

Validation des méthodes d'analyse

Application à la microbiologie des eaux

**Protocole de validation
d'une méthode alternative commerciale
par rapport à une méthode de référence**

Révision 2 – Adoptée par AFNOR Certification le **17 mai 2013**
(suite à l'approbation du Bureau technique concerné)

SOMMAIRE

Introduction	4
1 Objet	4
2 Références normatives	4
3 Termes, définitions et symboles	4
3.1 Termes généraux	5
3.1.1 Méthode de référence.....	5
3.1.2 Méthode alternative	5
3.1.3 Validation d'une méthode alternative	5
3.1.4 Analyte	5
3.1.5 Méthode quantitative	5
3.1.6 Méthode qualitative.....	5
3.1.7 Laboratoire organisateur / laboratoire expert	5
3.1.8 Laboratoire collaborateur.....	6
3.2 Critères de validation pour les méthodes qualitatives	6
3.2.1 Déviation positive (PD)	6
3.2.2 Déviation négative (ND)	6
3.2.3 Sensibilité relative (SE)	6
3.2.4 Exactitude relative (AC).....	6
3.2.5 Spécificité relative (SP).....	7
3.2.6 Inclusivité et exclusivité (Sélectivité)	7
3.2.7 Niveau de détection relatif ou LOD ₅₀	7
3.2.8 Praticabilité	7
3.3 Critères de validation pour les méthodes quantitatives	7
3.3.1 Linéarité	7
3.3.2 Exactitude	7
3.3.3 Exactitude relative	7
3.3.4 Répétabilité.....	7
3.3.5 Reproductibilité	8
3.3.6 Limite de détection.....	8
3.3.7 Limite de quantification (ou limite de détermination)	8
3.3.8 Inclusivité et exclusivité (Sélectivité)	8
3.3.9 Praticabilité (voir 3.2.8).....	8
3.3.10 Limite d'acceptabilité	8
4 Principe	8
5 Étude préliminaire	9

5.1	Méthodes qualitatives	9
5.1.1	Protocole d'évaluation de l'inclusivité et de l'exclusivité (sélectivité)	9
5.1.2	Protocole de mesure de l'exactitude, sensibilité et spécificité relatives	10
5.1.3	Protocole de mesure du niveau de détection relatif	12
5.1.4	Praticabilité	13
5.2	Méthodes quantitatives	14
5.2.1	Principe et définitions	14
5.2.2	Protocole d'évaluation de l'inclusivité et de l'exclusivité (sélectivité)	14
5.2.3	Protocole de mesure de la linéarité et de l'exactitude relative	15
5.2.4	Limites de détection et de quantification	16
5.2.5	Praticabilité de la méthode alternative	17
6	Étude interlaboratoire	17
6.1	Méthodes qualitatives	17
6.1.1	Plan d'expérience	17
6.1.2	Calculs et interprétation	17
6.1.3	Interprétation	19
6.2	Méthodes quantitatives	19
6.2.1	Profil d'exactitude : principes généraux	19
6.2.2	Plan d'expérience	22
6.2.3	Calculs statistiques	23
6.2.4	Construction du profil d'exactitude	26
6.3	Synthèse de l'essai inter-laboratoire	27
6.3.1	Interprétation des résultats	27
6.3.2	Limite de quantification	27
6.4	Bibliographie relative au chapitre 6	28
Annexe 1. Classification par catégories d'eau		29
Annexe 2. Stress à appliquer par type d'eau		30
Annexe 3. Organisation des analyses réalisées en parallèle en fonction du type de méthode. ..		31
Annexe 4. Méthodes qualitatives. Test de rejet des résultats discordants		33
Annexe 5. Méthodes qualitatives. Étude interlaboratoire, critères d'accord, de concordance et ratio de chances de concordance		34
Annexe 6. Méthodes quantitatives. Calcul des écarts-types de répétabilité et reproductibilité. ..		39
Annexe 7. Exemple de calcul du profil d'exactitude		40
Annexe 8. Lignes directrices relatives à l'organisation et à la gestion des études inter-laboratoires		43

Introduction

Le présent document expose une procédure de caractérisation et de validation d'une méthode d'analyse microbiologique (qualitative ou quantitative) par rapport à une méthode de référence normalisée quand elle existe.

1 Objet

Le présent document définit des lignes directrices qui doivent mener à la validation, pour un domaine d'application spécifié (ex : eaux destinées à la consommation humaine et/ou eaux traitées et/ou eaux non traitées et/ou eaux de surface et/ou eaux profondes et/ou eaux de baignade et/ou eaux de piscines et/ou eaux minérales et/ou eaux embouteillées, etc.), d'une méthode d'analyse microbiologique « rapide » ou commerciale visant la recherche (méthode qualitative) ou le dénombrement (méthode quantitative) de microorganismes cibles par rapport à une méthode de référence normalisée ciblant les mêmes microorganismes.

La procédure s'organise de la façon suivante :

1. Une étude préliminaire réalisée par un laboratoire expert qui fournit une caractérisation initiale de la méthode en termes de critères classiques de validation (répétabilité, praticabilité, ...)
2. Une étude inter-laboratoire qui permet d'estimer divers critères pour les méthodes qualitatives (§ 6.1.3) et de construire le profil d'exactitude (§ 6.2.4), pour les méthodes quantitatives.
3. Une interprétation finale des résultats.

Les méthodes à valider peuvent être des méthodes de comptages directs, par exemple de colonies, ou de comptages indirects, par exemple le dénombrement du nombre le plus probable NPP.

2 Références normatives

Les documents normatifs suivants contiennent des dispositions qui, par suite de la référence qui y est faite, constituent des dispositions valables pour le présent document. Pour les références datées, les amendements ultérieurs ou les révisions de ces publications ne s'appliquent pas. Toutefois, les parties prenantes aux accords fondés sur le présent document sont invitées à rechercher la possibilité d'appliquer les éditions les plus récentes des documents normatifs indiqués ci-après. Pour les références non datées, la dernière édition du document normatif en référence s'applique.

ISO/TR 13843:2000, Qualité de l'eau – Lignes directrices pour la validation des méthodes microbiologiques.

EN ISO 16140:2003, Microbiologie des aliments – Protocole pour la validation des méthodes alternatives.

3 Termes, définitions et symboles

Pour les besoins du présent document, les termes et définitions suivants s'appliquent.

3.1 Termes généraux

3.1.1 Méthode de référence

Méthode d'analyse normalisée prescrite dans le but de détecter ou de dénombrer une espèce ou un groupe donné de micro-organismes.

3.1.2 Méthode alternative

Méthode commerciale permettant de détecter ou de dénombrer, pour un type d'eau donné, le(s) même(s) microorganisme(s) que celui (ceux) mesuré(s) par une méthode normalisée, et qui présente par exemple un ou plusieurs des critères suivants :

- Rapidité d'analyse et/ou de réponse
- Facilité d'exécution et/ou d'automatisation
- Caractéristiques analytiques (fidélité, limite de détection, etc....)

Le terme « méthode alternative » désigne l'ensemble produit, équipement et mode opératoire. Il inclut tous les ingrédients matériels ou non, nécessaires à la mise en œuvre de la méthode alternative.

3.1.3 Validation d'une méthode alternative

D'après VIM:2006 clauses 2.44 et 2.45

Fourniture de preuves tangibles que la méthode alternative satisfait à des exigences spécifiées, à savoir des résultats comparables aux résultats obtenus avec la méthode de référence, en prenant en compte toute incertitude de mesure et que ces exigences spécifiées sont adéquates pour un usage déterminé.

3.1.4 Analyte

Composant mesuré à l'aide de la méthode d'analyse (microorganismes).

3.1.5 Méthode quantitative

Méthode d'analyse dont la réponse est la quantité d'analyte mesurée soit directement (dénombrement masse volume), soit indirectement (absorption colorimétrique, impédance, etc.) dans une quantité d'échantillon donnée.

3.1.6 Méthode qualitative

Méthode d'analyse dont la réponse est qualitative (présence/absence d'analyte) et mesurée soit directement, soit indirectement (absorption colorimétrique, impédance, etc.) dans une quantité d'échantillon donnée.

3.1.7 Laboratoire organisateur / laboratoire expert

Laboratoire ayant l'expérience et les capacités nécessaires pour réaliser l'étude préliminaire et l'étude interlaboratoire. Il est sélectionné selon les critères définis par AFNOR Certification.

3.1.8 Laboratoire collaborateur

Laboratoire participant à l'étude interlaboratoire, selon les instructions fixées par le laboratoire organisateur /expert.

3.2 Critères de validation pour les méthodes qualitatives

Les différentes situations d'accord ou de déviation entre les deux méthodes sont résumées au Tableau 1.

Tableau 1. Types d'accord et déviation de résultats entre les méthodes de référence et alternative

Résultats	Méthode de référence positive	Méthode de référence négative	Total
Méthode alternative positive	<i>PA</i> Accord positif +/+	<i>PD</i> Déviation positive -/+	<i>A+</i>
Méthode alternative négative	<i>ND</i> Déviation négative +/-	<i>NA</i> Accord négatif -/-	<i>A-</i>
Total	$N+ = PA+ND$	$N- = NA+PD$	<i>N</i>

3.2.1 Déviation positive (PD)

La méthode alternative présente une déviation positive si elle donne un résultat positif lorsque la méthode de référence donne un résultat négatif. Une déviation positive devient un faux résultat positif lorsqu'il peut être démontré que le vrai résultat est négatif. Une déviation positive est considérée comme vrai positif lorsqu'il peut être démontré que le vrai résultat est positif.

3.2.2 Déviation négative (ND)

La méthode alternative présente une déviation négative si elle donne un résultat négatif lorsque la méthode de référence donne un résultat positif. Une déviation négative devient un faux résultat négatif lorsqu'il peut être démontré que le vrai résultat est positif.

3.2.3 Sensibilité relative (SE)

Capacité de la méthode alternative à détecter l'analyte lorsque la méthode de référence le détecte.

$$SE = 100 \times \frac{PA}{N+}$$

3.2.4 Exactitude relative (AC)

Exactitude relative

$$AC = 100 \times \frac{PA+NA}{N}$$

3.2.5 Spécificité relative (SP)

Capacité de la méthode alternative à ne pas détecter l'analyte lorsque la méthode de référence ne le détecte pas.

$$SP = 100 \times \frac{NA}{N -}$$

3.2.6 Inclusivité et exclusivité (Sélectivité)

L'inclusivité est la capacité de la méthode alternative à détecter l'analyte cible à partir d'un éventail approprié de souches.

L'exclusivité est la capacité de la méthode alternative à ne pas détecter d'analytes non cibles à partir d'un éventail approprié de souches.

3.2.7 Niveau de détection relatif ou LOD₅₀

Le niveau de détection relatif correspond au nombre le plus petit de microorganismes cultivables qu'il est possible de détecter dans l'échantillon, avec une probabilité de 50% à l'aide des méthodes alternative et de référence. Le LOD₅₀ est estimé à l'aide d'un modèle statistique mais jamais directement mesuré.

3.2.8 Praticabilité

Aptitude à l'usage d'une méthode, définie par un ensemble de critères renseignés par l'utilisateur.

3.3 Critères de validation pour les méthodes quantitatives

3.3.1 Linéarité

Aptitude de la méthode, pour une matrice donnée, à fournir des résultats proportionnels à la quantité d'analyte présente dans l'échantillon, c'est-à-dire qu'à une augmentation de l'analyte correspond une augmentation linéaire ou proportionnelle des résultats.

3.3.2 Exactitude

Étroitesse de l'accord entre le résultat d'essai et la valeur de référence acceptée.

3.3.3 Exactitude relative

Niveau de correspondance entre la réponse obtenue avec la méthode de référence et la réponse obtenue avec la méthode alternative sur les mêmes échantillons.

NOTE Le terme « exactitude relative » utilisé dans ce contexte vient compléter le 3.3.2. Pour les besoins du présent document, la valeur de référence acceptée est choisie comme étant la valeur obtenue par la méthode de référence. Le terme "relatif" implique donc que la méthode de référence ne fournit pas automatiquement la valeur de référence acceptée.

3.3.4 Répétabilité

Étroitesse de l'accord entre les mesurages successifs de la même quantité, effectués dans

les mêmes conditions de mesurage.

3.3.5 Reproductibilité

Étroitesse de l'accord entre les mesurages successifs de la même quantité, effectués dans les conditions de reproductibilité.

3.3.6 Limite de détection

(VIM. 4.18) Valeur mesurée, obtenue par une procédure opératoire donnée, pour laquelle la probabilité de déclarer faussement l'absence d'un constituant dans un matériau est β , étant donnée la probabilité α de déclarer faussement sa présence.

Ou

(Projet de norme XP T 90-465-1) Niveau de charge bactérienne le plus faible pour lequel la probabilité d'un résultat positif (observation d'au moins un germe) est supérieure ou égale à 95% ou 99%. Note : cette définition met en jeu une loi de probabilité théorique de type Poisson ou binomiale négative.

3.3.7 Limite de quantification (ou limite de détermination)

Concentration moyenne d'analyte la plus faible qui peut être mesurée et quantifiée avec une incertitude connue, dans les conditions expérimentales de la méthode considérée.

3.3.8 Inclusivité et exclusivité (Sélectivité)

L'inclusivité est la capacité de la méthode alternative à détecter l'analyte cible à partir d'un large éventail de souches.

L'exclusivité est l'absence d'interférences par un éventail approprié de souches non cibles de la méthode alternative.

NOTE Ce critère ne s'applique pas au dénombrement de la flore totale.

3.3.9 Praticabilité (voir 3.2.8)

3.3.10 Limite d'acceptabilité

Écart maximal acceptable entre la valeur assignée par la méthode de référence et un résultat obtenu par la méthode alternative. On le note λ . Dans le cas des comptages bactériens il est exprimé en unités log.

