



HAL
open science

Les essais inter-laboratoires en microbiologie des aliments Inter-laboratory studies in food microbiology

Bertrand Lombard

► **To cite this version:**

Bertrand Lombard. Les essais inter-laboratoires en microbiologie des aliments Inter-laboratory studies in food microbiology. Chemical Sciences. INAPG (AgroParisTech), 2004. English. NNT : 2004INAP0017 . pastel-00001258

HAL Id: pastel-00001258

<https://pastel.hal.science/pastel-00001258>

Submitted on 31 May 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



École Doctorale ABIÉS
INSTITUT NATIONAL AGRONOMIQUE PARIS-GRIGNON
Département Sciences pour les Industries Biologiques et Alimentaires

THÈSE
pour obtenir le grade de
Docteur de l'Institut National Agronomique Paris-Grignon

présentée et soutenue publiquement par
Bertrand LOMBARD

le 13 décembre 2004

Les essais inter-laboratoires en microbiologie des aliments
Inter-laboratory studies in food microbiology

Jury

Président

M. Christian Ducauze

Professeur à l'INA-PG

Rapporteurs

M. Maurice Claisse

Professeur à l'Université Paris VI

M. Claude Gaillardin

Chef de Département Microbiologie alimentaire, INRA

Examineurs

M. Vincent Carlier

Professeur à l'ENVA

M. Laurent Rosso

Directeur de laboratoire à l'AFSSA

Directeurs de thèse

M. Max Feinberg

Directeur de Recherches à l'INRA

Mme Cécile Lahellec

Directeur de Recherches à l'AFSSA

A Isabelle, et à tous les miens

Remerciements

Cette thèse s'appuie sur les travaux réalisés au sein du Laboratoire d'Etudes et de Recherches sur la Qualité des Aliments et sur les Procédés Agro-alimentaires de l'Agence Française de Sécurité Sanitaire des Aliments (AFSSA), que je remercie de m'avoir permis de mener à bien ce projet.

Je voudrais remercier tout particulièrement le Pr. Max FEINBERG qui a dirigé ces travaux. A l'issue de plusieurs années de collaboration fructueuse pour la normalisation, marquée par des relations chaleureuses, et lors de mon entrée au CNEVA, il m'a incité à entreprendre ce travail de thèse et il a tout de suite proposé de l'encadrer. Il m'a ensuite suivi avec rigueur, apportant son soutien à ce travail, notamment pour ses aspects bio-statistiques et informatiques et lors de la rédaction de ce mémoire. Il m'a en outre fait partager la démarche d'un chercheur. De même, je souhaite remercier très vivement Cécile LAHELLEC, qui a accepté d'encadrer avec M. Feinberg mes travaux dans leur valence microbiologique. Là aussi après de longues années d'une riche collaboration pour l'établissement de normes au niveau européen, elle m'a soutenu tout au long de ce parcours, tout particulièrement dans la phase finale de rédaction de ce mémoire, et elle m'a aidé à surmonter les obstacles de différentes natures qui se sont dressés. Elle m'a associé au projet européen de validation de méthodes microbiologiques, qu'elle a coordonné. Ce projet constitue la principale base expérimentale de mon travail, et j'ai pu par ce biais tisser des relations aux niveaux européen et international qui sont fécondes aujourd'hui encore.

J'exprime ma reconnaissance au Pr. Christian DUCAUZE, qui m'a fait l'honneur de présider le jury de cette thèse. C. Ducauze a suivi avec beaucoup de sollicitude mon parcours professionnel depuis mon année de spécialisation en chimie analytique à l'INA-PG, dont il était responsable.

Je suis également très honoré que les Pr. Claude GAILLARDIN et Maurice CLAISSE aient accepté de consacrer une partie de leur temps déjà très chargé à l'examen de mes travaux.

Ma reconnaissance va aussi au Pr. Vincent CARLIER et à Laurent ROSSO, qui ont bien voulu participer au jury de cette thèse. Les relations établies avec V. Carlier dans différents cadres, et depuis plusieurs années, ont été toujours enrichissantes. L. Rosso, depuis son arrivée à la direction du laboratoire, m'a vivement encouragé dans la finalisation de ce projet.

Je voudrais également remercier Françoise JANIN, qui, à la direction du laboratoire pendant la plus grande partie de mon travail, a d'emblée adhéré à ce projet, et m'a soutenu tout au long de mon parcours, notamment lors des passages difficiles.

La collaboration étroite avec les différents partenaires du projet européen SMT de validation de méthodes, en particulier avec Marie-Laure DE BUYSER, Marylène BOHNERT, Susan SCOTTER, Paul IN'T VELD, Phil FELDSINE, ainsi qu'avec les biostatisticiens Steve LANGTON, Niko NAGELKERKE et Remi CHEVENNEMENT, a été très fructueuse pour mener ce travail, et je les en remercie.

Le projet européen MicroVal a été à l'origine de la norme EN ISO 16140, mais on peut y trouver également la genèse de ce projet de thèse. Les échanges avec les différents partenaires de MicroVal et du groupe de préparation de la norme EN ISO 16140, notamment Marie-Pierre COPIN et Jeanmarc AESCHLIMAN, ont été très enrichissants, et je tiens à les

remercier vivement.

Les travaux de normalisation européenne et internationale sur les aspects statistiques de l'analyse microbiologique des aliments m'ont permis d'engager des échanges, notamment avec les statisticiens Hilko VAN DER VOET, les Pr Hartmut WEISS et Peter-Theodor WILRICH. Ces échanges ont enrichi les aspects de cette thèse relatifs à l'exploitation des résultats. Je tiens à leur exprimer ici toute ma gratitude.

Par ailleurs, les discussions avec le Pr. Michel CATTEAU, avec lequel j'ai eu le plaisir de collaborer pour la normalisation française et internationale pendant de nombreuses années, m'ont permis de bien saisir la portée et la signification des essais inter-laboratoires pour la microbiologie des aliments, et les limites des approches statistiques dans cette discipline.

Je remercie mes anciens collègues et mes interlocutrices à AFNOR, notamment Catherine GUISSÉ et Céline DRUEZ, Catherine GOMY et Valentine DIGONNET, avec lesquelles j'ai collaboré et collabore encore dans les travaux de normalisation et validation en microbiologie des aliments. Elles m'ont donné accès au fonds normatif qui constitue la base de ce travail.

Je voudrais exprimer ici ma reconnaissance aux Pr Gérard et Maryvonne MARTIN, qui, m'ayant accueilli dans leur laboratoire pour effectuer mon stage de DEA et de fin d'études à l'INA-PG, m'ont ainsi initié à la recherche. Ils m'ont ensuite incité à entreprendre puis à poursuivre cette thèse.

Je ne saurais oublier l'ensemble de mes collègues du laboratoire, entre autres Anne BRISABOIS et Martine POUMEYROL, Jean-Marc FREMY et Patrick FACH. Je tiens à les remercier toutes et tous pour l'amitié qu'il ont su me témoigner, ainsi que pour leur soutien et leur collaboration dans différents domaines de mon travail. Ma reconnaissance va tout spécialement à Elodie VOITOUX, qui a mis en œuvre avec beaucoup de compétence les différents essais inter-laboratoires pour les Laboratoires Nationaux de Référence « Lait », ainsi qu'à Marie CORNU, qui a bien voulu collaborer à la phase finale de mes travaux, et sans laquelle un projet de publication n'aurait pas abouti.

Je remercie enfin mon épouse Isabelle, mes parents, mon frère Philippe, Claudine et mes amis pour leur soutien dans la réalisation de ce projet.

Résumé

La validité des contrôles microbiologiques, réalisés dans l'objectif d'assurer la sécurité sanitaire des aliments, nécessite notamment l'obtention de résultats d'analyse fiables. La fiabilité des résultats implique l'utilisation de méthodes validées, mises en œuvre par un laboratoire compétent. Les essais inter-laboratoires permettent de s'assurer, du moins en partie, du respect de ces deux conditions. Cependant, en raison de limites expérimentales, ces essais ne sont pas aussi largement pratiqués dans le domaine de la microbiologie des aliments qu'ils ne le sont dans d'autres domaines analytiques.

Dans un premier temps, une revue des documents de référence permet d'établir un état des lieux. Cette revue concerne les trois objectifs que l'on peut assigner à des essais inter-laboratoires, à savoir l'évaluation de méthodes d'analyse, celle des laboratoires, et la caractérisation de matériaux de référence. Les documents de portée générale, puis ceux spécifiques de l'analyse des aliments, sont pris en compte, et leur degré d'applicabilité à l'analyse microbiologique des aliments est envisagé. Les référentiels et pratiques propres au domaine d'intérêt traité sont finalement présentés, et les déviations par rapport aux documents généraux analysées.

Sur cette base, sont présentées les conditions de mise en œuvre de deux types d'essais inter-laboratoires, soit la validation de méthodes dans le cadre d'un projet européen du 4^{ème} Programme Cadre de Recherche & Développement d'une part, et l'évaluation de laboratoires par le biais d'essais d'aptitude pour les Laboratoires Nationaux de Référence sur le lait d'autre part. Les difficultés relatives au protocole expérimental, et liées aux spécificités de la microbiologie, sont mises en exergue.

Les modes d'exploitation des résultats, en fonction des objectifs et de la nature, qualitative ou quantitative, de la détermination, sont expliqués. En ce qui concerne la caractérisation de la performance des méthodes d'analyse, l'utilisation de statistiques robustes pour estimer la fidélité des méthodes quantitatives est discutée, ainsi que la façon de caractériser la fidélité comme la justesse des méthodes qualitatives. Sur ces aspects, des perspectives d'amélioration sont envisagées.

L'intérêt de l'organisation des essais inter-laboratoires en microbiologie des aliments est ensuite abordé. Celui-ci réside dans l'utilisation que l'on peut faire de ces essais comme éléments incontournables de validation d'une méthode d'analyse et d'évaluation d'un laboratoire afin, d'une part, de crédibiliser ou d'améliorer les méthodes d'analyse normalisées au niveau international, et d'autre part d'estimer l'incertitude de mesure attachée aux résultats d'analyse.

Quant aux limites de ces essais, essentiellement d'ordre expérimental, elles tiennent surtout à la nature vivante de l'analyte, et concernent des questions de représentativité.

MOTS CLEFS : analyse microbiologique, aliments, essais inter-laboratoires, validation de méthode, méthode de détection, méthode de dénombrement, fidélité, répétabilité, reproductibilité, statistiques robustes, justesse, essais d'aptitude, matériaux de référence, incertitude de mesure.

Abstract

The validity of microbiological controls, performed for ensuring food safety, requires in particular reliable analytical results. Such reliability needs validated methods, implemented by a competent laboratory. The inter-laboratory studies enable to respect, at least partly, these two conditions. Meanwhile, given experimental limitations, these studies are not so largely performed in food microbiology than in other analytical fields.

In a first step, a review of reference documents enables to make a point on what exists. It covers the three objectives that can be attributed to inter-laboratory studies, that is the assessment of analytical methods, or of laboratories, or the characterisation of reference materials. The documents of general scope, then the ones specific to food analysis, are taken into account, and their applicability to the microbiological analysis of food is investigated. The reference documents and practices specific to the field of interest are finally presented, and the deviations to the general documents are considered.

On this base, the conditions for implementing two types of inter-laboratory studies are presented, that is on one side the validation of methods in the frame of a European project of the 4th Framework Programme, and on the other side the assessment of laboratories through proficiency testing for the National Reference Laboratories for milk. The difficulties linked to the experimental protocol, as well as to the specificity of microbiology, are highlighted.

The exploitation of studies' results is explained, given the objectives and the nature, qualitative or quantitative, of the determination. As regards to the characterisation of method performance, the use of robust statistics to estimate the precision of analytical methods is discussed, as well as the estimation of both precision and trueness of qualitative methods. For these aspects, ways of improvement are envisaged.

The interest to organise inter-laboratory studies in food microbiology is then covered. It consists in the use that can be made of these studies as essential tool for method validation or for laboratory assessment. The purpose is, on one side, to give credibility or to improve the analytical methods standardised at international level, and on the other side, to estimate the measurement uncertainty attached to analytical results.

