methods for drinking water bacteriology

This annex describes worked examples of trials comparing two methods following the procedures detailed in sections 9.6 - 9.8. The first example is based on counts from membrane filtration methods and contains real data that have been selected to illustrate relevant situations. Both parametric and non-parametric data analyses have been applied to demonstrate both approaches. In practice, only one approach might be used.

Example B1

Stage 1 evaluation

A total of 150 samples were each split into 2 sub-samples. One sub-sample was analysed by a reference method and the second sub-sample was analysed by a new (trial) method. Counts were reported as confirmed coliform bacteria counts per 100 ml. Samples were derived from five geographically distinct areas, all relevant to the routine work of the laboratory where the trial was undertaken. The data are listed in Table B1. Whilst the aim was to produce counts in the range 20 - 50, in practice, the range was wider. Summary data analysis of all counts in stage 1 reveals

	Median value	Range of counts
Reference method	30	5 - 180
Trial method	30	3 - 170

The data are plotted as paired counts (shown in Figure B1) and on a logarithmic (base 10) scale as log paired counts (shown in Figure B2). Also, the difference in the paired counts (count for trial method - count for reference method) is plotted on a frequency distribution plot or histogram, shown in Figure B3. Similarly, the difference in the log paired counts, ie log (count for trial method) - log (count for reference method) is plotted and shown in Figure B4.

From a casual study of the data in Table B1, it appears that the trial method gives counts similar in magnitude to counts produced by the reference method. Also, the trial method appears to be comparable to the reference method in that there are similar numbers of observations, as shown in Figures B1 and B2, above and below the theoretical line of equivalence. The paired differences, both in counts and as the log scale, appear clustered around zero as shown in Figures B3 and B4, but some differences are large in magnitude. This is not unusual in "real life" microbiology where occasional observations display variation much greater than random (Poisson) variation. The median difference in paired counts is 1 and the range is between - 61 to + 90.

From the data, it now needs to be ascertained whether the data are suitable for parametric analysis, ie are the data distributed in an approximately Gaussian (or normal) manner, see section 9.7.1.2.1.

The histogram of the frequency distribution of the difference in paired counts, as shown in Figure B3, appears symmetric but with long tails. The histogram based on the logarithmically converted scale, as shown in Figure B4, shows that the tails are reduced, but are still affected by possible "outlying" values. If cumulative Gaussian frequency distribution data are plotted on "normal probability paper", a straight line is produced.

Table B1   Data for example B1 - Stage 1

| Source 1		Source 2		Source 3		Source 4		Source 5	
Ref. method	Trial method	Ref. method	Trial method	Ref. method	Trial method	Ref. method	Trial method	Ref. method	Trial method
96	106	13	7	35	41	8	10	25	25
22	24	15	21	27	24	11	3	60	52
22	30	21	28	20	23	76	70	19	20
40	40	23	34	40	108	15	17	11	12
40	20	14	35	20	13	19	33	10	8
15	23	5	29	40	16	27	28	9	10
6	11	8	74	70	84	21	21	7	11
80	80	16	12	37	81	21	30	19	26
10	15	84	52	50	15	50	81	58	75
60	40	71	10	80	32	50	60	43	39
63	43	9	69	57	95	34	19	142	140
50	50	80	78	80	60	30	25	8	10
80	170	60	60	14	10	41	47	20	30
14	14	33	16	40	15	22	24	180	170
40	38	24	14	60	126	100	98	10	30
60	40					17	13	10	20
21	32					75	77	13	12
31	45					11	7	32	24
86	72					14	17	10	10
14	23					33	33	70	76
						41	37	15	17
						6	20	27	28
						30	17	21	21
						27	23	21	30
						26	18	50	81
						15	8	34	19
						18	15	25	30
						20	28	24	22
						58	30	100	98
						17	21	13	17
						58	56	75	77
						21	19	14	17
						30	38	41	37
						64	64	27	23
						41	44	39	50
						25	25	17	21
						17	7	58	56
						52	36	21	16
						67	63	30	38
						12	32	52	36
						44	34	63	67
						32	19	34	44
						64	68	25	25
						39	36	64	68
						57	32	39	33
						39	42	39	42
						40	51	40	51
						21	29	12	16
						24	35	41	80
						41	80	32	27

