Centre for Environment Fisheries & Aquaculture Science

Cefas

Protocol CV019/P002

Protocol for validation of methods for quantifying micro-organisms in wastewater using nucleic acid amplification-based techniques

Version 1.1

Author: David Walker

November 2022

© Crown copyright 2022

This information is licensed under the Open Government Licence v3.0. To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/

www.cefas.co.uk

Cefas Document Control

Submitted to:	Matthew Wade (UKHSA), Andrew Zealand (UKHSA) and Robin Cutler (Defra)
Date submitted:	16/09/2022
Project Manager:	Jane Heywood
Report compiled by:	David Walker
Quality control by:	James Lowther & Mickael Teixeira Alves
Approved by:	Michelle Price-Hayward
Version:	1.1

Version control history

Version	Author	Date	Comment
0.1	David Walker	09/08/2022	First draft
0.2	James Lowther	12/08/2022	Comments on v0.1
0.3	Mickael Teixeira Alves	25/08/2022	Comments on v0.1
0.4	David Walker	07/09/2022	Corrections from v0.2 and v0.3
0.5	James Lowther	08/09/2022	Comments on v0.4
0.6	David Walker	08/09/2022	Corrections from v0.5
1.0	David Walker	16/09/2022	Final version integrating all comments received and final typesetting.
1.1	David Walker	07/11/2022	Added minor edits as suggested by Defra

Contents

1. Introduction 5
1.1. Scope of this document 5
1.2. The need for method validation 5
1.3. Criteria to evaluate in a validation study 7
1.4. A note about wastewater definitions and variability 8
1.5. A note about qNAA inhibition 9
2. Terms and definitions 11
3. Specification of requirements 14
3.1. Intended use of the Method 14
3.2. Performance criteria 14
4. Technical protocol for validation 16
4.1. General considerations 16
4.2. Determination of precision characteristics 18
4.3. Linearity 20
4.4. Recovery 21
4.5. Impact of Variable Matrix 23
4.6. Selectivity 24
4.7. Limit of detection 27
4.8. Limit of quantification 30
5. Validation report 31
6. References 32

1. Introduction

1.1. Scope of this document

This document has been prepared to provide guidance on the validation of new methods for quantifying micro-organisms (including viruses, bacteria, fungi and protists) in wastewater by quantitative nucleic acid amplification (qNAA) based methods such as qPCR, dPCR (and their variants) and isothermal methods. It is intended that this guidance will be followed by laboratories that intend to produce long-term wastewater data for surveillance programmes. This guidance document covers validation studies conducted by a single laboratory (intralaboratory rather than interlaboratory validation), using a conventional study design (rather than a factorial study design). At the time of writing this guidance, there is no validated reference method available for this type of analysis, and so it is assumed that the validation study will not be conducted in comparison with a reference method. Future versions of this guidance should be produced that include comparisons against a reference method when one exists.

While no specific International Organization for Standardization (ISO) standards currently exist for microbial surveillance of wastewater using qNAA based methods, an ISO standard for determination of SARS-CoV-2 in wastewater using PCR is in development which may cover method validation. Much of the information in that ISO standard is likely to be of relevance to other target micro-organisms. However, in the current absence of a specifically relevant ISO standard, the basic principles of method validation for other matrices are mostly covered by existing ISO standards. Therefore, this document has adopted the principles laid out in existing ISO standards where possible so that they can specifically be used to validate methods for quantifying micro-organisms in wastewater by qNAA.

This guidance should be used to develop a validation protocol for new methods intended for producing data for wastewater surveillance programmes. This document does not cover methods that are already validated, and which need to be adopted into a new laboratory; in this case laboratories will need to use a method verification protocol that will allow laboratories to demonstrate that they are able to carry out a validated method within the performance characteristic parameters as defined in the validation study. A protocol for method verification with specific guidance for wastewater methods may be produced in the future. However, in the current absence of such specific guidance on method verification, laboratories should refer to ISO 16140-3 (ISO, 2021a), or the UK Health Security Agency (UKHSA) protocol for the Verification and Validation of Methods (UKHSA, 2022).

This guidance document assumes that the qNAA element within the Method has been validated against the requirements of ISO 20395 (ISO, 2019a), if undertaken before the publication of ISO 16099 (ISO, 2022) (still under development at the time of writing this guidance). This is necessary to ensure that the target can be reliably quantified by the Method and is stipulated in ISO 16140-4:2020 section 4.1 (ISO, 2020a).

The output of any method validation study will be a study report that outlines the performance characteristics for the Method and provides a recommendation as to whether the Method has been demonstrated to perform within the expected requirements, or whether further method development is needed before the Method is adopted.

1.2. The need for method validation

When a new analytical method is developed, it usually undergoes a period in which new developments are researched and implemented in order to improve the method incrementally. During this period, it may be difficult to compare analytical results from one set of samples to

another if the methodology has been changed between analyses. It is therefore important that when developing a long-term surveillance programme, the analytical methods remain stable, and that any impact on the results brought about by necessary methodological changes is well understood. It is also important that the results generated by a method can be trusted and relied upon to make sound decisions based on scientifically robust evidence. In order to achieve this, it is vital to understand how well the method performs against the expectations of the requirements for the data that it must produce. This means that the method must be studied so that a set of performance characteristics can be created. The performance of the method within a laboratory can then be continually monitored to ensure that the results can be relied upon.

The International Organization for Standardization (ISO) defines method validation as “establishment of the performance characteristics of a method and provision of objective evidence that the performance requirements for a specified intended use are fulfilled” (ISO, 2016a). In order to carry out a method validation, a study must be designed, and investigations carried out to demonstrate that the method yields repeatable, consistent results and establish the performance characteristics of the method. The performance characteristics of a method may be predefined based on existing precedent or knowledge. Alternatively, where a method or field of study is still relatively new, and no precedents exist, a method validation study may be used to demonstrate what those characteristics are. Future validation studies either for new methods, or new applications of the validated method, can then be evaluated against those established performance characteristics.

Statutory monitoring (monitoring of an analyte as required by law) requires a reference method which is used as the standard by which specific legal definitions can be derived for writing legislation. There is currently no legal framework for Wastewater Based Epidemiology (WBE) in the UK. However, if there was a need for legislation in the future, then it is vital that validated reference methods exist. In the area of food and feed law, the requirements for official controls are set out in Retained Regulation (EU) 2017/625 (Regulation (EU) 2017/625 of the European Parliament and of the Council, 2017), which covers food and feed law, rules on animal health and welfare, plant health and plant protection products. A key requirement of this legislation is that official control laboratories must be accredited to use analytical methods according to ISO/IEC 17025 (ISO, 2017c). In turn a key requirement of ISO/IEC 17025 is that methods used by testing laboratories must be validated.

Even where there is no legislative requirement for monitoring or surveillance, the data generated by analytical techniques are often used to enable rational, evidence-based decision making by organisations from all sectors. It is therefore important that any analytical method that is being used to generate data for these purposes is validated accordingly.

It should be noted that while it is important to consider the variability in laboratory methods for surveillance of micro-organisms in wastewater, other sources of uncertainty exist externally to the laboratory. The sources of uncertainty and variability in wastewater surveillance were reviewed by (Wade et al., 2022). These sources of uncertainty and variability may have a greater over all impact to surveillance than variability in the method, and so they are important to address for wastewater surveillance programmes in addition to variability introduced by the laboratory methods. However, this guidance only deals with the variability introduced by laboratory methods.

1.3. Criteria to evaluate in a validation study

According to Retained Regulation (EU) 2017/625, which sets out the law for official control testing for food and feed law in the UK, analytical methods should be characterised by the following criteria (the definitions of which are shown in Terms and definitions):

- Accuracy (trueness and precision)
- Applicability (matrix and concentration range)
- Limit of detection
- Limit of quantification
- Precision
- Repeatability
- Reproducibility
- Recovery
- Selectivity
- Sensitivity
- Linearity
- Measurement uncertainty
- Other criteria that may be selected as required

It should be noted that there is some overlap between these criteria. For example, repeatability and reproducibility are both measures of precision. Applicability is not a numerical measure, but rather a definition of which types of samples a method is to be validated against. Additionally, for methods that require a high degree of sample preparation prior to analysis, recovery and trueness are closely interlinked. For the purposes of this guidance, trueness and recovery will be treated as the same measure. This is because it is assumed that there is no exiting reference method to which to compare trueness to the method under consideration. Therefore, the only measure of the trueness of the sample result that is available at this point is the relative recovery of target analyte from a spiked sample. This is discussed further in section 4.3.

