



guardians of drinking water quality

# Drinking Water Inspectorate

**STANDARD OPERATING PROTOCOL**

**FOR THE MONITORING OF *CRYPTOSPORIDIUM* OOCYSTS**

**IN TREATED WATER SUPPLIES TO SATISFY**

**THE WATER SUPPLY (WATER QUALITY) REGULATIONS**

**2000, SI No. 3184 ENGLAND**

**THE WATER SUPPLY (WATER QUALITY) REGULATIONS**

**2001, SI No. 3911 (W.323) WALES**

**Part 3 – Validation of New Methods or Parts of Methods for Sampling and Analysis**

**16 June 2005**

## CONTENTS

1.	INTRODUCTION TO THE STANDARD OPERATING PROTOCOL.....	3
2.	APPROVAL PROCESS .....	4
3.	APPROACH .....	6
4.	PERFORMANCE PARAMETERS .....	8
4.1	Relevant performance parameters.....	8
4.2	Performance criteria .....	8
4.3	Spiking sources and levels .....	9
4.4	Positive and negative controls .....	9
4.5	Phase 1 – In-house study.....	9
4.5.2	Sampling Equipment Suitability.....	9
4.5.3	Laboratory Method Performance Characteristics .....	9
4.6	Phase 2 – Inter-laboratory comparison .....	10
4.7	Performance criteria involving a variation of the SOP .....	10
5.	PHASE 1 .....	11
5.1	Validation criteria .....	11
5.1.1	Identification and selectivity.....	11
5.2	Sampling equipment suitability .....	11
5.2.1	Compatibility with existing sampling points .....	11
5.2.2	Capability of processing a suitable sample size within a 24 hour period 11	
5.2.3	Ability to recover oocysts from water samples.....	12
5.2.4	Modifications to the analytical technique and the detection of oocysts using microscopy .....	12
5.2.5	Verification of the use of an approved product against another approved product.....	13
5.3	Statistics to be used in phase 1 .....	14
5.4	Documentation.....	14
6.	PHASE 2 .....	14
6.3	Suitability of sampling equipment .....	15
6.4	Validation criteria for the <i>alternative</i> method or modified approved method 15	
6.5	Phase 2: Equivalence .....	15
	ANNEX A: ABBREVIATIONS, DEFINITIONS AND TERMINOLOGY .....	17
	ANNEX B: STATISTICAL TREATMENT OF VALIDATION DATA .....	19
	Statistical analysis of validation data .....	19
	Data presentation.....	19
	EXAMPLE 1: FICTITIOUS TRIAL VERSUS APPROVED METHODS .....	22
	Phase 1 - data from the comparison of a trial and approved methods for recovering and enumerating <i>Cryptosporidium</i> oocysts - single laboratory study. ....	22
	Phase 2 - data from the comparison of a trial and an approved method for recovering and enumerating <i>Cryptosporidium</i> oocysts - inter-laboratory study .....	26
	EXAMPLE 2: STUDY OF ‘TRIAL’ VERSUS ‘APPROVED’	
	IMMUNOMAGNETISABLE BEADS .....	29
	Phase 1 - data from the trial method IMS-1 and approved method IMS-2 for the recovery of <i>Cryptosporidium</i> oocysts - single laboratory trial .....	29
	Phase 2 - data from the trial method IMS-1 and approved method IMS-2 for the recovery of <i>Cryptosporidium</i> oocysts - inter-laboratory study.....	33

## 1. INTRODUCTION TO THE STANDARD OPERATING PROTOCOL

- 1.1 This Standard Operating Protocol (SOP) provides guidance from the Drinking Water Inspectorate (DWI) on behalf of the Secretary of State and the Welsh Assembly Government on the sampling and analysis requirements associated with the Water Supply (Water Quality) Regulations 2000 (for England) and the Water Supply (Water Quality) Regulations 2001 (for Wales), (The Regulations).
- 1.2 This Standard Operating Protocol is published in four parts:
- Part 1 Sampling and Transportation of Samples.
  - Part 2 Laboratory and Analytical Procedures.
  - Part 3 Validation of New Methods, or Parts of Methods for Sampling and Analysis.
  - Part 4 Requirements for the Inter-laboratory Proficiency Schemes.
- 1.3 **Wherever the terms *Cryptosporidium* or *Cryptosporidium* oocysts, or oocysts are used in this SOP, they refer to all species (active or inactive) of that genus within the size range 4-6µm, [i.e. *Cryptosporidium* spp].**
- 1.4 The conditions of sampling and examination must be such as to permit the use of the analytical results as admissible evidence in a Court of Law. The Standard Operating Protocol therefore describes a system that would ensure both the scientific quality of the results, and that a chain of evidence is maintained in compliance with the Police & Criminal Evidence Act 1984. The Protocol covers the minimum requirements to satisfy this requirement.
- 1.5 Any proposed deviation from the requirements contained in parts 1 and 2 of this SOP will require a full laboratory appraisal to be carried out in accordance with part 3 of this SOP. Results of such appraisal must be submitted to the DWI for consideration of approval. This approval must be obtained in writing before any changes are made to the relevant sections of the SOP. Any approved changes to this SOP will be circulated to all water companies and DWI approved laboratories.
- 1.6 This part of the Protocol provides guidance on validating of the sampling equipment and methods of analysis as well as validation of changes to parts of the analytical method. The validation is split in two phases:

Phase 1: a single laboratory validation to determine that the method produces results that are statistically equal to, or superior to, the performance of the approved methods detailed in the SOP<sup>1</sup>;

Phase 2 a multi laboratory/Company exercise to determine the method produces results that are statistically equal to, or superior to, the performance of the current methods detailed in the SOP in a representative number of approved locations.

---

<sup>1</sup> Full details of the approved methods detailed in the SOP are found on the DWI *Cryptosporidium* website [www.dwi.gov.uk](http://www.dwi.gov.uk).

## 2. APPROVAL PROCESS

2.1 This protocol applies to modification of the approved methods specified in the SOP<sup>2</sup>. In addition it applies to products which have already been approved but have not been approved for all the equipment or reagents specified in the SOP Methods used for the determination of *Cryptosporidium* oocysts in treated water supplies to meet the requirements of 'The Regulations' must be approved by the Drinking Water Inspectorate (DWI). Before approval can be granted, the fitness-for-purpose of the method must be established using the validation process described in this part of the Standard Operating Protocol (SOP).

2.2 The approval process must be as follows:

i) For this part of the SOP, relating to the approval of new or modified methods, the action limit is 100 oocysts in 1000 litres of water over 24 hours;

ii) The proposer appoints an independent DWI approved laboratory<sup>3</sup> to compare the candidate method and equipment with the SOP; using phase 1 of this document. The proposer may wish to submit trial protocols to the DWI or DWI's approved agents for examination prior to the evaluation starting;

iii) The independent laboratory reviews data from phase 1 and decides with the proposer whether the method is suitable for proceeding to phase 2. The data, together with a modified SOP and any additional data generated by the proposer, for example, stability of reagents, should be forwarded to the DWI or DWI's approved agents for approval to proceed to phase 2. If the data is not suitable and the proposer wishes to develop the method further before repeating phase 1, any further changes to the method should be documented and included in the retrial submission to the DWI or the DWI's approved agents;

iv) If data collected from phase 1 indicate the trial (an alternative or modified method) appears equivalent (or better) than the approved method for selectivity and recovery of oocysts then the proposer will be allowed to initiate phase 2 (the inter-laboratory study) using DWI approved laboratories. At this stage 'equivalent' will be interpreted as 'not significantly worse'.

v) Approved laboratories carry out the inter-laboratory study using the new SOP under the co-ordination of the proposer. The laboratory undertaking phase 1 trials may be included in phase 2;

vi) Collated and processed data from phase 2 and the new SOP are submitted, in an approved format, to DWI or DWI's approved agents;

vii) DWI or DWI's approved agents assess data and method;

viii) DWI issues either:

- written notice of approval;

---

<sup>2</sup> Full details of the approved methods detailed in the SOP are found on the DWI *Cryptosporidium* website [www.dwi.gov.uk](http://www.dwi.gov.uk).

<sup>3</sup> A list of approved laboratories can be found on the DWI *Cryptosporidium* website [www.dwi.gov.uk](http://www.dwi.gov.uk)

- written notice of approval subject to specified changes;
  - written notice of rejection.
- ix) DWI publishes notice of approval by letter to all water companies and approved laboratories and a copy will be posted on the *Cryptosporidium* website. DWI revises the SOP, which will become an approved method when posted on the DWI website.

### 3. APPROACH

These validation studies, parts 1 and 2, must collectively demonstrate the comparability of the alternative or modified method with the approved methods. The recommended sample types and material must be used to give representative information. The required performance of the trial method is that it should be at least equivalent (or better) than the standard method. 'Equivalent' will be interpreted as either 'significantly better' or as 'not significantly different', but the latter must be to a level of confidence acceptable to DWI. Thus, if the trial method is not finding significantly larger numbers of oocysts than the average difference between results from the two methods should be compatible with the null hypothesis of zero difference (with 95% probability). This statistical analysis will automatically lead to a statement about the likely range of the 'true average difference' between the methods, which is the average you would get from an infinite number of samples of the same types of water. This range is usually expressed as a 95% confidence interval and the lower end of this should rule out unacceptably worse recovery.