As to the limitations of these studies, mainly of experimental nature, most of them are linked to the living nature of the analyte, and they refer to questions of representativity.

KEYWORDS: microbiological analysis, foods, inter-laboratory studies, method validation, detection method, enumeration method, precision, repeatability, reproducibility, robust statistics, trueness, proficiency trials, reference materials, measurement uncertainty.

Table des matières
(jusqu'au 3^{ème} niveau de sub-division)

Introduction	19
1. Première partie. Essais inter-laboratoires : normes et pratiques actuelles	23
Introduction	25
Terminologie et définitions	25
Objectifs des essais inter-laboratoires	26
1.1 Établir les performances d'une méthode	29
1.1.1 Présentation des normes de la série ISO 5725	30
1.1.2 ISO 5725. Partie 1	31
1.1.3 ISO 5725. Partie 2	37
1.1.4 ISO 5725. Partie 4	40
1.1.5 ISO 5725. Partie 5	42
1.1.6 ISO 5725. Partie 6	46
1.1.7 Lignes directrices IUPAC/AOAC	50
1.1.8 Norme FIL 135 B	53
1.1.9 Norme EN ISO 16140	54
1.1.10 Lignes directrices AOAC	61
1.1.11 Procédures de validation des méthodes de référence en microbiologie des aliments	68
1.1.12 Procédures de validation des méthodes alternatives en microbiologie des aliments	73
1.2 Vérifier la compétence d'un réseau de laboratoires	77
1.2.1 Guide ISO 43-1	78
1.2.2 Protocole harmonisé IUPAC/AOAC	88
1.2.3 Le projet de norme ISO/DIS 13 528	94
1.2.4 Mise en œuvre des référentiels en microbiologie des aliments	100
1.3 Caractériser un matériau de référence	107
1.3.1 Généralités	107
1.3.2 Définitions et typologie	108
1.3.3 Matériaux de référence en agroalimentaire	110
Conclusion de la première partie	115
2. Deuxième partie. Mise en œuvre d'essais inter-laboratoires en microbiologie des aliments	117
Introduction	119
2.1 Études européennes de validation de méthodes de référence	120
2.1.1 Protocoles expérimentaux	120
2.1.2 Méthodologie d'exploitation des résultats pour les méthodes de dénombrement	126
2.1.3 Méthodologie d'exploitation des résultats pour les méthodes de détection	127
2.1.4 Résultats des études menées	130
2.2 Fidélité des méthodes de dénombrement	139
2.2.1 Représentation graphique des données	140
2.2.2 Transformation des données en logarithmes décimaux	145
2.2.3 Comparaison des estimateurs robustes et non robustes	149
2.2.4 Perspectives	158
2.3 Performance des méthodes de détection	159
2.3.1 Fidélité des méthodes de détection	159
2.3.2 Justesse des méthodes de détection	164
2.4 Essais d'aptitude des Laboratoires Nationaux de Référence « Lait »	167
2.4.1 Contexte	167
2.4.2 Matériel et méthodes	168
2.4.3 Méthodes de traitement des données	170

2.4.4	Dénombrement des coliformes dans le lait _____	171
2.4.5	Détection des <i>Salmonella</i> _____	175
2.4.6	Conclusion sur les deux essais d'aptitude _____	175
2.5	Valorisation des résultats d'essais inter-laboratoires en microbiologie des aliments	177
2.5.1	Essais inter-laboratoires visant à établir les performances d'une méthode _____	178
2.5.2	Essais d'aptitude des laboratoires _____	180
2.5.3	Essais inter-laboratoires de caractérisation des matériaux de référence _____	181
2.5.4	Estimation de l'incertitude de mesure _____	181
2.6	Limites expérimentales des essais inter-laboratoires en microbiologie des aliments	185
2.6.1	Limites inhérentes aux essais inter-laboratoires en général _____	185
2.6.2	Limites liées aux spécificités de la microbiologie des aliments _____	187
	Conclusion de la deuxième partie _____	195
	Conclusion générale _____	197
3.	Annexes _____	201
3.1	Annexe 1. Publication de l'essai de validation sur le dénombrement de <i>L.monocytogenes</i> _____	203
3.2	Annexe 2. Publication de l'essai de validation sur le dénombrement des staphylocoques _____	213
3.3	Annexe 3. Publication de l'essai de validation sur la détection de <i>L.monocytogenes</i> _____	225
3.4	Annexe 4. Rapport de l'essai de validation sur la détection de <i>Salmonella</i> _____	239
3.5	Annexe 5. Etude des estimateurs de fidélité _____	281
3.5.1	Approche non robuste selon l'ISO 5725-2 _____	283
3.5.2	Approche robuste selon l'EN ISO 16140 _____	286
3.5.3	Approche robuste selon l'ISO 5725-5 _____	287
3.6	Annexe 6. Publication sur la fidélité des méthodes qualitatives _____	291
3.7	Annexe 7. Publication sur la fidélité des méthodes de comptage _____	297
3.8	Annexe 8. Rapport de l'essai d'aptitude sur le dénombrement des coliformes _____	325
3.9	Annexe 9. Rapport de l'essai d'aptitude sur la détection de <i>Salmonella</i> _____	341
4.	Bibliographie _____	363
4.1	Textes normatifs _____	365
4.2	Publications et rapports _____	367

Liste des tableaux

Tableau 1 : Termes et définitions relatifs à l'exactitude d'une méthode d'analyse.....	32
Tableau 2 : Limites de répétabilité et de reproductibilité.....	46
Tableau 3 : Critères de sensibilité, spécificité et exactitude pour les méthodes qualitatives, selon la Norme EN ISO 16140.....	58
Tableau 4 : Critères de performance des méthodes qualitatives, selon AOAC	65
Tableau 5 : Graphiques de combinaison de scores de performance pour un seul essai d'aptitude, selon l'ISO/DIS 13528.....	98
Tableau 6 : Graphiques de combinaison de scores de performance pour plusieurs essais d'aptitude d'un même système, selon l'ISO/DIS 13528.....	99
Tableau 7 : Calcul de l'écart-type relatif géométrique.....	104
Tableau 8 : Valeurs limites d'acceptabilité de répétabilité pour les essais d'aptitude CECALAIT	104
Tableau 9 : Valeurs limites d'acceptabilité de biais pour les essais d'aptitude CECALAIT.	105
Tableau 10 : Valeurs certifiées des matériaux de référence BCR en microbiologie des aliments	114
Tableau 11 : Niveaux de contamination choisis pour le projet SMT.....	122
Tableau 12 : Préparation des échantillons pour le projet SMT	123
Tableau 13. Procédures de contrôle de la qualité des matériaux soumis à l'essai dans le projet SMT.....	124
Tableau 14. Nombre d'échantillons et de répétitions en fonction du type de méthodes, dans le projet SMT	125
Tableau 15 : Calcul du degré d'accord selon l'EN ISO 16140 (avec p le nombre de laboratoires participants, n le nombre de répétitions, et k_i le nombre de résultats positifs trouvés par le laboratoire i)	128
Tableau 16 : Calcul de la concordance selon l'EN ISO 16140 (avec les notations utilisées pour le degré d'accord).....	129
Tableau 17 : Valeurs générales de fidélité exprimées en \log_{10} ufc/g, pour le dénombrement de <i>L.monocytogenes</i>	131
Tableau 18 : Valeurs générales de fidélité exprimées en \log_{10} ufc/g, pour le dénombrement des staphylocoques à coagulase positive.....	132
Tableau 19 : Comparaison des normes ISO 7937 : 1987 et EN 13401 : 1999 pour le dénombrement de <i>C.perfringens</i>	133
Tableau 20. Valeurs de fidélité exprimées en \log_{10} ufc/g, pour le dénombrement de <i>C. perfringens</i> avec deux techniques de confirmation (LS / MN+LG)	134
Tableau 21 : Valeurs de justesse et de fidélité, exprimées en pourcentages, pour la détection de <i>L. monocytogenes</i>	135
Tableau 22 : Valeurs de justesse et de fidélité, exprimées en pourcentages, pour la détection de <i>Salmonella</i>	136

Tableau 23 : Statistique de cohérence inter-laboratoires h de Mandel (laboratoire i , niveau j)	140
Tableau 24 : Statistique de cohérence intra-laboratoires k de Mandel (laboratoire i , niveau j , p_j laboratoires au niveau j)	141
Tableau 25 : Comparaison des intervalles d'incertitude de mesure (IM) du laboratoire 1, avec et sans transformation en \log_{10} des données, lors de l'essai sur le dénombrement de $L.$ <i>monocytogenes</i>	147
Tableau 26 : Calcul de l'estimateur d'échelle Q_n , selon [Wilrich, 2004a] (p : nombre de laboratoires, \bar{x}_i : moyennes des répétitions du laboratoire i)	157
Tableau 27 : Mode de calcul de la concordance dans un cadre aléatoire (n : nombre de répétitions, k_i : nombre de résultats positifs)	160
Tableau 28 : Application aux méthodes qualitatives du modèle de l'ISO 5725-2 pour la fidélité des méthodes quantitatives (résultats y_{ij} du laboratoire i , n répétitions ; résultats notés '1' pour présence et '0' pour absence).....	162
Tableau 29 : Variances théoriques de répétabilité et de reproductibilité pour une méthode qualitative (K laboratoires).....	163

Liste des figures

- Figure 1 : Histogramme de justesse, selon la statistique h de Mandel – Exemple des résultats de l’essai inter-laboratoires sur le dénombrement de *L.monocytogenes* 143
- Figure 2 : Histogramme de fidélité, selon la statistique k de Mandel – Exemple des résultats de l’essai inter-laboratoires sur le dénombrement de *L.monocytogenes* 144
- Figure 3 : Représentation des intervalles d’incertitude de mesure (IM), avec ou sans transformation en \log_{10} des données. Exemple du laboratoire 1, essai sur le dénombrement de *L.monocytogenes*. 148
- Figure 4 : Ecarts-types de reproductibilité s_R pour l’essai inter-laboratoires sur le dénombrement de *L. monocytogenes*, calculés selon ISO 5725-2 (\diamond), ISO 5725-5 (\square) et EN ISO 16140 (\triangle), sur des données transformées en \log_{10} 153
- Figure 5 : Coefficients de variation de reproductibilité RSD_R pour l’essai inter-laboratoires sur le dénombrement de *L. monocytogenes*, calculés selon ISO 5725-2 (\diamond), ISO 5725-5 (\square) et EN ISO 16140 (\triangle), sur des données non transformées en \log_{10} 153
- Figure 6 : Représentation graphique de la performance des laboratoires – Essai d’aptitude des LNR Lait sur le dénombrement des coliformes dans le lait – Exemple du lot à taux moyen de contamination (4,06 log ufc/ml) 173

Liste des principales abréviations

AFSSA	Agence Française de Sécurité Sanitaire des Aliments
AOAC	<i>Association of Official Analytical Chemists</i>
AFNOR	Association Française de NORmalisation
BAM	<i>Bacteriological Analytical Manual</i> /Manuel de bactériologie analytique (de la FDA)
CECALAIT	Centre d'Expertise et de Contrôle des Analyses LAITières
CE	Commission Européenne
CEN	Comité Européen de Normalisation
CEN/TC 275/WG 6	Groupe de travail 6 « Contaminants microbiens » du Comité Technique 275 « Analyse des denrées alimentaires – Méthodes horizontales » du CEN
COFRAC	Comité Français d'Accréditation
CSL	<i>Central Science Laboratory</i> (RU)
d.d.l.	degrés de liberté
FDA	<i>Food and Drug Administration</i> /Administration des aliments et des médicaments (des Etats-Unis)
FIL-IDF	Fédération Internationale Laitière, <i>International Dairy Federation</i>
IRMM	<i>Institute for Reference Materials and Measurements</i> /Institut pour pour les Matériaux de Référence et pour les Mesures
ISO	Organisation Internationale de Normalisation
ISO/TC 34/SC 9	Sous-Comité 9 « Microbiologie » du Comité Technique 34 « Produits alimentaires » de ISO
ISO/TC 34/SC 9/WG 2	Groupe de Travail 2 « Statistiques » de l'ISO/TC 34/SC 9 « Produits alimentaires – Microbiologie »
ISO/TC 69/SC 6	Sous-Comité 6 « Méthodes et résultats de mesure » du Comité Technique 69 « Application des méthodes statistiques » de ISO
IUPAC	<i>International Union of Pure & Applied Chemistry</i> /Union Internationale de Chimie Pure et Appliquée
NMKL	Comité Nordique sur l'Analyse des Aliments
RAEMA	Réseau d'Analyses et d'Echanges en Microbiologie des Aliments
RIVM	<i>Rijksinstituut voor Volksgezondheid en Milieu</i> /Institut National de Santé Publique et d'Environnement (Pays-Bas)
SMT	<i>Standards, Measurements & Testings</i> /Normes, Mesures & Essais
ufc	unités formant colonies
USDA	<i>United States Department of Agriculture</i> /Ministère de l'agriculture des Etats-Unis