Figure B1 **Example B1 - Stage 1 - Plot of paired counts for trial and reference methods**
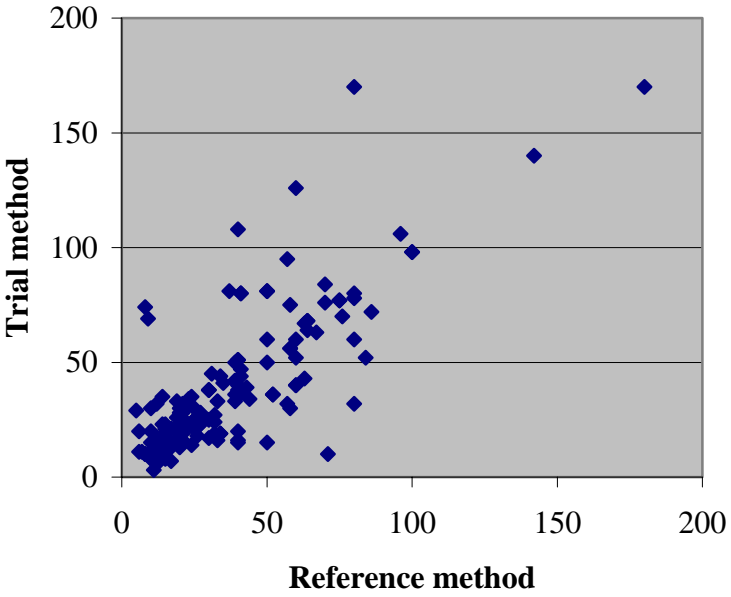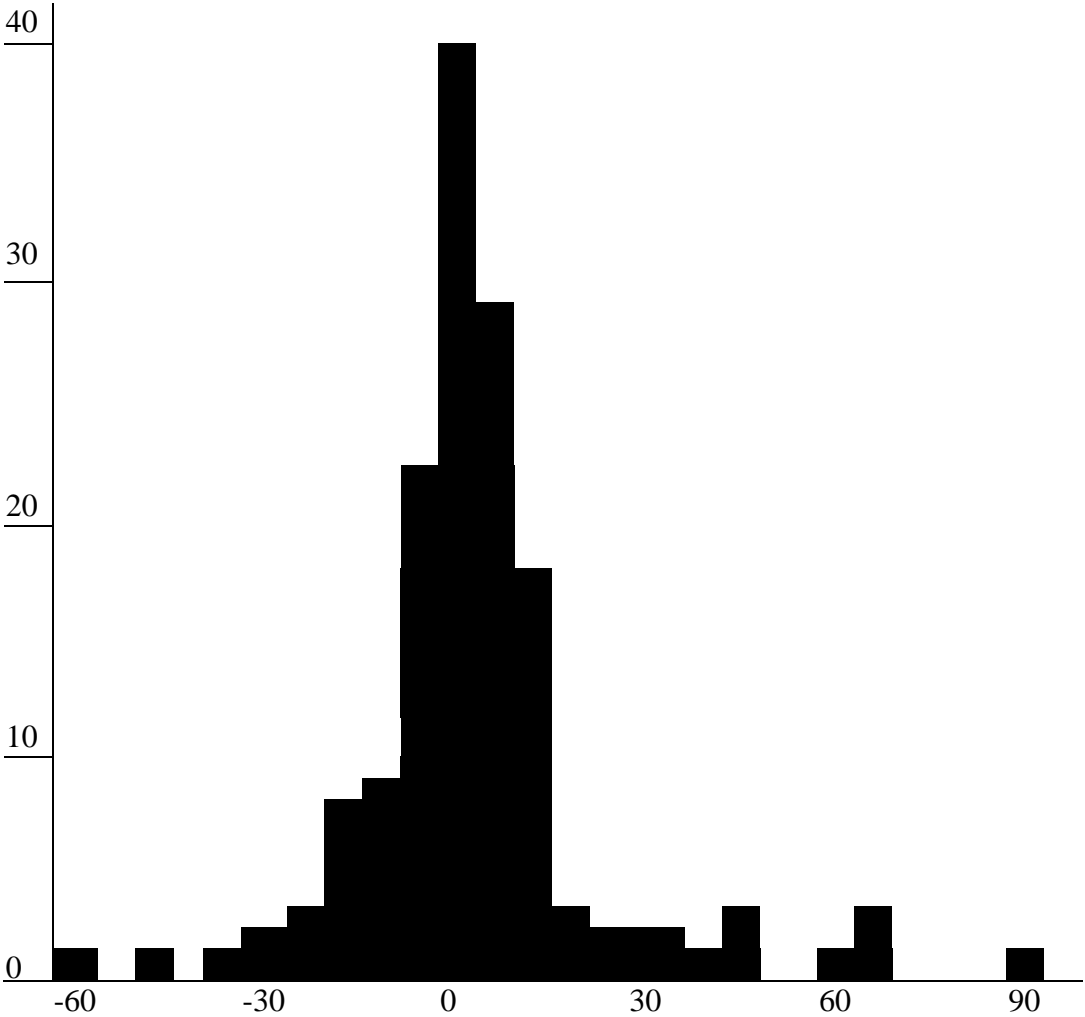