4 Principe

Le protocole de validation comprend deux étapes :

1. une étude préliminaire réalisée par un seul laboratoire dit « laboratoire expert » qui vise à estimer les performances de la méthode alternative, par rapport à la méthode de référence, et qui est réalisée sur des échantillons naturels, artificiels ou des matériaux de référence ;
2. une étude interlaboratoire organisée par le laboratoire expert avec un nombre minimum fixé de laboratoires impliquant l'utilisation d'échantillons artificiellement dopés et leur analyse en parallèle par la méthode alternative et la méthode de référence.

Avant de démarrer l'étude, il convient de définir si le champ d'application de la méthode alternative est compatible avec celui de la méthode de référence.

Cela inclut notamment :

- l'identification du ou des microorganismes ciblés ;
- les catégories d'eau concernées et la gamme de concentration visée.

Les différentes catégories d'eaux sur lesquelles l'étude peut porter sont résumées à l'Annexe 1. Classification par catégories d'eau.

5 Étude préliminaire

5.1 Méthodes qualitatives

5.1.1 Protocole d'évaluation de l'inclusivité et de l'exclusivité (sélectivité)

Pour étudier la sélectivité de la méthode alternative, analyser des souches selon la liste du Tableau 2.

Tableau 2. Nombres minimums de souches pour l'étude de la sélectivité

	Types de souches	
Inclusivité	Souches cibles (espèce) (par ex. <i>E. coli</i> ou <i>Legionella pneumophila</i>)	20
	Souches cibles appartenant à un genre, un groupe ou une famille (par ex. coliformes, <i>Legionella</i> spp, ou entérocoques intestinaux)	30
Exclusivité	Souches non cibles connues pour être interférentes avec les souches cibles (genre-espèce)	20
	Souches non cibles connues pour être interférentes avec les souches cibles appartenant à un groupe ou une famille	30

NOTE Le nombre de souches à tester pourra être augmenté en fonction de la cible (germe) de la méthode alternative à valider (exemple : 50 souches cibles en inclusivité dans le cas d'une recherche de salmonelles).

Choisir les souches cibles et non cibles qui contaminent régulièrement la ou les classes et/ou catégories d'eaux sélectionnées. Par ailleurs, les souches doivent si possible majoritairement provenir d'échantillons naturels représentatifs du domaine d'application visé d'eau. L'origine de chaque souche doit être connue, renseignée et traçable.

Le niveau d'inoculation de chaque microorganisme est défini comme suit :

- Microorganismes cibles (inclusivité) : il doit être de l'ordre de 10 à 100 fois le niveau de détection relatif minimal de la méthode alternative. Le protocole complet de la méthode alternative – sans y inclure la prise d'essai – doit être mis en œuvre, mais y compris le pré-enrichissement s'il existe. Si le principe de la méthode de référence est différent de celui de la méthode alternative, les deux méthodes seront mises en œuvre en parallèle.

- Microorganismes non-cibles (exclusivité) : il doit être de l'ordre de 1000 à 100000 fois le niveau de détection relatif minimal de la méthode alternative. Si le milieu d'enrichissement est un bouillon sélectif, le remplacer par un bouillon non sélectif. Lorsque la méthode alternative donne des résultats positifs ou douteux avec les microorganismes non ciblés, l'essai doit être répété en suivant le protocole complet. Le laboratoire pourra réaliser en parallèle la méthode de référence s'il le souhaite.

En cas de discordance de résultats entre la méthode alternative et la méthode de référence, le laboratoire devra investiguer pour expliquer les résultats.

L'ensemble de ces résultats doit être consigné dans le rapport final et leur interprétation doit faire l'objet d'un commentaire de la part du laboratoire expert.

5.1.2 Protocole de mesure de l'exactitude, sensibilité et spécificité relatives

5.1.2.1 Sélection et préparation des échantillons

Il est important de trouver des échantillons d'eau naturellement contaminés. D'une façon générale, il faut étudier l'ensemble des catégories/types pour lesquelles la validation est demandée par le fabricant. Par conséquent, si la validation est demandée pour toutes les catégories, il faut les étudier toutes. Se référer à l'Annexe 1. Classification par catégories d'eau.

Dans le cadre de l'analyse d'échantillons naturellement contaminés, il convient que la gamme des niveaux de contamination des échantillons et leur répartition soient représentatives des niveaux normalement relevés dans ces classes d'eaux.

S'il est impossible de réunir un nombre suffisant d'échantillons naturellement contaminés dans chacune des catégories, il est autorisé de recourir à une contamination artificielle. Il convient que la méthode de dopage et les niveaux ciblés permettent d'obtenir des échantillons dont le comportement simule au mieux celui d'échantillons naturellement contaminés (par exemple, choisir une faible contamination pour les eaux de production ou de distribution). Ainsi, les microorganismes utilisés pour le dopage doivent notamment avoir été soumis à des stress dont la description sera renseignée dans le rapport final.

Le stress appliqué aux suspensions microbiennes qui serviront à doper les échantillons artificiels devra être approprié au type d'eau concerné et à la souche choisie. Les types de stress applicables par type d'eau sont définis en Annexe 1.

Les types et modalités de stress, ainsi que l'intensité du stress appliqué (à minima 0,5 log) devront être renseignés dans le rapport d'étude préliminaire du laboratoire expert.

Le nombre total d'échantillons à analyser est de 60 par catégorie d'eaux. Il est préférable d'obtenir environ 50% de résultats positifs et 50 % de résultats négatifs. Pour chaque catégorie d'eau, tous les types d'eau référencés en Annexe 1. Classification par catégories d'eau doivent être testés, avec une répartition équitable des échantillons entre les types représentatifs du domaine d'application de la méthode à valider.

Les méthodes de référence et alternative doivent toujours être mises en œuvre à partir de deux prises d'essai du même échantillon. Ensuite, chaque prise d'essai subit la totalité du mode opératoire de la méthode. Cependant, il arrive fréquemment que les premières étapes des deux méthodes soient communes ; dans ce cas, les deux prises d'essai sont réalisées à l'issue de la dernière étape commune (voir Annexe).

Remarque Dans le cas d'une contamination artificielle, ensemercer directement l'échantillon puis homogénéiser avant les prises d'essai.

5.1.2.2 Interprétation des résultats

Rassembler les couples de résultats obtenus avec les méthodes alternative et de référence sous la forme d'un tableau, comme le Tableau 3, en utilisant les conventions décrites au Tableau 1.

Tableau 3. Calcul de l'exactitude relative, de la sensibilité relative et de la spécificité relative

Catégorie	PA	NA	ND	PD	N	Exactitude relative AC	N+	Sensibilité relative SE	N-	Spécificité relative SP
Eau 1										
Eau 2										
...										
Total										

Pour chaque catégorie d'eau, calculer les paramètres suivants :

$$AC = 100 \times \frac{PA + NA}{N} \quad \text{Exactitude relative}$$

$$SE = 100 \times \frac{PA}{N+} \quad \text{Sensibilité relative}$$

$$SP = 100 \times \frac{NA}{N-} \quad \text{Spécificité relative}$$

L'étude des résultats discordants se fait à l'aide du décompte de PD et ND, en appliquant, par exemple, le test de McNemar tel qu'il est décrit à l'annexe F de la norme EN ISO 16140.

Quand les valeurs de PD et ND sont élevées et proches, aucune différence statistique entre les deux méthodes ne peut être mise en évidence par le test McNemar. Dans ce cas, le laboratoire expert doit porter une attention particulière quant à la raison de ces valeurs PD et ND élevées.

5.1.2.3 Interprétation

La capacité de la méthode alternative à donner un nombre de résultats positifs supérieur ou inférieur à la méthode de référence est évaluée, en tenant compte du nombre de déviations positives et du nombre de déviations négatives.

Le rapport de l'étude doit présenter de manière distincte les résultats obtenus sur des échantillons naturellement et artificiellement contaminés.

À titre d'illustration des valeurs prises par les critères de validation, la Figure 1 présente 10 cas extrêmes afin de mieux comprendre comment interpréter les résultats. L'échelle de gauche donne les comptages pour 100 échantillons et celle de droite les valeurs des critères en %.

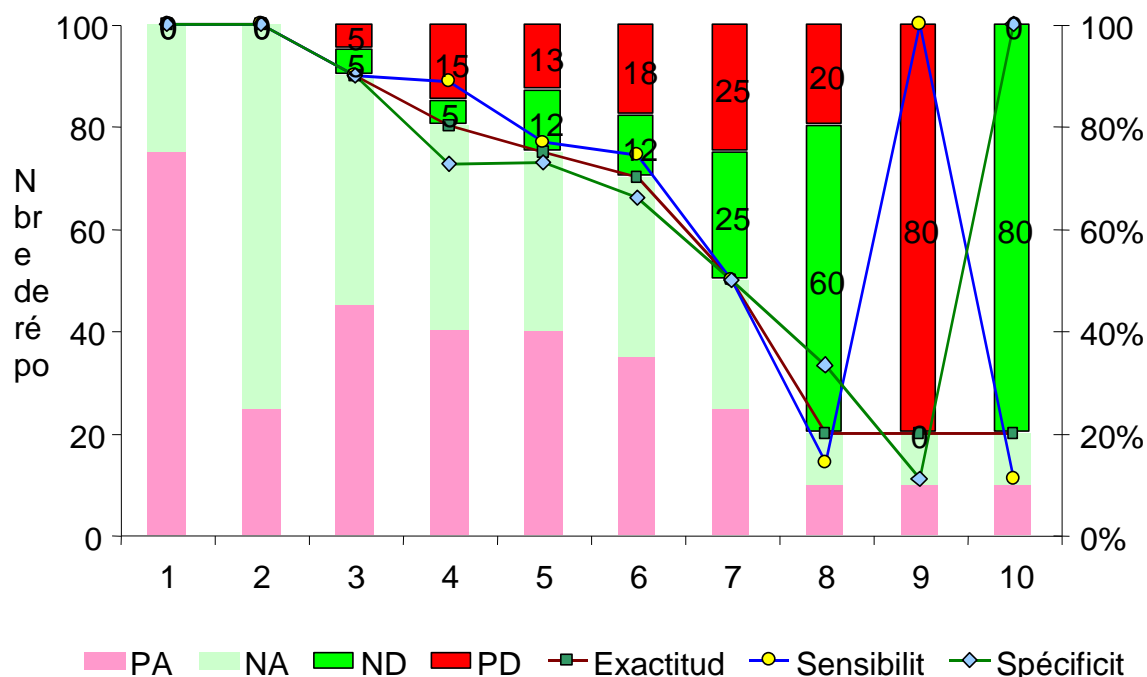


Figure 1. Exemples de calcul des critères de validation pour différentes situations.

5.1.3 Protocole de mesure du niveau de détection relatif

5.1.3.1 Plan d'expérience

Le niveau de détection relatif est défini en 3.2.7. Pour le mesurer, tester en parallèle avec la méthode de référence et la méthode alternative selon le plan d'expérience suivant. Pour chaque catégorie d'eau, choisir une souche habituellement présente. Cependant, veiller à diversifier les souches, de façon à ce que deux catégories d'eau ne soient pas associées à la même souche.

Déterminer initialement, le niveau de contamination en flore associée de chaque catégorie d'eau choisie.

Pour chaque combinaison « souche/catégorie », préparer indépendamment 6 échantillons dopés pour, au minimum 4 niveaux de contamination, y compris un contrôle négatif ; soit un nombre total d'au moins de 24 échantillons dopés par combinaison. Le choix des niveaux de contamination suit la recommandation suivante :

1. le Niveau 1 correspond à l'absence de contamination ;
2. le Niveau 2 correspond à une contamination qui donne moins de 50% de résultats positifs ($\leq 3/6$) avec l'une des deux méthodes ;
3. le Niveau 3 doit donner plus de 50% des résultats positifs ;
4. le Niveau 4 sera tel que 100 % des résultats seront positifs.

Remarque Il est possible que le laboratoire expert soit amené à réaliser un inoculum à un niveau inférieur à 1 cellule par prise d'essai. La maîtrise des faibles contaminations passe par un étalonnage précis et une expérience acquise sur les souches testées.

En fonction des méthodes, organiser les mesures selon les recommandations de l'Annexe .

5.1.3.2 Calcul et interprétation

Pour chaque niveau de contamination et chaque combinaison « catégorie/souche », estimer le niveau de détection relatif au moyen du point limite à 50 % appelé LOD₅₀. Ce calcul est réalisé selon la méthode de Spearman-Kärber qui fournit aussi un intervalle de confiance associé. Le LOD₅₀ peut aussi s'interpréter comme le niveau de contamination qui correspondrait théoriquement à un taux de récupération de 50%.

Le niveau de détection relatif global de la méthode s'exprime en prenant le minimum et le maximum de tous les intervalles de confiance des divers LOD₅₀.

5.1.4 Praticabilité

Les 13 critères suivants sont à renseigner par le laboratoire expert.