Introduction

Lors de ces dernières décennies, les questions liées à la *sécurité sanitaire des aliments* ont pris une importance croissante dans l'esprit des consommateurs et ceci dans l'ensemble des pays industrialisés. Pour assurer cette sécurité, les aliments sont produits dans des conditions qui doivent permettre la maîtrise de l'hygiène par les opérateurs économiques. La mise en place de procédures de type HACCP (Analyse des Dangers et Points Critiques pour leur Maîtrise) est un outil fondamental de cette maîtrise. Les matières premières, ainsi que les aliments en cours et après fabrication, sont soumis à des contrôles réguliers pour vérifier la bonne maîtrise de l'hygiène, qui consiste notamment à vérifier que les points critiques d'un procédé sont effectivement maîtrisés (soit la validation des systèmes HACCP). Les résultats de ces contrôles doivent être accessibles à l'ensemble des parties, en particulier lors des contrôles officiels, et reconnus par l'ensemble des parties.

Or, la *reconnaissance des résultats de contrôles* suppose que deux étapes intervenant lors de leur production soient validées :

1. l'étape d'échantillonnage, qui correspond à la constitution et au prélèvement de l'échantillon à partir d'un lot de production, et dont les paramètres (taille, nombre, conservation,...) doivent être parfaitement définis, si possible sur des bases statistiques ;
2. l'étape d'analyse, qui suppose non seulement qu'il existe une méthode d'analyse validée, mais aussi que le laboratoire prestataire soit compétent pour la mettre en œuvre.

Dans ce but, nous verrons dans quelle mesure les essais inter-laboratoires permettent de définir les performances métrologiques d'une méthode d'analyse ou d'un laboratoire, donc de remplir partiellement les deux conditions évoquées pour valider l'étape d'analyse. Cependant, il convient de prendre compte deux éléments :

1. Un essai inter-laboratoires ne permet pas, à lui seul, de valider une méthode.
2. La compétence d'un laboratoire ne peut s'apprécier en totalité qu'au travers d'un ensemble de critères, dont la participation à un essai d'aptitude ne constitue qu'un élément.

Dans le domaine de l'analyse physico-chimique des aliments, les essais inter-laboratoires sont largement répandus [Bugner, 1992]. Ils s'appuient généralement sur un ensemble de textes pratiques bien établis et reconnus, tels que la norme ISO 5725 et le Guide IUPAC/AOAC [IUPAC, 1995] pour les essais visant à établir les performances d'une méthode, le guide ISO 43 et le guide IUPAC/AOAC [Thompson & Wood, 1993] pour les essais d'aptitude et enfin le guide ISO 30 pour les matériaux de référence.

Par contre, dans le domaine de l'analyse microbiologique, on ne disposait d'aucun texte normatif adapté jusqu'à un passé récent. Des difficultés spécifiques peuvent expliquer ce retard. Ainsi, en matière d'organisation de ces essais, la stabilité et l'homogénéité des échantillons utilisés sont difficiles à maîtriser en raison de la nature vivante de l'analyte. Il était de plus nécessaire de déterminer la façon d'exploiter les résultats de dénombrements qui sont des variables discrètes, de distribution souvent non Normale. Enfin, n'a été introduite que récemment une caractérisation de la fidélité des déterminations qualitatives, de type présence-absence, qui ne suivent pas non plus de statistiques gaussiennes.

Pourtant, la validation des méthodes d'analyse revêt une importance toute particulière en microbiologie des aliments. En effet, la validation s'impose d'autant plus qu'elle porte sur des méthodes dont la variabilité est très grande. Cette variabilité peut être principalement attribuée aux deux aspects suivants :

1. la nature intrinsèque du microorganisme vivant, qui possède une grande capacité de variation génétique et de mutation ;
2. les principes de détection et de dénombrement de microorganismes par les techniques conventionnelles de la microbiologie pasteurienne, qui impliquent l'utilisation de substances biologiques variables en elles-mêmes, tels que les milieux de culture, les anticorps ou les réactifs biochimiques.

Par ailleurs, nous venons de voir comment les analyses microbiologiques peuvent constituer un moyen de preuve de la maîtrise de l'hygiène des aliments. Dans ce cadre, l'obtention d'un résultat d'analyse incorrect avec une méthode d'analyse non fiable peut alors avoir de lourdes conséquences, tant pour l'opérateur économique (retrait du marché de marchandises, image de marque) que pour la santé des consommateurs (mise sur le marché de produits considérés à tort comme ne représentant pas de risques pour la santé humaine).

Dans ce contexte, l'**objectif** de ce travail a été de déterminer l'intérêt et les limites de l'organisation d'essais inter-laboratoires dans trois buts différents, à savoir la validation d'une méthode, l'évaluation des laboratoires et la caractérisation de matériaux de référence, avec une finalité commune d'amélioration de la qualité des analyses en microbiologie des aliments. Pour ce faire, sur la base d'une revue des référentiels applicables ainsi que des pratiques existantes dans notre domaine d'intérêt, nous avons mis en œuvre deux types d'essais inter-laboratoires (validation de méthodes et essais d'aptitude) en microbiologie des aliments, avant d'envisager les possibilités d'améliorer la caractérisation de la performance inter-laboratoires des méthodes microbiologiques, et plus largement l'intérêt et les limites de ces démarches.

Ces travaux concernent les bactéries, à l'exclusion des autres types de microorganismes. Ils portent essentiellement sur les méthodes conventionnelles d'analyse microbiologique, qui se fondent sur l'aptitude des microorganismes à se développer et à former des colonies sur des milieux de culture. Cependant, une grande partie des considérations peut s'appliquer à des méthodes alternatives d'analyse microbiologique, plus récentes dans leur conception.

Précisions sur la présentation du mémoire

Un résumé figure au début de chaque chapitre. Par ailleurs, les références des textes normatifs (normes AFNOR et ISO, ou documents de référence AOAC, IUPAC et NMKL) sont rassemblées en 4.1, alors que les publications figurent en 4.2. Les normes sont référencées dans le texte uniquement par leur numéro.

Plusieurs publications auxquelles nous avons été associées ne sont pas annexées à ce mémoire, car elles ne portent pas spécifiquement sur les essais inter-laboratoires. Elles figurent néanmoins dans les références bibliographiques, documentant la présentation de l'existant et alimentant la discussion sur l'intérêt et les limites de ces démarches [Gnanou-Besse *et al*, 2004 ; Gomy et Lombard, 1992 ; Leclercq *et al*, 2000 ; Lombard *et al*, 1996 ; Lombard, 2001 ; Lombard, 2004 ; Voitoux *et al*, 2002c]. Par ailleurs, seuls deux rapports d'essai d'aptitude des Laboratoires Nationaux de Référence ont été joints en annexe, considérés comme représentatifs des quatre autres effectués [Lafarge et Lombard, 1999 ; Le Barillec *et al*, 2003 ; Voitoux *et al*, 2001 ; Voitoux et Lombard 2002b].

1. Première partie.
Essais
inter-laboratoires :
normes et pratiques
actuelles

Introduction

Terminologie et définitions

Alors que plusieurs documents de référence portent sur les essais inter-laboratoires, seul un document émanant du Codex Alimentarius propose à proprement parler une définition de l'essai inter-laboratoires, reprise plus loin [Codex Alimentarius, 2004]. C'est d'autant plus surprenant que cette structure des Nations Unies, placée sous la responsabilité mixte de la FAO et de l'OMS, n'est pas un organisme purement normatif et reconnu comme tel, comme l'Organisation Internationale de Normalisations (ISO) ou l'Association Française de Normalisation (AFNOR).

La terminologie ne fait déjà pas l'unanimité parmi les différents documents de référence. La norme ISO 5725-1, traitant de l'exactitude (justesse et fidélité) des résultats et méthodes de mesure, emploie le terme « expérience inter-laboratoires collective ». Le guide ISO 43-1, portant sur les essais d'aptitude des laboratoires, retient le terme général « essai par intercomparaison » pour des expériences aussi diverses que la comparaison de mesures par transmission d'un laboratoire à l'autre de l'objet soumis à l'essai ou l'envoi simultané d'un ensemble d'échantillons aux laboratoires. L'AOAC International¹ et l'IUPAC (ou Union Internationale de Chimie Pure et Appliquée), quant à elles, préfèrent « étude collaborative » (*collaborative study*).

Nous avons retenu le terme « essai inter-laboratoires » car c'est celui employé par le guide ISO 43-1 pour la situation que nous allons traiter, à savoir l'envoi simultané d'un ensemble d'échantillons à plusieurs laboratoires, et parce qu'il est très proche de celui employé par la norme ISO 5725. Le mot « expérience » a été remplacé par « essai », car ce dernier évoque plus directement dans notre domaine analytique la réalisation d'essais ou d'analyses dans différents laboratoires. L'adjectif « collective » a été supprimé, apparaissant redondant avec « inter-laboratoires » ; d'ailleurs, la norme ISO 5725 elle-même ne l'emploie pas systématiquement. La terminologie retenue par AOAC et IUPAC semble être moins précise puisque « étude » a un sens plus large que « essai », ainsi que « collaborative » par rapport à « inter-laboratoires ».

Le Codex Alimentarius, en particulier le Comité sur les Méthodes d'Analyse et d'Echantillonnage (CCMAS), est donc la seule instance à avoir défini l'étude inter-laboratoires de la façon suivante :

« Étude dans laquelle plusieurs laboratoires mesurent une quantité dans une ou plusieurs parties identiques de matières homogènes et stables dans des conditions documentées et dont les résultats sont regroupés dans un seul document. »

¹ Association d'analystes de portée internationale, mais localisée aux Etats-Unis (siège : Gaithersburg, MD), qui valide et officialise des méthodes d'analyse des produits alimentaires, pharmaceutiques et agrochimiques. Elle sert d'organisme de normalisation aux Etats-Unis dans ces domaines.

Pour nos besoins, nous reprendrons cette définition d'un essai inter-laboratoires en la précisant de la façon suivante :

« Étude dans laquelle plusieurs laboratoires détectent ou quantifient le ou les analyte(s) objet de l'essai dans un ou plusieurs échantillons identiques de matières homogènes et stables au regard de l'analyte considéré. Les essais sont menés dans des conditions prédéfinies et communes à l'ensemble des laboratoires participant à l'essai. Les résultats de chaque participant sont rassemblés dans un seul document et exploités en fonction de l'objectif de l'essai. »

Objectifs des essais inter-laboratoires

Trois objectifs principaux peuvent être assignés à un essai inter-laboratoires :

1. Procéder à la **validation** inter-laboratoires d'une **méthode d'analyse** (ou de plusieurs), par l'étude du critère de fidélité et/ou du critère de justesse ;
2. **Évaluer la compétence des laboratoires** participant à l'essai, par le biais de leur aptitude (i) à trouver la valeur attendue pour chaque matériau d'essai et (ii) à répéter l'analyse dans des conditions acceptables de similitude des résultats ;
3. Assigner une valeur de niveau de concentration pour un analyte dans un matériau, sur la base des résultats obtenus par les laboratoires participant à l'essai, afin de **caractériser un matériau de référence**, éventuellement un matériau de référence certifié.