3.1 The performance of the alternative or modified method is evaluated using a reference material. In parallel with this, the approved Standard Operating Protocol is checked to provide a comparison. The reference material is a known number of oocysts in suspension obtained from an approved DWI supplier. The reference material can be prepared by:

- (a) using a flow cytometer; or
- (b) purchasing ready made standards containing known numbers of oocysts in a specified volume.

In each case the target number of oocysts in each aliquot to be used to compare the performance of the modified and standard methods shall be 100 oocysts in most cases (see paragraph 4.1.1).

3.2 In the event of modification of the approved method by the use of different equipment, suitable equipment qualification checks will be necessary to ensure that the proposed modifications will allow processing of the appropriate amount of sample and that chain of custody requirements in SOP part 1 are met.

3.3 The general approach is therefore to:

- check equipment suitability, (where the modification involves a change of equipment);
- determine specified equivalent performance parameters for the candidate *alternative* or modified approved method;
- compare these with the required performance.

3.4 The validation study will be in two phases. The first phase will be conducted in-house by a single laboratory (approved by DWI<sup>4</sup> and appointed by the proposer). This will include any appropriate confirmation of equipment suitability. Phase 1 will

---

<sup>4</sup> A list of approved laboratories can be found on the DWI *Cryptosporidium* website [www.dwi.gov.uk](http://www.dwi.gov.uk)

involve sampling at **three** sites. The second phase will involve a full field-performance study using a suitable group of **five** DWI approved laboratories, each sampling at **two** sites.

- 3.5 At various points in this part of the SOP reference is made to the use of *different* waters or water from *different sources*. It is intended that the validation study should evaluate the performance of an *alternative* or *modified* method over a range of water types, 'hard' and 'soft', 'treated' and 'untreated', 'surface and ground' in order to confirm its suitability for use throughout England and Wales. Details of the chemical composition of the water used shall be provided in the report for each phase of the trials. The chemical analysis must be undertaken such that the requirements in the Water Supply (Water Quality) Regulations 2000 and 2001 are met. Analytical results from the relevant public register may be used.
- 3.6 Following the validation studies, the SOP for the new method must be documented in a suitable format (consistent with the existing parts 1 and 2 of the SOP) to ensure consistent implementation once in use. Using data derived from the validation studies, this format must incorporate suitable quality control requirements and guidance on the correct expression of test results obtained by the method.

## **4. PERFORMANCE PARAMETERS**

### **4.1 Relevant performance parameters**

4.1.1 The following criteria for method performance are considered to be relevant and it is expected that these will have been studied (in preliminary validation) before the alternative or modified method is put forward for this validation within DWI approved laboratories:

- correct identification of oocysts;
- selectivity;
- recovery rate at the limit concentration (100 oocysts);
- recovery rate above and below the limit concentration;
- precision (in terms of repeatability within a laboratory);

#### **The outcome of phase 1 and 2 trials will also assess**

- usability of the method (ease of application and effort required to learn its use) assessed on the basis of inter-laboratory participants' comments
- direct comparison with the approved method.

4.1.2 Limits of detection and quantification have not been considered necessary for this SOP. Performance at the limit concentration (100 oocysts) and blank concentration provide sufficient information for comparison with the approved method.

### **4.2 Performance criteria**

4.2.1 Performance criteria have been set which are considered reasonable using the methods and equipment considered to be the best available at present. With improvements in methodology and equipment, DWI will review and revise the criteria from time to time:

- selectivity requirements will be evaluated by DWI by review of the evidence presented.
- recovery >30% (DWI is looking for improvements in percentage recoveries). It is, however recognised that different water types may give different recoveries when sampling large volumes of water. The 30% recovery was not determined on a scientific basis but on what was considered reasonable when the Regulations were implemented and will be subject to review as necessary.
- comparison of results from the test and approved methods for each laboratory and a variety of water sources to estimate equivalence (or improvement) of the test method.

### **4.3 Spiking sources and levels**

- 4.3.1 This SOP involves the spiking of filtration devices with oocysts suspensions, usually at the limit concentration, to test various method parameters. The filtration devices are used to sample a minimum of 1000 litres of water and the oocysts recovered. *Cryptosporidium* oocyst suspensions, from a DWI approved supplier, must be used (See paragraph 3.1).
- 4.3.2 The suspension should be vortex mixed prior to spiking. In the event of commercial suspensions being used, the suppliers' instructions should be followed.

### **4.4 Positive and negative controls**

- 4.4.1 Sufficient positive and negative controls should be prepared and analysed to determine that all the equipment and/or reagents used during the validation of new or part methods are demonstrated to be fit for purpose.
- 4.4.2 The positive and negative controls should be prepared according to Part 2 of the SOP.

### **4.5 Phase 1 – In-house study**

- 4.5.1 The within-laboratory study may examine new reagents (including oocysts) or new equipment for the recovery of oocysts from water. The new equipment may include new sampling equipment.

#### ***4.5.2 Sampling Equipment Suitability***

- 4.5.2.1 The following details must be studied with respect to the suitability of sampling equipment:
- compatibility of the sampling equipment with existing approved sampling points and equipment together with approved sampling arrangements (such as equipment size, connections and chain of evidence considerations);
  - capability of processing a suitable sample size within the period allowed by the Regulations and looking to the future, consideration should be given to sampling over a longer period e.g. 48 hours or until the headloss reaches four bar;
  - ability to recover oocysts from water samples;
  - compatibility with existing arrangements for transport of samples to the testing laboratory.

#### ***4.5.3 Laboratory Method Performance Characteristics***

- 4.5.3.1 The following within-laboratory performance characteristics must be evaluated and reported:
- confirmation of identity and selectivity;
  - confirmation that the method recovers oocysts from clean water samples;

- confirmation that the method is able to recover oocysts from dirty samples known to contain interfering material.
- assessment of false negative and positive results at the limit concentration;
- main study at limit concentration and at levels above and below – these to be studied in parallel with an identical examination using the standard method (thus giving paired results).

#### **4.6 Phase 2 – Inter-laboratory comparison**

4.6.1 The inter-laboratory comparison will determine:

- within and between laboratory equivalence under field conditions;
- recovery at the limit concentration under field conditions;
- performance on blank controls under field conditions.

4.6.2 The inter-laboratory comparison will also serve as the main confirmation of the sampling equipment suitability and the main performance parameters (as determined in phase 1). It will also serve to show that the method can produce consistent, reproducible results across a group of laboratories.

#### **4.7 Performance criteria involving a variation of the SOP**

4.7.1 A full study involving a phase 1 or phase 2 study is not required when a variation of the SOP does not require a change in the sampling or analytical procedure. Examples of this are:

- (i) A change from a manual process to an automated process at a particular stage of analysis.
- (ii) A variation in the type of sampling pump from that quoted in the SOP.
- (iii) Comparison of the performance of an approved product with another approved product where these have not previously been compared.

Further advice on this can be obtained from the DWI to determine if any amendment can be used without phase 1 and phase 2 trials.

4.7.2 The required performance of the trial method is that it should be at least equivalent (or better) than the standard method. 'Equivalent' will be interpreted as either 'significantly better' or as 'not significantly different', but the latter must be to a level of confidence acceptable to DWI. Thus, if the trial method is not finding significantly larger numbers of oocysts then the average difference between results from the two methods should be compatible with the null hypothesis of zero difference (with 95% probability). This statistical analysis will automatically lead to a statement about the likely range of the 'true average difference' between the methods, which is the average you would get from an infinite number of samples of the same types of water. This range is usually expressed as a 95% confidence interval and the lower end of this should rule out unacceptably worse recovery.

## **5. PHASE 1**

### **5.1 Validation criteria**

#### **5.1.1 Identification and selectivity**

5.1.1.1 The proposed method needs to be able to allow an analyst or an instrument to identify oocysts correctly and discriminate between them and interferents that might be reasonably encountered in real samples. In the case of microscopic methods this means structures of a comparable size and similar appearance to oocysts. However, the types of interference encountered may vary according to the principle of isolation and detection. For this reason it is not appropriate to specifically list typical interferences.

5.1.1.2 The proposer will be expected to show that for any *alternative* method or modified approved method, the interferences that might interfere with the enumeration of oocysts have been identified and the ability of the method to correctly enumerate oocysts in the presence of those interferences has been established. The following studies in a single laboratory analysing several different water types (see 3.5) and likely interferents, must be undertaken. It is important to select one water type that is likely to contain a significant amount of interferents which will challenge new filtration devices and modifications to the analytical technique.

5.1.1.3 Interferents may be generated from surface water using wound polypropylene or pleated membrane cartridge filters with a nominal pore size of 1 µm (for example Cuno filters – Microwynd) to collect the material. Water can be pumped through the filter in its housing without the need for flow measurement or flow restriction. Where necessary a 12 volt leisure battery and submersible pump can be used. Alternatively, a filter can be attached to the raw water tap at a water treatment works. Water is filtered until a suitable amount of particulate material has been collected. Filters can be run for between one and 24 hours or longer if required.