The term "Sensitivity" is often mistakenly used interchangeably with "Limit of Detection" and has several definitions depending on the context it is used. In the context of foodborne microbiology (and therefore the context of the legislation in question), sensitivity may be defined as the "number of samples found to be positive divided by the total number of samples tested at a given level of contamination" (ISO, 2017b). However, this is specific to quantitative methods and in the case of the analytical sensitivity of a qNAA based method, the sensitivity can be defined as the slope of the calibration curve that is used to calculate the concentration of the target analyte from the data generated by the qNAA equipment. This parameter will form part of the validation of the qNAA method itself, which must be validated prior to carrying out the procedures laid out in this guidance. During a validation study that is designed according to this guidance document, it is important therefore that any data generated conforms to the performance characteristics of the validated qNAA method (e.g. amplification efficiency, and linearity of any standard curves).

It should also be noted that not all of these criteria are relevant to all methods, and so the specific criteria to be evaluated must be chosen based on relevance to the method under investigation. For those micro-organisms that are expected to always be present at relatively high levels in wastewater (such as faecal indicator organisms), it may not be necessary (or possible) to characterise the limit of detection for the target analyte. However, in the case of surveillance of specific pathogens or other micro-organisms which may not occur ubiquitously within wastewater (such as SARS-CoV-2 and *Salmonella*), the working assumption is that the

aim is to be able to detect and quantify very low levels of the target analyte. This type of analysis is often referred to as "Trace Analysis" and requires the characterisation of the limit of detection and limit of quantification.

Additionally, not all of these criteria are relevant to methods in use by single laboratories. The criteria listed above are for a reference method, which will be used by multiple laboratories. However, this guidance is for validation of methods for use in a single laboratory. Reproducibility is a measure of the variability (precision) in results between laboratories using the same method. It will therefore not be covered in this guidance. However, the principles used for measuring intermediate precision for a single laboratory (section 4.2), will be similar to measuring inter-laboratory reproducibility.

Evaluation of measurement uncertainty is an important component of verifying a method's performance in a laboratory. Measurement uncertainty should be built into the reporting of results generated by the Method. However, the requirements for measuring measurement uncertainty will be dependent on individual methods under investigation and so will not be covered in this guidance. Evaluation of measurement uncertainty should follow the ISO/JCGM guidance on the topic (JCGM, 2008). Users of this guidance may also find guides on the interpretation of the ISO/JCGM guidance useful (Barwick & Ellison, 2005).

An additional parameter to evaluate for the purposes of wastewater is the impact of the variation in the wastewater matrix. This is discussed further in section 1.4. For the purposes of this guidance, this with be referred to as the Impact of Variable Matrix, which is distinct from the Matrix Effect which is often measured in other validation studies.

For the purposes of this guidance, the list of criteria that the method will be validated against can be simplified to:

- Precision (including repeatability and intermediate precision)
- Linearity
- Recovery
- Impact of Variable Matrix
- Selectivity
- Limit of detection
- Limit of quantification

1.4. A note about wastewater definitions and variability

The definitions of the words "wastewater" and "sewage" are different in different contexts. The word "wastewater" has been adopted for the matrix that is evaluated for the purposes of tracking infectious diseases in human populations in a field broadly known as Wastewater Based Epidemiology (WBE). On the other hand, according to ISO 24513 (ISO, 2019b), wastewater is "water arising from any combination of domestic, institutional, commercial or industrial activities, surface runoff and any accidental sewer inflow/infiltration water and which can include collected stormwater, discharged to the environment or sewer". This definition does not explicitly require that wastewater contains faecal material, which is usually a key requirement of the matrix used for WBE.

Sewage on the other hand is defined in ISO 8099-1 (ISO, 2018) as "human body wastes and the wastes, including flushing water, from toilets and other receptacles intended to receive or retain these wastes". For some contexts, this definition may not be satisfactory especially where no sewerage infrastructure is in place.

For the purposes of this document, the term wastewater refers to water containing high concentrations of faecal and/or non-faecal bodily wastes discharged from residential, commercial and industrial premises. This water may be sampled at any point in a sewerage network, or outside of a sewerage network where such infrastructure is lacking.

However, samples of wastewater taken from different points within a sewerage network or beyond will likely have very different properties and so a single method for quantifying micro-organisms is unlikely to be suitable for all types of wastewater sample. For example, near-source wastewater samples are likely to be relatively un-homogenised, while influent wastewater samples are likely to be homogenised to a much greater degree. For this reason, it should be made clear during the development of the validation plan, what type of wastewater the Method is intended for.

Another major source of variability in wastewater is the changing composition of the matrix both temporally in the same sampling location, and spatially across multiple sampling locations (e.g. multiple sewage treatment works across different sewerage catchments). The composition of wastewater taken from a single sampling location will vary depending on the inputs received to a sewerage system, and this varies greatly throughout the day and throughout the week. For example, faecal input is likely to be low during the night when most people are sleeping but is likely to increase during the day when most people are awake. The composition of one sewerage catchment's sewage is also likely to be different from another catchment's due to the differences in other inputs such as industrial waste and the relative contribution of combined sewers vs. foul sewers.

It is not feasible to expect any new wastewater method to be fully validated for every wastewater composition possible. It is nonetheless important to consider this level of complexity associated with wastewater when designing the validation study. Most of the experiments detailed in this guidance use artificial samples created with a negative matrix sample spiked with reference material. To represent the complexity of wastewater, the negative matrix samples must be taken from as broad a range of locations as possible. Details of how to use these negative matrix samples effectively are outlined in section 4.1.1.

1.5. A note about qNAA inhibition

Due to the nature of the wastewater matrix, there is a possibility that sample analysis may be complicated by inhibition of qNAA due to the presence in the matrix of chemicals that interfere with the enzymes used in the amplification. It should be noted that some forms of qNAA are less susceptible to inhibition than others, and so may be less important to consider in some cases than others. However, inhibition of this type may lead to falsely low or negative results for individual samples or may increase the level of variability (i.e. reduce precision) to unacceptable levels. For this reason, all assays that are validated for quantification of micro-organisms in wastewater must include controls for inhibition, unless a lack of inhibition has been demonstrated during the qNAA validation. These controls may include internal amplification controls (i.e. extraneous nucleic acid added to each qNAA and tested in multiplex in parallel with the target), external amplification control (i.e. excess target or extraneous nucleic acid added to separate qNAA reactions containing sample nucleic acids and tested in parallel with the unspiked sample nucleic acids in separate reactions) or dilution of sample nucleic acids (i.e. qNAA is carried out on both diluted and undiluted sample nucleic acids).

In some cases, it may be impossible to completely eradicate qNAA inhibition. In those cases, it would be impossible to validate a qNAA according to the guidance in ISO 20395. It may therefore be acceptable to establish a quality control criterion for qNAA inhibition. For example,

ISO 15216-1-2017 (ISO, 2017a) uses a quality threshold of 75% inhibition. In this case, qNAA reactions found to have >75% inhibition are rejected and the samples retested.

The specific inhibition threshold will need to be evaluated for each method to ensure that it is fit for purpose. This will be based on the expected level of inhibition within samples. However, it is recommended that the threshold does not exceed 75%. If a method frequently gives inhibition levels of >75%, then this suggests that further optimisation is required. Additionally, due to inherent variability in qNAA even between sample replicates, it may be very difficult to reliably quantify low levels of inhibition, and so this also needs to be considered when selecting an appropriate inhibition threshold (i.e. it is not feasible to use a threshold of 0% inhibition).

2. Terms and definitions

For the purposes of this guidance, the terms and definitions are defined in the ISO Online Browsing Platform (https://www.iso.org/obp/ui/#search). Terms that are either important to include for easy reference in this document or that are not defined in the Online Browsing Platform are defined below. Where references have been cited for the definition, the exact definition text may have been adapted from the original for consistency of terminology within this document.

2.1.1. The Method

The detailed procedure used for quantifying a target analyte in a wastewater sample. This includes any processes to concentrate, purify and measure the analyte as well as any calculations to convert machine data to reportable quantities. It also includes any quality control procedures for those steps. For the purposes of this guidance, it does not include sample collection, unless the rest of the Method is dependent on the sampling procedures.