Figure B2 **Example B1 - Stage 1 - Plot of paired counts (converted to $\log_{10}$ scale) of trial and reference methods**
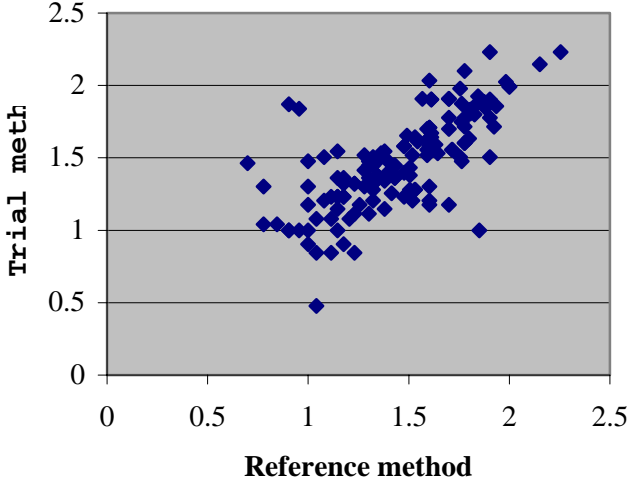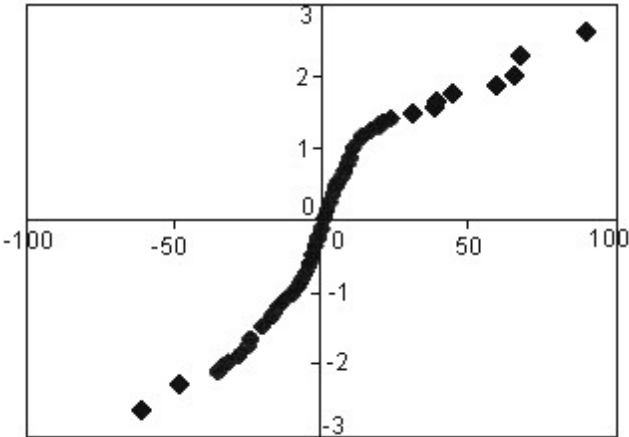
Figure B3   Example B1 - Stage 1 - Difference in paired counts between trial and reference methods plotted as a histogram