Quelque soit l'objectif poursuivi, une des caractéristiques essentielles d'un essai inter-laboratoires est de pouvoir garantir que les échantillons, au moment de leur analyse par les laboratoires participant à l'essai, seront effectivement *identiques*, principalement au regard de l'analyte considéré. Cette remarque reste vraie même si plusieurs analytes sont impliqués dans l'essai. En d'autres termes, pour un même matériau, le niveau (ou la concentration) de l'analyte doit être identique pour l'ensemble des échantillons distribués aux laboratoires et ce jusqu'au moment où débute l'analyse. Ainsi, il est nécessaire de s'assurer non seulement de l'*homogénéité* entre les échantillons au regard de l'analyte et de la méthode d'analyse utilisée, mais aussi de la *stabilité* de l'analyte dans les échantillons, au moins dans les conditions de transport des échantillons vers les laboratoires, telle que la durée du transport et toute autre condition pouvant affecter le niveau de l'analyte. En fait, il est impossible de prétendre à une homogénéité parfaite ; c'est l'hétérogénéité qui doit être négligeable par rapport à la variabilité de la méthode d'analyse considérée. Nous verrons plus loin que ces conditions de stabilité et d'homogénéité sont particulièrement difficiles à respecter lorsque le matériau est un aliment et l'analyte un microorganisme vivant.

La comparaison de mesures telle que la définit le guide ISO 43, à savoir la transmission d'un laboratoire à l'autre d'un ensemble d'échantillons analysés successivement, est certes en théorie idéale pour assurer la similitude du matériau analysé dans chaque laboratoire. Cependant, elle n'est pas envisageable pour l'analyse microbiologique des aliments, en raison du manque de stabilité de l'analyte dans les échantillons.

Dans cette première partie, nous présenterons les divers référentiels développés pour chacun des trois objectifs distingués ci-dessus. Nous avons préféré centrer cette revue bibliographique sur des documents de référence, et non sur des publications scientifiques, car la plupart des publications portant sur les essais inter-laboratoires dans notre domaine d'intérêt décrivent des essais organisés selon un modèle fixé, faisant l'objet d'un document de référence. Certaines

pratiques actuelles en matière d'organisation d'essais inter-laboratoires seront également discutées. Une revue exhaustive de tous les textes traitant des essais inter-laboratoires n'a pas été menée, mais un certain nombre de textes de référence ont été sélectionnés, certains faisant l'objet de normes ISO ou de normes du Comité Européen de Normalisation (CEN), et d'autres étant utilisés en référence par les acteurs du secteur agroalimentaire. De même, nous n'avons pas recensé toutes les procédures mises en œuvre, mais certaines qui nous ont paru les plus illustratives.

Dans cette présentation, nous nous attacherons à mettre en valeur les analogies et différences qui peuvent exister entre les différents documents. De plus, nous discuterons de l'applicabilité des référentiels et des notions qu'ils définissent au domaine de la microbiologie des aliments. Cette applicabilité sera bien évidemment surtout discutée pour les référentiels de portée générale. Enfin, nous envisagerons dans quelle mesure les référentiels spécifiques soit à l'analyse des aliments, soit à l'analyse microbiologique des aliments, sont conformes aux référentiels de portée générale. En cas de divergence entre référentiels spécifiques et référentiels de portée générale, nous nous interrogerons sur la possibilité d'expliquer la divergence par la spécificité du domaine analytique considéré.

Les référentiels de portée générale seront d'abord présentés : en premier lieu ceux s'appliquant à tous les types d'analyses, puis ceux s'appliquant à l'analyse des aliments, l'analyse microbiologique des aliments représentant à chaque fois un cas particulier. Les référentiels spécifiques à l'analyse microbiologique des aliments seront finalement abordés.

1.1 Établir les performances d'une méthode

Résumé

Ce premier chapitre introduit, par le biais de l'ISO 5725, une série de normes de portée générale, les notions de base pour la conception et l'exploitation des trois types d'essais inter-laboratoires : non seulement ceux qui font l'objet de ce chapitre, mais aussi ceux qui permettent d'évaluer la compétence d'un réseau de laboratoires, ou ceux visant à caractériser un matériau de référence. Ces notions portent tout d'abord sur l'organisation des essais inter-laboratoires et sur la conception de leur protocole expérimental. Dans un deuxième temps, sont définis les critères permettant, à l'issue d'un essai inter-laboratoires, de caractériser la performance d'une méthode d'analyse, mais aussi par la suite la performance de laboratoires, ainsi que les matériaux de référence. Il s'agit d'une part de la fidélité, composante aléatoire de la variabilité des résultats, avec deux situations extrêmes, la répétabilité et la reproductibilité, et d'autre part de la justesse, composante systématique de cette variabilité.

Deux référentiels propres à l'analyse des aliments sont ensuite analysés : les lignes directrices IUPAC/AOAC et la norme FIL 135 B. Ces deux protocoles sont d'orientation plus pratique que le premier, et comportent des aspects propres au domaine considéré, notamment en terme de protocole expérimental.

Lors de la présentation du protocole de portée générale et des protocoles applicables aux aliments, leur degré d'applicabilité à la microbiologie des aliments est envisagé. Les notions de base s'appliquent correctement, alors que certains aspects des protocoles expérimentaux nécessitent des adaptations. Par ailleurs, la performance des méthodes qualitatives n'est pas traitée.

Les deux protocoles spécifiques à la microbiologie des aliments, la norme EN ISO 16140 et les lignes directrices AOAC, sont abordés de façon plus détaillée, et comparés entre eux. Si de nombreuses similitudes apparaissent, il subsiste des différences notables, en matière de protocole expérimental –nombre de matrices– et d'exploitation des résultats –statistiques robustes/non robustes pour les méthodes quantitatives, mode de calcul de la justesse pour les méthodes qualitatives, et caractérisation de la fidélité de ces dernières–.

Enfin, plusieurs mises en pratique d'essai inter-laboratoires en microbiologie des aliments sont présentées. Certaines concernent la caractérisation de la performance de méthodes de référence dans les différents contextes de AOAC, NMKL, CEN/ISO, des contrôle officiels au Royaume-Uni, ainsi que du projet SMT. D'autres permettent de valider des méthodes considérées comme alternatives à ces mêmes méthodes de référence, dans le cadre de AOAC, AFNOR, NordVal ou du projet FOOD-PCR.

1.1.1 Présentation des normes de la série ISO 5725

A notre connaissance, un seul référentiel s'appliquant à tous types d'analyse existe à ce jour. Il s'agit de la norme ISO 5725, divisée en six parties et établie par la structure de l'ISO traitant des méthodes statistiques relatives à l'analyse, le Sous-Comité 6 « *Méthodes et résultats de mesure* » du Comité Technique 69 « *Application des méthodes statistiques* », appelé ISO/TC 69/SC 6. L'unicité de ce référentiel présente l'avantage d'éviter les risques de divergence « à la base », même si cela n'évite pas des modifications lorsque tel ou domaine analytique décide d'établir un référentiel sur les essais inter-laboratoires, référentiel qui lui sera spécifique.

Cette norme a le mérite de définir clairement les concepts de base et de traiter d'une grande variété de situations, ce qui a d'ailleurs justifié son découpage en six parties. Il ne s'agissait pas d'ailleurs de la première version de la norme, mais d'une profonde révision de la version précédente, qui datait de 1986. Outre une actualisation, la nouvelle version, parue pour cinq de ses parties en 1994, a permis d'aborder des notions complémentaires, comme en partie 4 la justesse d'une méthode, ou en partie 3 les conditions intermédiaires de fidélité, autres que la répétabilité et la reproductibilité. Par ailleurs, des méthodes alternatives à la méthode de base d'estimation des écarts-types de fidélité, ont été introduites ultérieurement en partie 5, parue en 1998, avec notamment l'utilisation de statistiques robustes.

Avant d'aller plus avant, il importe de se poser la question de l'applicabilité de cette norme à la microbiologie qualitative et quantitative des aliments. En effet, la norme stipule qu'elle s'applique aux méthodes produisant des mesures sur une échelle continue, soit une mesure pouvant prendre un nombre infini de valeurs. Cependant, une note prévoit que les notions de répétabilité et reproductibilité peuvent également s'appliquer à des résultats discrets, qui ne prennent qu'un nombre fini de valeurs. De ce fait, nous pouvons considérer que la norme peut être utilisée pour les dénombrements bactériens. Même si les concepts de base et leurs définitions restent toujours valables, il est clair que cette norme ne peut en revanche pas être utilisée, du moins dans son intégralité, pour caractériser les méthodes qualitatives, qui établissent la présence/absence d'un germe dans une prise d'essai.

Nous ne présenterons pas la partie 3, qui ne traite pas de l'établissement de l'exactitude d'une méthode par le biais d'un essai inter-laboratoires.

1.1.2 ISO 5725. Partie 1

1.1.2.1 Définitions et concepts

La partie 1 introduit une caractéristique globale et essentielle de performance d'une méthode d'analyse, l'*exactitude* qui combine une composante systématique, la *justesse* et une composante aléatoire, la *fidélité*. Les définitions des termes relatifs à l'exactitude sont présentées et commentées, notamment en ce qui concerne leur applicabilité à la microbiologie, dans le Tableau 1.

Tableau 1 : Termes et définitions relatifs à l'exactitude d'une méthode d'analyse

Terme	Définition (selon ISO 5725-1)	Commentaires
Exactitude	Étroitesse de l'accord entre le résultat d'essai et la valeur de référence acceptée	Dépend de la définition de la valeur de référence acceptée
Justesse	Étroitesse de l'accord entre la valeur moyenne obtenue à partir d'une large série de résultats d'essais et une valeur de référence acceptée	Dépend également de la définition de la valeur de référence acceptée
Valeur de référence	Valeur qui sert de référence, agréée pour une comparaison et qui résulte : a) d'une valeur théorique ou établie, fondée sur des principes scientifiques ; b) d'une valeur assignée ou certifiée, fondée sur les travaux expérimentaux d'une organisation nationale ou internationale ; c) d'une valeur de consensus ou certifiée, fondée sur un travail expérimental en collaboration et placé sous les auspices d'un groupe scientifique ou technique ; d) dans le cas où a), b) et c) ne sont pas applicables, de l'espérance de la quantité (mesurable), c'est-à-dire la moyenne d'une population spécifiée de mesures.	Applicabilité au domaine de la microbiologie envisagée dans le texte
Fidélité	Étroitesse d'accord entre des résultats d'essais indépendants obtenus sous des conditions stipulées.	Deux aspects à retenir : a) <i>Résultats indépendants</i> . Une note définit plus précisément que les résultats doivent être obtenus d'une façon non influencée par un résultat précédent sur le même matériau d'essai, ou sur un matériau similaire. b) <i>Conditions stipulées</i> . L'estimation de la fidélité en dépend entièrement. Deux conditions extrêmes particulières sont définies, la répétabilité et la reproductibilité, ainsi que des conditions intermédiaires.
Répétabilité	Fidélité sous des conditions de répétabilité, c'est-à-dire des conditions où les résultats d'essai indépendants sont obtenus par la même méthode sur des individus d'essai identiques dans le même laboratoire, par le même opérateur, utilisant le même équipement et pendant un court intervalle de temps.	Conditions de variabilité minimale
Reproductibilité	Fidélité sous des conditions de reproductibilité, c'est-à-dire des conditions où les résultats d'essai sont obtenus par la même méthode sur des individus d'essais identiques dans différents laboratoires, avec différents opérateurs et utilisant des équipements différents.	Conditions de variabilité maximale

Dans le Tableau 1, la définition de la valeur de référence offre un éventail de possibilités assez large pour la déterminer. En ce qui concerne l'analyse microbiologique, le premier cas semble exclu puisqu'il n'est pas possible à ce jour de déterminer de façon théorique et sur la base de principes scientifiques un nombre exact de microorganismes (voir ci-dessous). Les trois autres cas, de nature expérimentale, sont envisageables :

- Le 2^{ème} cas peut consister en une valeur fixée par les travaux d'un laboratoire de référence qui effectue une contamination artificielle d'un matériau par une suspension bactérienne très précisément définie, en matière de concentration, état physiologique,...
- Le 3^{ème} cas peut correspondre à :
 - la valeur d'un matériau de référence certifiée à la suite d'un essai inter-laboratoires ;
 - ou la valeur d'un matériau utilisé lors d'un essai inter-laboratoires de validation d'une méthode de référence, ou lors d'un essai inter-laboratoires d'aptitude. Le matériau peut alors être considéré, sous des conditions de stabilité suffisantes, comme matériau de référence externe, selon le Fascicule de Documentation AFNOR FD V03-115.
- Le 4^{ème} cas est la valeur obtenue par un laboratoire effectuant une large série de répétitions d'un dénombrement dans un matériau donné, lequel, suffisamment stable, peut alors être considéré comme matériau de référence interne selon le FD V03-115.