2.1.2. Analyte

Component represented in the name of a measurable quantity (ISO, 2016a) For this guidance, this means a micro-organism or group of micro-organisms targeted for quantification using the Method.

2.1.3. Quantitative Nucleic Acid Amplification (qNAA)

A group of quantitative analytical techniques that are based on the amplification of nucleic acids. This may include a reverse transcription (RT) step. It may be based on any nucleic acid amplification-based method such as real-time PCR, digital PCR or isothermal techniques.

2.1.4. Multiplex

A qNAA format in which multiple targets are detected and/or quantified simultaneously within a single reaction well/tube.

2.1.5. Wastewater

Water containing high concentrations of faecal wastes discharged from residential, commercial and industrial premises.

2.1.6. Influent

Untreated or minimally treated wastewater entering a wastewater treatment works.

2.1.7. Effluent

Treated wastewater discharged from a wastewater treatment works.

2.1.8. Near-source

A sampling location close to the source of contamination and upstream of a sampling location for influent.

2.1.9. Negative matrix

Wastewater that is representative of samples that would normally be analysed by the Method but does not contain the target analyte.

2.1.10. Sample volume

The volume of wastewater that makes up a single sample for analysis by the Method.

2.1.11. Accuracy

The closeness of agreement between a test result and the accepted reference value (ISO, 1994)

2.1.12. Trueness

Closeness of agreement between the expectation of a test result or a measurement result and a true value (ISO, 2006a).

2.1.13. Applicability

Analytes, matrices, and concentrations for which an analytical approach may be used satisfactorily (ISO, 2016c)

2.1.14. Limit of detection (LOD)

Lowest concentration of the target organism per defined amount of matrix that can be consistently detected under the experimental conditions specified in the method (ISO, 2005).

2.1.15. LOD_x

Measured analyte concentration, obtained by a given measurement procedure, for which the probability of detection is x (ISO, 2016a)

2.1.16. Limit of quantification

Lowest analyte concentration that can be quantified with an acceptable level of precision and trueness under the conditions of the test (ISO, 2016a).

2.1.17. Precision

Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions (ISO, 2016a).

2.1.18. Repeatability

Measurement precision under a set of repeatability conditions of measurement (ISO, 2016a). Also known as in-house repeatability.

2.1.19. Repeatability condition of measurement

Condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time (ISO, 2016a).

2.1.20. Intermediate precision

Measurement precision under a set of intermediate precision conditions of measurement (ISO, 2020b). Also known as in-house reproducibility.

2.1.21. Intermediate precision condition of measurement

Condition of measurement, out of a set of conditions that includes the same measurement procedure, same location, and replicate measurements on the same or similar objects over an extended period of time, but may include other conditions involving changes (ISO, 2020b)

2.1.22. Reproducibility

Measurement precision under reproducibility conditions of measurement (ISO, 2016a).

2.1.23. Reproducibility condition of measurement

Condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects (ISO, 2016a).

2.1.24. Recovery

Proportion of the amount of analyte, present in, added to, or present in and added to the [sample], which is presented for measurement following extraction from the matrix (ISO, 2016c)

2.1.25. Impact of Variable Matrix

The impact on results generated by the Method caused by the variable nature of the wastewater matrix.

2.1.26. Selectivity

Measure of the inclusivity (detection of the target micro-organism) and exclusivity (non-detection of non-target micro-organisms) (ISO, 2011)

2.1.27. Sensitivity

Change in the response divided by the corresponding change in the concentration of a standard (calibration) curve. i.e. the slope of the analytical calibration curve. (ISO, 2006b)

2.1.28. Linearity

Ability of a method of analysis, within a certain range, to provide results proportional to the quantity of nucleic acid target sequence to be determined in the sample (ISO, 2019a)

2.1.29. Measurement uncertainty

Parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the quantity intended to be measured (ISO, 2021b).

3. Specification of requirements

To determine the Method's fitness for purpose, it is necessary to define the intended use of the Method and the required performance criteria. These may change depending on the customer or user of the Method, and so these must be defined for each validation study. The intended use and required performance criteria will be part of the validation report and will form part of the scope of validation for the Method.

3.1. Intended use of the Method

Define the intended use of the Method by answering the following questions prior to designing the validation study.

- What is the target analyte(s)?
  - Define the taxonomic level to which the Method will be specific to the target analyte. For example, if the Method is used to quantify norovirus, will it quantify all noroviruses or specific genogroups or genotypes?
- How will the data generated by the Method be reported and used?
- What is the nature of the sample matrix?
  - For example, will the Method be used for both near-source samples and wastewater treatment works (WWTP) samples?
  - Will the Method be used for both influent and effluent samples?
  - Is the sample liquid, sludge or both?
- Has the qNAA been validated using the guidance in ISO 20395 (ISO, 2019a)?
- What is the likely concentration range of the analyte expected within samples?
- How many analysts will use the Method?
- Will samples always be analysed immediately upon arrival into the laboratory or will there be a delay between sample arrival and analysis?

3.2. Performance criteria

Prior to starting the validation study, benchmark criteria must be decided upon against which the success or failure of the validation will be assessed. For each of the performance parameters under investigation, this will take the form of a minimum/maximum level or a range of levels that the parameter under investigation must conform to. Table 1 lists the form in which these criteria should be listed in order to assess the validity of the Method.

If the Method does not conform to the validation criteria, it indicates that either the Method may require further optimisation/development. Alternatively, failure to meet performance criteria may also indicate that the benchmark criteria are unrealistic. While this should not be the case if those criteria were developed with sufficient care, this should be considered before embarking on further method development. As noted previously in section 1.2, for new methods or fields of study, there may be no precedent from which to decide upon benchmark criteria. It may therefore be necessary in the first instance to use the best judgement of the experts carrying out the validation study to select the most appropriate benchmark criteria and readjust them later if they are deemed to be unreasonable.

Table 1: Example performance criteria for each parameter evaluated using this guidance. Note, the example criteria are for illustrative purposes only and the actual criteria used must be decided upon based on the needs of the customer or user of the Method.

Performance parameter	Form	Example criteria	Explanation
Precision - Repeatability	$s_r < X$	$s_r \leq 0.1$	The repeatability standard deviation must be no greater than 0.1.
Precision – Intermediate precision	$s_I < X$	$s_I \leq 0.3$	The intermediate precision standard deviation should be no greater than 0.3.
Linearity	Linear between X and Y	Linear between 100 gc/L and 100,000 gc/L	The Method must give quantifiable results when samples contain between 100 and 100,000 gc/L of the target analyte.
Recovery	$\overline{R} > X$	$\overline{R} > 20\%$	The Method must have an average recovery greater than 20%
Impact of Variable Matrix	No impact of variable matrix on precision No impact of variable matrix on accuracy	N/A	The precision and accuracy should not be significantly impacted by the variable nature of the wastewater matrix as shown by F tests and ANOVA tests respectively.
Selectivity	>X% inclusive, >X% exclusive	100% inclusive, ≥95% exclusive	The Method must be able to detect all types or strains of the target analyte and no more than 5% of non-target micro-organisms can be detected.
Limit of detection	$LOD_{95} < X$	$LOD_{95} \leq 10$ gc/L	The maximum acceptable LOD_{95} for the Method is 10 gc/L.
Limit of quantification	LOQ <X	LOQ ≤100 gc/L	The Method must give quantifiable results when samples contain 100 gc/L or more.

4. Technical protocol for validation

4.1. General considerations

This validation protocol assumes that the qNAA method has been validated using the guidance from ISO 20395 (ISO, 2019a). This will allow confidence in the results from the analytical (i.e. data generation) element of the method. The qNAA validation will ideally include an inclusivity/exclusivity (selectivity) study in which 50 variants, strains or types of the target micro-organisms and 30 other non-target micro-organisms are included. This requirement is discussed further in section 4.6.2. If the qNAA method is not 100% selective for the target micro-organism, then this must be noted in the validation report and form part of scope of validation.

This protocol is for method validation using a conventional approach and assumes that there is no reference method against which the Method will be compared. This means that the validation study must rely mostly on artificially contaminated samples containing a known level of the target micro-organism because there will be no reference method against which to compare trueness of results.

It is assumed that prior to the validation of the Method, its performance characteristics have been studied to at least a basic level and so the approximate working range of concentrations is known for the method. Having this knowledge prior to embarking on a validation study will improve the probability the Method will meet the expected performance criteria during the formal validation study, and will allow samples to be used that contain appropriate concentrations of the target analyte.