Tableau 4. Liste des critères de praticabilité recensés pour caractériser la méthode alternative

	Critères à contrôler	Communication sur le critère	Méthode de contrôle du critère
1	Mode de conditionnement des éléments de la méthode	Emballage ou notice	Vérification par le laboratoire expert
2	Volume des réactifs	Emballage ou notice	Vérification par le laboratoire expert
3	Conditions de stockage des éléments (date de péremption des produits non ouverts)	Emballage ou notice	Vérification par le laboratoire expert que les conditions existent
4	Modalités d'utilisation après première utilisation (dates limites d'utilisation)	Emballage ou notice	Vérification par le laboratoire expert que les modalités existent
5	Équipements ou locaux spécifiques nécessaires	Notice	Vérification par le laboratoire expert de la véracité des écrits
6	Réactifs prêts à l'emploi ou à reconstituer (dans ce cas existence d'un mode opératoire)	Emballage ou notice	Vérification par le laboratoire expert de la véracité des écrits
7	Durée de formation de l'opérateur non initié à la méthode	Rapport	Mesurée par le laboratoire expert (possibilité de se servir des durées mises en œuvre par les laboratoires collaborateurs) et réparti dans une des 3 catégories suivantes: moins de 1 jour, entre 1 jour et une semaine, plus d'une semaine.
8	Temps réel de manipulation et flexibilité de la méthode en fonction du nombre d'échantillons à analyser.	Rapport	Temps de manipulation mesuré en comparaison avec la méthode de référence: inférieur, égal ou supérieur au temps de manipulation de la méthode de référence

	Critères à contrôler	Communication sur le critère	Méthode de contrôle du critère
9	Délai d'obtention des résultats	Rapport et attestation	Établissement de 2 cycles décrivant chaque étape de la méthode uniquement en terme de temps: -1er cycle: échantillons négatifs- 2ème cycle: échantillons positifs
10	Type de qualification de l'opérateur	Rapport	Précisé par le laboratoire expert par rapport au niveau requis pour la méthode de référence : niveau identique ou différent de celui nécessaire pour la méthode de référence (le laboratoire expert peut se servir des données des laboratoires collaborateurs)
11	Étapes communes avec la méthode de référence	Rapport	Vérification par le laboratoire expert
12	Moyens de traçabilité des résultats d'analyse pour l'utilisateur	Notice	Vérification par le laboratoire expert
13	Obligations de maintenance d'un appareillage spécifique pour l'utilisateur	Rapport	Durée et fréquence

5.2 Méthodes quantitatives

5.2.1 Principe et définitions

Pour les méthodes quantitatives, l'objectif de l'étude préliminaire est de mesurer les performances de la méthode alternative et, si besoin est, de la méthode de référence pour les paramètres suivants :

- sélectivité, spécificité (inclusivité/exclusivité)
- linéarité, exactitude relative
- limite de détection, limite de quantification
- praticabilité

5.2.2 Protocole d'évaluation de l'inclusivité et de l'exclusivité (sélectivité)

Pour étudier la sélectivité de la méthode alternative, analyser des souches selon la liste du Tableau 5.

Ce protocole ne s'applique pas aux méthodes non sélectives (par ex. le dénombrement de germes aérobies revivifiables).

Tableau 5. Nombres minimums de souches pour l'étude de la sélectivité

		Types de souches	
Inclusivité		Souches cibles (espèce) (par ex. <i>E. coli</i> ou <i>Legionella pneumophila</i>)	20
		Souches cibles appartenant à un genre, un groupe ou une famille (par ex. coliformes, <i>Legionella</i> spp., ou entérocoques intestinaux)	30
Exclusivité		Souches non cibles connues pour être interférentes avec les souches cibles (genre-espèce)	20
		Souches non cibles connues pour être interférentes avec les souches cibles appartenant à un groupe ou une famille	30

Choisir les souches cibles et non cibles qui contaminent régulièrement la ou les classes et/ou catégories d'eaux sélectionnées. Par ailleurs, les souches doivent si possible majoritairement provenir d'échantillons naturels représentatifs du domaine d'application visé d'eau. L'origine de chaque souche doit être connue, renseignée et traçable.

Le niveau d'inoculation de chaque microorganisme est défini comme suit :

- Microorganismes cibles (inclusivité) : dénombrer une suspension de germe cible par la méthode alternative et sur un milieu non sélectif. Si le principe de la méthode de référence est différent de celui de la méthode alternative, les deux méthodes seront mises en œuvre en parallèle.
- Microorganismes non-cibles (exclusivité) : utiliser des concentrations de germe non cible de l'ordre de 10^4 à 10^5 UFC / volume analysé.

En cas de discordance de résultats entre la méthode alternative et la méthode de référence, le laboratoire devra investiguer pour expliquer les résultats.

L'ensemble de ces résultats doit être consigné dans le rapport final et leur interprétation doit faire l'objet d'un commentaire de la part du laboratoire expert.

5.2.3 Protocole de mesure de la linéarité et de l'exactitude relative

NOTE La validation s'applique à l'ensemble du domaine d'application de la méthode alternative. Les niveaux de contamination choisis doivent inclure les limites réglementaires, si elles existent, et le maximum du domaine d'application. Si le calcul de la limite de quantification est important, il est conseillé de fixer un niveau à une fois la LOQ supposée et un autre à deux fois la LOQ supposée.

5.2.3.1 Linéarité

Pour vérifier la linéarité, pour chaque catégorie d'eau étudiée telle que décrite en Annexe 1, préparer J échantillons dont le niveau de concentration est connu, tels que des matériaux de référence ou des échantillons naturels dopés. Le nombre J est appelé nombre de niveaux, il dépend du domaine d'application visé par la méthode alternative et doit être ≥ 3 .

À chaque niveau et pour chaque catégorie d'eau, effectuer k répétitions complètes et indépendantes avec chaque méthode ($k \geq 2$). On notera y_{jk} le résultat obtenu avec la méthode alternative et x_{jk} celui obtenu avec la méthode de référence. Ces deux mesures sont apparées.

La linéarité peut généralement être étudiée de façon graphique en reportant sur un diagramme les couples de valeurs obtenues avec les 2 méthodes (référence et alternative). Si besoin est, on peut calculer par une méthode de régression linéaire la droite $y_{jk} = a_0 + a_1 x_{jk}$ qui doit passer entre les points du diagramme. Sa pente a_1 doit être proche de 1 et son or-

donnée à l'origine a_0 proche de 0. Cependant, il n'est pas utile de réaliser des tests statistiques de conformité.

Les valeurs brutes obtenues pour les échantillons analysés sont à présenter sous la forme de graphiques bidimensionnels par catégories d'eau et éventuellement par type d'eau si nécessaire. Les résultats des comptages par catégorie sont à présenter sur les graphiques bidimensionnels en log N et en UFC.

5.2.3.2 Exactitude relative

Pour chaque catégorie d'eau testée, à minima 20 échantillons seront analysés en duplicat avec chacune des méthodes, soient 80 analyses au total par catégorie, avec une répartition homogène pour chaque type d'eau et un minimum de 6 à 7 échantillons par type.

Le stress appliqué aux suspensions microbiennes qui serviront à doper les échantillons artificiels devra être approprié au type d'eau concerné et à la souche choisie. Les types de stress applicables par type d'eau sont définis en Annexe 2.

Les types et modalités de stress ainsi que l'intensité du stress appliqué (à minima 0,5 log) devront être renseignées dans le rapport d'étude préliminaire du laboratoire expert.

L'exactitude relative entre la méthode alternative et la méthode de référence est évaluée en calculant pour chaque couple de valeurs les différences $d_{jk} = y_{jk} - x_{jk}$. Pour chaque niveau j regrouper les valeurs de d_{jk} et calculer une exactitude relative du niveau par la médiane :

$$D_i = \text{mediane}(d_{jk})$$

La relation entre les exactitudes relatives D_j et les valeurs des niveaux (x_j) peut être modélisée. Une valeur nulle ($D = 0$) pour chaque niveau est attendue. Si les valeurs D sont trop grandes pour tout ou partie du domaine d'application, on peut conclure que la méthode alternative diffère trop de la méthode de référence et ne mérite pas une validation plus approfondie. Par ailleurs, les valeurs de D permettront de commencer à définir une valeur pour la limite d'acceptabilité (λ) qui servira à interpréter le profil d'exactitude (chapitre 6.2.1). Cependant, il faudra aussi tenir compte de la répétabilité et de l'objectif de la méthode pour complètement fixer la valeur de λ .

Les valeurs brutes obtenues pour les échantillons analysés sont à présenter sous la forme de graphiques bidimensionnels par catégories d'eau et éventuellement par type d'eau si nécessaire. Les résultats des comptages par catégorie sont à présenter sur les graphiques bidimensionnels en log N et en UFC.

En cas de discordance de résultats entre la méthode alternative et la méthode de référence, le laboratoire devra investiguer pour expliquer les résultats.

5.2.4 Limites de détection et de quantification

NOTE Le calcul doit prendre en considération le type de matrice et les valeurs des seuils réglementaires.

Les limites de détection et de quantification sont vérifiées, en fonction des préconisations du fabricant et de la méthode de référence, conformément au 6.2.2.4 de la norme EN ISO 16140.

Les limites de détection et de quantification sont déterminées en analysant une culture pure de la souche cible. Au minimum, 6 répétitions pour chaque niveau sont testées dans de l'eau stérile. Au minimum, 3 à 4 niveaux doivent être testés (en plus du niveau 0) dont au moins un niveau avec moins de 50% de réponse positive et un niveau avec plus de 50% de réponse positive.

5.2.5 Praticabilité de la méthode alternative

Les 13 critères qui sont à renseigner par le laboratoire expert sont listés au tableau 4.

6 Étude interlaboratoire

L'étude interlaboratoire doit permettre une évaluation des performances de la méthode dans plusieurs laboratoires dans les conditions réelles de son application en routine. Elle est organisée sous le contrôle du laboratoire expert et suit généralement l'étude préliminaire. Son interprétation est orientée « méthode » et non pas « laboratoire ».

6.1 Méthodes qualitatives

6.1.1 Plan d'expérience

L'étude interlaboratoire applicable aux méthodes qualitatives doit impliquer au minimum 8 laboratoires collaborateurs dont les résultats ne présentent aucune valeur aberrante.

Il est nécessaire que le laboratoire collaborateur fasse la preuve de ses compétences quant à l'utilisation de la méthode alternative et de la méthode de référence pour pouvoir participer à l'étude.

- une matrice d'eau appropriée est utilisée pour préparer les échantillons d'essai ;
- chaque échantillon doit êtreensemencé individuellement ;
- un minimum de 3 niveaux de contamination différents doit être utilisé : un contrôle négatif (L_0), un niveau légèrement supérieur au niveau de détection relatif de la méthode alternative (L_1), et un niveau égal à 10 fois environ le niveau de détection (L_2) (par exemple 0, 3 et 30 cellules / 25 mL) ;
- la stabilité du microorganisme dans la matrice eau utilisée doit être démontrée ;
- répétitions en aveugle de chaque niveau de contamination sont analysés avec la méthode de référence et avec la méthode alternative dans chacun des laboratoires collaborateurs, soit $X=24$ échantillons à analyser par laboratoire ;
- si la méthode alternative et la méthode de référence n'ont pas d'étape commune, le laboratoire expert prépare et envoie à chaque laboratoire collaborateur les 24 échantillons en double ($2X$) ;
- l'analyse des échantillons doit être réalisée dans chaque laboratoire à la date stipulée ;
- la température de l'échantillon témoin ne doit pas dépasser 10°C pendant le transport, jusqu'à réception des échantillons par le laboratoire collaborateur. L'utilisation d'un enregistreur de température est obligatoire pour tracer l'évolution de la température pendant l'expédition (par exemple établir les profils thermiques).

Le laboratoire expert exploitant tous les données consignées doit déterminer quels résultats sont corrects et lesquels sont des valeurs aberrantes pour le calcul des données de fidélité.

6.1.2 Calculs et interprétation

Classer les résultats positifs obtenus à chaque niveau par chacune des méthodes comme illustré dans le Tableau 6 et le Tableau 7.

Tableau 6. Résultats positifs obtenus avec la méthode de référence

Laboratoires	Niveau de contamination		
	L_0	L_1	L_2
Laboratoire 1	/8	/8	/8
Laboratoire 2	/8	/8	/8
Laboratoire 3	/8	/8	/8
Etc.	/8	/8	/8
Total	FP	$TP1r$	$TP2r$

FP : faux positif obtenu avec la méthode de référence

$TP1r$: vrai positif obtenu au niveau 1 avec la méthode de référence

$TP 2r$: vrai positif obtenu au niveau 2 avec la méthode de référence

Tableau 7. Résultats positifs obtenus avec la méthode alternative

Laboratoires	Niveau de contamination		
	L_0	L_1	L_2
Laboratoire 1	/8	/8	/8
Laboratoire 2	/8	/8	/8
Laboratoire 3	/8	/8	/8
Etc.	/8	/8	/8
Total	FP	$TP1r$	$TP2r$

FP : faux positif obtenu avec la méthode alternative

$TP1r$: vrai positif obtenu au niveau 1 avec la méthode alternative

$TP 2r$: vrai positif obtenu au niveau 2 avec la méthode alternative

Calculer le pourcentage de spécificité SP pour le niveau L_0 et pour chaque méthode à l'aide l'équation suivante :

$$SP = 100 \times \left(1 - \frac{FP}{N-} \right)$$

Avec $N-$ est le nombre total de tous les essais L_0 ;

FP est le nombre de faux positifs.

Calculer le pourcentage de sensibilité SE pour chaque niveau de contamination positif (L_1 et L_2) à l'aide de l'équation suivante :

$$SE = 100 \times \left(\frac{TP}{N+} \right)$$

Avec $N+$ est le nombre total des essais L_1 ou L_2 respectivement ;

TP est le nombre de vrais positifs.

Comparer la méthode alternative et la méthode de référence pour chaque niveau de contamination et pour la totalité des résultats afin de calculer l'exactitude relative et d'étudier les résultats discordants.

Les couples de résultats d'un échantillon mesuré avec la méthode alternative et la méthode de référence doivent être consignés comme illustré au Tableau 8.

Tableau 8. Couples de résultats de la méthode alternative et de la méthode de référence dans le cadre de l'étude collaborative.

Méthode alternative	Méthode de référence		Total
	+	-	
+	PA	PD	
-	ND	NA	
Total	N+	N-	N

Calculer l'exactitude relative AC exprimée en pourcentage, à l'aide de l'équation suivante :

$$AC = 100 \times \left[\frac{PA + NA}{N} \right]$$

Calculer les intervalles de confiance pour chaque proportion (Cf. 5.1.2.2.2 de la norme EN ISO 16140).