Figure B5   Example B1 - Stage 1 - Normalised probability plot for differences in paired counts between trial and reference methods
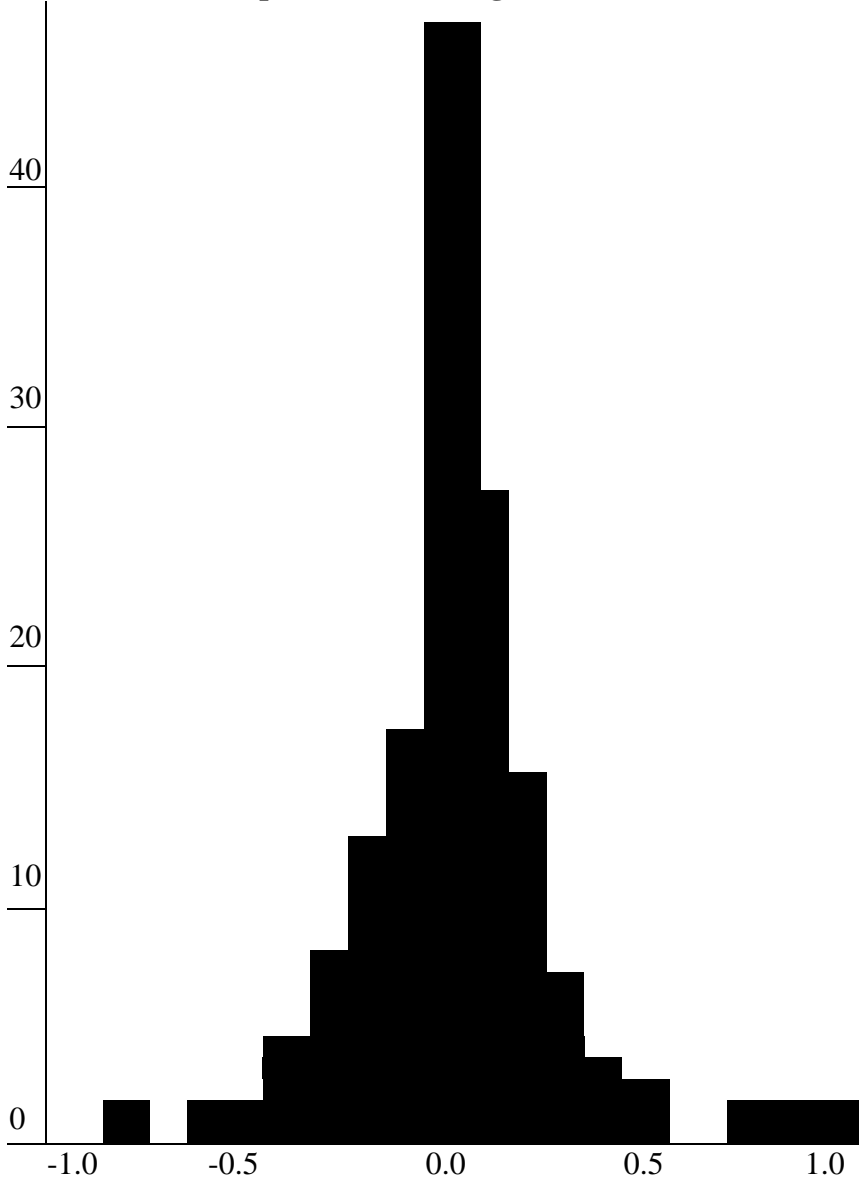
Figure B4   Example B1 - Stage 1 - Difference in log paired counts between trial and reference methods plotted as a histogram
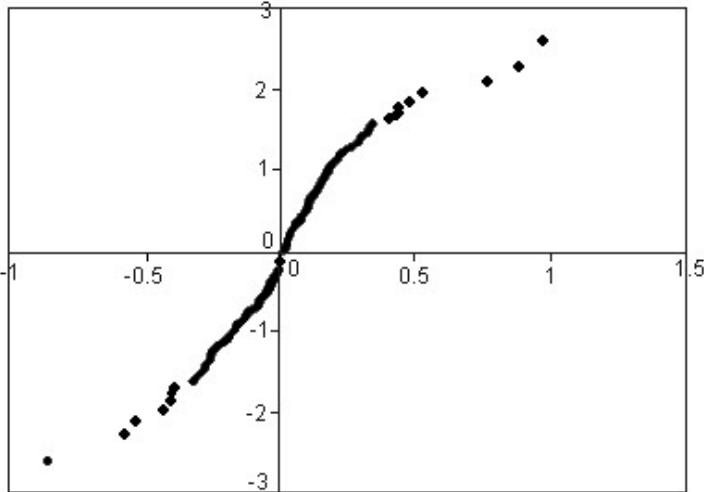

Figure B6   Example B1 - Stage 1 - Normalised probability plot for differences in log paired counts between trial and reference methods

When this procedure is carried out (as shown in Figures B5 and B6 for both histograms shown in Figures B3 and B4 respectively) it can be seen that the probability plots produce lines that are approximately linear. This is good evidence that the data are distributed in an approximately Gaussian or normal manner. Normal probability values for the differences, using both scales, were computed and plotted against the differences as normal plots as shown in Figures B5 and B6. This was performed using the NSCORES function of the Minitab statistical software package (Minitab Inc.,1989, release 7, State College, PA, USA) but other packages are also available. As already stated a plot for perfectly Gaussian or normal data is linear. The data based on original paired counts, as shown in Figure B5, produce more of an S-shape plot, but approach linearity when logarithmically converted data, as shown in Figure B6, are used. The affects of the "outlying" data can, however, still be observed.

Whilst a parametric analysis based on logarithmically converted data might be acceptable, a cautious approach might also be to use a non-parametric analysis.

Parametric analysis of stage 1 of example B1

Question (I) should now be considered. Namely, is the relative performance of the two methods similar for all the sources or categories of origin used in the study? The differences were compared between the sources or categories of origin using a one way ANOVA test. The output from a Minitab software package is shown in Table B2. As indicated, there is no significant difference between the sources. A significant difference (due to one or more sources showing higher counts from the trial method) would not affect this outcome, but a significant difference (due to one or more sources showing significantly lower counts from the trial method) would lead to significant differences. As no significant differences between the sources are shown, Question (II) should now be considered. (See section 9.7.1.2.2). The ANOVA test of differences between log counts using data from Table B1 is shown in Table B2.

Table B2   Example B1 - Stage 1 - Parametric analysis

	Degrees of freedom	Sum of squares	Mean square	F ratio	p-value
Source	4	0.2393	0.0598	1.12	0.349
Error	145	7.7427	0.0534		
Total	149	7.9821			

The 2-sided tabulated F-test value at the 95% confidence level with the associated degrees of freedom is approximately 2.8, hence, since this value is greater than the F ratio calculated value of 1.12, there is probably no statistical difference between the data from all sources.