4.1.1. Spiked sample preparation

4.1.1.1. Negative matrix

Studies conducted using this guidance require the use of a negative matrix spiked with reference material. Negative matrix must be sourced from as wide a range of locations as feasible for the method under investigation. Wastewater for use as negative matrix must be taken from at least three sampling locations where possible. In all studies other than the Impact of Variable Matrix study, the Negative matrix from different sources must be combined prior to use to form a pooled representative matrix. For the Impact of Variable Matrix study, the different sources of negative matrix will be tested separately. Negativity of the matrix will be confirmed using the Method under investigation or another method that is known to be more sensitive than the Method.

The negative matrix and the standard reference material must be relatively stable over the timeframe of the study. A stable negative matrix will not degrade significantly over the period of the study which may otherwise impact the results of the study. The negative matrix will be split into aliquots with a volume that is equal to the sample volume and stored at <-15°C until required.

If it is not possible to obtain negative matrix due to the ubiquity of the target analyte, then the following options exist:

- Use samples with a low level of the target analyte instead of a negative sample. In this case, it will not be possible to define the true LOD and may not be possible to define the LOQ. However, this may be acceptable depending on the requirements of application of the Method. This must be noted in the validation report and will form part of the scope of validation. It should also be noted that the use of samples with low levels of the target

analyte may impact the Impact of Variable Matrix study. The options to overcome this are discussed further in section 4.5.1.

- Use an artificial matrix that is representative of the wastewater that will be evaluated using the Method. At the time of writing this guidance, we were not aware of any published recipes for a suitable artificial wastewater matrix. However, if a suitable alternative is developed this may be an acceptable alternative to negative matrix.

4.1.1.2. Standard reference material

The material that is used to spike the negative matrix must be representative of the target analyte. Where possible and when it exists, certified reference material (CRM) will be used. Where no CRM is available, individual laboratories must source their own reference material and determine its intra-batch concentration and variability before use.

The standard reference material must be stable over the course of the study and stored as independent aliquots that can be used without the need to disturb the other aliquots (e.g. a batch of frozen viral aliquots, Lenticule discs, lyophilised cultures or some other stable form of the analyte).

4.1.1.3. Preparation of samples

Calculate the required concentration for a high-concentration reference material suspension (HRMS) using Equation 1.

Equation 1: $$C_s = \frac{V_f C_f}{V_s}$$

where

C_s	is the concentration of the HRMS
V_f	is the sample volume
C_f	is the concentration of the target analyte required for the study
V_s	is the volume of the HRMS that will be spiked into the negative matrix (V_s must be ≤1% of V_f)

Create HRMS by diluting CRM or other reference material in phosphate buffered saline, or another osmotically balanced buffer suitable for the target analyte and mixing thoroughly. The volume of HRMS created will be enough to spike the appropriate number of samples for the study, with enough excess to allow triplicate direct nucleic acid extractions from the HRMS in those studies that require it. The aliquot for direct nucleic acid extraction will be taken at this stage and stored a <-70°C until needed.

Thaw enough negative matrix aliquots for the study. To account for natural variation in the wastewater matrix, negative matrix aliquots for all studies except the Impact of Variable Matrix study must consist of a negative matrix pool made from at least three different sources. For the Impact of Variable Matrix study, the uncombined negative matrix from the different locations will be used separately.

Add a volume of HRMS equal to V_s to each negative matrix aliquot. Mix these thoroughly before using them in the validation studies.

4.1.2. Data handling

All data generated must meet the quality criteria stipulated in the qNAA validation study. Examples of qNAA quality criteria are (for illustrative purposes):

- The correlation coefficient (R^2) of $\log_{10}$ analyte concentration vs C_q for the standard curve (for qPCR) is ≥0.98.
- The slope of $\log_{10}$ analyte concentration vs C_q for the standard curve (for qPCR) is between -3.1 and -3.6, which is the equivalent of approximately 90% to 110% PCR amplification efficiency.
- qNAA inhibition levels are <75%.

It should be noted that these are just examples of qNAA quality criteria, and the justification for selecting qNAA quality criteria applied to the Method should be documented rather than just using these example values.

Any data not meeting the qNAA quality criteria must be removed from the dataset and the qNAA for those samples repeated. Where replicate qNAA reactions have been conducted for each sample, the calculated quantities should be averaged on a per-sample basis before proceeding with data analysis.

4.1.3. Calculation of F critical values

Several of the evaluations of the performance criteria outlined in this guidance require the use of F-tests to determine the significance of results. Where F critical values are shown, these were calculated using the qf() function in the R programming language (R Core Team, 2022) with the following code:

```
qf(p=α, df1=A, df2=B, lower.tail=FALSE)
```

where α is the significance level (e.g. 0.05), A is the degrees of freedom for the numerator and B is the degrees of freedom for the denominator.

Where multiple F tests comparisons are carried out, the significance level is adjusted using the Bonferroni correction method (α/number of comparisons).

4.2. Determination of precision characteristics

For single laboratory methods, precision will be determined under two sets of conditions; repeatability conditions (i.e. the same analyst and same conditions on the same day) and intermediate precision conditions (i.e. different analysts over several days). Each of the types of precision characteristics will also be measured for all matrix types (if using more than one) at a target analyte concentration close to the lowest level required for the Method's performance.

To measure both types of precision characteristics (repeatability and intermediate precision), measurements must be made on similar samples over eight days. This excludes the possibility of using naturally contaminated samples for this study. Samples must therefore be made up from negative matrix artificially contaminated with standard reference material on each day of testing.

4.2.1. Sample preparation

The concentration at which precision is measured must be representative of the lower concentrations that are expected to occur naturally in real samples and so must be close to the minimum level expected to occur naturally.

For each day of testing, prepare five samples at this concentration according to section 4.1.1. To minimise variability introduced at the sample preparation stage, the daily sample must be

prepared by the same analyst on all eight days using the same batches of negative control matrix and standard refence material. This analyst must be trained and competent for sample preparation.

4.2.2. Sample Analysis

All sample replicates must be analysed as independent samples according to the laboratory standard operating procedure for the Method.

Replicates must be tested over eight separate days by multiple analysts according to Table 2. If only one analyst is available, the Method is only validated for that analyst. Note that on any given day, only one analyst will test samples for the precision study. This means that testing will always take place over at least eight days regardless of the number of analysts. On each day of testing, each analyst will test five replicate samples.

Table 2: The allocation of the number of days each analyst must independently test the Method to measure repeatability and intermdiate precision where between 1 and 4 analysts are available.

Number of analysts	Number of days testing			
	Analyst 1	Analyst 2	Analyst 3	Analyst 4
1	8	-	-	-
2	4	4	-	-
3	3	3	2	-
4	2	2	2	2

4.2.3. Data analysis

The repeatability (repeatability standard deviation; s_r) is calculated according to Equation 2 (ISO 16140-4 formula 7 (ISO, 2020a)):

Equation 2:
$$s_r = \sqrt{\frac{1}{\mathrm{N} - \mathrm{J}} \sum_{j=1}^{\mathrm{J}} \sum_{i-1}^{\mathrm{n}} \left(y_{ij} \bar{y}_j\right)^2}$$

where

s_r	is the repeatability standard deviation
y_{ij}	is the $\log_{10}$ measured target concentration of the ith replicate on day j.
$\bar{y}_j$	is the average $\log_{10}$ measured target concentration of five replicates on day j.
N	is the total number of samples at each dilution (40)
J	is the number of days over which testing was conducted (8)
n	is the number of samples tested on each day (5)

The intermediate precision (in-house reproducibility standard deviation; s_l) is calculated according to Equation 3 (ISO 16140-4 formula 8 (ISO, 2020a)):

Equation 3: $$s_l = \sqrt{{s_A}^2 + {s_r}^2}$$

where

s_l is the in-house reproducibility standard deviation

$$ {s_A}^2 = \frac{1}{\mathrm{J}-1}\sum_{j=1}^{\mathrm{J}}\left(\left(\bar{y}_j - \bar{y}\right)^2 - \frac{1}{\mathrm{n}}{s_r}^2\right) $$

and

$$ \bar{y} = \frac{1}{\mathrm{N}}\sum\nolimits_{i=1}^{\mathrm{n}}\sum\nolimits_{j=1}^{\mathrm{J}} y_{ij} $$

Where the precision of the Method does not meet the stated performance requirements of the Method, further work should be conducted the improve its performance before the validation study is repeated.