Étudier les résultats discordants (cf. 5.1.2.2.3 de la norme EN ISO 16140).

6.1.3 Interprétation

Comparer AC, SE et SP avec leurs valeurs correspondantes, obtenues dans le cadre de l'étude préliminaire. Ces critères ne traitent pas réellement de la variabilité de la méthode au sein d'un laboratoire et entre laboratoires.

Les critères degré d'accord, concordance et odds ratio sont utilisés pour aider à approcher la cette notion de variabilité (répétabilité et reproductibilité) (cf. Annexe 5 du présent document)

6.2 Méthodes quantitatives

6.2.1 Profil d'exactitude : principes généraux

Le profil d'exactitude permet une approche statistique de la validation. Il est basé sur l'application directe des principes décrits dans les normes de la série ISO 5725, pour les parties 1, 2 et 3. Ces normes décrivent une méthode statistique pour estimer l'exactitude (justesse et fidélité) d'une méthode ou de résultats. Dans ce cas, le modèle qui est appliqué pour expliquer un mesurage z d'un mesurande Z provenant d'un laboratoire est du type :

$$z = m + B + e$$

où m est la moyenne générale de l'échantillon homogène envoyé aux laboratoires, B la composante du biais du laboratoire sous condition de répétabilité et e l'erreur aléatoire survenant dans chaque mesurage sous condition de répétabilité. La composante du biais provient du laboratoire, dans le cas d'une étude inter-laboratoires, mais elle peut aussi provenir d'une autre source d'incertitude, comme le jour, l'opérateur, l'instrument... dans le cas d'une étude intralaboratoire. L'application la plus connue de la norme ISO 5725 est le calcul de critères de fidélité dans le cadre d'une étude inter-laboratoires qui débouche sur les écarts-types de répétabilité et de reproductibilité. Outre les méthodes de calcul des critères d'exactitude, ces normes précisent aussi en détail l'organisation de la collecte et les précautions à respecter.

À partir de l'écart-type de reproductibilité calculé selon les recommandations de la norme ISO 5725, on construit un intervalle dit de tolérance (ou de prédiction) qui contient une proportion β de futurs résultats. **Il est important de noter que l'intervalle de tolérance ne doit pas être confondu avec l'intervalle de confiance.** Un intervalle de confiance décrit la dis-

persion de résultats autour d'un paramètre, comme la moyenne, à partir d'expériences passées. Ainsi, l'intervalle de confiance de la moyenne s'exprime comme suit, avec k_{IC} son facteur d'élargissement et $s_{\bar{x}}$ l'écart-type de la moyenne :

$$\bar{x} \pm k_{IC} \times s_{\bar{x}}$$

Un intervalle de tolérance décrit la dispersion de **futurs** résultats. Il tient compte des expériences passées et de l'incertitude sur la moyenne. Il s'exprime avec k_{IT} un autre facteur d'élargissement calculé de façon complètement différente de k_{IC} et s_R l'écart-type de reproductibilité :

$$\bar{x} \pm k_{IT} \times s_R$$

Ce calcul est répété pour chaque niveau de concentration j afin d'obtenir une série d'intervalles de tolérance qui, réunis, couvriront l'ensemble du domaine d'application de la méthode et formeront ce qu'on appelle le profil d'exactitude.

La Figure 2 donne un exemple de profil d'exactitude construit avec 3 niveaux de concentration, entre 2.2 et 4.2 unités log, qui définissent le domaine de validation (ou d'application) de la méthode à valider.

L'axe horizontal représente la concentration théorique des niveaux et l'axe vertical la différence entre la concentration théorique et la concentration retrouvée exprimée en log, soit le biais absolu. Les limites des intervalles de tolérance définissent un domaine où se situe une proportion β de futurs résultats (dans ce cas 80%).

Ensuite, le profil d'exactitude peut être comparé à un intervalle d'acceptabilité, défini en fonction de l'objectif de la méthode. Les limites des intervalles d'acceptabilité, notées $\pm\lambda$, sont elles aussi exprimées par une différence de log. La stratégie d'interprétation est décrite en détail à l'Annexe 7. Exemple de calcul du profil d'exactitude. Elle est définie comme informative car elle fait appel à une limite d'acceptabilité λ qui dépend du contexte d'utilisation de la méthode et de la proportion β choisie.

Dans le domaine délimité par les traits verticaux discontinus la méthode est capable de produire une proportion β de résultats compris entre les limites d'acceptabilité. La méthode est alors dite valide dans ce domaine indiqué par la double flèche. Le domaine d'application représente le domaine initialement choisi pour conduire la validation.

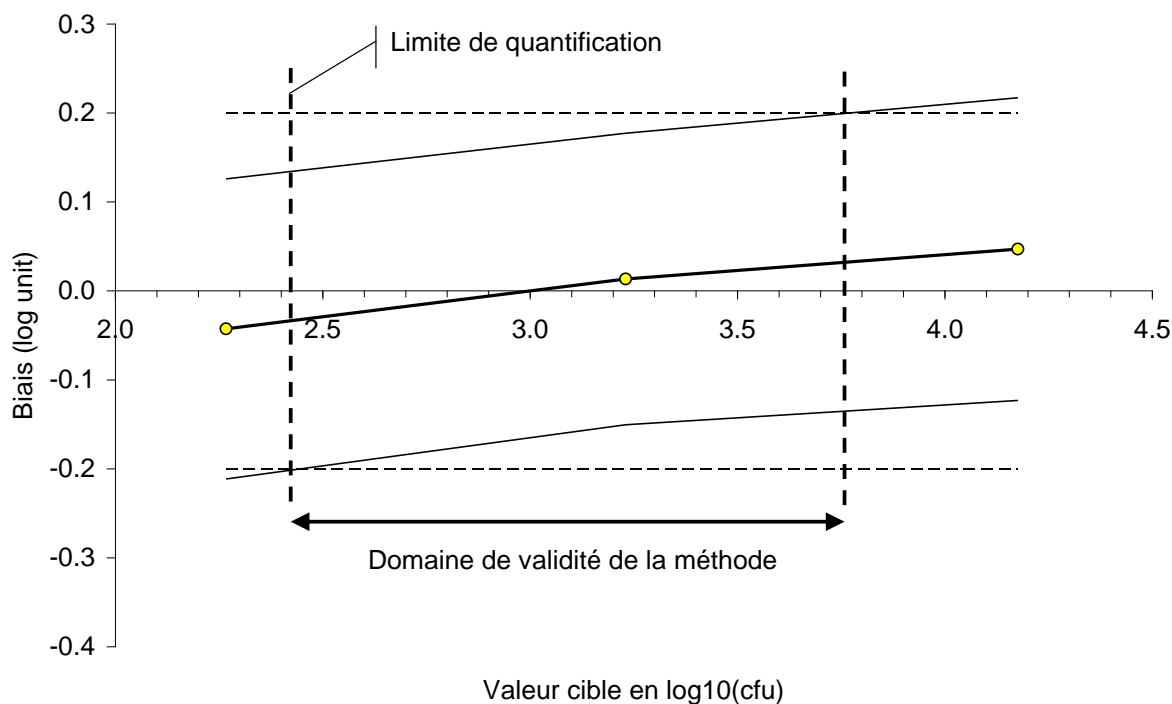


Figure 2. Profil d'exactitude calculé pour 3 niveaux de concentration. Chaque cercle gris représente la concentration moyenne retrouvée pour le niveau et quantifie la justesse. Les lignes pointillées délimitent l'intervalle d'acceptabilité et les lignes continues l'intervalle de tolérance calculé à partir des écarts-types de fidélité intermédiaire pour chaque niveau. Les traits verticaux délimitent le domaine de validité, dans lequel la méthode est capable de produire une proportion connue (80%) de résultats acceptables.

Les opérations qui conduisent à la construction du profil d'exactitude peuvent se résumer comme suit :

1. Organiser la collecte des données selon le plan d'expérience décrit au chapitre 6.2.2 ;
 2. Transformer les comptages en log₁₀ ;
- NOTE IMPORTANTE** Présenter les profils d'exactitude en log et en nombre de microorganismes, sauf cas particulier validé par le Bureau Technique.
3. Pour chaque niveau (bas, moyen et haut), calculer l'écart-type de reproductibilité avec les répétitions Z de la méthode alternative, selon la norme ISO 5725 ;
 4. Pour chaque niveau, calculer la médiane des répétitions X de la méthode de référence ; ces médianes seront les **valeurs cibles** qui devront être atteintes par la méthode alternative, en tenant compte des limites d'acceptabilité.
 5. Pour chaque niveau, calculer la moyenne des répétitions Z de la méthode alternative.
 6. Pour chaque niveau, calculer les limites de l'intervalle de tolérance selon les formules de la publication (Hubert, 2006). Rappelons que l'intervalle de tolérance définit un domaine où le microbiologiste peut garantir qu'il est capable de produire pourcentage connu (80% dans l'exemple) de futures mesures.
 7. Résumer tous ces calculs dans un tableau qui réunit les données du Tableau 10 et du Tableau 11 et une figure conçue comme la Figure 2.

6.2.2 Plan d'expérience

6.2.2.1 Organisation de l'étude interlaboratoire

L'étude interlaboratoire sert à l'évaluation de la méthode alternative relativement à la méthode de référence. Elle doit être organisée par le laboratoire expert qui sélectionne un nombre suffisant de laboratoires (au moins 10 de préférence) afin d'obtenir un minimum de 8 laboratoires n'ayant aucun mesurage rejeté comme aberrant. Elle est basée sur l'envoi d'échantillons qui seront analysés en double et en parallèle par la méthode alternative et la méthode de référence. Ces essais ont pour but de permettre le calcul des critères de performance de la méthode alternative qui serviront à construire le profil d'exactitude et prendre une décision sur sa validité. Les lignes directrices et les exigences relatives à l'organisation, à la distribution et à la gestion des études inter-laboratoires sont indiquées en Annexe 8.

Le laboratoire expert est chargé de préparer le protocole d'essai et la fiche d'enregistrement des mesurages permettant de reporter toutes les mesurages et les conditions expérimentales critiques utilisées par chacun des laboratoires. L'analyste de chaque laboratoire participant doit faire la preuve de ses compétences quant à l'utilisation de la méthode alternative et de la méthode de référence pour pouvoir participer à l'étude.

Le protocole est le suivant :

- Choisir une matrice appropriée généralement définie à l'issue de l'étude préliminaire (voir 6.2.2.3) ;
- Choisir des niveaux de concentration des échantillons artificiellement ou naturellement contaminés et qui couvrent au minimum les niveaux inférieur, moyen et supérieur du domaine d'application attendu de la méthode alternative. Il est également possible d'inclure un contrôle négatif mais les mesures réalisées n'entreront pas dans les calculs ;
- Le laboratoire expert doit démontrer l'homogénéité et la stabilité des trois niveaux de concentration testés.
- Utiliser les mêmes échantillons avec chaque méthode ;
- Pour chaque laboratoire, le laboratoire organisateur prépare 4 aliquotes pour chaque niveau (soit deux aliquotes mesurés chacun avec les deux méthodes). Il code ces aliquotes à l'aveugle, de façon à ce que deux d'entre eux soient mesurés avec la méthode de référence et deux autres avec la méthode alternative. Dans certains cas particuliers, subdiviser les aliquotes, juste avant la mesure avec les deux méthodes (voir Annexe 3. Organisation des analyses réalisées en parallèle en fonction du type de méthode) ;
- Chaque laboratoire participant doit effectuer l'analyse des aliquotes à une date stipulée en utilisant des lots courants de milieux, réactifs et/ou de kits ;
- Le laboratoire expert doit définir les règles d'arrondi des résultats.
- Le laboratoire expert regroupe les résultats des laboratoires et les trie par niveau j et par laboratoire i selon le modèle du Tableau 9.

Tableau 9. Présentation des résultats de l'étude interlaboratoire

Laboratoire <i>i</i>	Niveau <i>j</i>	Méthode de référence <i>X</i>		Méthode alternative <i>Z</i>	
		Répétition 1	Répétition 2	Répétition 1	Répétition 2
Labo 1	1	x_{111}	x_{112}	z_{111}	z_{112}
Labo 2	1				
Labo <i>i</i>	1				
Labo <i>l</i>	1				
Labo <i>i</i>	<i>J</i>				
Labo <i>l</i>	<i>J</i>				

6.2.2.2 Choix des nombres de niveaux, laboratoire participants et répétitions

Les conditions minima consistent à choisir un nombre de niveaux $J = 3$ et un nombre de répétitions $K = 2$. Il est indispensable d'avoir $J \geq 3$ car il faudra vérifier la linéarité entre les valeurs de référence et les valeurs retrouvées. On a besoin de 3 niveaux pour cette vérification. Pour la valeur du nombre de laboratoires l on peut se référer à la partie 1 de la norme ISO 5725. Pour compenser les faibles valeurs de J et K , on a décidé de prendre un nombre de laboratoires $l \geq 8$. De préférence, ces laboratoires doivent être accrédités selon la norme EN ISO 17025 mais surtout posséder une bonne expérience des méthodes microbiologiques et enfin accepter les règles de participation et d'assurance qualité (traçabilité) décrites par le laboratoire organisateur.

En final, le nombre de mesures recueillies doit donc être au moins égal à $8 \times 2 \times 2 \times 3 = 96$.

6.2.2.3 Type d'échantillons

L'étude inter-laboratoire ne porte que sur une seule matrice. Le choix de la catégorie à laquelle elle appartient doit être défini suite à l'étude préliminaire. Il est nécessaire de préparer les niveaux de concentration par des dopages pour atteindre les valeurs requises.

6.2.3 Calculs statistiques

NOTE L'exclusion des résultats d'un laboratoire doit rester exceptionnelle et doit être systématiquement documentée (enquête du laboratoire organisateur auprès du laboratoire collaborateur).