The particulate material is eluted from the filter by cutting the filter from its core, teasing out the fibres and washing in 0.1% polyoxyethylene sorbitan mono-oleate (Tween 80) in reagent water. Washing can be done manually or in an appropriate machine (for example a Stomacher - Seward Medical). The particulate material is concentrated from the eluate by centrifugation at an appropriate speed (for example 1,500 x g for 30 minutes). The final pellet should be washed in reagent water and centrifuged as above. The pellet volume is measured and the material suspended in reagent water such that a measured aliquot of the suspension provides a pellet volume of 0.5 ml or suitable required alternative. The material should be stored at 5 ± 3 °C.

### **5.2 Sampling equipment suitability**

#### **5.2.1 Compatibility with existing sampling points**

5.2.1.1 Alternative sample collection device systems must be compatible with the approved sampling equipment or approved collection device without significant leaks or risks of back-siphonage into the supply line. The alternative systems must be compatible with security considerations for approved sampling equipment in order to preserve chain of evidence requirements. In addition, it must satisfy the requirements set down in Part 1 of the SOP.

#### **5.2.2 Capability of processing a suitable sample size within a 24 hour period**

5.2.2.1 Although the suitability of the collection devices used in the approved method has already been established. When testing any new collection device the proposer must also run an approved filter in parallel with the new filtration device.

5.2.2.2 If alternative collection devices are proposed, either in *alternative* methods or as a modification to the approved method, the alternative collection devices need to be evaluated to confirm their ability to sample volumes, a minimum of 40 litres per hour for at least 24 hours (but could be up to 48 hours) without blockage, [the present regulations specify a filter change after 24 hours but a current review of the regulations may extend the sampling period, see 4.5.2.1], leakage or breakthrough. The study should be carried out on a minimum of three different sites, which must be sites with different water characteristics (see 3.5). The evaluation must be done on a water treatment works using the final disinfected water. The following procedure for checking filter types is to be used:

- (i) Connect the filter to a suitable sampling point on the water supply. A flow-meter is connected downstream of the filter unit to measure the volume sampled and a flow restrictor (nominally 1 litre/min at 1 – 10 bar) is fitted downstream of the flow meter. A pump may be fitted on the inlet side of the filter to ensure sufficient inlet pressure (maximum 8 bar). The differential pressure across the filter must be monitored continuously over the 24 hours of each run.
- (ii) On each of 30 days at each of the proposed sites, run water through the filter for a 24 hour period and measure the volume filtered. A fresh filter is used for each 24 hour period. An approved filtration device must also be run in parallel. Negative control filters must also be included on each day of testing.
- (iii) Additionally, at the start and finish of each 24 hour period measure the initial and final flow-through rates by timing how quickly a 250 ml measuring cylinder is filled from the filter outflow.
- (iv) Two additional runs at each site should be undertaken over a 48 hour period.

5.2.2.3 The collection device must be pressure tested to 8 bar water pressure for 1 hour. This shall be repeated on 10 randomly selected devices.

### **5.2.3 *Ability to recover oocysts from water samples***

5.2.3.1 The filters must be able to retain oocysts at a minimum specified recovery after spiking at the limit concentration and sampling a minimum of 1000 litre of water, whilst at the same time removing as much of the interferences as possible. This is tested through the comparison study described below and does not require a separate study.

### **5.2.4 *Modifications to the analytical technique and the detection of oocysts using microscopy***

5.2.4.1 Depending on the modifications eg. alternative IMS, evaluation of a new monoclonal antibody etc., spiked filters and controls should be used to establish the selectivity of the whole method and the detection system. The study should be carried out at a minimum of three different sites which must be sites with different water characteristics (see 3.5). Sites within the distribution system may be used. The following procedure for checking the modifications is to be used:

- (i) Seed the two filters with oocysts, each at the limit concentration;

- (ii) Connect the filters to a suitable sampling point on the water supply using a manifold. A flow-meter is connected downstream of the filter unit to measure the volume sampled and a flow restrictor (nominally 1 litre/min at 1 – 10 bar) is fitted downstream of the flow meter. A pump may be fitted on the inlet side of the filter to ensure sufficient inlet pressure (maximum 8 bar).
- (iii) An additional unseeded filter should be connected to the manifold upstream of the seeded filters to act as a negative control;
- (iv) On each of five days at each of the proposed sites, run water through the filters for a 24 hour period and measure the volume filtered. Fresh filters are used for each 24 hour period.
- (v) Checks should also be made on **five** occasions above and below the limit concentration of oocysts, at 0.5X and 2X the limit concentration, using a reduced study at one site. Paired samples plus controls should be used to assess recovery at each spiking concentration. These multilevel checks do not need to be done on the same day. If recovery with the modified analytical technique appears to be concentration dependent, the study should be repeated using further sets of samples.
- (vi) Analyse the seeded filters using the approved and modified methods. Elute the control filter and analyse half of the eluate by each method.

Note: Most control filter measurements in the spiking studies used for the approved method gave a count of zero. Exceptions were encountered with genuine *Cryptosporidium* oocysts in the water supply used for the tests. This suggests the selectivity, in terms of not giving false positive results, is good. However, *alternative* methods might appear to show positives in their controls. This could happen for three reasons:

- genuine *Cryptosporidium* contamination of the water supply used in the testing;
- inadvertent contamination of the water supply used for the control caused by poor technique during the spiking, sampling or analysis stages;
- bias in the detection system, assuming an alternative detection system is under evaluation.

Any occurrence of oocysts in the controls must be investigated. If it is confirmed that the oocysts have not originated from the water supply, the data must be discarded and the study repeated. If the oocysts are traced to the water supply and their level is  $\leq 10\%$  of the limit concentration, the data may be accepted but corrected for the level in the control. If the oocysts are traced to the water supply and their level is  $>10\%$  of the limit concentration, the data must be discarded and the study repeated.

### **5.2.5 Verification of the use of an approved product against another approved product**

As more products become approved for regulatory use, many will not have been approved against other products that have passed through the above validation procedure. For example, since the SOP part 2 contains more than one approved

filtration device, an IMS bead may have been approved for use with one of the devices through this protocol but not with the other. Under these circumstances approval with the alternative device will not require a full phase 1 and phase 2 validation (see 4.7.1).

Using a single water site, the two filters should be tested in parallel using the following procedure for checking the verification:

- (i) seed the two filters with oocysts, each at the limit concentration. Where a product other than a filter is being tested then two filters of the same type should be seeded.
- (ii) connect the filters to a suitable sampling point on the water supply using a manifold. A flow-meter is connected downstream of each filter unit to measure the volume sampled and a flow restrictor (nominally 1 litre/min at 1 – 10 bar) is fitted downstream of the flow meter. A pump may be fitted on the inlet side of the filter to ensure sufficient inlet pressure (maximum 8 bar).
- (iii) on each of twelve days at the proposed site, run water through the filters for a 24 hour period and measure the volume filtered. Fresh filters are used for each 24 hour period.
- (iv) It is not required that a negative control is run with these filters.
- (v) The data generated must demonstrate that one approved product is equivalent (or better) than the other approved product. 'Equivalent' will be interpreted as either 'significantly better' or as 'not significantly different', but the latter must be to a level of confidence acceptable to DWI. Should the data fail to demonstrate equivalence then the DWI may require more replicates for this to be achieved.

### **5.3 Statistics to be used in phase 1**

- 5.3.1 In phase 1, the laboratory checking the *alternative* or modified approved method needs to establish equivalence between the two methods. The statistics for this together with two worked examples are given in Annex B. It may also be apparent that there is a difference in recovery of oocysts between the water sources used for the validation. This difference may not be apparent between the *alternative* or modified approved method and the SOP.

### **5.4 Documentation**

- 5.4.1 Assuming that at the conclusion of phase 1, the performance of the *alternative* or modified approved method is considered to be sufficiently promising, to warrant proceeding to phase 2, a full report of the phase 1 findings should be submitted to the DWI. Adequate documentation of the procedure will also be required to ensure that other laboratories can perform the procedure with the minimum of interpretation. The standard format for documentation, as used in part 2 of this SOP must be used.

## **6. PHASE 2**

- 6.1 The purpose of phase 2 is to establish the precision and recovery of the whole method that can be expected across a group of laboratories, when applied to samples in which oocysts are present at the limit concentration. These parameters are evaluated by means of an inter-laboratory study. The inter-laboratory comparison will determine:

- within and between laboratory precision under field conditions;
- recovery at the limit concentration under field conditions;
- performance on blank controls under field conditions.

6.2 Phase 2 must consist of an inter-laboratory study, in which at least **five** laboratories repeat the comparison between the *alternative* or modified approved method and the SOP. This applies to new filtration devices, modifications to the analytical technique and microscopy.

### **6.3 Suitability of sampling equipment**

6.3.1 Sampling equipment will be specified from phase 1 with no provision to make variations. Specific study of sampling equipment suitability is not required in phase 2, as a satisfactory study under phase 1 effectively endorses the suitability of the equipment. Participating laboratories are however required to validate the new device and to comment in the report to the DWI on the general performance and usability of the equipment which has been tested.