4.3. Linearity

For the purposes of this guidance, linearity of a method is the target concentration range within which a measurement is proportional to the quantity of the target analyte in a sample.

While the linearity of the qNAA will already have been established, this is not enough to infer linearity of the Method. This is because additional factors may affect the linearity of the Method that will not affect the linearity of the qNAA. It is therefore important that the linearity study is carried out using representative samples created using spiked negative matrix as described in section 4.1.1.

4.3.1. Sample preparation

Select six target analyte concentrations that span the range of the Method's performance requirements with approximately even distribution between each concentration level on a $\log_{10}$ scale. For example, if the Method must be able to quantify an analyte between 100 and 100,000 gc/L, then you may select concentrations according to Table 3. Prepare three samples for each of these concentrations according to section 4.1.1.

Table 3: Example concentrations over which to test linearity of the Method where the requirements are to be able to quantify between 100 and 100,000 gc/L from wastewater samples.

Concentration (gc/L)	Log_{10} concentration (gc/L)
100	2.0
400	2.6
1,600	3.2
6,300	3.8
25,100	4.4
100,000	5.0

4.3.2. Sample Analysis

All spiked sample replicates must be analysed as independent samples according to the laboratory standard operating procedure for the Method.

4.3.3. Data Analysis

For the purposes of this guidance, it is assumed that there will be a linear positive relationship between the $\log_{10}$-observed and the $\log_{10}$-expected results. Perform a linear regression test of the $\log_{10}$ observed results for each of the samples against the $\log_{10}$ expected results. If the slope is between 0.9 and 1.1, accept the data as linear.

Where the data from the linearity study do not meet the assumptions of linearity above, one of two options are recommended:

1. Repeat the study with a narrower range of analyte concentrations to allow six levels to be tested over a linear range.
2. If the first option does not meet the requirements of the Method (i.e. the Method must be linear over the tested range), then the Method should be improved to expand the linear range.

4.4. Recovery

It is not possible to prove the trueness of a microbiological technique, particularly one that requires processing of a sample prior to analysis. In the case of qNAA based techniques for wastewater analysis, there are several steps in which there may be losses of the target analyte and the final result will therefore not be an accurate measure of the true levels of analyte in the sample. For example, bacteria or virus particles may be lost during a concentration step, or the nucleic acid extraction may not be 100% efficient. A measure of the proportion of the analyte from the original sample that is detected during the analysis is known as the recovery (or extraction efficiency). If the recovery is too low, then this will negatively impact the limit of detection. If the recovery is too variable, then this will reduce the precision of the Method.

Where existing reference methods exist against which the Method can be tested, recovery would be assessed as the recovery relative to the reference method. However, this guidance assumes that no reference method exists. It is therefore necessary to measure the recovery of a known concentration of the target analyte that has been spiked into a representative matrix sample.

In addition to measuring recovery during the validation study, recovery is frequently measured as part of the routine use of methods for quantifying micro-organisms from environmental samples. If the Method includes the routine measurement of recovery, it is still necessary to carry out the recovery study independently to inform the validation report.

4.4.1. Sample preparation

If a LOD study will be carried out according to the instructions in section 4.7, then the data from the LOD study can be used to negate the need to carry out additional sample analyses for the Recovery Study. However, the high-concentration reference material suspension (HRMS) used to create the samples must also be analysed at this stage as detailed below. This will only be possible if these stocks are stored during the LOD study as outlined in section 4.1.1

If no LOD study will be carried out, or the samples from the LOD study are otherwise deemed to be unsuitable, samples will be prepared for the Recovery Study with at least three concentrations of the target analyte according to section 4.1.1.

Recovery must be measured for at least three different target analyte concentrations to account for potential variability in recovery at different concentrations. The concentrations at which recovery is measured must be representative of the concentrations that are expected to occur naturally in real samples. The lowest concentration must be close to the minimum level expected to occur naturally, the highest concentration must be close to the highest level expected to occur naturally, and an intermediate concentration at some point between these values (e.g. close to the median or geometric mean result for real samples).

The recovery study can be carried out over multiple days if necessary. For example, each of the three analyte concentrations tested on separate days. However, if this is the case, then this will affect the choice of terms used in the F test as described in section 4.4.3.

Prepare ten samples for each of these concentrations according to section 4.1.1. Aliquots of the standard reference material used for spiking the samples will be retained for analysis.

4.4.2. Sample Analysis

All spiked sample replicates must be analysed as independent samples according to the laboratory standard operating procedure for the Method. Additionally, three replicate direct nucleic acid extractions must be carried out using the HRMS used for spiking the samples. The nucleic acids must be extracted from HRMS using the same technique as used for the Method. The volume of HRMS used must be suitable for the nucleic acid extraction technique used as part of the Method. Quantify the concentration of each of these nucleic acid extracts using the same qNAA technique as used for the spiked samples.

4.4.3. Data Analysis

Check that the data generated for the recovery study fall within the expected precision of the Method using an F test for each of the three concentrations tested. Calculate the F statistic using Equation 4.

Equation 4: $$\mathrm{F}_i = \frac{{s_i}^2}{{s_r}^2}$$

where

F_i is the F statistic for the dilution i

s_i Is the standard deviation of the $\log_{10}$ results (gc/L) for dilution i

If F >2.752[*] for any of the concentrations, this indicates that the Method did not perform well during the recovery study. A single anomalous result can be removed for each of the dilutions in this case and the F test repeated with a F critical value of 2.181. If the precision of the recovery study results remains lower than the expected precision for the Method, new samples must be tested for any of the concentrations affected.

Assuming the results for the recovery study meet the precision criteria, calculate the percentage recovery for each concentration as the proportion of target analyte detected in the spiked samples relative to the concentrated standard reference material samples according to Equation 5.

[*] This F critical value is based on 9 degrees of freedom (10-1) for the $\log_{10}$ recovery study result standard deviation and 32 degrees of freedom (40-8) for the repeatability standard deviation at the 1.67% significance level (significance level adjusted for multiple tests).

Equation 5:

$$R_{ij} = \frac{S_{ij}}{(\overline{m_j} v_m)/v_s} \times 100$$

where

R_{ij}	is recovery for the ith replicate of dilution j
S_{ij}	is the concentration of target analyte measured in the ith replicate sample of dilution j (gc/l)
$\overline{m_j}$	is the average concentration of target analyte measured in dilution j of concentrated reference material stock (gc/l)
v_m	is the volume of the concentrated reference material stock spiked into the sample (l)
v_s	is the sample volume (l)

Using the recovery data for each replicate and dilution, carry out a one-way ANOVA test to determine whether there are significant differences in recovery at each of the three concentrations at the 5% significance level.

If there are significant differences in recovery between dilutions, this indicates that the Method does not perform equally well when there are different levels of the target analyte within samples. This may mean that the Method should be developed further to improve recovery performance at all levels.

If there are no significant differences in recovery between dilutions, calculate the overall recovery as the mean recovery for all dilutions and the overall variation in recovery as the coefficient of variation between all recovery data according to Equation 6 and Equation 7.

Equation 6:

$$\bar{R} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} R_{ij}}{N}$$

where

$\bar{R}$	is the mean recovery for all recovery data
n	is the number of samples tested at each dilution (10)
m	is the number of dilutions tested (3)
N	is the total number of samples tested in the recovery study (30)

Equation 7:

$$C_R = \frac{1}{\bar{R}} \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (R_{ij} - \bar{R})}{N - 1}}$$

where

C_R	is the coefficient of variation for all recovery data

4.5. Impact of Variable Matrix

Due to the variable nature of wastewater as outlined in section 1.4, there is a possibility that the Method will not perform equally well for samples taken across a broad range of geographic locations. It is therefore important to measure the extent of the impact that this will have on the Method. For the purposes of this guidance, the impact on precision and accuracy will be tested.

4.5.1. Sample preparation

Select a target analyte concentration that is well within the linear range of the Method (such as one of the mid-levels used in the Linearity study). For this study, use a minimum of three and a

maximum of five negative matrix sources. For each of the negative matrix sources, prepare five samples according to section 4.1.1.

In cases where it is not possible to obtain negative matrix samples (see section 4.1.1.1), the presence of the target analyte may impact the results of the Impact of Variable Matrix Study. To overcome this, the concentration of target analyte spiked into the samples should be high enough to mask any impact of the intrinsic target analyte. Alternatively, the target analyte may be replaced with a suitable, representative surrogate that is known to not be present in the negative matrix samples.