Deux fichiers Excel sont mis à disposition de façon contrôlée par AFNOR Certification et doivent être utilisés pour la construction du profil d'exactitude, l'un pour l'interprétation des données en UFC et l'autre en Log.

6.2.3.1 Vérification de la linéarité

Avant tout traitement numérique, il est toujours judicieux de procéder à une illustration graphique préalable et à un examen visuel des données afin de détecter une erreur flagrante, comme une donnée mal enregistrée. Ces graphiques peuvent être mis en annexe du dossier d'évaluation. La présence d'une donnée anormale amène à suspecter une mauvaise mise au point de la méthode et peut invalider la procédure de validation ou un laboratoire en particulier.

En microbiologie, les mesures X et Z ne se distribuent pas toujours selon la loi normale, en particulier on observe une forte dissymétrie de la distribution. En général, pour obtenir une distribution symétrique, il suffit de transformer les dénombrements en leurs logarithmes. D'autres transformations peuvent également être utilisées pour les petits nombres qui se rencontrent en microbiologie des eaux.

6.2.3.2 Calcul des valeurs cibles de référence par niveau de concentration

Utiliser les données X du Tableau 9 pour calculer la valeur cible de référence par niveau. La solution la plus simple consiste à utiliser la médiane :

$$\tilde{x}_j = \text{mediane}(x_{i,k}) \quad (1)$$

Ces médianes serviront à calculer la justesse de la méthode et l'exactitude des mesures. On peut remarquer qu'il pourrait ne pas être utile de faire autant de répétitions avec la méthode de référence, dans la mesure où elles ne sont pas toutes utilisées comme telles.

6.2.3.3 Calcul de écarts-types de fidélité par niveau de concentration

Utiliser les données Z du Tableau 9 pour calculer les écarts-types de répétabilité s_r , inter-séries s_B et de reproductibilité s_R . Ce calcul s'effectue indépendamment pour chaque niveau de concentration j (avec $1 \leq j \leq 3$) selon le principe de la norme ISO 5725-2 décrit à l'Annexe . Réunir les données dans un tableau de la forme du Tableau 10.

Tableau 10. Critères de fidélité et de justesse par niveau

Niveau		1	...	J
Nombre de laboratoires	I			
Nombre de répétitions	K			
Concentration cible théorique moyenne	\tilde{x}_j			
Concentration retrouvée moyenne	$\bar{\bar{z}}_j$			
Écart-type de répétabilité	$s(j)_r$			
Écart-type inter-séries	$s(j)_B$			
Écart-type de fidélité intermédiaire	$s(j)_R = \sqrt{s(j)_r^2 + s(j)_B^2}$			
Biais moyen absolu	$\bar{\bar{z}}_j - \tilde{x}_j$			

Il peut être délicat d'obtenir des estimations correctes de la moyenne ou de l'écart-type en présence de valeurs aberrantes. La norme ISO 5725 propose des tests de rejet des moyennes ou des variances aberrantes (tests de Grubbs et Cochran respectivement) qui permettent d'exclure les valeurs ayant une forte influence et donc d'établir une meilleure estimation. Cela a néanmoins pour inconvénient de réduire le nombre de valeurs utiles pour l'analyse statistique et il n'est pas conseillé d'y recourir pour ce type d'étude.

6.2.3.4 Calcul de l'intervalle de tolérance

Rappel. L'intervalle de tolérance définit un intervalle dans lequel on s'attend à trouver en moyenne une proportion β de futurs résultats obtenus en appliquant la méthode en routine, c'est-à-dire dans des conditions de fidélité appliquées (dans ce cas de reproductibilité). Pour cette raison, on l'appelle aussi intervalle de prédiction.

La méthode proposée par Mee [Mee, 1984] est celle qui a été choisie pour de protocole. Elle a aussi été l'objet de plusieurs publications [Hubert, 2003, Hubert 2006a, Hubert, 2006b]. Le calcul se fait à partir des données Z du Tableau 10 pour chaque niveau de concentration j (avec $3 \leq j \leq J$). Pour des raisons de simplification des formules, on a omis l'indice j dans les formules suivantes.

L'intervalle de tolérance est symétrique autour de la concentration retrouvée moyenne \bar{z} et s'exprime par la formule :

$$\bar{z} \pm k_{tol} \times s_R \quad (2)$$

où k_{tol} est un facteur d'élargissement qui vaut :

$$k_{tol} = t_{v, \frac{1+\beta}{2}} \sqrt{1 + \frac{1}{KIB^2}} \quad (3)$$

avec :

$$B = \sqrt{\frac{R+1}{IR+1}} \quad (4)$$

$$R = \frac{s_B^2}{s_r^2} \quad (5)$$

Par ailleurs, $t_{v, \frac{1+\beta}{2}}$ est le quantile de la distribution t de Student pour v degrés de liberté, calculés selon une méthode d'approximation proposée par Satterthwaite [Mee, 1984] :

$$v = \frac{(R+1)^2}{\left(R + \frac{1}{I}\right)^2 + \frac{1 - \frac{1}{I}}{KI}} \quad (6)$$

Pour chaque niveau j , calculer le facteur d'élargissement k_{tol} et les limites basse et haute de l'intervalle de tolérance. La valeur choisie pour β doit être au moins de 80%. L'ensemble des calculs est rassemblé dans un tableau de la forme du Tableau 11.

Tableau 11. Limites des intervalles de tolérance par niveau

Niveau		1	...	J
Concentration théorique moyenne	\tilde{x}			
Limite de tolérance basse	$\bar{z} - k_{tol} \times s_{FI}$			
Limite de tolérance haute	$\bar{z} + k_{tol} \times s_{FI}$			
Limite de tolérance basse différentielle	$(\bar{z} - k_{tol} \times s_{FI}) - \tilde{x}$			
Limite de tolérance haute différentielle	$(\bar{z} + k_{tol} \times s_{FI}) - \tilde{x}$			

REMARQUE Le nombre v est rarement un nombre entier et il convient d'utiliser des tables de la loi de Student qui accepte des nombres de degrés de liberté non entiers. Cependant, il est possible d'approcher la valeur du quantile par interpolation linéaire entre les 2 nombres de degrés de liberté entiers qui encadrent v .

La valeur de v dépend du rapport $R = \frac{s_B^2}{s_r^2}$ entre les variances inter-séries et de répétabilité.

Ce rapport traduit l'importance relative de l'effet de la série. Par exemple, dans le cas où l'effet jour est l'effet série, si la méthode est très stable d'un jour à l'autre et est capable de fournir des résultats très proches pour un même échantillon, la correction introduite par R sera faible. Par contre, elle augmentera avec ce rapport. En effet, plus le nombre de degrés de liberté est petit, plus le quantile de la distribution t de Student est élevé et élargit l'intervalle de tolérance.

Le tableau suivant illustre ce mécanisme pour un plan V avec $K = 3$ et $I' = 3$ et s_{FI} constant et égal à 1. Lorsque le rapport des variances R passe de 0 à 9, le facteur d'élargissement k_{tol} et la largeur de l'intervalle de tolérance passent de 1,97 à 3,07.

Tableau 12. Influence du rapport des variances sur la largeur de l'intervalle de tolérance ($p = 0.90$)

K	I'	s_r	s_B	s_{FI}	R	v	$t_{v; \frac{1+\beta}{2}}$	k_{tol}
3	3	1.00	0.00	1.0	0.0	7.714	1.869	1.970
3	3	0.71	0.71	1.0	1.0	4.154	2.109	2.332
3	3	0.58	0.82	1.0	2.0	3.219	2.290	2.569
3	3	0.50	0.87	1.0	3.0	2.842	2.408	2.722
3	3	0.45	0.89	1.0	4.0	2.642	2.489	2.826
3	3	0.41	0.91	1.0	5.0	2.518	2.549	2.902
3	3	0.38	0.93	1.0	6.0	2.434	2.594	2.959
3	3	0.35	0.94	1.0	7.0	2.374	2.629	3.004
3	3	0.33	0.94	1.0	8.0	2.328	2.658	3.041
3	3	0.32	0.95	1.0	9.0	2.292	2.681	3.070

6.2.4 Construction du profil d'exactitude

Le profil d'exactitude peut être construit de différentes façons, en fonction du type de données traité. La méthode la plus classique, lorsqu'on a à faire à des concentrations relatives, est celle représentée par la Figure 2 où les résultats sont exprimés par le taux de recouvrement ou le biais relatif. Mais si les données sont des comptages exprimés en logarithmes, il vaut mieux utiliser le biais absolu dans la mesure où une différence entre 2 quantités exprimées par leurs logarithmes est équivalente à leur rapport.

Pour construire le profil d'exactitude sélectionner dans le Tableau 10 et le Tableau 11, les 5 lignes suivantes :

1. Concentration théorique moyenne (axe horizontal)
2. Concentration retrouvée moyenne
3. Limite de tolérance basse relative
4. Limite de tolérance haute relative
5. Taux de recouvrement moyen

Reporter les données sur un graphique en utilisant la concentration théorique moyenne pour dessiner l'axe des abscisses.

Construire au moins deux profils avec $\beta=80\%$ et $\beta=90\%$, à des fins d'interprétation.

6.3 Synthèse de l'essai inter-laboratoire

6.3.1 Interprétation des résultats

Interpréter le graphique selon les règles suivantes :

- Propose une valeur pour le critère d'acceptabilité λ . Justifier ce choix en fonction du problème traité et de l'expérience antérieure. Le choix final revient au Bureau Technique.
- En fonction de la valeur de λ , la méthode alternative est dite valide dans tout le domaine où l'intervalle de tolérance est compris entre les limites d'acceptabilité.
- Enfin, définir la limite de quantification comme le point où l'intervalle de tolérance coupe une des deux limites d'acceptabilité. C'est la limite au-delà de laquelle, le microbiologiste ne peut plus garantir un pourcentage β de résultats obtenus par la méthode alternative qui soient acceptables.

6.3.2 Limite de quantification

Elle est calculée par simple interpolation linéaire entre les points où l'intervalle de tolérance coupe la limite d'acceptabilité. La Figure 3 sert à illustrer ce calcul.

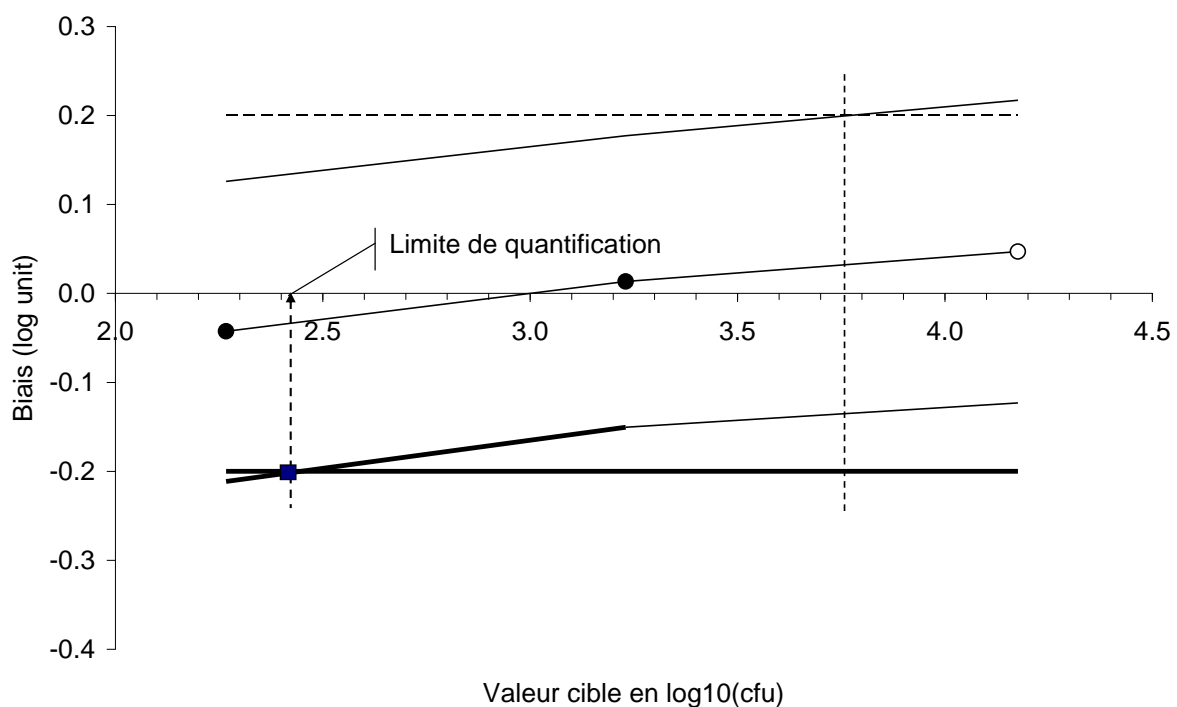


Figure 3. Calcul de la limite de quantification par interpolation

On va donc considérer la droite formée par la limite inférieure de l'intervalle de tolérance (segment indiqué en gras). Si on pose que son équation avec pour abscisse x la valeur des niveaux de la cible et pour ordonnée y la valeur de la limite de inférieure de l'intervalle de tolérance, on a :

$$y = c_0 + c_1x$$

Puisqu'on connaît les deux couples de valeurs $\{x_1, y_1\}$ et $\{x_2, y_2\}$, on peut facilement calculer la pente et l'ordonnée à l'origine par des règles de trois et prédire la valeur de l'abscisse pour la valeur $-\lambda$. Dans le cas où l'intersection se fait avec la limite supérieure, il convient de remplacer $-\lambda$ par $+\lambda$:

$$\begin{aligned} c_1 &= \frac{y_2 - y_1}{x_2 - x_1} \\ c_0 &= y_1 - c_1 x_1 \\ LOQ &= \frac{-\lambda - c_0}{c_1} \end{aligned} \quad (7)$$

Dans l'exemple présenté à la Figure 3, les valeurs trouvées sont les suivantes :

Points	1	2
Valeur de l'abscisse x (concentration cible)	2.267	3.230
Valeur de l'ordonnée y (limite inférieure de l'IT)	-0.211	-0.150
Pente	0.0633	
Origine	-0.3546	
Limite d'acceptabilité (lambda)	0.2	
LOQ	2.44	

La LOQ, d'après les choix faits pour l'acceptabilité λ et la proportion β , est de 2.4 unités log.