### **6.4 Validation criteria for the *alternative* method or modified approved method**

6.4.1 The following procedure shall be used:

- (1) The study shall consist of five laboratories, each sampling two different sites supplied from two different sources. The sites may consist of suitable taps in distribution (see 3.5);
- (ii) On each of three days at the proposed sites, two filters seeded at the limit concentration and one unseeded filter are used to sample 1000 litres of water;
- (iii) The recovery of oocysts from each filter is determined;

6.4.2 Different laboratories may not sample the same site or source. The three days used do not need to be consecutive – in fact wider spacing will give a better examination of any time dependent effects. The concentrated material from the control filter must be split into two portions and be examined by both methods. For the validation of new filtration devices, the differential pressure across the filter must be measured continuously for each 24 hour period and the flow through the filter measured at the start and completion of the 24 hour period.

6.4.3 Any occurrence of oocysts in the controls at any site must be investigated. If it is confirmed that the oocysts have not originated from the water supply, the data must be discarded and the study for that site/day repeated. If the oocysts are traced to the water supply and their level is  $\leq 10\%$  of the limit concentration, the data may be accepted but corrected for the level in the control. If the oocysts are traced to the water supply and their level is  $>10\%$  of the limit concentration, the data must be discarded and the study for that site/day repeated. All discarded paired samples and controls must be reported to the DWI.

### **6.5 Phase 2: Equivalence**

- 6.5.1 Individual laboratory recoveries may be analysed by laboratory and by site. Equivalence between the *alternative* or modified approved method and the SOP may be determined for the pooled data for all the trials, between laboratories and between sites.
- 6.5.2 At the conclusion of phase 2, a full report of the findings of the inter-laboratory trial shall be submitted to the DWI.

## **ANNEX A: ABBREVIATIONS, DEFINITIONS AND TERMINOLOGY**

The following definitions are taken from relevant international documents which are listed at the end of this annex. 'Explanatory notes' are specific to parts 3 and 4 of this SOP. Additional informative notes may be found in the source documentation.

### **Approved Laboratory**

A laboratory approved and licenced by the Secretary of State to undertake analysis of regulatory samples and is listed on the DWI *Cryptosporidium* web site.

### **C. parvum**

*Cryptosporidium parvum*.

### ***Cryptosporidium***

This covers all species of the genus *Cryptosporidium*.

### **DWI**

Drinking Water Inspectorate.

### **Equivalent**

The required performance of the trial method is that it should be at least equivalent (or better) than the standard method. 'Equivalent' will be interpreted as either 'significantly better' or as 'not significantly different', but the latter must be to a level of confidence acceptable to DWI. Thus, if the trial method is not finding significantly larger numbers of oocysts then the average difference between results from the two methods should be compatible with the null hypothesis of zero difference (with 95% probability). This statistical analysis will automatically lead to a statement about the likely range of the 'true average difference' between the methods, which is the average you would get from an infinite number of samples of the same types of water. This range is usually expressed as a 95% confidence interval and the lower end of this should rule out unacceptably worse recovery.

### **Fitness for Purpose**

Degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose.

### **Limit Concentration**

The legal concentration of oocysts in drinking water required to be met in the Regulations (less than 1 oocyst per 10 litres of water based on a 1000 litre sample). For the purposes of validation 100 oocysts in 1000 litres using the sampling method in the SOP.

### **Official (Approved) Method**

The method, approved by the DWI, and documented in parts 1 and 2 of this SOP.

**Proposer**

Laboratory or organisation proposing an *alternative* method or changes to the *official* method.

**Recovery**

The number of oocysts added to a test sample prior to analysis, which is recovered by the method.

**Reference Material**

The reference material is a known number of oocysts in suspension either Moredun or Iowa strain of *Cryptosporidium* from an approved supplier who is listed on the DWI web site.

**Selectivity**

The ability of a method to determine accurately and specifically the analyte of interest in the presence of other components in a simple matrix under the stated conditions of the test.

**Standard deviation**

The square root of the variance.

## **ANNEX B: STATISTICAL TREATMENT OF VALIDATION DATA**

This annex discusses statistical aspects of the data analyses and of the presentation of the reports, both for phase 1 and phase 2. In those reports full information on the data and comments on unusual or problematic findings should be provided so that DWI can make a fair assessment of the trials. Statistical methods should be chosen in accordance with the individual study, but reasons for the choice should be given in the report. The methods recommended in this Annex and used in the worked examples are for guidance only.

### ***Statistical analysis of validation data***

Statistical analysis of the data is used to determine that the results produced by the *alternative* or modified method of analysis or equipment is at least as good as or better than the approved method of analysis or approved equipment.

**The number of samples tested and the quality of the data will determine the significance and the confidence of the conclusions drawn. The numbers of sample tests quoted in part 3 of the SOP are expected to be the minimum for demonstrating that the alternative method is probably fit for purpose. If no satisfactory conclusion can be drawn from these results then DWI may request more evidence.**

The advice detailed in this annex should therefore, be regarded as guidance. It is up to the proposer to ensure that the data is produced in the best format and the correct statistical test is applied using the most appropriate statistical package.

The Inspectorate places great importance on the correct statistical analysis of the data because it is the only way to show equivalence of the new method or part of method with the existing approved method. Much time and effort will be saved by spending time determining the correct statistical package to be used.

**NB. It is important to remember that in statistics, numbers should not be rounded up or down. Any rounding off of decimal places must only be undertaken on the result at the end of the statistical analysis. Any rounding of numbers during the statistical analysis could render the result invalid.**

In the paragraphs detailed below is guidance on the statistical tests to be used and suitable statistical approaches for use with these tests. It is up to the proposer to determine the correct statistical test(s) and the most appropriate statistical methods/approach.

### ***Data presentation***

The following format for the presentation of data is recommended and data not in a suitable format for analysis and assessment will be returned to the proposer.

All the raw data should be included in both phase 1 and phase 2 reports, including any data not used in the statistical analysis (outliers).

Data summaries should appear in tables and the key comparisons between the trial and approved methods should be plotted. All tables and figures should be clearly labelled with stand-alone titles and comprehensive footnotes.

Statistical results should be presented in an informative way rather than as raw output from a statistical package. For example, statistics such as averages and test parameters should be given with the number of digits which are significant rather than meaningless strings of

decimal places. (The decimal places are needed during calculations to avoid cumulating inaccuracies. The rounding off is done only for the final answer which is then presented in the report.)

The main analysis should test the null hypothesis that there is no difference in the number of oocysts which the two methods detect. Therefore the numerical difference between paired results should be the key observation analysed. A clear distinction should be made between counts and recovery estimates, and both may have their place. If exact and identical numbers of oocysts were presented to the two methods then the analyses of counts and of recoveries should be equivalent. If there is an unknown variation in numbers of oocysts in the two portions examined by the methods then the recovery 'rates' become ratios of two variables. This introduces complication in the theoretical application of parametric statistical methods. Therefore initial analyses should concentrate on counts.

The null hypothesis of zero average difference in paired counts can be tested by parametric methods (usually the paired *t*-test) or by a non-parametric equivalent (e.g. the sign test). Parametric analysis can be used if the differences are approximately 'Normally' distributed. If the data are not normal, then non-parametric statistics provide a more valid result and are more efficient at detecting true differences between methods.

If, overall, the test method comes out as significantly better or significantly worse then the trial is conclusive, with the proviso discussed below concerning consistency. If there is 'no significant difference' between the two methods then the 95% confidence interval of the estimated average difference must be presented and discussed. If the lower end of this interval implies that the trial method could be very much worse than the approved method then this will affect the conclusions. The researchers may wish to recommend that more samples are examined to increase the power of the study and therefore narrow the confidence interval and clarify the worst-case scenario.

Consistency of the method comparison between sites (in Phase 1) or between laboratories (in Phase 2) should be apparent from scrutinising the plot of the data. Concerns about inconsistencies should be commented upon. Statistical tests (such as one-way analysis of variance of the differences) can be used to test the null hypothesis of consistency, but the planned study sizes are not large enough to detect small differences. Significant inconsistencies must be discussed as they may affect the prudence of approving the new method for general use.

- a) Preliminary data and data complementary to the main comparison, for example data generated by the proposer prior to the phase 1 study.
- b) Tabulations should give details of the data collected and characteristics of the waters from the different sites, giving dates, numbers of oocysts seeded, volumes, controls etc. A table of the numbers of oocysts found by both methods should be given and the differences shown.

The counts from the two methods should be plotted, with distinction made between results from different sites and, where appropriate, levels of contamination. A second plot of recoveries (or estimated recovery ratios) should be drawn. The plots should show the line of equality. Comments on the graphs should be presented.

Summaries of the counts and the differences should be presented (e.g. means, medians, ranges).

A statistical test of the null hypothesis of zero average difference should be done - a paired t-test if the differences are approximately 'Normal' or a non-parametric equivalent (e.g. the sign test) if they are not.

The 95% confidence interval for the estimated average difference must be calculated and reported.

A one-way analysis of variance of the differences to test the null hypothesis of 'same average difference for each site' should be done, or a non-parametric equivalent. If appropriate, for the site with 3 contamination levels, the average difference in results from the two methods can be compared between the levels.

In order to clarify the statistical approach outlined above, two examples follow. The first is fictitious, but uses data adapted from real trials. It illustrates a straight-forward study where no significant difference between two products is demonstrated. The second is of a previously undertaken study comparing two different immunomagnetisable (IMS) beads and illustrates more complex outcomes. Despite much lower recoveries being obtained by both methods at the beginning of the trial, the data has been included in the calculations.