L'intérêt de la notion de justesse dépend donc de la possibilité de concevoir une valeur vraie de la propriété mesurée, qui correspondrait au premier cas de la définition.

Comme nous venons de l'évoquer, déterminer un nombre « vrai » de microorganismes n'apparaît pas possible actuellement, avec les techniques à notre disposition pour le dénombrement bactérien. En effet, ce dénombrement est encore aujourd'hui réalisé essentiellement par comptage de colonies formées sur une boîte de Petri. Or le choix et la composition du milieu de culture, les durée et température d'incubation ont une influence directe et majeure sur le nombre de colonies se développant sur la boîte de Petri [Lombard *et al*, 1996]. Même si une technique alternative est utilisée, elle est toujours étalonnée par rapport à une méthode classique qui demeure la référence. Ce lien intime entre méthode et germe cible est encore plus important lorsque sont dénombrées des populations bactériennes « floues », c'est-à-dire non précisément définies d'un point de vue taxonomique, telles que la flore totale mésophile, les coliformes, les bactéries psychrotrophes... Ces méthodes de dénombrement correspondent alors à ce que l'on peut appeler des méthodes « empiriques », où l'analyte est strictement défini par la méthode d'analyse utilisée.

A défaut d'une valeur vraie, il est donc possible de retenir une **valeur de référence acceptée** pour la propriété mesurée, tel qu'envisagé dans les trois autres cas de la définition. Il s'agit d'une valeur fournie par un matériau de référence ou une méthode de référence ; valeur qui est estimée soit par un seul laboratoire, soit par consensus de plusieurs laboratoires.

La fidélité correspond à la variabilité entre des mesures répétées, dues à des erreurs aléatoires. Cinq facteurs principaux de variabilité des résultats d'une méthode d'analyse sont distingués, correspondant aux 5 M du diagramme d'analyse des procédés, établi initialement pour la production industrielle :

1. *Main d'œuvre* (opérateurs) ;
2. *Milieu* (installation, temps entre les mesures, environnement, température, humidité,...) ;
3. *Méthode* (méthode d'analyse) ;
4. *Moyen* (équipement, réactifs, milieux de culture) ;

5. Matière (constitution et conservation de l'échantillon, prise d'essai).

Dans cette liste, la correspondance de chacun des 5 M avec le processus particulier de l'analyse est donnée entre parenthèses, notamment pour l'analyse microbiologique.

1.1.2.2 Modèle statistique

Seul le modèle de base est présenté ; c'est celui qui sera utilisé dans la partie 2 de la norme.

Pour l'estimation de l'exactitude d'une méthode de mesure, on pose que chaque résultat y pour le matériau soumis à essai est la somme de trois composantes :

$$y = m + B + e \quad \text{Eq. 1}$$

Où :

m est la moyenne générale (espérance) de tous les résultats obtenus par les laboratoires pour un niveau donné. Le terme « niveau », emprunté à la théorie de plans d'expériences, désigne dans ce cas une combinaison entre un matériau donné et une concentration donnée de l'analyte. m ne correspond pas forcément à la valeur vraie μ qui n'est quasiment jamais accessible en microbiologie des aliments, comme nous venons de le souligner. Si l'on compare le résultat à une spécification qui se réfère à la valeur vraie μ , ou si l'on compare différentes méthodes, le biais de la méthode ne pourra pas être ignoré.

B est la composante laboratoire du biais sous des conditions de répétabilité, et est supposé constant lors de répétitions. La distribution générale, entre laboratoires, de B est à la base du concept de reproductibilité. La partie 2 de la norme suppose que cette distribution est normale et de moyenne nulle, mais la plupart des distributions unimodales sont acceptables. La variance de B est appelée variance inter-laboratoires, notée σ_L^2 . Elle inclut la variance entre équipements, réactifs, milieux de culture, opérateurs, ... Dans la partie 2, ces composantes ne sont pas séparées, à la différence de la partie 3.

e est l'erreur aléatoire survenant dans chaque mesure sous des conditions de répétabilité. Tout comme pour le terme B , on suppose que sa distribution est approximativement normale et de moyenne nulle, mais des distributions unimodales sont également acceptables. Au sein d'un laboratoire, la variance de e est appelée variance intra-laboratoire, notée σ_W^2 . On pourrait s'attendre à ce que les valeurs de variance intra-laboratoire diffèrent d'un laboratoire à l'autre, en raison par exemple de différences dans la compétence des techniciens. Cependant, la norme suppose que, pour une méthode correctement normalisée et décrite à un niveau de détail suffisant, ces différences sont limitées, et qu'il est justifié d'établir une valeur commune, la moyenne de toutes les variances intra-laboratoires, représentant une estimation non biaisée de la *variance de répétabilité*, $\sigma_r^2 = \overline{\sigma_W^2}$

Finalement, les valeurs de fidélité consistent en un écart-type de répétabilité $\sigma_r = \sqrt{\sigma_W^2}$ et un écart-type de reproductibilité $\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$.

1.1.2.3 Organisation d'un essai inter-laboratoires

1.1.2.3.1 Méthode d'analyse

Pour pouvoir établir les valeurs de fidélité d'une méthode, il est nécessaire qu'elle soit robuste, c'est-à-dire que de petites variations dans le mode opératoire n'entraînent de variations importantes des résultats et que le biais ait été réduit au maximum. La norme demande également que la méthode soit décrite de la façon la plus détaillée et univoque possible.

1.1.2.3.2 Laboratoires participants

Afin d'obtenir des estimations satisfaisantes des grandeurs définies que sont la justesse, la répétabilité et la reproductibilité, les laboratoires participant à l'essai doivent former un *échantillon représentatif* de la population des laboratoires qui sont susceptibles de mettre en œuvre la méthode d'analyse. En d'autres termes, les laboratoires participants doivent être choisis au hasard dans cette population. C'est pourquoi il ne faut pas sélectionner les laboratoires sur la base de leur expertise, ne retenant par exemple que les laboratoires possédant une maîtrise très approfondie de la méthode. Par ailleurs, recruter les seuls laboratoires volontaires n'est également pas satisfaisant en théorie. En effet, les volontaires peuvent ne pas être représentatifs de la population générale des laboratoires, par exemple par un niveau plus élevé de maîtrise de la méthode. Cependant, des considérations pratiques engagent le plus souvent à retenir ce dernier mode de sélection.

Contrairement à ce que l'on pourrait attendre, la norme ne fixe pas de nombre minimum de participants. En revanche, elle indique, en fonction du nombre de participants, l'incertitude dans les estimations des écarts-types de répétabilité et de reproductibilité. Puisque les écarts-types calculés à partir d'essais inter-laboratoires (s) ne sont que des estimations des valeurs vraies (σ), on déduit, en utilisant la loi du χ^2 , l'intervalle de confiance de σ à partir de s et du nombre de résultats utilisés. Une démarche similaire est suivie pour l'estimation de la justesse.

Il revient donc à l'utilisateur de choisir le niveau d'incertitude acceptable pour l'estimation des paramètres de fidélité/justesse de la méthode, en fonction des objectifs poursuivis ; ce qui conditionnera le nombre minimal de participants. Cependant, deux considérations pratiques guident ces choix :

1. Les estimations s'éloignent nettement de la valeur vraie si le nombre de participants est inférieur ou égal à 5.
2. Si $\sigma_L > \sigma_r$ (le cas le plus fréquent), l'incertitude sur les estimations est réduite de façon nettement plus efficace en augmentant le nombre de laboratoires qu'en augmentant le nombre de répétitions.

1.1.2.3.3 Echantillons

Les exigences relatives au choix des échantillons apparaissent difficilement compatibles entre elles dans le domaine qui nous intéresse :

- La norme prescrit que le matériau utilisé soit choisi de façon à être *représentatif* des échantillons analysés *en routine* avec la méthode d'analyse considérée. Nous verrons ultérieurement que cette condition, essentielle pour la validité des valeurs de fidélité

obtenues, est difficile à réaliser en microbiologie alimentaire, compte tenu des exigences de stabilité et d'homogénéité relatives aux échantillons.

- L'homogénéité et la stabilité des échantillons sont requises pour assurer une des caractéristiques de base de tout essai inter-laboratoires, à savoir : envoyer, pour un niveau donné, des échantillons identiques à l'ensemble des laboratoires participants. Nous discuterons plus en détail cet aspect, mais nous pouvons d'ores et déjà retenir qu'il sera particulièrement délicat de vérifier ces deux qualités si l'analyte est un microorganisme vivant, sachant que les matrices alimentaires sont souvent favorables à leur développement. La norme indique une façon intéressante de s'affranchir de cette exigence d'homogénéité. Elle consiste à prendre en compte l'hétérogénéité dans l'exactitude totale de la méthode, la variation entre échantillons devenant alors partie inhérente de la variabilité de la méthode de mesure. Cette approche pourrait s'appliquer à notre domaine, mais elle présente une limite importante : les valeurs de fidélité obtenues ne sont alors strictement applicables qu'au matériau testé. Quant à la stabilité, pour les matériaux instables (ce qui nous concerne), la norme préconise logiquement d'une part de prévoir des modes de transport limitant cette instabilité et d'autre part de fixer très précisément une date de début de l'analyse commune à l'ensemble des laboratoires.

D'une façon générale et toujours dans un souci de représentativité, la norme recommande que les matériaux choisis couvrent le domaine d'application de la méthode à caractériser. Elle suggère de retenir au moins 5 niveaux, soit 5 combinaisons matériau/concentration. Nous aborderons ultérieurement les recommandations propres à notre domaine analytique.

Enfin, la norme indique un moyen de réaliser l'indépendance entre les répétitions effectuées à un niveau donné : utiliser des échantillons répétés en aveugle. Chaque échantillon est analysé une seule fois et les échantillons sont codés par l'organisateur de façon à ce que le laboratoire ne sache pas quels sont les échantillons identiques. Ce mode d'organisation risque d'aller à l'encontre de la nécessité de respecter un court intervalle de temps entre les répétitions. Il convient de respecter un équilibre entre les deux exigences.

1.1.2.3.4 Utilisation des données d'exactitude

La partie 1 discute d'une hypothèse qui n'est pas automatiquement vérifiée. La caractérisation d'une méthode d'analyse par sa fidélité suppose que la fidélité soit :

- indépendante du matériau testé lors de l'essai inter-laboratoires ;
- ou dépendante du matériau de façon prévisible.

Si aucune de ces deux conditions n'est vérifiée, la fidélité ne doit être exprimée qu'en fonction des différentes classes de matériaux testés. On se trouve généralement dans ce cas en microbiologie des aliments, puisque le type d'aliment va influencer l'homogénéité de la distribution du microorganisme cible, ainsi que la flore annexe qui interfèrera plus ou moins avec la cible au cours de l'analyse, notamment lors de la croissance des colonies sur la gélose de dénombrement.

De même, la fidélité est rarement indépendante du niveau de concentration de l'analyte. C'est le cas de notre domaine d'application, comme nous l'illustrerons plus loin, dans le § 2.2. La norme préconise alors de chercher à établir une relation entre fidélité et niveau de concentration.