S'il n'y a pas d'intersection, comme le montre la Figure 4, la LOQ correspond à la valeur la plus basse du domaine de validation.

6.4 Bibliographie relative au chapitre 6

- Hubert P., Nguyen-Huu J.J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L. (2003) Validation of quantitative analytical procedure, Harmonization of approaches. Part I. STP Pharma Pratiques 13(3) 101-138.
- Hubert P., Nguyen-Huu J.J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L. (2006a) Quantative analytical procedures: Harmonization of the approaches. Part II – Statistics STP Pharma Pratiques 16(1) 28-58.
- Hubert P., Nguyen-Huu J.J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L. (2006b) Quantitative analytical procedures validation: harmonization of the approaches. Part III. Examples of application. STP Pharma Pratiques 16(2) 87-121.
- Mee R.W. (1984) β -Expectation and β -Content Tolerance Limits for Balanced One-Way ANOVA Random Model. *Technometrics* 26:3, 251-254

Annexe 1. Classification par catégories d'eau

Eaux de consommation humaine :

Eaux de puits, de source et de forage

Eaux conditionnées (bouteille ou bombonne) : eaux de source, minérales et autres

Eaux de réseau et fontaine

Eaux de baignade (non traités) :

Eaux de mer

Eaux douce

Eaux de piscine

Eaux industrielles :

Eaux de tours aéro-réfrigérantes (TAR)

Eaux chaudes sanitaires

Eaux de rejet, résiduaires brutes et traitées (STEP, ...)

Autres (Demandes spécifiques du fabricant)

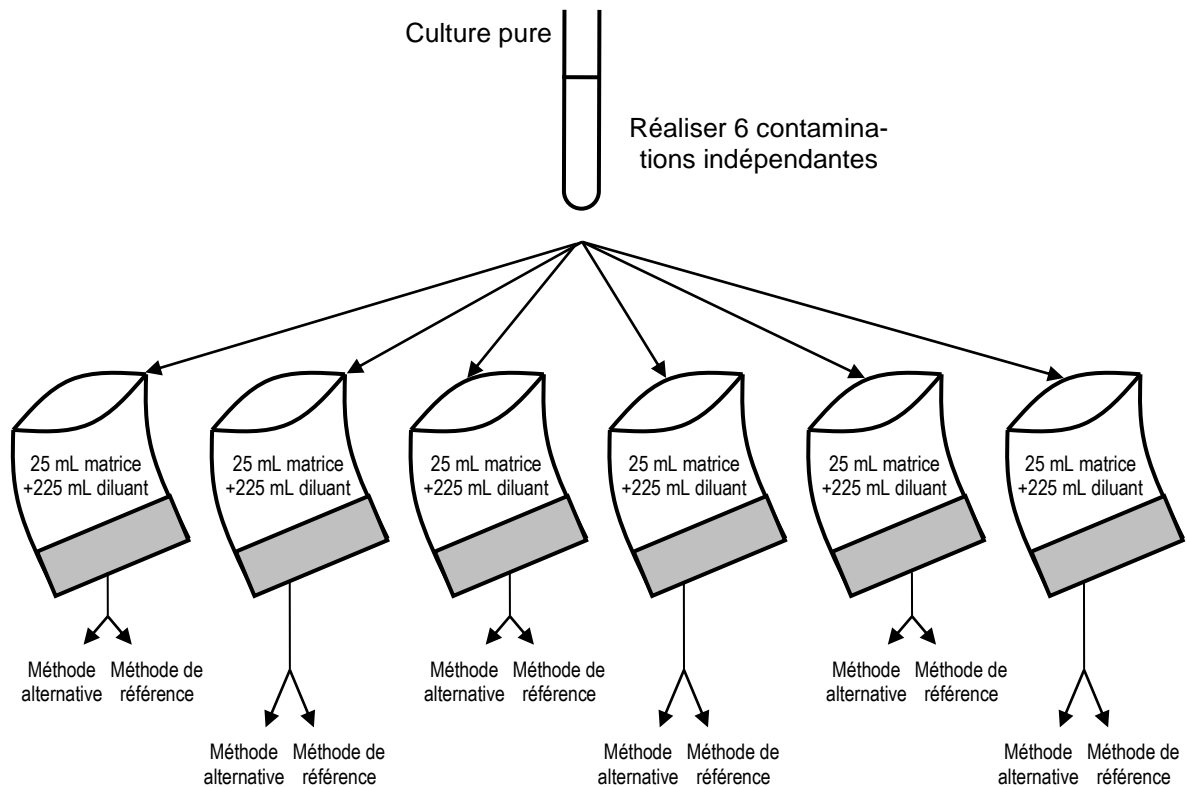
Exemples : eaux thermales (spa ...)

Annexe 2. Stress à appliquer par type d'eau

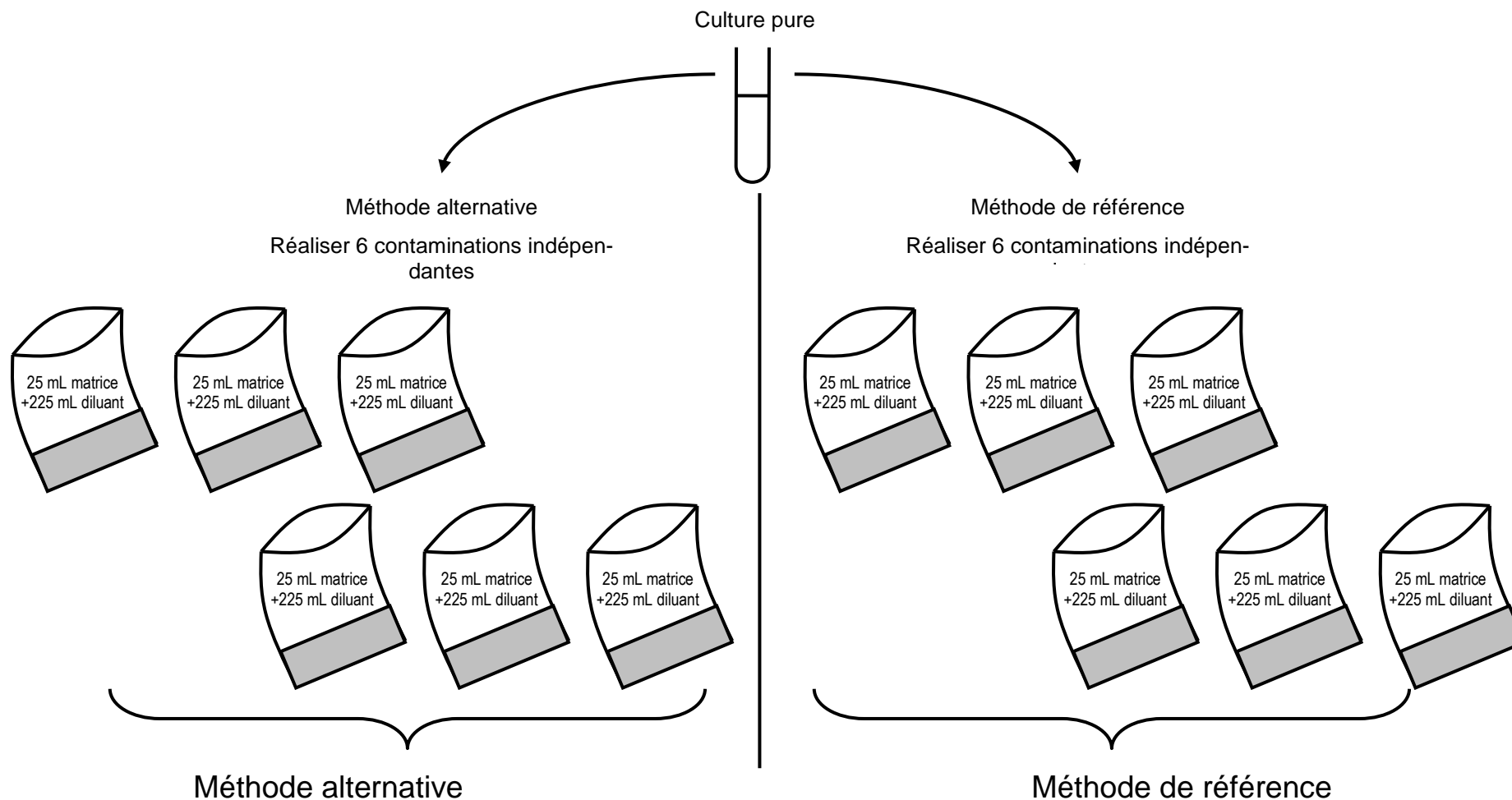
Catégorie	Type	Stress
Eaux de consommation humaine	Eaux de puits, de source et de forage	aucun
	Eaux conditionnées (bouteille ou bombonne) : eaux de source, minérales et autres	thermique (chaud et/ou froid positif)
	Eaux de réseau et fontaine	chloration, thermique (chaud et froid positif)
Eaux de baignade (non traités)	Eaux de mer	thermique (froid positif), osmotique
	Eaux douce	thermique (froid positif)
Eaux de piscine		chloration, thermique (chaud)
Eaux industrielles	Eaux de tours aéro-réfrigérantes (TAR)	biocide
	Eaux chaudes sanitaires	chloration, thermique (chaud)
	Eaux de rejet, résiduaires brutes et traitées (STEP, ...)	chimique (pH acide et/ou basique, ...)
Autres (Demandes spécifiques du fabricant)	par exemples : Eaux thermales (spa ...)	chloration, thermique (chaud)

Annexe 3. Organisation des analyses réalisées en parallèle en fonction du type de méthode

Méthodes avec une étape commune



Méthodes sans étapes communes



Annexe 4. Méthodes qualitatives. Test de rejet des résultats discordants

Procédure

Compter le nombre total de résultats discordants RD :

$$RD = PD + ND$$

Par exemple, si $PD = 2$ et $ND = 10$, alors $RD = 12$)

Si $RD < 6$ (moins de 6 résultats discordants), aucun test n'est disponible ;

Si $6 \leq RD \leq 22$ (entre 6 et 22 résultats discordants), déterminer m comme la plus petite des deux valeurs de PD et de ND (par exemple $m = PD = 2$, parce que $PD < ND$) et utiliser la loi binomiale conformément au Tableau 13.

Si $m \leq M$ (M défini dans le Tableau 13) pour un Y donné, les deux méthodes sont différentes à $\alpha < 0,05$ (bilatéral).

Tableau 13. Valeurs de M pour Y résultats discordants ($6 \leq Y \leq 22$)

Résultats discordants $RD = PD + ND$	6-8	9-11	12-14	15-16	17-19	20-22
$M = \text{Max}(m)$ avec $\alpha < 0,05$	0	1	2	3	4	5

Par exemple, avec $Y = 12$ résultats discordants et $m = 2$, $M = 2$ et $m \leq M$: donc les deux méthodes sont différentes avec $\alpha < 0,05$.

si $Y > 22$ (plus de 22 résultats discordants), utiliser le test de McNemar avec la distribution de chi deux pour 1 degré de liberté :

$$\chi^2 = d^2/Y, \text{ avec } d = |PD - ND| \text{ et } Y = PD + ND$$

Les deux méthodes sont différentes pour $\alpha < 0,05$ (bilatéral) si $\chi^2 > 3,841$

Ce test de χ^2 correspond au d minimal pour chaque Y du Tableau 14 ci-dessous avec $\alpha < 0,05$ (c'est-à-dire, pour un Y donné, d doit être supérieur ou égal à la valeur donnée dans le Tableau 14 pour conclure que les deux méthodes sont différentes).

Tableau 14. Valeurs de d pour Y résultats discordants ($Y > 22$)

Résultats discordants $Y = PD + ND$	22-26	27-31	32-37	38-44	45-51	52-58
$d = PD - ND \geq$	10	11	12	13	14	15

Annexe 5. Méthodes qualitatives. Étude interlaboratoire, critères d'accord, de concordance et ratio de chances de concordance

Inspirée de l'Annexe L de la norme EN ISO 16140

Généralités

Les critères d'exactitude, de sensibilité et de spécificité (voir 5.2.2) ne traitent pas réellement de la variabilité de la méthode au sein d'un laboratoire et entre laboratoires (fidélité de la méthode).

Cette annexe introduit des critères supplémentaires (degré d'accord, concordance et *odds ratio*) qui peuvent aider à approcher cette variabilité.

Les critères de répétabilité et de reproductibilité mesurent la différence attendue entre deux échantillons envoyés soit au même laboratoire, soit à des laboratoires différents. Comme la différence de résultats qui ne sont pas quantitatifs ne peut pas être utilisée, les statistiques pour les méthodes qualitatives sont plutôt fondées sur la probabilité (exprimée par un pourcentage) que deux échantillons identiques donneront le même résultat.

Ces critères ont été développés par le projet européen SMT4 CT 96 2098 financé par la Commission Européenne/DG XII et visant à valider les six principales méthodes normalisées utilisées en microbiologie des aliments (coordinateur : Dr C. Lahellec, AFSSA, France).

Les calculs seront illustrés avec les données suivantes (voir Tableau 15).