In the worked examples the graphics were produced in Excel, making a black and white dot-plot with the results from the approved method plotted on the x-axis and the paired results by the trial method on the y-axis. The ranges on both axes are the same and the plot area is made approximately square with the diagonal line of equality added. Groups of samples (e.g. from different sites in phase 1 and from different laboratories in phase 2) are identified by using the 'data series' option.

The statistical analyses were done in Minitab. Differences between paired results were calculated and the TTEST and TINT functions were applied to the difference column to provide a paired t-test and 95% confidence interval. Non-parametric equivalents were obtained with the STEST and WINT functions. Minitab also allows one way analysis of variance to compare differences between sites or laboratories when the data are Normal and the Kruskal-Wallis analysis when a non-parametric method is more appropriate.

## EXAMPLE 1: FICTITIOUS TRIAL VERSUS APPROVED METHODS

### Phase 1 - data from the comparison of a trial and approved methods for recovering and enumerating *Cryptosporidium* oocysts - single laboratory study.

Three separate treated water sources were used for the trial. In a full study report, the characteristics of these waters would be described, as would the methodology of the study. This worked example concentrates on the presentation and analysis of results.

Five samples were used from the first two sites, prepared on five different days, and each seeded with 100 oocysts using flow cytometry. Nine samples were used from the third site: three were seeded with 50, three with 100 and three with 150 oocysts. The results are summarised in Table 1.1.

**Table 1.1 Summary of counts from the comparison of a trial method with an approved method using samples from 3 sites. All 19 samples were seeded with known numbers of oocysts**

Site	Oocyst Concentration	Trial Method Count	Approved Method Count	Difference	Recovery (%)	
					Trial	Approved
1	100	44	49	-5	44	49
1	100	52	50	2	52	50
1	100	66	60	6	66	60
1	100	60	49	11	60	49
1	100	49	54	-5	49	54
2	100	38	36	2	38	36
2	100	30	43	-13	30	43
2	100	37	33	4	37	33
2	100	53	30	23	53	30
2	100	44	35	9	44	35
3	50	24	11	13	48	22
3	50	12	22	-10	24	44
3	50	30	30	0	60	60
3	100	65	53	12	65	53
3	100	55	47	8	55	47
3	100	43	52	-9	43	52
3	150	77	83	-6	51	55
3	150	80	90	-10	53	60
3	150	95	70	25	63	47

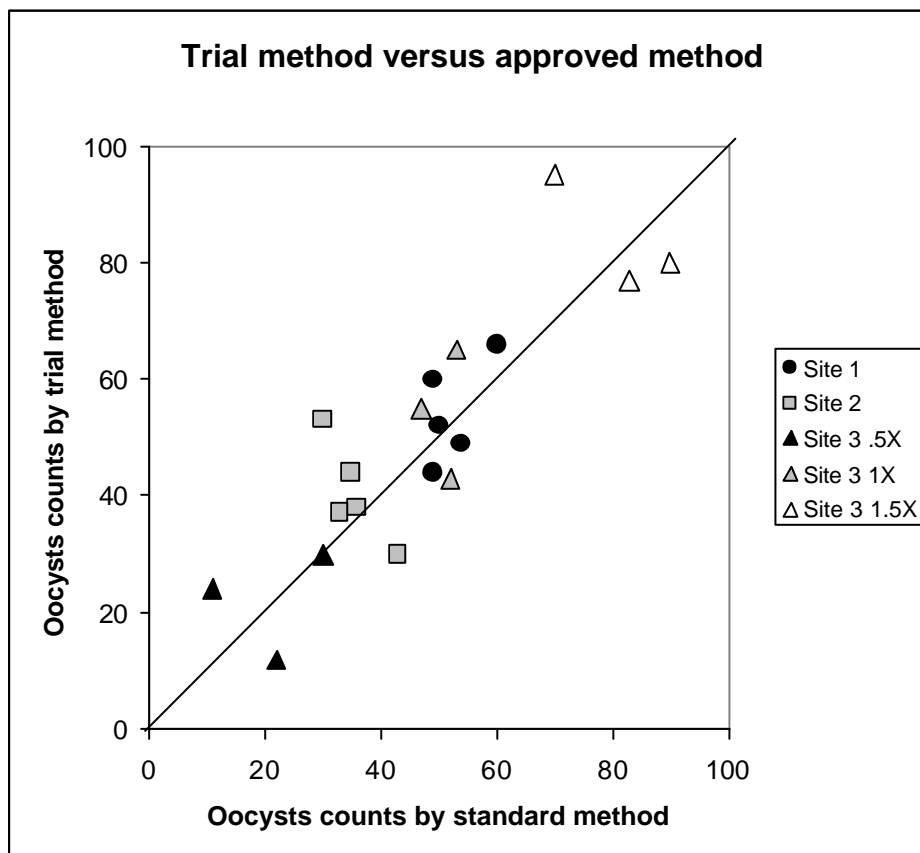
**Table 1.2 Summary of the statistical analysis of all the data from phase 1**

	<b>Trial method</b>	<b>Approved method</b>	<b>Difference</b>
<b>Number</b>	19	19	19
<b>Mean</b>	50.2	47.2	3.0
<b>Median</b>	49	49	2
<b>Standard deviation</b>	20.58	19.64	10.95
<b>Range</b>	12 - 95	11 - 90	-13 - +25

No oocysts were found in any control tests.

Figure 1.1 shows the plot of oocyst counts by the two methods and includes the line of equality. Eleven points lie above the line where the trial method gave the higher count, 7 points lie below where the approved method gave the higher count and one point is on the line where the difference was zero. The differences are widely distributed but there is no obvious non-symmetry which would invalidate use of the t-test, which is in fact robust for small departures from 'Normality'. (For 'non-Normal' data the non-parametric analysis will be more efficient but for approximately 'Normal' data they should be equivalent. Both approaches will be used in this example, for illustration. The statistical package Minitab has been used.)

**Figure 1.1 Comparison of trial and approved methods using oocyst counts (3 sites, 3 levels of seeding at site 3)**



**Figure 1.2 Comparison of trial and approved methods using percentage recoveries (3 sites, 3 levels of seeding at site 3)**

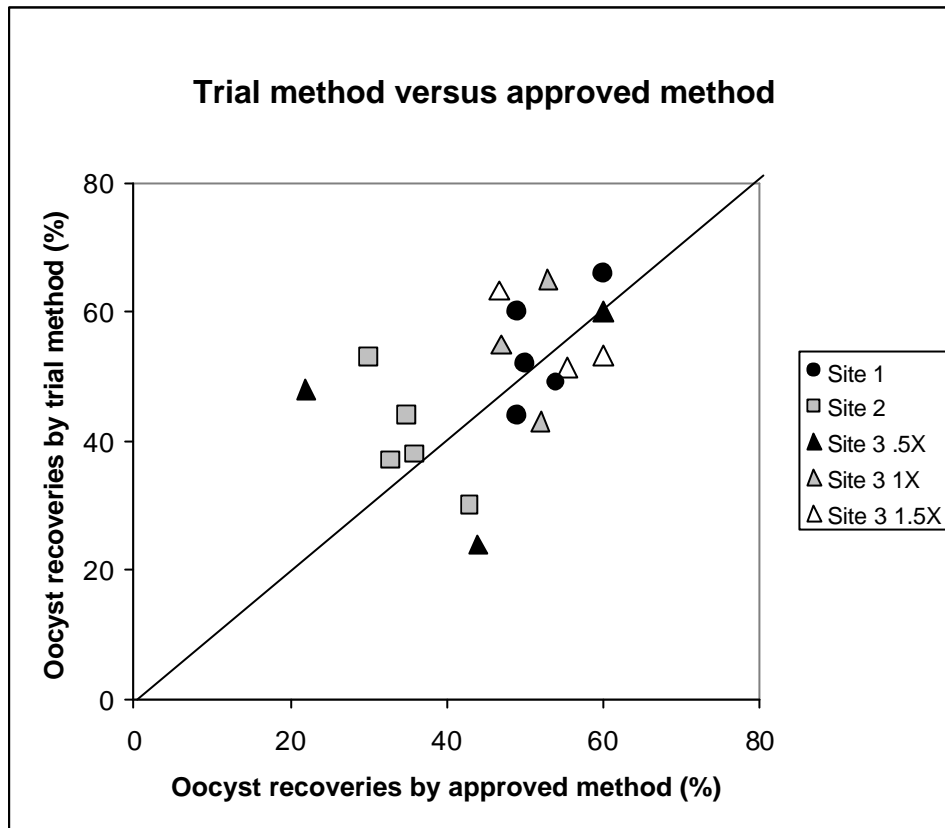


Figure 1.2 shows the plot of recovery percentages which gives similar evidence, except that the scatter of differences looks slightly larger for the three samples seeded with 50 oocysts, but this may be the effect of using lower numbers to seed.

**Testing the null hypothesis of no difference in average counts between the methods for all the sites**

- (i) paired t-test (n = 19):  
mean difference = +3, t = 1.19, p = 0.25
- (ii) sign test of the median:  
median = +2, 11 results were positive and four results were negative, p = 0.48 (which is the probability of tossing a coin 18 times and getting 7 or fewer 'heads' or 7 or fewer 'tails')

**95% confidence for the average difference**

- (i) from the t-test the interval is given by mean  $\pm$  t<sub>95</sub> (s/ $\sqrt{n}$ ), where t<sub>95</sub> is the t-table entry for 95% probability and (n-1) degrees of freedom and s is the standard deviation. This gives:  
-2.3 to +8.3
- (ii) the Wilcoxon confidence interval for the median difference is:

-2.5 to +8.5

This study has shown no overall significant difference between the methods.