4.5.2. Sample Analysis

All spiked sample replicates must be analysed as independent samples according to the laboratory standard operating procedure for the Method.

4.5.3. Data Analysis

Compare the variability among each of the negative matrix sources against the expected precision of the Method using F tests for each of the negative matrix sources tested. Calculate the F statistic using Equation 4 in section 4.4.3, but where:

F_i is the F statistic for the ith negative matrix source

s_i Is the standard deviation of the $\log_{10}$ results (gc/L) for the ith negative matrix source

Select an F critical value from Table 4 depending on the number of negative matrix sources used. If F >F critical for any of the negative matrix sources, this indicates that the precision of the Method is impacted by the variability of the matrix.

Table 4: F critical values used for determining significance of differences in the precision of the Method due to the variability of wastewater. F critical values were calculated using the qf() function in the R programming language (R Core Team, 2022).

Number of negative matrix sources	Bonferroni corrected α	F critical
3	0.0167	2.752
4	0.0125	2.903
5	0.0100	3.021

Compare the mean $\log_{10}$ results for each of the each of the negative matrix sources against each other using a one-way ANOVA test at the 5% significance level. If a significant difference is found between the means of the negative matrix sources, this indicates that the accuracy of the Method is impacted by the variability of the matrix.

If precision or accuracy are impacted by the negative matrix source, this may mean that the Method should be developed further to improve recovery performance at all levels and using a broader and more representative variety of matrix examples.

4.6. Selectivity

Selectivity is an important measure of a method’s ability to inclusively detect and/or quantify the entire diversity of the target analyte(s) while excluding non-target analytes. In the case of this guidance, it is assumed that initial studies of selectivity will have been carried out during the validation of the qNAA in line with ISO 20395 (ISO, 2019a). It should be noted that in ISO 20395 (ISO, 2019a), the relevant synonym for selectivity is “specificity”. However, the guidance

in ISO 20395 (ISO, 2019a) on determining selectivity/specificity is not detailed enough to allow full validation for selectivity of the Method. Of particular note is a lack of guidance in ISO 20395 (ISO, 2019a)on the selection of a suitable set of target and non-target organisms against which to test selectivity/specificity. Fortunately, guidance on this is given in ISO 16140-2 (ISO, 2016b), and this guidance document uses the guidance in that ISO for this study. Additionally, relying solely on the selectivity of the qNAA is not adequate in the case of wastewater testing. This is because the steps preceding the qNAA may also have an impact on selectivity. Specifically, if genotypic variations within a target organism result in phenotypic variation that modifies its ability to be concentrated and extracted from a wastewater sample, the Method will not be fully inclusive. This cannot be determined by validation of qNAA alone, and so further selectivity testing is required for validation of the Method.

4.6.1. A note on primer and probe design

Design of primers and probes used in the Method is assumed to have followed the guidance in ISO 20395 (ISO, 2019a). However, it should also be noted that the sequences of the primers and probes must be reviewed intermittently against genome data from a public sequence library such as GenBank. This is to check for any mutations in the primer and probe binding regions, which may impact the reverse transcription and/or amplification of the target analyte and therefore impact the trueness of results.

4.6.2. Selection of test analytes for selectivity

The selection of the test analytes for selectivity will depend on the taxonomic level at which identification of the target analyte is needed. The target and non-target analytes must be related to the target analytes at the appropriate taxonomic level. For example, if the Method is for quantification of a viral species, then the selectivity tests must include a range of variants, strains, genogroups and/or genotypes of that species for inclusivity testing. For exclusivity testing, a range of non-target organisms should be tested that represent the closest taxonomic relations to the target as possible.

It is recommended that when selecting the target and non-target analytes, a phylogenetic tree is created based on the type genome of the target organism using genome data from a public sequence library such as GenBank. This phylogenetic tree should show the relationship of the target type genome against other non-target analytes at higher taxonomic levels. This will indicate whether there are cases of non-conformity between genetic-based phylogeny and traditional phenotypic-based phylogenies. If this is the case, it may be necessary to include non-target analytes from more distantly related taxonomic groupings than would otherwise be necessary if the determination of the target analyte was by phenotypic characterisation (e.g. selective culture).

According to ISO 16140-2 (ISO, 2016b), at least 30 pure cultures of non-target analytes must be tested to inform the exclusivity of the Method(ISO, 2016b) and at least 50 pure cultures of the target analytes must be tested to inform the inclusivity of the Method. In the case of emergent and novel micro-organisms, there may not be 50 strains or variants to test inclusivity. In this case, the Method will be validated against the maximum number of strains or variants available at the time of validation. However, continual testing of inclusivity must be carried out intermittently throughout the lifetime of the Method to ensure that it remains relevant. ISO 16140-2 (ISO, 2016b) acknowledges that for some micro-organisms, it will be difficult or impossible to obtain pure cultures. This is particularly the case where those organisms cannot be cultured (e.g. many viruses). In this case, pure (or as close to pure as possible) suspensions of the test analytes, containing single strains or variants of the target micro-organism, must be used.

It is assumed that the laboratory conducting the Method validation will conform to at least biosafety 2 (containment level 2) requirements. In the case of qNAA based quantification of biosafety level 3 micro-organisms, it will not be possible to handle pure cultures in biosafety level 2 laboratories. These cultures must therefore be inactivated before handling, using validated inactivation techniques or the study must be conducted in a biosafety level 3 or above laboratory. In these cases, the representativeness of inactivated cultures must be considered.

Once the list of target and non-target analytes is created, *in-silico* specificity tests will be carried out according to ISO 20395 (ISO, 2019a) before confirming the *in-silico* results by empirical testing outlined below.

4.6.3. Exclusivity tests

It is adequate for exclusivity to be evaluated solely by qNAA. This is because even if non-target analytes are co-extracted by the processes upstream of qNAA, if the non-target analytes are not detectable by qNAA, then this will give adequate discrimination between target and non-target analytes.

It should be noted however, that if there is a desire to use the Method for other downstream analyses such as characterisation by high-throughput sequencing, then testing exclusivity of the whole of the Method may be desirable in some cases. This is beyond the scope of the current guidance and so will not be discussed further.

It may not be necessary or possible for a Method to be completely exclusive of non-target analytes. This is the case in some bacterial detection methods, where traditional taxonomic designations may not be compatible with genetic sequence-based phylogenies (for example, *Escherichia* and *Shigella* species). If this is the case for the Method, this must be noted in the validation report and form part of the scope of validation.

4.6.3.1. Sample Analysis

Obtain pure cultures or similar material containing the 30 non-target analytes for the exclusivity test. Extract and purify the nucleic acids from each of these samples independently. Alongside this, nucleic acids must be extracted from a pure culture of target analyte and a sample of deionised water to act as positive and negative extraction controls respectively. Carry out qNAA in at least duplicate for each of the nucleic acid extracts using the usual procedure used for detecting the target analyte. This must include the appropriate standard dilution series, negative controls and inhibition controls.

4.6.3.2. Data Analysis

Ensure that the qNAA meets all of the performance characteristics required for that method. Record the results for each of the samples and control qNAA. The negative extraction control must have no qNAA result and the positive extraction control must have a strong qNAA result. If either one of these assumptions is not met, the test will be repeated.

If replicate qNAA for individual non-target analytes give conflicting results (e.g. one positive and one negative), then the qNAA test must be repeated for that sample. If it continues to give conflicting results, and the possibility of sample contamination has been ruled out, then the result will be regarded as positive.

4.6.4. Inclusivity tests

Unlike the exclusivity tests, inclusivity tests must be carried out using the whole of the Method to test inclusivity. However, a qNAA based experiment for each of the target analytes will be carried out first as described for exclusivity testing in section 4.6.3. This will potentially reduce

the work required for any target strains or variants that are not detectable by qNAA, as they will be excluded from further testing.

4.6.4.1. Sample preparation

The inclusivity of the Method will be characterised by measuring the recovery of each of the target analytes at a single concentration in two replicates for each target analyte.

For each target analyte, prepare two replicate samples by spiking negative matrix at a single concentration level according to section 4.1.1. The concentration used will be an intermediate level for the Method rather than an extreme (high or low) level. If testing all 50 target analytes, this will be a total of 100 samples.