Tableau 15. Exemple de valeurs numériques

Labo	1	2	3	4	5	Nombre de positifs (sur 5)
1	+	+	+	+	+	5
2	+	+	+	+	+	5
3	+	+	+	+	+	5
4	+	+	+	+	+	5
5	-	-	+	+	+	3
6	+	+	+	+	+	5
7	-	-	+	+	+	3
8	+	+	+	+	+	5
9	+	+	+	+	+	5
10	+	+	+	+	+	5

Les valeurs données dans le tableau correspondent seulement à un niveau d'un type d'aliment. En pratique un essai inter-laboratoire serait plus grand, mais un ensemble de données plus petit permet d'expliquer plus facilement les calculs.

Degré d'accord

Définition

Le degré d'accord (*accordance* en anglais) est le pourcentage de chances de trouver le même résultat (c'est-à-dire tous les deux positifs ou tous les deux négatifs) pour deux échantillons identiques analysées dans le même laboratoire, dans des conditions de répétabilité (c'est-à-dire un seul opérateur utilisant le même appareillage et les mêmes réactifs dans l'intervalle de temps le plus court possible).

Le degré d'accord est ainsi l'équivalent de la répétabilité pour les méthodes quantitatives.

Calculs

Pour déduire le degré d'accord des résultats d'une étude inter-laboratoires, la probabilité que deux échantillons identiques donnent le même résultat est calculée pour chaque laboratoire participant à son tour, et cette probabilité est ensuite moyennée sur l'ensemble des laboratoires.

Pour les laboratoires (comme le laboratoire 1) pour lesquels tous les répétitions ont été trouvés positifs, la meilleure estimation de la probabilité d'obtenir le même résultat est clairement 1,00 ou 100 %.

Pour les autres laboratoires (dans l'exemple, les laboratoires 5 et 7), la probabilité qu'une répétition soit positive est $3/5 = 0,60$. Élever au carré cette probabilité pour obtenir la probabilité qu'une paire de répétitions soit positive ($0,6^2 = 0,36$). Faire de même pour la probabilité qu'une paire de répétitions soient tous les deux négatifs ($0,4^2 = 0,16$). Ensuite ajouter ces deux nombres pour obtenir la probabilité globale que deux répétitions donnent le même résultat ($0,36 + 0,16 = 0,52$). Le faire pour tous les laboratoires (voir Tableau 16).

Tableau 16. Calcul du degré d'accord

Labo	Nombre de positifs	Probabilité de positifs	Probabilité de paire de Positifs	Probabilité de négatif	Probabilité de paire de négatifs	Probabilité de paire de résultats identiques
1	5	1,00	1,00	0,00	0,00	1,00
2	5	1,00	1,00	0,00	0,00	1,00
3	5	1,00	1,00	0,00	0,00	1,00
4	5	1,00	1,00	0,00	0,00	1,00
5	3	0,60	0,36	0,40	0,16	0,52
6	5	1,00	1,00	0,00	0,00	1,00
7	3	0,60	0,36	0,40	0,16	0,52
8	5	1,00	1,00	0,00	0,00	1,00
9	5	1,00	1,00	0,00	0,00	1,00
10	5	1,00	1,00	0,00	0,00	1,00
					Moyenne :	0,904 =90,4%

Le degré d'accord est la moyenne des probabilités que deux répétitions donnent le même résultat pour chaque laboratoire : 90,4 % dans ce cas.

NOTE Comme la répétabilité mesure une différence alors que le degré d'accord mesure une similitude, des valeurs élevées de degré d'accord indiquent une méthode fiable, à la différence des méthodes quantitatives pour lesquelles de faibles valeurs de répétabilité sont souhaitables.

Concordance

Définition

La concordance (*concordance* en anglais) est le pourcentage de chances de trouver le même résultat pour deux échantillons identiques analysés dans deux différents laboratoires.

La concordance est donc l'équivalent de la reproductibilité pour les méthodes quantitatives.

Calculs

Pour calculer la concordance à partir des résultats d'une étude inter-laboratoires, prendre tour à tour chaque répétition dans chaque laboratoire participant, l'apparier avec les résultats identiques de tous les autres laboratoires.

La concordance est le pourcentage de toutes les paires donnant les mêmes résultats sur toutes les paires possibles de résultats.

EXEMPLE (voir Tableaux L.1 et L.3 de la norme EN ISO 16140)

Prendre successivement chaque répétition dans chaque laboratoire, en commençant avec la première répétition du laboratoire 1 qui est positive. Celle-ci peut être appariée avec l'une des 45 répétitions des autres laboratoires, et toutes sauf 4 de ces paires (celles contenant les deux répétitions des laboratoires 5 et 7) concordent (c'est-à-dire une paire avec deux positifs), 41 paires donnent donc le même résultat.

Ceci s'applique aux quatre autres répétitions du laboratoire 1, il y a ainsi un total de 225 (5 x 45) paires inter-laboratoires de répétitions impliquant le laboratoire 1, desquelles 205 (5 x 41) donnent le même résultat.

Ceci s'applique à tous les autres laboratoires dont tous les répétitions sont positives.

Pour le laboratoire 5, avec 3 répétitions positifs sur 5, les 2 répétitions négatifs correspondent chacun avec seulement 2 autres répétitions négatifs du laboratoire 7, alors que les 3 répétitions positifs correspondent chacun avec 43 répétitions positifs. Ainsi le nombre total de paires ayant le même résultat est 133 (2 x 2 + 3 x 43).

Tableau 17. Calcul de la concordance

Labo	Nombre de positifs	Paires inter-laboratoires avec le même résultat	Nombre total de paires inter-laboratoires
1	5	205	225
2	5	205	225
3	5	205	225
4	5	205	225
5	3	133	225
6	5	205	225
7	3	133	225
8	5	205	225
9	5	205	225
10	5	205	225
Total		1906	2250

La concordance est le pourcentage de toutes les paires de répétitions donnant le même résultat ; dans cet exemple, il est 84,7 % (1906/2250 x 100).

NOTE Comme la reproductibilité mesure une différence alors que la concordance mesure une similitude, des valeurs élevées de concordance indiquent une méthode fiable, à la différence des méthodes quantitatives pour lesquelles de faibles valeurs de reproductibilité sont souhaitables.

Odds ratio

Généralités et définition

Une concordance plus petite que le degré d'accord signifie que deux échantillons identiques ont plus de chances de donner le même résultat s'ils sont analysés par le même laboratoire que s'ils sont analysés par des laboratoires différents, suggérant ainsi qu'il peut y avoir une variabilité prédominante de performance entre laboratoires. C'est la même situation lorsque la reproductibilité est plus grande que la répétabilité pour une méthode quantitative.

Malheureusement, l'amplitude de la concordance et du degré d'accord est fortement dépendante du niveau d'exactitude, rendant difficile l'accès direct au degré de variation interlaboratoires.

Il est ainsi utile de calculer le *odds ratio* (COR) défini comme suit :

$$COR = \frac{\text{degré d'accord} \times (100 - \text{concordance})}{\text{concordance} \times (100 - \text{degré d'accord})}$$

Tests de signification

Une valeur pour le *odds ratio* de 1,00 serait attendue si le degré d'accord et la concordance étaient égaux, et plus le ratio est élevé, plus la variation interlaboratoire est prédominante.

Cependant, des valeurs supérieures à 1,00 peuvent résulter de variations dues au hasard, et ainsi il convient d'utiliser un test de signification statistique pour confirmer que la variation interlaboratoire est bien prédominante. Le « test exact » est le test le plus recommandé pour cet objectif¹. La philosophie sous-tendant de tels tests est que les probabilités de survenue sont calculées pour toutes les combinaisons de résultats de répétitions qui pourraient avoir été obtenus avec la totalité des positifs et des négatifs.

EXEMPLE (voir Tableau L.1 de la norme EN ISO 16140)

Avec un total de 46 positifs et de 4 négatifs, les combinaisons possibles sont données par colonnes dans le Tableau L.4. de la norme EN ISO 16140.

Tableau 18. Combinaisons possibles de répétitions positifs entre laboratoires (colonnes) pour donner un total de 46 positifs

4	3	3	2	1
4	4	3	4	5
4	4	5	5	5
4	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5

1) Ce test peut être conduit en utilisant des logiciels statistiques comme SAS®.

La combinaison véritable est indiquée en gras (troisième colonne) dans le Tableau M.4. Le test ajoute les probabilités P de toutes les combinaisons possibles qui dénotent une variation inter-laboratoires au moins aussi importante que celle de la combinaison véritable –ici elles correspondent aux permutations des deux colonnes de droite. Si cette somme de probabilités est inférieure à la valeur conventionnelle de 0,05 ou 5 %, il est improbable que ce degré de variation inter-laboratoires puisse avoir été obtenu par hasard et ainsi il est conclu qu'il y a une variation significative de performance entre laboratoires.

Dans l'exemple du Tableau M.4, P = 0,039 indique que la variation entre laboratoires est significative au niveau de 5 %.

Lorsqu'un logiciel pour le « test exact » n'est pas disponible, une analyse ordinaire de χ^2 pour les tables de contingence peut être utilisée comme alternative. Les résultats de ce test sont moins fiables que ceux du « test exact » avec le nombre de répétitions utilisé habituellement dans les études collaboratives, mais des simulations suggèrent que les résultats donnent une bonne idée de la signification des différences inter-laboratoires.

Avec l'un ou l'autre de ces tests, il faut se rappeler que la capacité de détecter des différences inter-laboratoires dépend du nombre de laboratoires et du nombre de répétitions analysés par chaque laboratoire. Un test non significatif ne devrait pas être interprété comme une absence de variation de performance entre laboratoires, mais plutôt que de telles différences n'ont pas pu être prouvées ; ceci est particulièrement vrai lorsque la valeur de P est juste au-dessus de 0,05. La solution idéale serait d'indiquer le *odds ratio* accompagné de l'écart-type ou des limites de confiance, mais la distribution du *odds ratio* est très différente de la normalité ; l'estimation de limites fiables est alors rendue très délicate.

Annexe 6. Méthodes quantitatives. Calcul des écarts-types de répétabilité et reproductibilité

Cette annexe reprend la clause 7 de la norme ISO 5725-2. Les calculs sont conduits selon le principe décrit dans la norme ISO 5725. Ils sont appliqués à chaque niveau j sur les données z . De ce fait, on considère que le nombre de répétitions par niveau est constant et égal à K .

Calculer pour chaque niveau j :

- la variance de répétabilité, notée $s(j)_r^2$;
- la variance inter-séries, notée $s(j)_B^2$;
- la variance de fidélité intermédiaire $s(j)_{FI}^2 = s(j)_r^2 + s(j)_B^2$.

Pour simplifier les notations, l'indice j du niveau est supprimé dans les formules qui suivent.

Le point de départ consiste à écrire l'égalité suivante qui décompose la somme des écarts à la moyenne générale en deux sommes de carrés d'écarts. n_k représente...

$$\sum_{k=1}^K \sum_{i=1}^I (z_{ik} - \bar{z})^2 = \sum_{k=1}^K (\bar{z}_k - \bar{z})^2 + \sum_{k=1}^K \sum_{i=1}^I (z_{ik} - \bar{z}_k)^2 \quad (8)$$

Cette équation est traditionnellement écrite sous une forme abrégée, faisant appel à trois sommes de carrés d'écarts (SCE). Équation générale de l'analyse de variance

$$SCE_t = SCE_B + SCE_r$$

Où chacune des sommes est définie comme suit pour faciliter l'interprétation :

SCE_t Somme totale des écarts à la moyenne générale du niveau

SCE_B Somme des écarts inter-séries

SCE_r Somme des écarts intra-série

Comme l'explique la norme ISO 5725, on développe les sommes de carrés pour effectuer plus facilement les calculs ; calcul de SCE_B se fait toujours par différence. Cette méthode peut poser des problèmes si le résultat est négatif : on force alors la valeur de SCE_B à 0.

$$SCE_B = SCE_t - SCE_r$$

si $SCE_B > 0$

si $SCE_B \leq 0$ alors $SCE_B = 0$

NOTE Une alternative à cette méthode de calcul consiste à utiliser un algorithme d'estimation de s_B^2 utilisant la méthode du maximum de vraisemblance restreint (REML) disponible dans la plupart des logiciels statistiques.