1.1.3 ISO 5725. Partie 2

Cette partie de la norme ISO 5725 détaille les recommandations générales d'organisation des essais inter-laboratoires, ainsi que la méthode de base pour l'exploitation statistique des résultats. Les principes d'organisation posés dans la partie 1 sont développés, ainsi que le traitement statistique des résultats.

1.1.3.1 Plan d'expérience d'un essai inter-laboratoires

Le schéma général du plan d'expérience d'un essai inter-laboratoires est le suivant :

- q lots d'échantillons, correspondant à q niveaux,
- envoyés à p laboratoires
- chaque laboratoire effectuant n répétitions dans des conditions de répétabilité, à chacun des q niveaux.

En ce qui concerne les répétitions, ce schéma diffère donc de celui utilisé habituellement en microbiologie des aliments. Dans ce dernier cas, les n répétitions ne sont pas effectuées par le laboratoire sur le même échantillon, mais le laboratoire analyse une fois n échantillons différents du même niveau. De fait, la norme privilégie le respect des conditions de répétabilité pour effectuer les répétitions, par rapport à l'indépendance des répétitions.

Au sein d'un laboratoire, le même opérateur doit effectuer toutes les répétitions en condition de répétabilité. Il doit être représentatif de ceux qui effectuent les mesures en routine. Le terme « opérateur » peut désigner une équipe de techniciens se répartissant les différentes étapes de l'analyse, mais toujours de la même façon.

On entend par « laboratoire participant » une combinaison <opérateur, équipements-réactifs, emplacement d'essai>. Donc un même emplacement d'essai, ou laboratoire au sens habituel, peut fournir plusieurs laboratoires participants s'il dispose de plusieurs opérateurs, chacun utilisant des moyens différents, en matière d'équipement, de milieux de culture,... Cette précision peut représenter une opportunité intéressante pour augmenter le nombre de laboratoires participants ; ce qui représente comme nous l'avons déjà vu un critère important de qualité d'un essai inter-laboratoires.

La norme propose de répartir les responsabilités pour :

- L'organisation de l'essai, qui est confiée à une commission d'experts de la méthode d'analyse, dont un expert chargé de l'analyse statistique et un autre chargé de la supervision de l'essai.
- Dans chaque laboratoire participant, on nomme d'une part un coordinateur, responsable des analyses qui sans y prendre part s'occupera de l'établissement du rapport d'essai, et d'autre part un opérateur.

Cette organisation apparaît un peu compliquée et n'est en général pas intégralement appliquée. L'organisateur de l'essai inter-laboratoires est indispensable, de même que la personne ayant des compétences suffisantes en statistiques pour en premier lieu donner un avis sur le protocole expérimental de l'essai et en deuxième lieu effectuer l'analyse statistique. La commission d'experts est habituellement composée de représentants des laboratoires participants, souvent les coordinateurs. Il est effectivement souhaitable, même si la pratique peut être différente, de toujours séparer le rôle de supervision de celui d'exécution des essais. De plus cette disposition s'intègre bien à un système d'assurance qualité.

1.1.3.2 Analyse statistique

Trois étapes successives peuvent être distinguées :

1. *l'examen critique des données*, consistant à identifier les valeurs aberrantes et à tester l'adéquation du modèle ;
2. le *calcul par niveau* des valeurs préliminaires de fidélité et des moyennes ;
3. le *calcul des valeurs finales* de fidélité et des moyennes, y compris l'établissement d'une relation entre fidélité et niveau.

La structure de base pour l'analyse statistique est la combinaison <laboratoire, niveau>, que la norme appelle « cellule ». Rappelons que le niveau est lui-même la combinaison entre un matériau et un niveau de concentration. Dans le cas idéal où les résultats sont complets, le tableau final comporte pq cellules (p laboratoires, q niveaux), chacune contenant n résultats ou répétitions. En pratique, ce tableau présentera souvent des données redondantes, manquantes ou aberrantes, définies comme suit :

- Données redondantes. Un participant renvoie plus de n résultats pour une cellule et s'ils sont tous valides, on sélectionnera au hasard l'effectif correct.
- Données manquantes. Les cellules complètement vides seront ignorées dans les calculs, alors que les cellules partiellement vides pourront être prises en compte.
- Données aberrantes. Ce sont des valeurs qui s'éloignent tellement des valeurs comparables qu'elles sont considérées comme incompatibles avec celles-ci. Un laboratoire pourra être également considéré comme aberrant s'il renvoie plusieurs résultats anormaux à différents niveaux. Certains ou tous les résultats seront alors écartés.

1.1.3.3 Technique graphique de cohérence

Pour vérifier la cohérence des données, deux approches peuvent être suivies : une technique graphique de cohérence et des tests numériques des valeurs aberrantes. Dans notre domaine, les tests seuls sont utilisés. Pourtant, nous verrons ultérieurement l'intérêt des techniques graphiques.

La norme introduit les graphes de Mandel, de deux types : le graphe h de cohérence inter-laboratoires, et le graphe k de cohérence intra-laboratoire. La description de ces graphiques, ainsi que des exemples, figurent dans la deuxième partie, en § 2.2.1.

1.1.3.4 Tests numériques pour les valeurs aberrantes

Le principe de ces tests est de supposer que les données pour un même niveau sont distribuées selon une loi Normale. Les valeurs s'éloignant significativement de cette distribution seront considérées comme douteuses ou aberrantes, selon le niveau de signification des tests. Nous décrirons dans la deuxième partie, en § 2.1.2, les tests préconisés par la norme, puisque nous les avons directement appliqués dans notre travail expérimental.

La norme recommande la façon d'utiliser le résultat de ces tests :

- Rechercher si les valeurs aberrantes peuvent être expliquées par des erreurs techniques, comme une erreur de calcul, une erreur de transcription ou l'analyse du mauvais échantillon. Si l'erreur a été identifiée, corriger le résultat.
- Si aucune source d'erreur n'a été trouvée, exclure le résultat aberrant.

En d'autres termes, la norme déconseille une exclusion directe des résultats considérés

aberrants par les tests statistiques.

1.1.3.5 Moyenne générale, variances de répétabilité et de reproductibilité

Comme pour les tests statistiques d'identification des valeurs aberrantes, nous ne présenterons pas dans cette partie les calculs, puisque nous les utiliserons pour traiter nos données.

La norme reconnaît qu'il n'est pas toujours possible d'établir une relation régulière entre fidélité et moyenne m à chaque niveau, en particulier lorsque l'hétérogénéité du matériau forme une partie inséparable de la variabilité des résultats. C'est à priori notre cas en microbiologie des aliments, où les interactions entre matrice et microorganisme cible sont différentes en fonction de la nature de la matrice. En effet, chaque partie d'une matrice hétérogène présente une interaction différente avec la cible microbienne.

Seuls trois types de relations entre écart-types de répétabilité/reproductibilité et moyenne sont envisagés :

1. Une droite passant par l'origine ;
2. Une droite avec ordonnée à l'origine positive ;
3. Une relation exponentielle.

D'un point de vue statistique, l'ajustement à une droite est compliqué du fait que moyennes et écart-types sont à la fois estimés et donc sujets à erreur. Comme la valeur de la pente est généralement petite et proche de 0,1, l'erreur sur la moyenne aura une faible influence et l'erreur sur l'estimation des écarts-types prédominera. Une régression pondérée est recommandée pour estimer les paramètres de la droite de régression.

En microbiologie, on pourrait s'attendre à utiliser une relation exponentielle. De fait, ce n'est pas le cas, puisque ce sont les données brutes – avant les calculs de moyennes et écarts-types –, qui sont habituellement transformées en logarithme, aspect que nous aborderons plus en détail en 2^{ème} partie.

1.1.3.6 Discussion des résultats

La norme définit un modèle de présentation pour le rapport de l'expert statisticien et recommande que ce rapport soit soumis à la commission qui :

- discute du rapport ;
- décide de la prise en compte ou non des résultats/laboratoires aberrants, et de la nécessité qui en découle d'améliorer le mode opératoire ou la description de la méthode ;
- décide de l'établissement de valeurs de fidélité pour la méthode et de leur domaine d'application.

Ce dernier point est particulièrement important en microbiologie, où nous avons vu qu'il était impossible d'établir des valeurs de fidélité identiques à tous niveaux et applicables à l'ensemble des denrées alimentaires.

1.1.4 ISO 5725. Partie 4

La partie 4 définit une méthode générale d'établissement de la justesse d'une méthode de mesure. Elle précise d'abord que le biais d'une méthode peut être décelé en chimie si la méthode d'analyse n'arrive pas à extraire la totalité de l'analyte (taux de récupération) ou si la présence d'un élément interfère avec la détermination de l'autre. De fait, contrairement à ce qui est suggéré, une situation comparable est rencontrée en microbiologie. Souvent, la méthode ne permettra pas de mettre en évidence la totalité des microorganismes cibles présents dans l'échantillon, dont une partie pourra ne pas se développer suffisamment dans les milieux de culture choisis. Une telle croissance partielle peut être notamment imputable à un état physiologique de stress, ou à la présence dans la matrice de substances inhibitrices, ou à la nature même des milieux de culture, sur lesquels certaines espèces/sérotypes de la population ciblée peuvent ne pas se développer. En effet, les milieux de culture des méthodes de référence normalisées sont conçus de façon à pouvoir s'appliquer de façon générique à l'ensemble des denrées alimentaires, aliments pour animaux et échantillons d'environnement agroalimentaire. Leur composition ne sera donc pas forcément optimale pour le développement des germes ciblés dans une matrice donnée.

Nous avons déjà discuté de la notion de valeur de référence (§ 1.1.2.1) et cette notion est prolongée ici par celle de matériau de référence, qui peut être utilisé pour définir cette valeur de référence. La norme laisse le choix entre :

- des matériaux de référence certifiés ;
- des matériaux préparés pour l'essai inter-laboratoires, avec des propriétés connues ;
- des matériaux dont les propriétés ont été établies par des mesures utilisant une méthode alternative dont le biais est connu pour être négligeable.

Nous retrouvons les 3^{ème} et 4^{ème} cas envisagés plus haut pour les valeurs de référence (§ 1.1.2.1).

La norme prévient le lecteur que cette partie n'est applicable que si le biais dans la mesure d'une propriété n'est pas affecté par un autre facteur. Cette condition représente une **sévère limite** à l'application de cette approche à la microbiologie des aliments et on peut même se demander s'il est envisageable de l'appliquer, puisque le niveau du microorganisme cible est influencé par le niveau des flores annexes, présentes naturellement dans la matrice. Seules les matrices stériles et contaminées artificiellement peuvent remplir au sens strict cette condition.

1.1.4.1 Biais d'une méthode d'analyse par essai inter-laboratoires

Pour calculer la fidélité, le modèle de base (voir page 34) est adapté de la façon suivante. Le terme moyen général m est remplacé par $\mu + \delta$, où μ représente la valeur de référence acceptée et δ le biais de la méthode. On en déduit le biais du laboratoire $\Delta = \delta + B$, où B est la composante laboratoire du biais.

Il s'agit de déterminer l'amplitude du biais de la méthode et de déterminer s'il est statistiquement significatif. Si le biais est non statistiquement significatif, son amplitude maximale sera déterminée.

L'organisation est identique à celle d'un essai inter-laboratoires pour la fidélité (§ 1.1.2.3). La norme définit néanmoins une approche complémentaire pour déterminer le nombre de laboratoires/essais nécessaire pour détecter avec une forte probabilité une ampleur

prédéterminée du biais.

L'analyse statistique comporte deux étapes :

1. Le contrôle de la fidélité. Il consiste à vérifier la stabilité de la variance de répétabilité conformément à la norme ISO 5725-2 (§ 1.1.3). Puis il faut comparer les écarts-types de répétabilité et de reproductibilité estimés à partir de l'essai inter-laboratoires qui estime le biais de la méthode aux mêmes écarts-types déterminés lors d'un essai précédent, qui avait établi la fidélité de la méthode.
2. L'estimation du biais de la méthode : $\hat{\delta} = \bar{y} - \mu$, où \bar{y} est la moyenne générale pour le niveau considéré.