### Comparison of the sites

The mean differences for the three sites are:

Site 1 mean = 1.8 (from 5 paired samples)

Site 2 mean = 5.0 (from 5 paired samples)

Site 3 mean = 2.6 (from 9 paired samples)

**Table 1.3 One way analysis of variance between the three sites**

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between sites	2	29	14	0.11	0.90
Residual	16	2129	133		
TOTAL	18	2158			

There is no significant difference between sites. This is also the finding from doing the non-parametric Kruskal-Wallis test ( $p = 0.95$ ).

### Conclusion

The graphs show no obvious difference between the methods, but with slightly more samples where the trial method gave the higher count. The statistical comparison of the methods shows no significant difference.

The 95% confidence interval for the mean difference give the lower end of the range as -2.3 and upper end as 8.3, estimated from 19 samples where the approved method gave a mean of 47.2 oocysts. This suggests an estimated worse-case scenario of the trial method, on average, finding 95% as many oocysts as the standard method (*i.e.* 5% fewer) and the best-case of 118% as many oocyst in similar situations (calculated by  $(47.2-2.3)/47.2$  and  $(47.2+8.3)/47.2$  expressed as percentages).

The findings are very similar for non-parametric methods, because in this example the differences are approximately Normal. Both methods are given for illustration. The 95% confidence interval for the median difference is -2.5 to 8.5 from 19 samples giving a median of 49 oocysts by the standard method. This suggests that the test method would find between 95% and 117% as many oocysts from similar waters.

There is no evidence that the method comparison was affected by the water sites.

This phase was satisfactory and phase 2 went ahead.

**Phase 2 - data from the comparison of a trial and an approved method for recovering and enumerating *Cryptosporidium* oocysts - inter-laboratory study**

Five laboratories took part, each examining three samples from each of two different sources, totalling six samples per laboratory and 30 overall. Each sample was seeded with 100 oocysts. The resulting counts are shown in Table 1.4. (Recovery percentages are the same as the counts, because of the original concentration of 100 oocysts.)

**Table 1.4 Data from the inter-laboratory phase 2 study: counts from the trial and the approved methods from 30 samples each seeded with 100 oocysts.**

<b>Lab (Site)</b>	<b>Approved Method</b>	<b>Trial Method</b>	<b>Difference</b>
1(1)	44	52	8
1(1)	37	39	2
1(1)	55	51	-4
1(2)	55	60	5
1(2)	40	49	9
1(2)	60	50	-10
2(1)	37	68	31
2(1)	38	55	17
2(1)	43	54	11
2(2)	45	47	2
2(2)	50	60	10
2(2)	50	50	0
3(1)	33	38	5
3(1)	35	35	0
3(1)	31	40	9
3(2)	57	49	-8
3(2)	41	57	16
3(2)	63	56	-7
4(1)	50	49	-1
4(1)	41	55	14
4(1)	54	65	11
4(2)	65	44	-21
4(2)	46	66	20
4(2)	45	59	14
5(1)	66	70	4
5(1)	61	65	4
5(1)	59	64	5
5(2)	70	76	6
5(2)	49	59	10
5(2)	68	55	-13

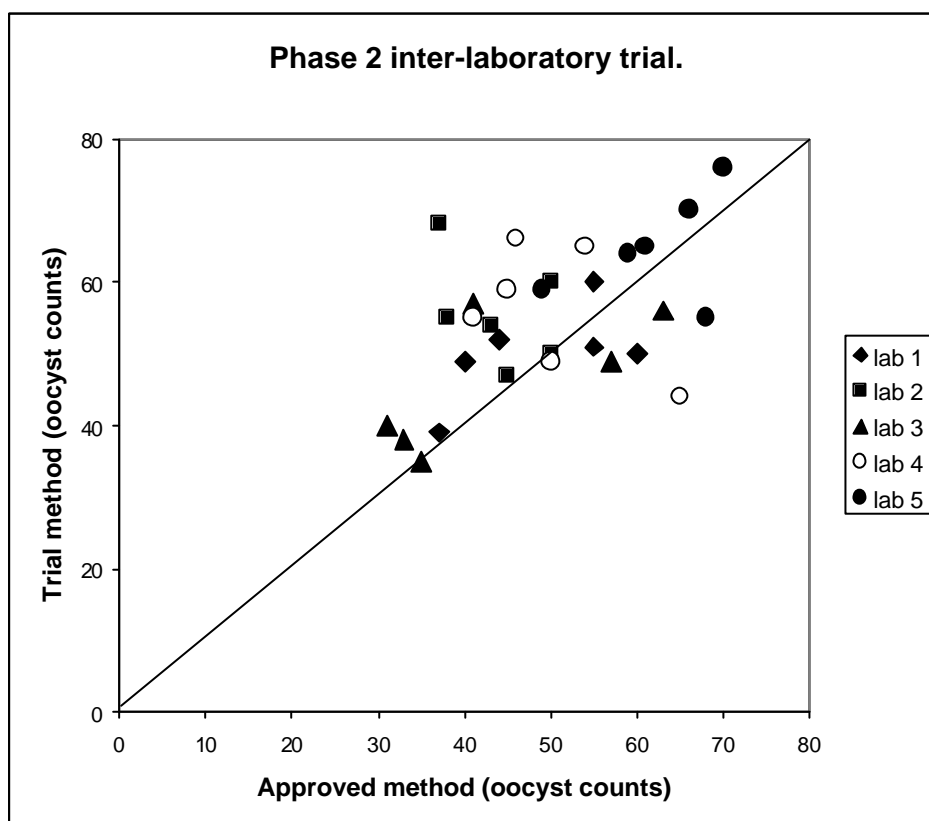
**Table 1.5 Summary of the statistical analysis of all the data from phase 2**

	Approved Method	Trial Method	Difference
<b>Mean</b>	49.6	54.6	5.0
<b>Median</b>	50	55	5
<b>Standard deviation</b>	11.98	9.93	10.53
<b>Range</b>	31 to 70	35 to 76	-21 to +31

No oocysts were found in any control tests.

The data are plotted in Figure 1.3.

**Figure 1.3 Oocysts recovered during the phase 2 inter-laboratory trial plotted as counts**



This shows the plot of the oocyst recoveries by the two methods and includes the line of equality. Twenty-one points lie above the line, where the trial method gave the higher count and 7 lie below the line where the approved method gave the higher count. Two points lie on the line, where the difference in counts was zero.

The differences are widely distributed but there is no obvious non-symmetry which would invalidate use of parametric statistics.

**Testing the null hypothesis of no difference in average counts between the methods for all sites:**

Paired t-test (n = 30)

Mean difference = +5.0, t = 2.58, p = 0.015

Thus, overall, there were significantly higher counts on average from the trial method as compared with the approved method. This is confirmed by looking at the:

**95% confidence interval for the average difference**

From the t-test, the interval is given by mean  $\pm t_{95} (s/\sqrt{n})$ , where  $t_{95}$  is the t-table entry for 95% probability and (n-1) degrees of freedom and s is the standard deviation.

This gives +1.03 to +8.90, which lies entirely above zero.

It is important to look at the individual laboratories to see whether there is agreement. Figure 1.3 shows that four laboratories had a mixture of points above and below the line (with the majority of samples above, in all cases) but laboratory 2 had no point below the line - with five samples giving the higher count by the trial method and the sixth giving zero difference.

Individual t-tests of each set of six results show:

Lab 1: mean difference = 1.7, t = 0.5, p = 0.6

Lab 2: mean difference = 11.8, t = 2.6, p = 0.05

Lab 3: mean difference = 2.5, t = 0.7, p = 0.5

Lab 4: mean difference = 6.2, t = 1.0, p = 0.4

Lab 5: mean difference = 2.7, t = 0.8, p = 0.5

All the laboratories showed a positive average difference but this was not significantly different from zero, except for lab 2 where the probability of the null hypothesis was exactly 5%.

**Table 1.5. One way analysis of variance between the five laboratories**

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between labs	4	425	106	0.95	0.45
Residual	25	2792	112		
TOTAL	29	3217			

There is no significant difference between laboratories.

It can be seen from Figure 1.3 that there was variation in numbers of oocysts recovered from some sites. In particular laboratory 3's first site gave low recoveries for all three samples but this did not affect the comparison between methods.

**Conclusion**

The phase 2 study showed an overall significantly higher recovery of oocysts from the trial method as compared with the approved method. None of the participating laboratories was in significant disagreement.

**EXAMPLE 2: STUDY OF ‘TRIAL’ VERSUS ‘APPROVED’ IMMUNOMAGNETISABLE BEADS**

**Phase 1 - data from the trial method IMS-1 and approved method IMS-2 for the recovery of *Cryptosporidium* oocysts - single laboratory trial**

Three treated waters were used for the phase 1 trial. These were an upland river source, a lowland river source and a borehole. Treated water was sampled in distribution as opposed to at water treatment works using paired Idexx Filita-Max<sup>R</sup> filters. Four paired samples were used at the first site two pairs seeded with 78 oocysts and the other two seeded with 124 oocysts. Five paired filters were used at the second site, each seeded with 124 oocysts and four filters were used at the third site each seeded with 124 oocysts. A single additional filter was used as the negative control. All the negative controls were zero. Oocyst concentrations were determined by staining 10 x 20 µl of the seed suspension in accordance with SOP 2. The studies were conducted in 2000.