Calculer la variance de répétabilité (intra-série) pour le niveau :

$$s_r^2 = \frac{1}{(I-1)K} \sum_{i=1}^I \sum_{k=1}^K (z_{ik} - \bar{z}_k)^2 = \frac{1}{(I-1)K} SCE_r \quad (9)$$

Calculer la variance inter-séries de la façon suivante :

$$s_B^2 = \frac{(K-1) \left(\frac{SCE_B}{K-1} - s_r^2 \right)}{I} \quad (10)$$

Finalement la variance de fidélité intermédiaire pour le niveau est égale à : $s_R^2 = s_r^2 + s_B^2$ (11)

Annexe 7. Exemple de calcul du profil d'exactitude

Labo	Séries	Niveau	Alternative (ufc)	Référence (ufc)	Log Alter.	Log Réf.
A	1	1	60	120	1.778	2.079
A	1	1	40	75	1.602	1.875
B	2	1	70	95	1.845	1.978
B	2	1	90	85	1.954	1.929
C	3	1	80	90	1.903	1.954
C	3	1	50	80	1.699	1.903
D	4	1	90	90	1.954	1.954
D	4	1	170	120	2.230	2.079
E	5	1	40	90	1.602	1.954
E	5	1	70	130	1.845	2.114
F	6	1	50	130	1.699	2.114
F	6	1	120	65	2.079	1.813
G	7	1	90	95	1.954	1.978
G	7	1	130	110	2.114	2.041
H	8	1	90	100	1.954	2.000
H	8	1	110	110	2.041	2.041
I	9	1	140	70	2.146	1.845
I	9	1	90	80	1.954	1.903
J	10	1	90	120	1.954	2.079
J	10	1	110	110	2.041	2.041
L	11	1	90	140	1.954	2.146
L	11	1	120	110	2.079	2.041
M	12	1	110	95	2.041	1.978
M	12	1	130	130	2.114	2.114
N	13	1	90	95	1.954	1.978
N	13	1	150	80	2.176	1.903
O	14	1	110	65	2.041	1.813
O	14	1	50	95	1.699	1.978
A	1	2	810	1300	2.908	3.114
A	1	2	520	1200	2.716	3.079
B	2	2	940	1000	2.973	3.000
B	2	2	950	1100	2.978	3.041
C	3	2	1000	1000	3.000	3.000
C	3	2	1100	1000	3.041	3.000
D	4	2	1000	1000	3.000	3.000
D	4	2	900	940	2.954	2.973
E	5	2	940	930	2.973	2.968
E	5	2	920	1100	2.964	3.041
F	6	2	810	1000	2.908	3.000
F	6	2	1000	1100	3.000	3.041
G	7	2	910	910	2.959	2.959
G	7	2	1000	900	3.000	2.954
H	8	2	1200	1100	3.079	3.041
H	8	2	1100	1000	3.041	3.000
I	9	2	1100	1000	3.041	3.000
I	9	2	980	1100	2.991	3.041
J	10	2	1200	950	3.079	2.978

J	10	2	1100	1100	3.041	3.041
L	11	2	830	1000	2.919	3.000
L	11	2	1300	1000	3.114	3.000
M	12	2	1100	1000	3.041	3.000
M	12	2	830	1200	2.919	3.079
N	13	2	660	940	2.820	2.973
N	13	2	810	980	2.908	2.991
O	14	2	640	820	2.806	2.914
O	14	2	940	1200	2.973	3.079
A	1	3	9300	9700	3.968	3.987
A	1	3	9900	9700	3.996	3.987
B	2	3	7800	9400	3.892	3.973
B	2	3	9700	9000	3.987	3.954
C	3	3	11000	10000	4.041	4.000
C	3	3	11000	12000	4.041	4.079
D	4	3	9500	11000	3.978	4.041
D	4	3	10000	11000	4.000	4.041
E	5	3	11000	9600	4.041	3.982
E	5	3	9400	10000	3.973	4.000
F	6	3	8900	11000	3.949	4.041
F	6	3	9500	10000	3.978	4.000
G	7	3	12000	13000	4.079	4.114
G	7	3	13000	12000	4.114	4.079
H	8	3	9400	12000	3.973	4.079
H	8	3	9500	9400	3.978	3.973
I	9	3	11000	12000	4.041	4.079
I	9	3	12000	12000	4.079	4.079
J	10	3	12000	12000	4.079	4.079
J	10	3	11000	9500	4.041	3.978
N	13	3	9500	12000	3.978	4.079
N	13	3	10000	13000	4.000	4.114
O	14	3	9000	7600	3.954	3.881
O	14	3	9500	8500	3.978	3.929

Tableau 19. Résultats de l'étude interlaboratoire

Critères	Niveau 1	Niveau 2	Niveau 3
Valeur cible	1.95	2.97	4.00
Moyenne méthode alternative	0.018	0.027	0.016
Limite de tolérance inférieure	-0.196	-0.117	-0.109
Limite de tolérance supérieure	0.232	0.171	0.140
Limite inférieure d'acceptabilité	-0.3	-0.3	-0.3
Limite supérieure d'acceptabilité	0.3	0.3	0.3

Construire le profil d'exactitude comme celui illustré à la Figure 4.

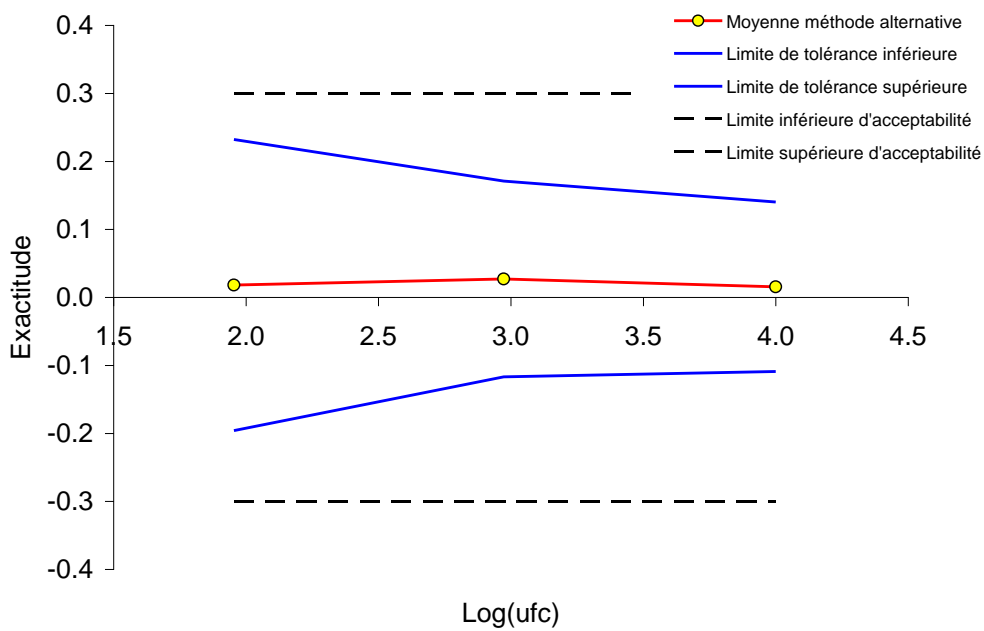


Figure 4. Profil d'exactitude. Les limites d'acceptabilité ont été choisies à ± 0.3 log

Pour montrer l'importance du critère d'acceptabilité, la Figure 5 illustre la situation où les limites d'acceptabilité ont été choisies à ± 0.2 log. La LOQ est alors de 2.51 unités log.

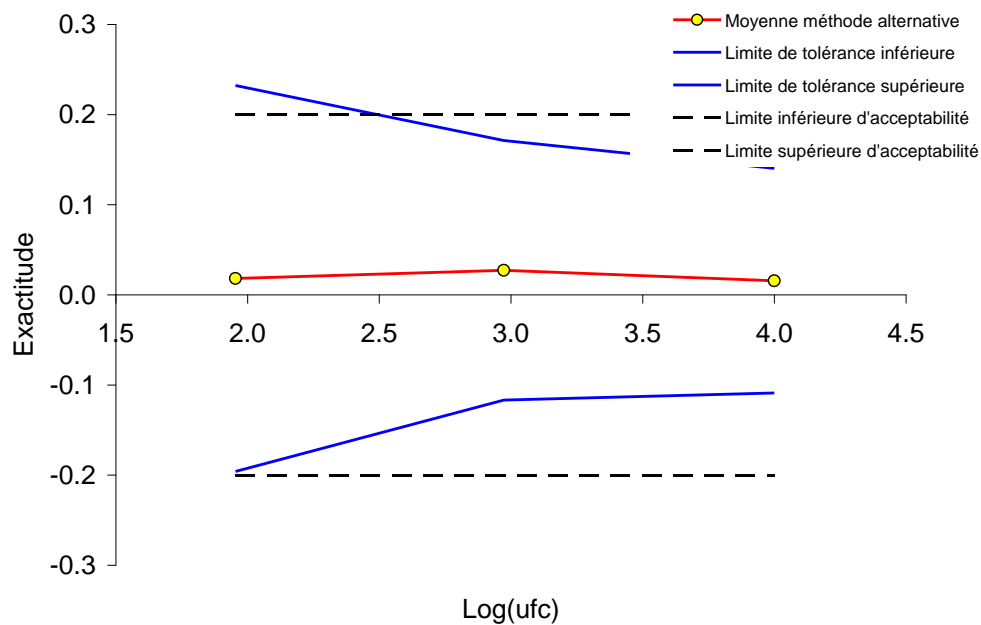


Figure 5. Profil d'exactitude II. Les limites d'acceptabilité ont été choisies à ± 0.2 log

Annexe 8. Lignes directrices relatives à l'organisation et à la gestion des études inter-laboratoires

Cette annexe est inspirée de l'annexe H (normative) de la norme EN ISO 16140.

Préparation des échantillons d'eau

Échantillons négatifs (contrôles)

Le germe cible ne doit pas être présent dans la matrice d'eau. Choisir un nombre d'échantillons approprié conformément au 6.1 ou au 6.2 du présent document et les analyser pour le confirmer. Des contrôles négatifs doivent également être utilisés au cours des essais quand ils sont spécifiés.

Échantillons positifs contenant le germe cible.

Pour les études **qualitatives** et **quantitatives**, les échantillons doivent être contaminés artificiellement comme indiqué dans les options 2 et 3 de l'annexe C de la norme EN ISO 16140. Lorsque des matériaux de référence sont utilisés, le laboratoire organisateur est tenu d'envoyer les matériaux de référence ainsi que les détails du mode opératoire d'ensemencement à chaque laboratoire participant.

L'échantillon et les micro-organismes doivent pouvoir être homogénéisés et il convient qu'ils restent stables pendant le transport et la durée des analyses. Le laboratoire organisateur doit procéder à des études de l'homogénéité et de la stabilité avant de distribuer le matériau d'essai.

Il convient que l'échantillon d'eau contienne une microflore de fond représentative ou des composés interférents qui doivent également demeurer stables pendant le transport et la durée des analyses.

Une méthode alternative doit être testée avec au moins un type d'eau, choisi parmi les catégories définies. La qualité et le caractère adéquat des échantillons d'eau doivent être établis avant l'étude.

En se fondant sur l'étude comparative, les caractéristiques de la souche microbienne choisie utilisée pour valider une méthode, doivent être représentatives du genre/de l'espèce recherché(e), par exemple taux de croissance, caractéristiques des antigènes et sensibilité aux agents défavorables, etc.

Transport des échantillons

Le laboratoire organisateur définit le type d'eau à utiliser et vérifie que l'emballage de transport est adapté au transport prévu.

Les échantillons individuels doivent être emballés dans des emballages à double paroi étanche pour éviter que la fuite d'un échantillon n'affecte l'intégrité de tous les autres.

Chaque laboratoire participant est tenu de fournir au laboratoire organisateur tous les détails utiles afin de garantir que la distribution des échantillons destinés à l'étude soit conforme aux réglementations nationales et internationales en matière de courrier.

Concernant la réfrigération, les conditions suivantes s'appliquent : la température des échantillons pendant le transport doit être inférieure ou égale à 10°C et, à l'arrivée au laboratoire, comprise entre 0°C et 10°C.

Les conditions d'emballage et la meilleure méthode de transport doivent être déterminées par le laboratoire organisateur. Il est souhaitable de mettre en œuvre des moyens appropriés permettant de surveiller les échantillons pendant le transport. Il faut également prévoir un paquet supplémentaire, identique aux échantillons pour essai, afin de mesurer leur température lors de la réception.

Organisation de l'étude interlaboratoire

protocoles opératoires : des modes opératoires doivent être diffusés aux collaborateurs afin qu'ils retournent leurs commentaires et qu'ils puissent se familiariser avec l'analyse avant de commencer l'essai. Un protocole final doit être envoyé avant l'évaluation ;

confirmation de la qualité de l'échantillon : il convient que chaque collaborateur procède au dénombrement de la flore bactérienne totale sur un échantillon supplémentaire spécifique préparé par le laboratoire expert. Le laboratoire organisateur doit remettre à chaque collaborateur une méthode de dénombrement normalisée ;

méthode alternative et méthode de référence : le collaborateur chargé de réaliser l'analyse dans chaque laboratoire participant doit être préalablement formé (si nécessaire) et doit s'être complètement familiarisé avec la méthode alternative et avec la méthode de référence avant de commencer l'essai ;

réactifs et conditions opératoires : certains facteurs supplémentaires, tels que la qualité et la composition des milieux de culture et des réactifs, la surveillance des températures d'incubation, etc., peuvent avoir une influence significative sur les résultats de l'essai. Cette variabilité doit être minimisée (par exemple en distribuant des milieux / réactifs à tous les collaborateurs) ou doit être prise en compte lors de l'interprétation des données ;

le niveau de tolérance autorisé pour les différents aspects du germe (par exemple durées, températures, masses, dénombrement total des boîtes et jour d'analyse) doit être précisé dans le protocole. Un avertissement explicite doit être prévu si aucune tolérance n'est autorisée.

lignes directrices : le laboratoire organisateur doit être disponible pour conseiller ou fournir des lignes directrices aux collaborateurs pendant toute la période de l'étude ;

collecte des données : le laboratoire organisateur doit préparer un questionnaire permettant de recueillir des informations sur les points critiques du mode opératoire. Il convient de demander à tous les collaborateurs de consigner les détails, tels que le pH des milieux, les durées d'incubation, les températures lors du démarrage et de l'arrêt de l'incubateur, les données relatives au contrôle de qualité, l'état des échantillons à leur réception, l'heure de leur réception, la température à réception, les conditions / durée de conservation, etc. Le questionnaire doit faire l'objet d'un accord entre le laboratoire organisateur et le fabricant afin de garantir son adéquation avec les étapes critiques de l'essai ou du mode opératoire d'essai de la méthode alternative ;

confirmation de la qualité des échantillons pour essai : Il convient que le laboratoire organisateur analyse des parties aliquotes représentatives des échantillons le jour de la distribution afin de confirmer la présence et l'homogénéité du germe cible. Il convient que des échantillons, préalablement conservés dans les conditions idéales, soient également analysés par le laboratoire organisateur le jour du début de l'analyse par les laboratoires participants. Ces analyses doivent être utilisées à la seule fin de confirmer le mode opératoire d'ensemencement et la stabilité du germe. Il convient que les résultats ne soient pas utilisés dans le cadre de l'évaluation statistique de la méthode soumise à essai.

Le fait d'utiliser des milieux de marque et de lot différents introduit une variabilité supplémentaire.