In a similar manner, it can be shown that using a pooled or combined standard deviation of 0.23109, the range of the individual 95 % confidence intervals for each source includes zero, ie, the value expected for methods giving similar results. The pooled or combined standard deviation, s_c, is calculated from

$$s_c = \sqrt{\frac{s_1^{\,2}(n_1-1)\ldots+\ldots s_2^{\,2}(n_2-1)\ldots s_3^{\,2}(n_3-1)\ldots+\ldots s_4^{\,2}(n_4-1)\ldots+\ldots s_5^{\,2}(n_5-1)}{(n_1+n_2+n_3+n_4+n_5-N)}}$$

Where s_i is the standard deviation for source i (where i = 1 to 5) and n_i is the number of samples for source i. N is the number of sources. As expected, it is noted that the combined standard deviation is almost identical to the overall standard deviation shown below. Thus,

Source	Number of samples	Mean difference	Standard deviation	Range of 95% confidence interval
1	20	0.040485	0.16596	- 0.067 to + 0.148
2	15	0.09601	0.492393	- 0.031 to + 0.223
3	15	-0.04865	0.302399	- 0.176 to + 0.079
4	50	-0.0109	0.191284	- 0.077 to + 0.055
5	50	0.042763	0.123882	- 0.023 to + 0.108

Addressing Question (II) now, namely, on average, does the new (trial) method find similar numbers of target organisms, or significantly lower or significantly higher numbers of organisms? The overall mean of the 150 differences in the log paired results, ie log (count for trial method - log (count for reference method), is calculated as 0.02075 with a standard deviation, s, of 0.230681, and a standard error, se, of 0.0188. The standard error being calculated from

$$se \quad = \quad s / \sqrt{n}$$

where n is the number of paired results, ie in this case, 150.

If a new (trial) method were to find similar numbers of target organisms compared to a reference method, then it would be expected that over a period of time, the average difference of the methods would be zero. Using a 2-sided t-test, and comparing the overall mean of 0.02075 with a null hypothesis of zero, then t_{calc} = 0.02075 / 0.0188, ie 1.10. The 2-sided tabulated t-test value for 149 degrees of freedom at the 95 % confidence level is 1.97. Hence, since this value is greater than the corresponding calculated value of 1.10, there is probably no statistical difference between the mean of the differences and zero (p-value = 0.27). Thus, both methods find similar numbers of target organisms and there is no significant difference, on average, between the counts from the two methods. Question (III) should now be considered, see section 9.7.1.2.3, namely, does the 95 % confidence interval for the estimated difference between the average counts exclude the value where the new (trial) method finds greater than, or equal to, 10 % lower numbers of organisms than the reference method?

To address this question, the 95 % confidence interval, 95 CI, needs to be calculated from

$$95\ CI \quad = \quad mean \ \pm \ (t_{149,\ 0.05} \ x \ se)$$

where $t_{149,\ 0.05}$ from statistic tables is the 2-sided t-test value at the 95 % confidence level for 149 degrees of freedom, namely 1.97. Hence

$$
\begin{aligned}
95\ CI \quad &= \quad 0.02075 \ \pm \ (1.97 \ x \ 0.0188) \\
&= \quad 0.02075 \ \pm \ 0.037036
\end{aligned}
$$

ie, the overall 95 % confidence interval range is - 0.016 to + 0.058.

Question (III) seeks to determine whether the lower end of the overall 95 % confidence interval for the average difference includes the situation where the trial method finds 10 % fewer organisms than the reference method. In other words, whether an unacceptable situation occurs when, on average, trial count, TC, is 90 % of the reference count, RC, ie

$$TC = 0.9 \times RC$$

Using the logarithmically converted data, the equation becomes

$$\log(TC) = \log(0.9) + \log(RC)$$

Thus, the difference in the log paired results becomes

$$\log(TC) - \log(RC) = \log(0.9) + \log(RC) - \log(RC)$$
$$= \log(0.9)$$
$$= -0.04576$$

The overall 95 % confidence interval for the average difference, - 0.016 to + 0.058, lies entirely above the value of - 0.04576, hence, the data for the trial method appear acceptable and stage 2 of the comparison should be undertaken.

Non-parametric analysis of stage 1 of example B1

Question (I) should be considered. Namely, is the relative performance of the two methods similar for all the sources or categories of origin used in the study? The differences were compared between the sources or categories of origin as shown in Table B3 where the count from the trial method is compared with the count from the reference method using two-tailed, binomial probability.