The results for phase 1 are summarised in Table 2.1.

**Table 2.1 Summary of counts from the comparison of the trial method IMS-1 with the approved method IMS-2 using samples from 3 sites using filters seeded with known numbers of oocysts**

Site	Oocyst Concentration	IMS-1 Recovery	IMS-2 Recovery	Difference	Recovery (%)	
					IMS-1	IMS-2
1	78	15	15	0	19	19
1	78	18	24	-6	23	31
1	124	71	67	4	57	54
1	124	71	56	15	57	45
2	124	89	59	30	72	48
2	124	79	62	17	64	50
2	124	71	67	4	57	54
2	124	77	62	10	62	54
2	124	99	62	37	80	50
3	124	64	69	-5	52	56
3	124	75	72	3	60	58
3	124	43	62	-19	35	50
3	124	63	67	-4	51	54

**Table 2.2 Summary of the statistical analysis of all the data from phase 1**

	IMS-1	IMS-2	Difference
<b>Mean</b>	64.2	57.6	6.6
<b>Median</b>	71	62	4
<b>Standard deviation</b>	25.0	17.6	15.3
<b>Range</b>	15 to 99	15 to 72	-19 to +37

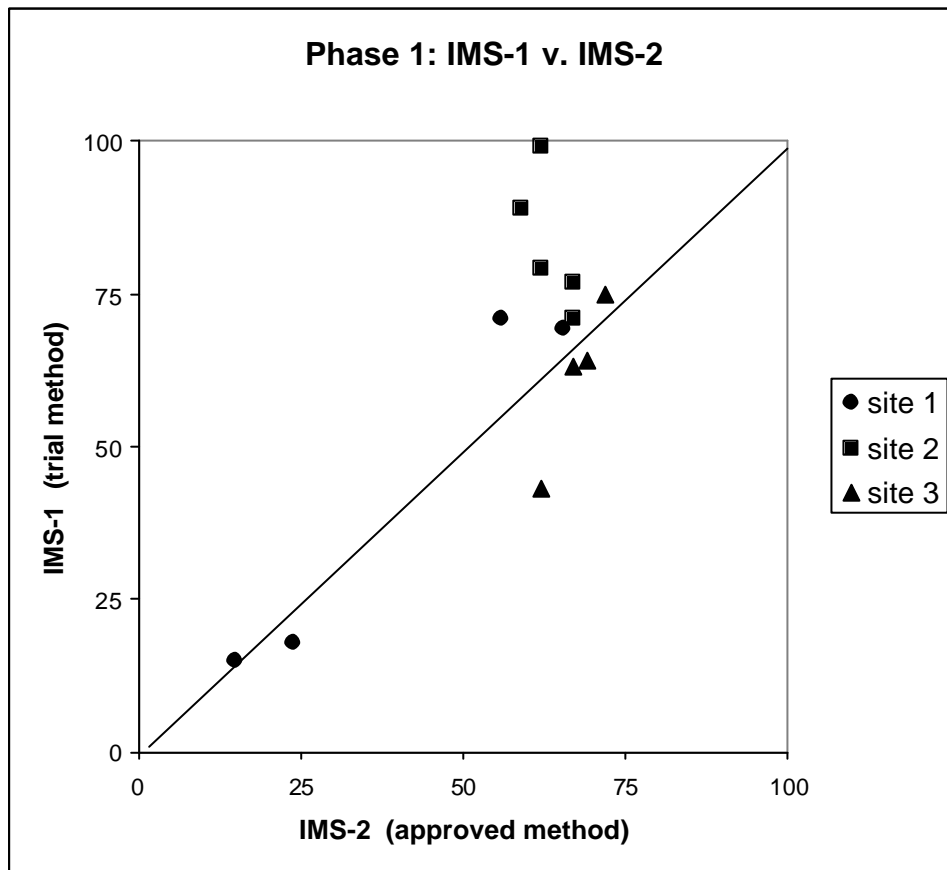
The data has also been plotted in Figure 2.1. This shows the plot of the oocyst counts by the two methods and includes the line of equality. Eight points lie above the line

where the test method (IMS-1) gave higher counts and four points lie below the line where the approved method gave higher counts. One point is on the line where the difference in counts was zero.

The differences (IMS-1 - IMS-2) are widely distributed but there is no obvious non-symmetry which would invalidate the use of the t-test, which is, in fact, robust for small departures from 'Normality'. (For non-normal data the non-parametric analysis will be more efficient but for approximately normal data they should be equivalent). Both approaches have been used for this example, for illustration. The statistical package Minitab has been used in this example.

The data plotted in Figure 2.1 is the individual counts. Plots of the percentage recoveries gives a similar picture.

**Figure 2.1 Oocysts recovered during the phase 1 trial**



**Testing the null hypothesis of no difference in average counts between the methods for all the sites**

- (i) paired t-test (n = 13):  
mean difference = +6.6, t = 1.56, p = 0.14
- (ii) sign test of the median:  
median = +4, 8 results were positive and four results were negative, p = 0.38

### 95% confidence for the average difference

- (i) from the t-test the interval is given by  $\text{mean} \pm t_{95} (s/\sqrt{n})$ , where  $t_{95}$  is the t-table entry for 95% probability and (n-1) degrees of freedom and s is the standard deviation. This gives:  
-2.6 to +15.8
- (ii) The Wilcoxon confidence interval for the median difference is:  
-2 to +16.5

### Comparison of the sites

Figure 1 appears to show inconsistency between the sites in that site 2 has all five samples above the line, *i.e.* there was a higher recovery of oocysts with IMS-1. The other two sites had a mixture of positive and negative differences, with site 3 having three of the four points below the line.

The average difference (IMS-1 - IMS-2) in counts for the three sites are:

- Site 1 mean = 3.3, median = 2 (from four paired samples)
- Site 2 mean = 19.6, median = 17 (from five paired samples)
- Site 3 mean = -6.3, median = -4.5 (from four paired samples)

**Table 2.3. One way analysis of variance between the three sites**

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between sites	2	1550	775	6.25	0.017
Residual	10	1243	124		
TOTAL	12	2793			

There is therefore significant difference between the sites. This is also the finding from the non-parametric Kruskal-Wallis test. If the three sites are looked at separately, site 2 shows significantly higher counts by IMS-1 than by IMS-2. (paired t-test,  $t = 3.19$ ,  $p = 0.03$ ). Sites 1 and 3 show no significant difference from zero average difference (site 1,  $p = 0.5$ , 95% confidence interval for the mean difference is -11 to + 17; site 3,  $p = 0.3$ , 95% confidence interval for the mean difference is -21 to + 8)

### Conclusion from the phase 1 study

The graph shows more samples where the test method gave the higher count. The statistical comparison overall of the methods shows no significant difference with a 95% c.i. for the mean difference of -2.6 to +15.8. This lower bound implies a worst case of IMS-1 giving mean counts of  $(2.6/57.6) \times 100$  percent worse than IMS-2 and at best  $(15.8/57.6) \times 100$  per cent better: a range of - 4.5% to + 27%.

However there is statistically significant variation between sites from which the samples were taken. This might not affect the overall conclusion that IMS-1 was as good or better than IMS-2 if there can be confidence that it was not significantly worse at any particular site. In fact, with only four samples from two of the sites it is unlikely that significant differences, even if they were large, would be detected.

So at this stage the evidence for IMS-1 looked promising and Phase 2 went ahead. There may be some question mark about how IMS-1 compares with IMS-2 with the sample types coming from Site 3. In this scenario, DWI might ask for more samples to be taken from site 3 to establish whether there is a significant difference at this site.

**Phase 2 - data from the trial method IMS-1 and approved method IMS-2 for the recovery of *Cryptosporidium* oocysts - inter-laboratory study**

Five laboratories, each using two sites for source water took part. Each laboratory examined six paired samples, three from each of their sites. The results are summarised in Table 2.4.