4.6.4.2. Sample analysis

Samples and concentrated stocks will be analysed as outlined in section 4.4.2.

4.6.4.3. Data analysis

Recoveries must be calculated using Equation 5, but the term j will refer to the target analyte strain or variant rather than the dilution.

Using the recovery data for each replicate and target, calculate the average recovery and the coefficient of variation for recovery for each target.

Using the recovery data for each replicate and target, carry out a one-way ANOVA test to determine whether there are significant differences in recovery between targets.

If there are significant differences in recovery between targets, this indicates that the Method does not perform equally well for all strains or variant of the target analyte. This may mean that the Method should be developed further to improve inclusivity.

4.7. Limit of detection

Characterisation of the limit of detection (LOD) may not be necessary for methods that are used for quantifying micro-organisms that are expected to always be present in wastewater at high levels. This includes faecal indicator organisms such as faecally associated bacteriophages (e.g. coliphages and phages of some *Bacteroides*) and human gut bacteria (e.g. *Escherichia coli*, intestinal *Enterococci*).

For micro-organisms that are expected to be intermittently present in wastewater, then characterisation of the LOD is necessary. This includes micro-organisms whose concentration ranges within wastewater may vary over time from very high levels to very low levels between samples (e.g. SARS-CoV-2 and *Salmonella* spp.).

Many different approaches to determine the LOD are available; here we describe a method using a $\log_2$ dilution series of target analyte in wastewater samples. Multiple subsamples at each level of the dilution series are tested and used to determine LOD_{95} (the lowest concentration of target virus that can be consistently detected in 95% of samples tested under routine laboratory conditions). A probability of detection function is used to determine the LOD_{95} characteristics as applied for various International Standard methods in food microbiology including ISO 15216-1-2017 (ISO, 2017a) the method for quantification of viruses in foods.

4.7.1. Sample replication

For the purposes of the LOD_{95} study, it is necessary to use a representative matrix that does not contain the target(s) of interest (negative matrix), and which can be artificially contaminated at a large range of levels. Artificially contaminated samples must therefore be prepared using the guidance outlined in section 4.1.1. The measurement of LOD_{95} is carried out using replicate

samples on a $\log_2$ dilution series of the target analyte. A minimum of 10 replicate samples will be analysed for each dilution to generate LOD_{95} characteristics using the methods described below.

4.7.2. Sample preparation

The recommended number of samples to test for determining LOD_{95} is 90 (10 replicates of 9 dilutions, plus any additional controls required by the Method). It is recommended that those laboratories that are capable analysing this number of samples under repeatability conditions (see Repeatability condition of measurement) on a single day do so. Those laboratories not capable of analysing more than 90 samples under repeatability conditions (see Repeatability condition of measurement) on a single day will need to perform the LOD_{95} study over the course of several days, testing all samples from a single concentration in one day.

The highest target analyte concentration to test must give a result of approximately 100 gc/reaction in the qNAA step. Eight other concentrations of target analyte must then be selected representing a descending $\log_2$ dilution series (e.g. 100 gc/reaction, 50 gc/reaction, 25 gc/reaction etc.). Nine concentrations in total must be selected with expected concentrations ranging from 100 gc/reaction to 0.39 gc/reaction in a $\log_2$ series. Prepare ten samples for each of these concentrations according to section 4.1.1.

4.7.3. Sample analysis

All spiked sample replicates must be analysed as independent samples according to the laboratory standard operating procedure for the Method.

4.7.4. Data analysis

For calculation of LOD_{95} quantification data are only needed for the Neat Sample. For all other dilutions, only the number of samples with positive results is required. However, quantification for all dilutions is required for the limit of quantification study. All data used for the LOD95 test must conform to the data quality parameters for the qNAA. The results of up to two samples at each dilution can be removed from the dataset if they do not conform to the data quality parameters. If more than two results at each dilution do not conform to the data quality parameters, new replicates must be tested at the dilutions affected.

Check that the data generated for the Neat Sample fall within the expected precision of the Method using an F test as outlined in section 4.4.3. If the precision of the Neat Sample results remains lower than the expected precision for the Method, the results of up to two of the Neat Samples can be removed and the F test repeated (with the relevant adjusted degrees of freedom). Note, at least eight Neat Sample results should be used to determine precision. If samples are removed in this way, those results must be omitted for the rest of the analyses. If the precision of the Neat Sample results remains lower than the expected precision following the removal of two data points, new replicates must be tested for the Neat Sample.

Calculate the anticipated values for each dilution as the geometric mean of the observed results for all replicates of the Neat Sample multiplied by the dilution factor.

Example:

Where the 10 Neat Sample replicates give results of 161.6, 120.8, 128.1, 141.5, 139.2, 130.1, 115.3, 142.2, 152.8 and 156.5 gc/L respectively (geometric mean = 138.04 gc/L), the anticipated results for the different dilutions are shown in Table 5.

Table 5: Example of anticipated results for a log_2 dilution series for a Neat Sample with an observed geometric mean concentration of 138.04 gc/L.

Dilution	Anticipated value (gc/L)
Neat RNA	138.04 x 1 = **138.04**
1:2	138.04 x 1/2 = **69.02**
1:4	138.04 x 1/4 = **34.51**
1:8	138.04 x 1/8 = **17.26**
1:16	138.04 x 1/16 = **8.63**
1:32	138.04 x 1/32 = **4.31**
1:64	138.04 x 1/64 = **2.16**
1:128	138.04 x 1/128 = **1.08**
1:256	138.04 x 1/256 = **0.54**

Determine the LOD_{95} for the data using the approach developed by Wilrich and Wilrich (2009). There is no limitation on the tools that can be used to analyse the data using the approach by Wilrich and Wilrich, but this can most easily be achieved using the Excel based calculator, PODLOD_ver10.xls, which is available online from: Wilrich • Forschungsschwerpunkt Statistik und Ökonometrie • Fachbereich Wirtschaftswissenschaft (fu-berlin.de)

To use the calculator, open the file in Excel and then use the following instructions (correct for Excel version 16 on 31st May 2022):

- Ensure that macros are enabled by clicking the "Enable Content" button near the top of the Window (if displayed).
- In the yellow "General information on the experiment" box, input:
    - Sample size A_0: **1**
    - No. of matrices: **1**
    - No. of inoculation levels as the number used: **9**
    - The Name of Experiment, Date of Experiment and Micro-organism fields are optional.

This will generate a yellow "Input data for the matrices" table. In this table input the following for each of the columns:

- Inoculation level in cfu/g or cfu/ml: **<leave this section blank – values will autofill when entering values in the next column>**
- Inoculation level in cfu/A_0: **The anticipated values for each dilution in which ever units you are using for this study (e.g. gc/L)**
- No of inoculated tubes: **10 (or the number of replicate reactions used)**
- No of positive tubes: **Number of positive reactions for a given dilution**
- Press Control + B or click on the "Calculate results" button in the top right of the window to start the calculation.

Two new boxes containing the results of the calculation will appear. If you input the data according to the previous instructions, both of these boxes will contain the same values. Record the LOD_{95} value (shown as "Detection limit $d_{0.95,i}$") in the table "Results of the PODLOD calculations – with the LOD relating to d in cfu/A_0". Note here that the units used in the PODLOD tool do not necessarily relate to the units relevant to your LOD_{95}. The units for the LOD_{95} will be

the same as were used to input data into the “Input data for the matrices” table. The upper and lower confidence limits may be recorded but are not required for the determination of the LOD_{95}). Report the LOD_{95} to three significant figures.

4.8. Limit of quantification

As with LOD, characterisation of the limit of quantification (LOQ) may not be necessary for methods that are used for quantifying micro-organisms that are expected to always be present in wastewater at levels within the linear range of the Method. For micro-organisms that are expected to be present at low levels in wastewater, then characterisation of the LOQ is necessary.

The LOQ study will be carried out after the LOD_{95} study and use the data that were generated therein. This means that for determining LOQ, it is not necessary to carry out more practical work than is required for determining LOD_{95}.

4.8.1. Data analysis

Using the LOD_{95} study dataset, discard the data points where the anticipated values are lower than the determined LOD_{95} value. If fewer than four anticipated values remain in the dataset at this point, then it will not be possible to accurately determine the LOQ and a new set of samples covering a higher range of levels above the LOD must be tested. Determine whether the remaining data are linear according to the procedure outlined in section 4.3.3.