1.1.4.2 Biais du laboratoire

L'approche proposée permet à un laboratoire d'estimer son biais à condition que la méthode qu'il utilise soit caractérisée par un écart-type de répétabilité estimé au cours d'un essai inter-laboratoires organisé au préalable. Alternativement, nous verrons que ce biais peut être aussi estimé pour un laboratoire participant à un essai inter-laboratoires d'aptitude (voir § 1.2.3).

Le schéma expérimental reprend le schéma général, avec un seul laboratoire. En matière de nombre de résultats d'analyse à obtenir, la norme donne une façon de déterminer un nombre minimal de résultats tels qu'ils soient capables de détecter, avec une forte probabilité, une amplitude prédéterminée du biais, compte tenu de la répétabilité de la méthode.

Comme pour le premier type de biais, l'analyse statistique comporte deux étapes :

1. Le contrôle de l'écart-type de répétabilité intra-laboratoire. Il consiste à rechercher les valeurs aberrantes dans les répétitions du laboratoire, et à comparer l'écart-type de répétabilité intra-laboratoire à celui de la méthode d'analyse, déterminé lors de l'essai inter-laboratoires préalable.
2. L'estimation du biais de laboratoire : $\hat{\Delta} = \bar{y}_w - \mu$, où \bar{y}_w est la moyenne des n répétitions du laboratoire.

1.1.5 ISO 5725. Partie 5

Alors que la partie 2 définit une méthode de base pour l'estimation de mesures de fidélité par essai inter-laboratoires, la partie 5 introduit des méthodes alternatives, en matière :

- de plans d'expérience :
 - plan à niveau fractionné à la place d'un plan à niveau uniforme,
 - plan pour les matériaux hétérogènes à la place d'une exigence d'identité des échantillons préparés à un niveau donné ;
- d'exploitation des résultats, en proposant l'utilisation de statistiques robustes à la place de statistiques non robustes.

Même s'il est compréhensible que l'établissement de modèles alternatifs ne se fasse qu'après le modèle initial, ces ajouts ont été adoptés non sans difficultés puisque la partie 5 de la norme n'a été publiée qu'en 1998, soit quatre années après les autres parties.

1.1.5.1 Plan à niveau fractionné

Nous avons vu précédemment qu'aucune solution idéale n'avait pu être trouvée pour effectuer les répétitions avec un plan à niveau uniforme (§ 1.1.2.3). En effet :

- Soit les répétitions sont effectuées par chaque laboratoire à partir d'un même échantillon ; ce qui permet de favoriser le respect des conditions de répétabilité entre répétitions, mais qui risque de sous-estimer la répétabilité de la méthode, l'opérateur pouvant être influencé par le résultat obtenu lors de la première analyse.
- Soit les répétitions sont effectuées en aveugle sur des échantillons identiques d'un même matériau ; ce qui assure une estimation plus juste de la répétabilité, mais éventuellement au détriment du respect des conditions de répétabilité entre les répétitions.

La partie 5 propose pour remédier à ces inconvénients d'utiliser un plan à niveau fractionné. Dans ce cas, chaque laboratoire reçoit deux échantillons provenant de matériaux **similaires** mais non identiques à chaque niveau. Les opérateurs les analysent consécutivement, donc dans des conditions satisfaisantes de répétabilité, mais ignorent l'amplitude de la différence entre ces échantillons. La norme ne présente que le cas simple de répétitions en double. Il est nécessaire de supposer que la méthode d'analyse possède les mêmes valeurs de fidélité pour ces échantillons appariés ; ce qui impose que les matériaux utilisés proviennent de la même matrice ou de matrices très proches. En pratique, dans notre domaine analytique, les échantillons appariés proviendront de la même matrice mais différeront par leurs niveaux de contamination qui devront néanmoins rester proches.

Le schéma général de la partie 2 est conservé, à l'exception du modèle statistique. En effet, la moyenne générale m_{jk} dépend ici du matériau k (a ou b) à l'intérieur du niveau j , pour le laboratoire i :

$$y_{ijk} = m_{jk} + B_{ij} + e_{ijk} \quad \text{Eq. 1}$$

On suppose ici que le biais de la méthode B_{ij} ne dépend pas du matériau ; d'où l'importance de la similarité entre les deux matériaux.

Les écarts-types de répétabilité et de reproductibilité de la méthode sont calculés à partir de :

- l'écart-type des différences entre échantillons appariés pour chaque laboratoire ;
- l'écart-type des moyennes des échantillons appariés.

Cette approche, qui présente un avantage certain par rapport à celles suivies habituellement pour effectuer les répétitions, n'est cependant pas utilisée de façon courante dans notre domaine d'intérêt même si, comme nous le verrons, l'AOAC la préconise sous la dénomination de paires de Youden. Aucune raison fondamentale ne semble expliquer cette situation.

1.1.5.2 Plan pour un matériau hétérogène

La norme distingue deux cas où l'hétérogénéité du matériau utilisé dans l'essai inter-laboratoires aura une influence sur la variabilité des résultats :

1. Un certain degré d'hétérogénéité ne peut pas être évité entre les échantillons du même niveau. La variabilité des résultats de l'essai inter-laboratoires inclura donc cette source de variabilité, qui n'est pourtant pas attribuable à la méthode d'analyse, d'où une surestimation de la répétabilité et de la reproductibilité de la méthode.
2. Certaines méthodes prévoient la réalisation d'une prise d'essai à partir de l'échantillon reçu au laboratoire. L'hétérogénéité du matériau d'essai peut être alors à l'origine de la principale source de variabilité des données. Si les échantillons utilisés pour l'essai inter-laboratoires sont préparés de telle façon à réduire cette hétérogénéité par homogénéisation, les paramètres de fidélité de la méthode seront cette fois-ci sous-estimés, à l'inverse du cas précédent.

Il se trouve justement que l'analyse microbiologique des aliments illustre ces deux cas à la fois ! En effet, l'homogénéité de la contamination est apparue particulièrement difficile à réaliser dans ce domaine. Par ailleurs, l'analyse microbiologique prévoit en général une prise d'essai à partir de l'échantillon reçu au laboratoire, et une des caractéristiques de la contamination microbienne des aliments sous forme solide est d'être hétérogène, avec des foyers de contamination localisés.

La norme propose une approche qui va permettre d'éliminer du calcul des écarts-types de répétabilité/reproductibilité la variabilité entre échantillons du même niveau, ainsi que la variabilité au sein des échantillons. Il s'agit d'un plan fractionné à trois facteurs, hiérarchisés de la façon suivante : facteur « laboratoires » > facteur « échantillons dans les laboratoires » > facteur « résultats d'essai dans les échantillons ». Le protocole expérimental de ce plan implique le doublement du nombre d'échantillons à chaque niveau, par rapport à l'ISO 5725-2.

Le modèle statistique de base est ainsi complété pour le résultat y_{ijk} du laboratoire i , au niveau j , pour l'échantillon t (a ou b), $k^{\text{ième}}$ répétition :

$$y_{ijk} = m_j + B_{ij} + H_{ijt} + e_{ijk} \quad \text{Eq. 2}$$

Le terme nouveau, H_{ijt} , représente la variation entre échantillons, supposée aléatoire et indépendante du laboratoire. Nous ne détaillerons pas ici les calculs d'écarts-types de fidélité selon ce modèle.

Comme nous l'avons mentionné, ce type de plan apparaît parfaitement adapté à l'analyse microbiologique des aliments. Cependant, les référentiels spécifiques ne l'ont pas repris et effectivement, nous n'avons pas trouvé d'essais inter-laboratoires organisés suivant ce plan. Les implications pratiques du doublement du nombre d'échantillons expliquent sans doute cette désaffection, dans un domaine où les méthodes d'analyse sont particulièrement longues et laborieuses.

1.1.5.3 Méthodes robustes d'estimation

1.1.5.3.1 Intérêt

La norme recense un certain nombre d'arguments en faveur de l'utilisation de statistiques robustes. Ces arguments se réfèrent essentiellement à l'étape d'exclusion de valeurs sur la base de tests statistiques :

- Même sur la base de tests statistiques, l'expérience montre qu'il est difficile d'exclure *objectivement* des données. Pour l'illustrer, la norme donne un exemple dans lequel un laboratoire a une moyenne de cellule aberrante et trois autres moyennes douteuses, sur 14 niveaux : doit-on écarter seulement les données de moyenne aberrante, ou toutes les données de ce laboratoire ? Le choix de l'une des deux options aura une influence non négligeable sur l'estimation des écarts-types de répétabilité et de reproductibilité. Or les statistiques robustes, recourant à la médiane comme statistique de position centrale, sont peu sensibles aux valeurs extrêmes et en, tout cas, beaucoup moins que la moyenne. Les valeurs extrêmes peuvent alors être conservées dans les calculs, sans avoir besoin de se poser la question de les exclure ou pas. Ces statistiques robustes permettent d'obtenir des estimations des écarts-types de répétabilité/reproductibilité valables pour des données de « bonne qualité », sans être affectées par les données de « mauvaises qualité », à condition que la proportion de ces dernières demeure minoritaire. Dans ce contexte, la qualité des données reflète la notion de donnée aberrante.
- La méthode de base repose sur l'hypothèse qu'on peut définir une valeur commune d'écart-type de répétabilité pour tous les laboratoires participants (§ 1.1.2.2). Or, en pratique, il n'est pas rare que certains laboratoires aient une moins bonne maîtrise de la méthode que d'autres participants, surtout lorsque cette méthode a été récemment développée. Leur répétabilité est alors significativement plus mauvaise. Les statistiques robustes représenteront alors une alternative appropriée, puisqu'elles ne requièrent pas la stabilité des variances intra-laboratoires.
- Les dénominateurs des statistiques de Mandel h et k sont des écart-types calculés avec les données brutes. Les valeurs aberrantes éventuelles augmenteront les dénominateurs et produiront un effet déformant dans les graphiques. Cette déformation est évitée avec l'utilisation d'estimateurs robustes des écarts-types et des moyennes générales.

Nous reviendrons en 2^{ème} partie sur les avantages et les limites de ces statistiques robustes.

1.1.5.3.2 Mise en oeuvre

Le plan d'expérience est identique à celui de la méthode de base. La norme insiste sur l'importance de la vérification de cohérence graphique et statistique avant de mener les calculs, même si les valeurs aberrantes ne sont pas exclues. De fait, il s'agit de rechercher les causes de toute déviation et de corriger éventuellement les erreurs manifestes.

Les deux algorithmes choisis pour une estimation robuste des écarts-types de répétabilité et de

reproductibilité, détaillés dans la 2^{ème} partie, se fondent sur les travaux du Comité des Méthodes d'Analyse de la *Royal Society of Chemistry* britannique [Analytical Method Committee, 1989a et b]. Ils ont été établis de façon à pouvoir s'appliquer à l'ensemble des plans d'expérience définis dans les différentes parties de la norme ISO 5725 et de façon à mettre en œuvre des calculs relativement simples. C'est pourquoi une approche robuste intégrale n'a pas été retenue en raison de sa complexité ; le point de départ des combinaisons robustes reste des moyennes, des écarts-types et des étendues de cellules.

1.1.6 ISO 5725. Partie 6

Après les cinq premières parties de la norme ISO 5725 portant sur les méthodes d'évaluation de l'exactitude, la dernière partie de la norme traite de l'*utilisation pratique* des données d'exactitude. Elle envisage une grande diversité d'utilisation des valeurs de fidélité et de justesse.

En particulier, la partie 6 propose une façon de prendre en compte la justesse et la fidélité dans l'interprétation des résultats d'analyse, pour évaluer la différence entre un résultat d'essai et une valeur spécifiée. C'est une utilisation fréquente des résultats d'analyse en microbiologie des aliments, puisque les analyses microbiologiques sont encore aujourd'hui souvent effectuées afin de vérifier la conformité d'un lot à des critères sanitaires microbiologiques. En fonction des risques, liés aux types de germe et d'aliment, ainsi qu'au stade de production/distribution, ces critères fixent des limites qualitatives, généralement exprimées sous la forme d'une absence dans 25 g ou 25 ml de prise d'essai, ou des limites quantitatives, comme un nombre de germes par ml ou par g.