Table B3   Example B1 - Stage 1 - non-parametric analysis

Source or category of origin	Number of results where trial method finds lower counts than reference method	Number of results where trial method finds an equal number of counts as the reference method	Number of results where trial method finds higher counts than reference method	Number of samples	p-value
1	6	4	10	20	0.4
2	7	1	7	15	0.6
3	8	0	7	15	0.5
4	23	4	23	50	0.6
5	17	4	29	50	0.052
All data	61	13	76	150	

The probability values test the null hypothesis that any discrepant paired result, where the counts were not identical, was as likely to give a higher count by the trial method as by the reference method. The values were calculated using exact binomial probabilities, parameter $p = 0.5$, two-tailed. Alternatively, McNemar's test can be used, but this becomes approximate when the number of discrepant pairs is less than about 15. The results show that there was no significant difference between methods in any of the sources or categories of origin and also, there was no significant difference between the sources or categories of origin.

Addressing Questions (II) and (III) the median of the 150 differences in the paired counts was compared with the null hypothesis of zero average difference and the 95 % confidence interval estimated. The Wilcoxon confidence interval was calculated using a Minitab software package.

The median difference (count for trial method - count for reference method) is 1, whilst the 95 % confidence interval range is - 0.5 to + 3. As already stated, the median count by the reference method was 30. The lower value of the confidence interval would represent an average of 1.7 % deficiency in counts by the trial method (ie 100 - [{(30 - 0.5) / 30} x 100]). Hence, the data for the trial method appear acceptable and stage 2 of the comparison should be undertaken.

Stage 2 evaluation

Stage 2 of the comparison is undertaken to ascertain if results complement the conclusions of stage 1. Six additional samples from each source or category of origin were obtained using the same five geographically distinct areas as used in stage 1 of the comparison. The 30 paired results are listed in Table B4 and plotted as shown in Figures B7 and B8. As can be seen, there is a scatter of results above and below the theoretical line of equivalence for all sources. These data appear to agree with the findings from stage 1.

Table B4   Data for example B1 - Stage 2

Source 1		Source 2		Source 3		Source 4		Source 5	
Ref. method	Trial method	Ref. method	Trial method	Ref. method	Trial method	Ref. method	Trial method	Ref. method	Trial method
4	0	0	3	5	9	9	6	1	1
2	1	4	1	3	3	1	0	3	3
8	7	0	6	1	0	6	11	13	3
1	4	14	10	1	1	2	2	0	2
0	2	13	7	3	2	0	1	2	3
2	4	4	7	7	14	6	7	0	6

Figure B7   Example B1 - Stage 2 - Plot of paired counts for trial and reference methods
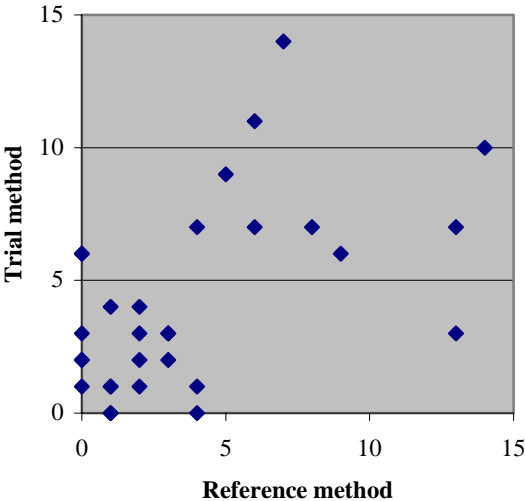
Figure B8   Example B1 - Stage 2 - Plot of paired counts (converted to $\log_{10}$ scale) of trial and reference methods
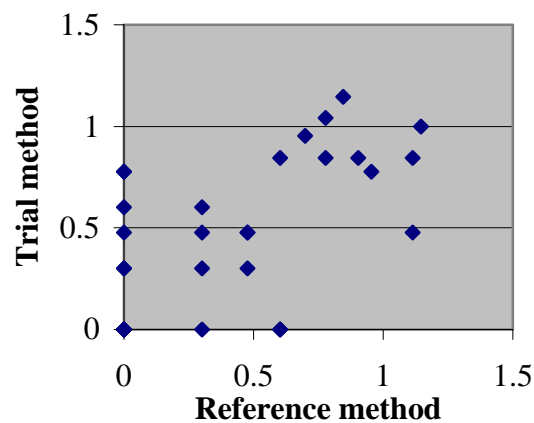
Parametric analysis of stage 2 of example B1

Summary data analysis of all counts in stage 2 reveals

	Mean	Median value	Range of counts
Reference method	3.8	2.5	0 - 14
Trial method	4.2	3	0 - 14

In addition, the median difference in paired counts is 0, and the range is - 10 to + 7. The mean difference in paired counts is calculated to be 0.3667 with a standard error of 0.6704. Using a 2-sided t-test, and comparing the mean difference of 0.3667 with a null hypothesis of zero, then $t_{calc} = 0.3667 / 0.6704$, ie 0.547. The 2-sided tabulated t-test value for 29 degrees of freedom at the 95 % confidence level is 2.045. Hence, since this value is greater than the corresponding calculated value of 0.547, there is probably no statistical difference between the mean of the differences and zero. Thus, both methods find similar numbers of target organisms and there is no significant difference, on average, between the counts from the two methods.