If the data are linear, retain all of the data points above the LOD_{95} to determine the LOQ. If data are not linear, exclude the data points corresponding to the lowest remaining anticipated value and test for linearity again. If the data are linear, the data used for this estimation can be used to determine LOQ. If the data continue to be non-linear, a new set of data will need to be generated. In this case, it is advised in the first instance to repeat the qNAA reactions using the existing nucleic acid extracts. If the data continue to be non-linear, then a new set of samples will need to be generated and analysed.

Using the data retained after determination of linearity, discard any individual negative qNAA results and determine the standard deviation (SD) for the $\log_{10}$ transformed **observed** results (i.e. those calculated relative to the standard curve for each reaction) for each anticipated value.

The LOQ is the lowest anticipated level where the SD is <0.33 and all higher anticipated levels’ SDs are also <0.33. If all of the values are <0.33, then the LOQ equals the lowest anticipated value in the data set retained after determination of linearity. Report the LOQ to three significant figures.

5. Validation report

Following the validation study, a report should be written that includes the following minimum information. Additional information that allows interpretation of the results and recommendations may also be included where required.

- Description of the specification of requirements for the Method.
- Description of the materials and methods used for the validation study.
- Statement of the performance characteristics of each of the parameters tested in the validation study.
- The results obtained in the validation study in a form that allows re-analysis of the data by a third party.
- A statement on the validity of the method, detailing its fitness for the intended use.
- A reference to the qNAA validation report.

6. References

Regulation (EU) 2017/625 of the European Parliament and of the Council, (2017). https://www.legislation.gov.uk/eur/2017/625/contents

Barwick, V. J., & Ellison, S. L. R. (2005). The evaluation of measurement uncertainty from method validation studies. In P. de Bièvre & H. Günzler (Eds.), *Validation in Chemical Measurement* (pp. 84–90). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-27034-5_16

ISO. (1994). *ISO 5725-1:1994(en) Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*. https://www.iso.org/obp/ui#iso:std:iso:5725:-1:ed-1:v1:en:term:3.6

ISO. (2005). *ISO 22174:2005(en) Microbiology of food and animal feeding stuffs — Polymerase chain reaction (PCR) for the detection of food-borne pathogens — General requirements and definitions*. https://www.iso.org/obp/ui#iso:std:iso:22174:ed-1:v1:en:term:3.1.8

ISO. (2006a). *ISO 3534-2:2006(en) Statistics — Vocabulary and symbols — Part 2: Applied statistics*. https://www.iso.org/obp/ui#iso:std:iso:3534:-2:ed-2:v1:en

ISO. (2006b). *ISO 24276:2006(en) Foodstuffs — Methods of analysis for the detection of genetically modified organisms and derived products — General requirements and definitions*. https://www.iso.org/obp/ui#iso:std:iso:24276:ed-1:v1:en:term:3.1.5

ISO. (2011). *ISO 22118:2011(en) Microbiology of food and animal feeding stuffs — Polymerase chain reaction (PCR) for the detection and quantification of food-borne pathogens — Performance characteristics*. https://www.iso.org/obp/ui#iso:std:iso:22118:ed-1:v1:en:term:3.5

ISO. (2016a). *ISO 16140-1:2016(en) Microbiology of the food chain — Method validation — Part 1: Vocabulary*. https://www.iso.org/obp/ui#iso:std:iso:16140:-1:ed-1:v1:en:term:2.36

ISO. (2016b). *ISO 16140-2:2016(en) Microbiology of the food chain — Method validation — Part 2: Protocol for the validation of alternative (proprietary) methods against a reference method*. https://www.iso.org/obp/ui#iso:std:iso:16140:-2:ed-1:v1:en

ISO. (2016c). *ISO 16577:2016(en) Molecular biomarker analysis — Terms and definitions*. https://www.iso.org/obp/ui#iso:std:iso:16577:ed-1:v1:en:term:3.13

ISO. (2017a). *ISO 15216-1:2017(en) Microbiology of the food chain — Horizontal method for determination of hepatitis A virus and norovirus using real-time RT-PCR — Part 1: Method for quantification*. https://www.iso.org/obp/ui#iso:std:iso:15216:-1:ed-1:v1:en

ISO. (2017b). *ISO 19020:2017(en) Microbiology of the food chain — Horizontal method for the immunoenzymatic detection of staphylococcal enterotoxins in foodstuffs*. https://www.iso.org/obp/ui#iso:std:iso:19020:ed-1:v1:en:term:3.3

ISO. (2017c). *ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories*. https://www.iso.org/standard/66912.html

ISO. (2018). *ISO 8099-1:2018(en) Small craft — Waste systems — Part 1: Waste water retention*. https://www.iso.org/obp/ui#iso:std:iso:8099:-1:ed-1:v1:en

ISO. (2019a). *ISO 20395:2019(en) Biotechnology — Requirements for evaluating the performance of quantification methods for nucleic acid target sequences — qPCR and dPCR*. https://www.iso.org/obp/ui#iso:std:iso:20395:ed-1:v1:en

ISO. (2019b). *ISO 24513:2019(en) Service activities relating to drinking water supply, wastewater and stormwater systems — Vocabulary*. https://www.iso.org/obp/ui#iso:std:iso:24513:ed-1:v1:en

ISO. (2020a). *ISO 16140-4:2020(en) Microbiology of the food chain — Method validation — Part 4: Protocol for method validation in a single laboratory*. https://www.iso.org/obp/ui#iso:std:iso:16140:-4:ed-1:v1:en

ISO. (2020b). *ISO/TS 22176:2020(en) Cosmetics — Analytical methods — Development of a global approach for validation of quantitative analytical methods*. https://www.iso.org/obp/ui#iso:std:iso:ts:22176:ed-1:v1:en:term:3.1.8

ISO. (2021a). *ISO 16140-3:2021(en) Microbiology of the food chain — Method validation — Part 3: Protocol for the verification of reference methods and validated alternative methods in a single laboratory*. https://www.iso.org/obp/ui#iso:std:iso:16140:-3:ed-1:v1:en

ISO. (2021b). *ISO 20836:2021(en) Microbiology of the food chain — Polymerase chain reaction (PCR) for the detection of microorganisms — Thermal performance testing of thermal cyclers*. https://www.iso.org/obp/ui#iso:std:iso:20836:ed-1:v1:en:term:3.4.4

ISO. (2022). *ISO/AWI TS 16099 Water quality — Polymerase chain reaction (PCR) for the detection and quantification of microorganisms — Quality control and validation of molecular methods*. https://www.iso.org/standard/84465.html

JCGM. (2008). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e.html?csnumber=50461

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

UKHSA. (2022). *Standard Method for Verification and Validation of Methods: Report number FNES61*.

Wade, M. J., lo Jacomo, A., Armenise, E., Brown, M. R., Bunce, J. T., Cameron, G. J., Fang, Z., Farkas, K., Gilpin, D. F., Graham, D. W., Grimsley, J. M. S., Hart, A., Hoffmann, T., Jackson, K. J., Jones, D. L., Lilley, C. J., McGrath, J. W., McKinley, J. M., McSparron, C., … Kasprzyk-Hordern, B. (2022). Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. *Journal of Hazardous Materials*, *424*, 127456. https://doi.org/https://doi.org/10.1016/j.jhazmat.2021.127456

Wilrich, C., & Wilrich, P.-T. (2009). Estimation of the POD function and the LOD of a qualitative microbiological measurement method. *Journal of AOAC International*, *92 6*, 1763–1772.

Centre for Environment Fisheries & Aquaculture Science

World Class Science for the Marine and Freshwater Environment

We are the government's marine and freshwater science experts. We help keep our seas, oceans and rivers healthy and productive and our seafood safe and sustainable by providing data and advice to the UK Government and our overseas partners. We are passionate about what we do because our work helps tackle the serious global problems of climate change, marine litter, over-fishing and pollution in support of the UK's commitments to a better future (for example the UN Sustainable Development Goals and Defra's 25 year Environment Plan).
We work in partnership with our colleagues in Defra and across UK government, and with international governments, business, maritime and fishing industry, non-governmental organisations, research institutes, universities, civil society and schools to collate and share knowledge. Together we can understand and value our seas to secure a sustainable blue future for us all, and help create a greater place for living.

© Crown copyright 2022

Pakefield Road, Lowestoft, Suffolk, NR33 0HT
The Nothe, Barrack Road, Weymouth DT4 8UB
www.cefas.co.uk | +44 (0) 1502 562244