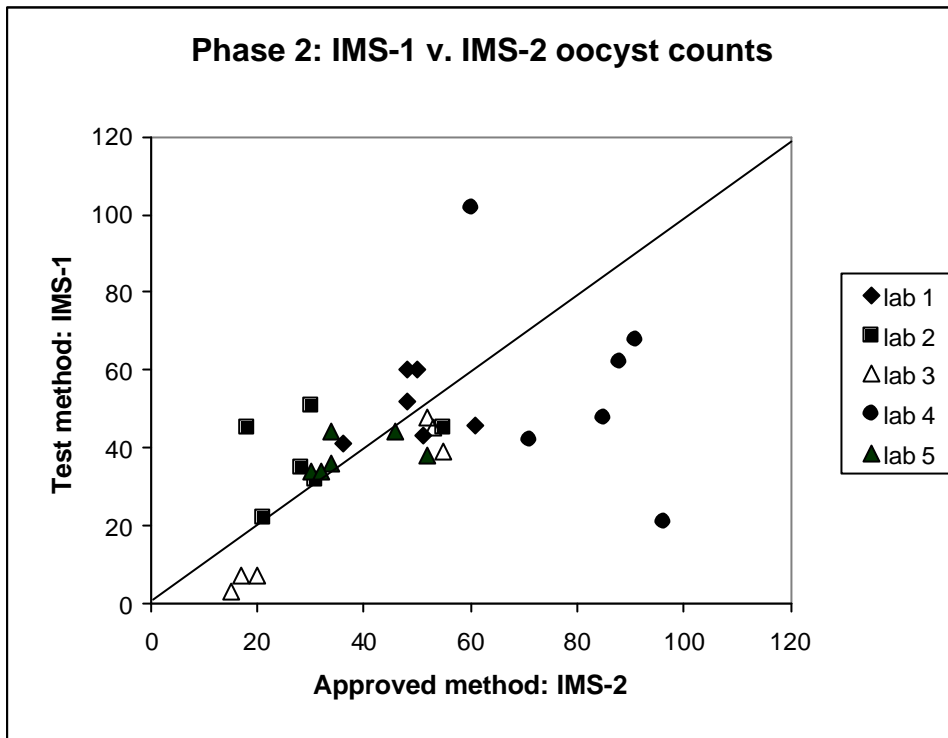
**Table 2.4 Data from the inter-laboratory trial**

Lab (Site)	Oocyst Concentration	IMS-1 Recovery	IMS-2 Recovery	Difference	Recovery (%)	
					IMS-1	IMS-2
1(1)	94	60	50	10	64	53
1(1)	94	60	48	12	64	51
1(1)	94	46	61	-15	49	65
1(2)	94	52	48	4	55	51
1(2)	94	43	51	-8	46	54
1(2)	94	41	36	5	44	38
2(1)	86	22	21	1	26	24
2(1)	86	45	18	27	52	21
2(1)	86	35	28	7	41	33
2(2)	86	45	55	-10	52	64
2(2)	86	51	30	21	59	35
2(2)	86	32	31	1	37	36
3(1)	113	48	52	-4	42	46
3(1)	113	39	55	-16	35	49
3(1)	113	45	53	08	40	47
3(2)	113	3	15	-12	3	13
3(2)	113	7	17	-10	6	15
3(2)	113	7	20	-13	6	18
4(1)	113	21	96	-75	19	85
4(1)	113	102	60	42	90	53
4(1)	113	48	85	-37	42	75
4(2)	113	62	88	-26	55	78
4(2)	113	68	91	-23	60	81
4(2)	113	42	71	-29	37	63
5(1)	98	34	30	4	35	31
5(1)	98	36	34	2	37	35
5(1)	98	34	32	2	35	33
5(2)	98	38	52	-14	39	53
5(2)	98	44	46	-2	45	47
5(2)	98	44	34	10	45	35

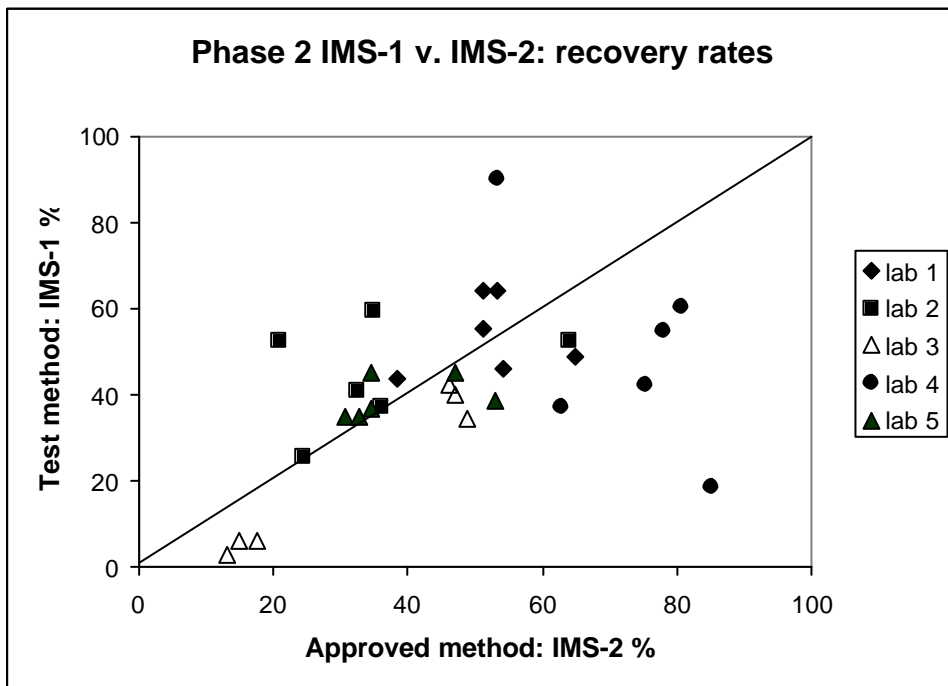
Figure 2.2 shows a plot of oocyst counts by the two methods and includes the line of equality. Fourteen points lie above the line where the test method IMS-1 gave the higher count and 16 points lie below the line where the approved method IMS-2 gave the higher count. The scatter of results for laboratory 4 appears larger than for the other laboratories.

Recovery percentages are also shown in Table 2.4 and are plotted in Figure 2.3. Figure 2.3 gives similar evidence to Figure 2.2.

**Figure 2.2** Oocysts recovered during the phase 2 inter-laboratory trial plotted as counts



**Figure 2.3** Oocysts recovered during the phase 2 inter-laboratory trial plotted as percentages



**Table 2.5 Summary of the statistical analysis of all the data from phase 2**

	<b>IMS-1 counts</b>	<b>IMS-2 counts</b>	<b>Difference</b>	<b>IMS-1 % Recovery</b>	<b>IMS-2 % Recovery</b>
<b>Mean</b>	41.8	46.9	-5.1	42	46
<b>Median</b>	43.5	48	-3	42	7
<b>Standard deviation</b>	19.3	22.5	21.2	18.2	19.4
<b>Range</b>	3 to 102	15 to 96	-75 to 42	3 to 90	13 to 85

**Testing the null hypothesis of no difference in average counts between the methods**

- (iii) paired t-test (n = 30):  
mean difference = -5.1, t = -1.33, p = 0.19
- (iv) sign test of the median:  
median = +4, 14 results were positive and 16 were negative, p = 0.8

**95% confidence interval for the average difference**

- (iii) from the t-test the interval is given by mean  $\pm t_{95} (s/\sqrt{n})$ , where  $t_{95}$  is the t-table entry for 95% probability and (n-1) degrees of freedom and s is the standard deviation. This gives:  
-13.0 to +2.8
- (iv) the Wilcoxon confidence interval for the median difference is:  
-11 to + 2.5

**Comparison of the laboratories**

Figure 2.2 appears to show inconsistency between the laboratories. Four of the five laboratories have a mixture of points above and below the line of equality and none of these four shows a significant difference between the methods (either by t-test or by sign test). Laboratory 3 has all six samples giving higher counts by IMS-2 and therefore their points in the figure are all below the line. This laboratory found significantly higher counts by IMS-2 than by IMS-1 (\* see below in Table 2.6).

**Table 2.6 Average differences (IMS-1 - IMS-2) in counts for the five laboratories**

<b>Lab No.</b>	<b>Mean</b>	<b>Median</b>	<b>Paired Samples</b>	<b>p-value from t-test</b>	<b>95% c.i.</b>
1	1.3	4.5	6	0.77	-10 to + 12
2	7.8	4	6	0.22	- 7 to + 22
3	- 10.5	- 11	6	0.001	- 15 to - 6*
4	- 24.7	- 27.5	6	0.17	- 64 to + 16
5	0.3	2	6	0.92	- 8 to + 9

**Table 2.7 One way analysis of variance between the five laboratories**

Source	df	Sum of Squares	Mean Square	F-value	p-value
<b>Between sites</b>	4	3901	975	2.69	0.055
<b>Residual</b>	25	9080	363		
<b>TOTAL</b>	29	12981			

This shows that the significant difference between laboratories is not quite significant, with the probability being 0.055, slightly greater than the arbitrary 0.05. The corresponding non-parametric Kruskal-Wallis test does show significant difference between laboratories ( $p=0.02$ ). Because of the varying scatter of observations from laboratories the non-parametric analysis may be more reliable. The data has already established that laboratory 3 found IMS-2 to be significantly better whereas the other laboratories found no significant difference between the methods.

### **Conclusion**

Figure 2.2 shows slightly more samples where the test method gave the lower count. The statistical comparison overall of the methods shows no significant difference with a 95% c.i. for the mean difference of  $-13$  to  $+3$ . This lower bound implies a worst case of IMS-1 giving mean counts of  $(13.0/46.9) \times 100$  percent worse than IMS-2 and at best  $(2.8/46.9) \times 100$  per cent better: a range of  $-28\%$  to  $+6\%$ .

However there is statistically significant variation between laboratories and for one laboratory the test method IMS-1 came out as significantly worse than the approved method IMS-2. In the others there was no significant difference. This makes it difficult to draw a conclusion which can be safely extrapolated to other laboratories. Under these circumstances, DWI may request that further samples are analysed by laboratory 3 to determine whether there is a significant difference.