Addressing Question (III), the overall 95 % confidence interval, 95 CI, needs to be calculated from

$$95 \text{ CI } = \text{ mean } \pm (t_{29, 0.05} \text{ x se})$$

where $t_{29, 0.05}$ from statistic tables is the 2-sided t-test value at the 95% confidence level for 29 degrees of freedom, namely 2.045. Hence

$$
\begin{aligned}
95 \text{ CI } &= 0.3667 \pm (2.045 \text{ x } 0.6704) \\
&= 0.3667 \pm 1.3710
\end{aligned}
$$

ie, the 95 % confidence interval range is - 1.00 to + 1.74.

Since the range includes the value zero, and is within the 20 % criterion (see 9.8.1), ie in this case 16 % lower (ie 100 - [{4.2 - 1.0} / 3.8} x 100}), the trial method, therefore, appears comparable to the reference method at stage 2.

Non-parametric analysis of stage 2 of example B1

The count from the trial method is compared with the count from the reference method using two-tailed, binomial probability.

Table B5   Example B1 - Stage 2 - non-parametric analysis

Source or category of origin	Number of results where trial method finds lower counts than reference method	Number of results where trial method finds an equal number of counts as the reference method	Number of results where trial method finds higher counts than reference method	Number of samples	p-value
All data	11	5	14	30	0.5

From the information available, it would seem reasonable to accept the trial method as being comparable to the reference method in this laboratory.

Example B2

Stage 1 evaluation

The results presented for this example represent a smaller study where the available evidence casts doubt on the suitability of a trial method when compared to a reference method. This illustrates how a large study is not always necessary to demonstrate whether a new trial method is satisfactory

Forty-five samples were each split into 2 sub-samples, and the first sub-sample analysed by the trial method and the second sub-sample analysed by the reference method. Counts were recorded as confirmed *E. coli* counts per 100 ml. The samples were derived from three geographically distinct areas, all relevant to the routine work of the laboratory. The data are listed in Table B6. Whilst the aim was to produce counts in the range 20 - 50, in practice, the range was wider. Summary data analysis of all counts in stage 1 reveals the median and range of counts were:

	Median	Range
Reference method	30	10 - 84
Trial method	28	3 - 78

The data are plotted as paired counts (shown in Figure B9) and on a logarithmic (base 10) scale as log paired counts (shown in Figure B10). The median difference in paired counts is 1 and the range is - 41 to + 41. As in example B1, the difference in paired log counts was computed. The results of the parametric analysis of differences in paired log counts from Table B6 using the ANOVA test is shown in Table B7. The 2-sided tabulated F-test value at the 95% confidence level with the associated degrees of freedom is approximately 4, hence, since this value is less than the F ratio calculated value of 5.50, there is probably a statistical difference between the data from all sources. The results of the non-parametric approach are shown in Table B8, where the count for the trial method was compared with the count for the reference method using two-tailed, binomial probability. This approach shows that the counts for the second source displayed significant differences by the trial method compared to the

reference method. As indicated in Table B6, of the 15 paired results, 11 results were higher by the reference method, two paired results were the same and two results were lower (p-value = 0.02). In addition, the individual 95 % confidence interval for the results from source 2 indicate that the mean difference in paired log counts is entirely below zero, while the individual 95 % confidence intervals for the sources 1 and 3 includes zero.

Thus,

Source	Number of samples	Mean difference	Standard deviation	Range of 95% confidence interval
1	15	0.0503	0.2495	- 0.067 to + 0.148
2	15	- 0.2399	0.2407	- 0.031 to - 0.223
3	15	-0.0245	0.2564	- 0.176 to + 0.079

The results from sources 1 and 3 show no significant differences between the 2 methods. For source 2, the performance of the trial method compared to the reference method would suggest that it would not be advantageous to investigate further comparisons until, at least, further study resolved the indicative poor performance for source 2 and that this could be corrected. From the information available, it would seem reasonable to reject the trial method as being comparable to the reference method in this laboratory and that the comparison be terminated and a different trial method be investigated.

Table B6   Data for example B2 - Stage 1

Source 1		Source 2		Source 3	
Ref. method	Trial method	Ref. method	Trial method	Ref. method	Trial method
32	24	76	51	35	15
18	21	27	18	23	40
21	28	21	12	40	40
28	27	18	18	40	16
10	32	33	37	30	44
19	60	30	11	37	18
16	9	30	10	50	18
84	52	22	3	57	34
71	30	15	10	14	28
80	78	32	15	40	51
60	60	20	23	44	60
33	46	23	10	15	15
24	44	20	12	33	60
55	50	32	18	26	29
20	30	24	24	30	40

Figure B9   Example B2 - Stage 1 - Plot of paired counts for trial and reference methods
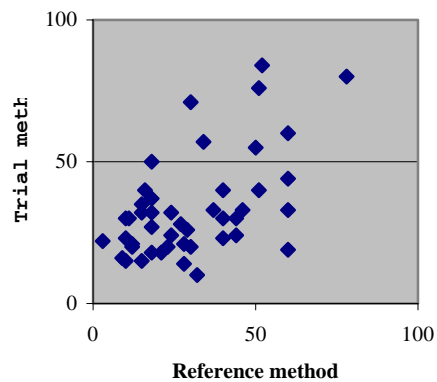
Figure B10   Example B2 - Stage 1 - Plot of paired counts (converted to $\log_{10}$ scale) of trial and reference method
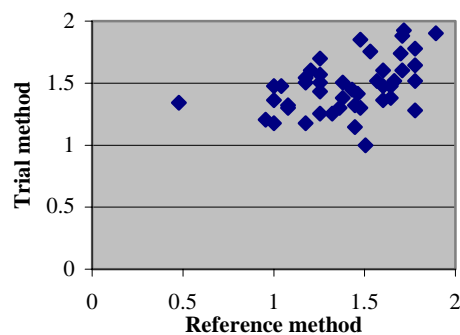
Table B7   Example B2 - Stage 1 - Parametric analysis

	Degrees of freedom	Sum of squares	Mean square	F ratio	p-value
Source	2	0.6812	0.3406	5.50	0.008
Error	42	2.6027	0.0620		
Total	44	3.2839			

Table B8   Example B2 - Stage 1 - non-parametric analysis

Source or category of origin	Number of results where trial method finds lower counts than reference method	Number of results where trial method finds an equal number of counts as the reference method	Number of results where trial method finds higher counts than reference method	Number of samples	p-value
1	7	1	7	15	0.5
2	11	2	2	15	0.02
3	5	2	8	15	0.3
All data	23	5	17	45	

Address for correspondence

However well procedures may be tested, there is always the possibility of discovering hitherto unknown problems. Analysts with such information are requested to contact the Secretary of the Standing Committee of Analysts at the address given below.

Secretary
Standing Committee of Analysts
Environment Agency
Wheatcroft Office Park
Landmere Lane, Edwalton
Nottingham
NG12 4DG

Environment Agency
Standing Committee of Analysts
Members assisting with this booklet

R A Barrell	N Lightfoot
C Benton	P Machray
D Blake	K Punter
P Boyd	H Roberts
S Cole	D Sartory
D Drury	H Tillett
C Fricker	M Waite
J V Lee	J Watkins

CONTACTS:

ENVIRONMENT AGENCY REGIONAL OFFICES

ANGLIAN
Kingfisher House
Goldhay Way
Orton Goldhay
Peterborough PE2 5ZR
Tel: 01733 371 811
Fax: 01733 231 840

MIDLANDS
Sapphire East
550 Streetsbrook Road
Solihull B91 1QT
Tel: 0121 711 2324
Fax: 0121 711 5824

NORTH EAST
Rivers House
21 Park Square South
Leeds LS1 2QG
Tel: 0113 244 0191
Fax: 0113 246 1889

NORTH WEST
PO Box 12
Richard Fairclough House
Knutsford Road
Warrington WA4 1HG
Tel: 01925 653 999
Fax: 01925 415 961

SOUTHERN
Guildbourne House
Chatsworth Road
Worthing
West Sussex BN11 1LD
Tel: 01903 832 000
Fax: 01903 821 832

SOUTH WEST
Manley House
Kestrel Way
Exeter EX2 7LQ
Tel: 01392 444 000
Fax: 01392 444 238

THAMES
Kings Meadow House
Kings Meadow Road
Reading RG1 8DQ
Tel: 0118 953 5000
Fax: 0118 950 0388

WALES
Rivers House/Plas-yr-Afon
St Mellons Business Park
Fortran Road
St Mellons
Cardiff CF3 0EY
Tel: 029 2077 0088
Fax: 029 2079 8555

ENVIRONMENT AGENCY
GENERAL ENQUIRY LINE
0845 9 333 111

ENVIRONMENT AGENCY
FLOODLINE
0845 988 1188

ENVIRONMENT AGENCY
EMERGENCY HOTLINE
0800 80 70 60

ENVIRONMENT
AGENCY