1.1.6.1 Détermination de limites

La partie 6 introduit la notion de différence critique entre deux résultats (ou plus). Si les deux résultats sont obtenus dans des conditions de répétabilité ou de reproductibilité, cette différence critique sera appelée limite de répétabilité ou limite de reproductibilité. Ces notions sont définies dans le Tableau 2.

Tableau 2 : Limites de répétabilité et de reproductibilité

Terme	Définition/mode de calcul
Limite de répétabilité r	Valeur au-dessous de laquelle est située, avec une probabilité de 95 %, la valeur absolue de la différence entre deux résultats d'essai, obtenus sous des conditions de répétabilité.
Limite de reproductibilité R	Valeur au-dessous de laquelle est située, avec une probabilité de 95 %, la valeur absolue de la différence entre deux résultats d'essai, obtenus sous des conditions de reproductibilité.
Différence critique de répétabilité/reproductibilité	$f\sigma\sqrt{2} = 1,96 \sqrt{2}\sigma \approx 2,8 \sigma$ où : f est le facteur d'étendue critique, qui dépend de la probabilité associée à la différence critique et de la forme de la distribution sous-jacente. Dans les conditions de la norme ISO 5725, la probabilité est de 95 % et la <u>distribution étant supposée normale</u> , $f = 1,96$; σ est l'écart-type de répétabilité ou de reproductibilité.

Pratiquement, deux répétitions effectuées dans des conditions de répétabilité peuvent être considérées comme satisfaisantes, en termes de compétence de l'opérateur et de maîtrise de la méthode, si la différence entre deux résultats ne dépasse pas $2,8 s_r$. De même, deux répétitions effectuées par deux laboratoires dans des conditions de reproductibilité peuvent être considérées comme satisfaisantes si la différence entre les deux résultats n'excède pas $2,8 s_R$.

L'hypothèse de normalité est elle-même critique, puisque nous verrons dans la 2^{ème} partie qu'elle est habituellement remise en cause en microbiologie. Il conviendra donc d'être très prudent avant d'utiliser les limites de fidélité de cette façon.

La norme applique ensuite la notion de différence critique à plus de deux valeurs : deux groupes de mesures dans le même laboratoire, ou dans deux laboratoires différents.

1.1.6.2 Comparaison à une valeur de référence

On compare la moyenne \bar{y} de n résultats obtenus dans un laboratoire dans des conditions de répétabilité à une valeur de référence ou à une spécification μ_0 , à l'aide de la différence critique (CD) :

$$CD = \frac{1}{\sqrt{2}} \sqrt{(2,8\sigma_R)^2 - (2,8\sigma_r)^2 \left(\frac{n-1}{n} \right)} \quad \text{Eq. 3}$$

Une généralisation à plusieurs laboratoires est proposée.

Dans le domaine de la microbiologie, la comparaison par exemple à un seuil limite pour un critère microbiologique ne semble pas envisageable de façon générale, car, en routine, les analyses sont rarement répétées, en raison du coût trop élevé et de la lourdeur des méthodes conventionnelles d'analyse microbiologique.

1.1.6.3 Acceptabilité des résultats et détermination du résultat final

L'acceptabilité des résultats dans un laboratoire présuppose que des répétitions aient été systématiquement réalisées, or nous venons de voir que c'est peu probable en microbiologie conventionnelle. Nous ne présenterons donc pas ici cette approche, pourtant séduisante en théorie.

Nous ne retiendrons que le cas où deux laboratoires obtiennent chacun un résultat (x_1, x_2) sur des échantillons identiques, R étant la limite de reproductibilité définie dans le Tableau 2 :

- Si $|x_1 - x_2| \leq R$, les deux résultats sont en accord et on prend comme résultat final $\frac{x_1 + x_2}{2}$
- Si $|x_1 - x_2| > R$, les résultats diffèrent et il convient d'en rechercher la cause, soit une différence entre les échantillons, ou un défaut de l'un (ou des deux) laboratoire(s).
- En cas de désaccord entre les résultats des deux laboratoires, la norme propose des voies de résolution de ce désaccord :
 - les deux laboratoires peuvent échanger les échantillons, ou obtenir un résultat supplémentaire sur un échantillon commun, de préférence un matériau de référence ;

- si le désaccord semble être dû à une différence entre échantillons, les deux laboratoires définissent une préparation commune de l'échantillon pour essai, ou se réfèrent à un troisième laboratoire.

1.1.6.4 Suivi de la justesse dans un laboratoire

Dans le cadre d'un système d'assurance qualité interne, le meilleur moyen de contrôle est de mettre en place une procédure de suivi de la stabilité des résultats du laboratoire, soit contrôler la justesse et la fidélité des mesures à intervalle régulier.

Le *suivi de la justesse* à intervalle régulier requiert l'utilisation de matériaux de référence, ou au moins de matériaux dont l'analyte reste stable. Or nous avons vu que cette propriété est difficile à respecter pour un analyte vivant. La norme indique par ailleurs une alternative intéressante à l'utilisation des matériaux de référence, surtout si ceux-ci ne sont pas disponibles. Un résultat d'essai obtenu par un opérateur expérimenté, dans un laboratoire bien équipé, suivant strictement la méthode normalisée, peut être utilisé comme valeur de référence à la place de la valeur du matériau de référence. Cette alternative à l'utilisation des matériaux de référence apparaît comme tout à fait appropriée pour l'analyse des aliments, domaine dans lequel il est usuel de définir et de faire appel à des laboratoires experts/laboratoires de référence, dans le cadre des contrôles réglementaires ou des échanges commerciaux. Ces laboratoires sont choisis et désignés justement sur la base de leur compétence à mettre en œuvre les méthodes de référence. Ils sont donc tout naturellement aptes à fournir des valeurs de référence.

La norme recommande ensuite l'utilisation de cartes de contrôle de Shewhart, mises au point initialement pour la maîtrise de processus de fabrication industrielle ou des cartes de contrôle à somme cumulée. Ces dernières sont plus efficaces pour contrôler une tendance à dériver, cas probable de la justesse. Ces cartes de contrôle font l'objet de la norme ISO 8258.

1.1.6.5 Évaluation des laboratoires

La norme envisage trois sortes d'évaluation :

1. Evaluation d'un laboratoire en utilisant des matériaux de référence.
2. En l'absence de matériaux de référence, évaluation indirecte d'un laboratoire par comparaison avec un laboratoire de référence, dont le biais doit être connu.
3. Evaluation collective en comparant les résultats de chaque laboratoire avec ceux des autres laboratoires ayant participé à un essai inter-laboratoires. Dans ce cas, la limite de reproductibilité R peut être utilisée pour évaluer le biais de chaque laboratoire.

Ce type d'utilisation des essais inter-laboratoires fera l'objet du chapitre suivant.

1.1.6.6 Comparaison de méthodes alternatives

Dans le contexte agroalimentaire, un objectif classique des essais inter-laboratoires, est de vouloir comparer la fidélité et la justesse de deux méthodes, l'une étant considérée comme méthode de référence (A) et l'autre comme méthode candidate à la reconnaissance comme méthode alternative à la méthode de référence (B).

Pour ce faire, un essai inter-laboratoires sera organisé sur les deux méthodes en parallèle, et les résultats seront interprétés de façon à pouvoir mettre en évidence une différence significative entre les deux méthodes :

- La méthode B est-elle aussi fidèle que la méthode A ? C'est-à-dire, est-ce que le rapport des mesures de fidélité de la méthode B sur la méthode A est supérieur à une valeur critique donnée ?
- La justesse de la méthode B est-elle égale à la justesse de la méthode A ? C'est-à-dire, est-ce que la différence entre les moyennes générales des résultats sur des échantillons identiques est statistiquement non significative ?

Ce cas de figure fait l'objet de plusieurs référentiels et pratiques dans notre domaine, comme nous le verrons dans les § 1.1.9, 1.1.10 et 1.1.12.

1.1.7 Lignes directrices IUPAC/AOAC

Après avoir détaillé le contenu de la norme de base, d'application générale, nous allons présenter plus succinctement deux référentiels s'appliquant aux méthodes d'analyse des aliments, les lignes directrices IUPAC/AOAC et la Norme FIL 135 B, en se concentrant sur leurs aspects spécifiques et sur les choix effectués par les auteurs parmi les différentes possibilités prévues par la norme ISO 5725.

1.1.7.1 Présentation

Un groupe de travail « Harmonisation des schémas d'assurance qualité pour les laboratoires d'analyse » de IUPAC a développé des lignes directrices pour l'organisation et l'exploitation des essais inter-laboratoires pour le domaine agro-alimentaire [IUPAC, 1995], respectant dans l'ensemble les prescriptions de la norme ISO 5725. Ce travail a été mené en collaboration avec AOAC qui a adopté ces mêmes lignes directrices [AOAC, 1995] comme référentiel de base à utiliser pour toute méthode candidate à une validation AOAC. C'est ce dernier document que nous allons examiner.

Le protocole commun IUPAC/AOAC est plus pratique que la norme ISO 5725, prodiguant un grand nombre de conseils pour réussir un essai inter-laboratoires. Le protocole a été initialement développé pour les méthodes chimiques, mais il apparaît que bon nombre de ses dispositions sont applicables de façon plus générale.

Le document actuel représente la révision d'une première version publiée par IUPAC et AOAC, respectivement en 1988 et 1989. Il est le fruit d'une expérience très riche, puisque AOAC est sans doute la structure qui organise le plus grand nombre d'essais inter-laboratoires visant à valider des méthodes dans le domaine alimentaire. En effet, à la différence des organismes de normalisation, l'obtention de valeurs de fidélité est une condition *sine qua non* pour la reconnaissance de méthodes par AOAC comme méthodes officielles. Cette reconnaissance suit le schéma de validation AOAC le plus complet et, historiquement, le seul jusqu'aux années récentes [Andrews, 1996].

1.1.7.2 Protocole expérimental

Le protocole expérimental inclut une première étape qui n'est pas prévue dans la norme ISO 5725 : une étude préliminaire effectuée par un laboratoire qui est normalement le laboratoire pilote, organisateur de l'essai inter-laboratoires. Cette étude préliminaire comporte les étapes suivantes :

1. *Sélection* de la méthode appropriée à l'objectif poursuivi et éventuellement remise en cause du choix initial.
2. *Optimisation* de la méthode, en fonction des interférences et de la robustesse de la méthode, testée pour identifier et maîtriser les variables critiques. Ici est rappelée l'importance d'optimiser suffisamment une méthode avant de la soumettre à un essai inter-laboratoires ; le risque étant d'invalider le résultat d'un tel effort collectif qui aurait été consacré à une méthode non suffisamment optimisée.
3. *Caractérisation* intra-laboratoire de la méthode optimisée : spécificité (interférences), variabilité (répétabilité et reproductibilité intra-laboratoire) et domaine d'application.

4. *Description* détaillée de la méthode.
5. *Étude pilote* incluant trois laboratoires, si les ressources et le temps le permettent. AOAC n'accorde pas une importance particulière à cette étape, également préconisée par la FIL.

En accord avec l'ISO 5725, mais avec plus de détails, la procédure AOAC recommande de sélectionner les participants qui ont une expérience dans les techniques de base requises, sans pour autant avoir forcément une expérience de la méthode elle-même. AOAC fait ici preuve d'une grande flexibilité, puisqu'il est possible de sélectionner les participants dans trois ensembles différents :

1. Un échantillon aléatoire de laboratoires qui utiliseront la méthode en pratique ;
2. Des laboratoires de référence ;
3. La population entière de laboratoires qui utiliseront la méthode en pratique (option de l'ISO 5725).

Le choix final se fait sur la base de la participation du laboratoire à des essais précédents et éventuellement en fonction de sa performance avec des échantillons de familiarisation. AOAC et IUPAC ont le mérite d'avoir introduit cette notion d'échantillon de familiarisation, qui présente un intérêt tout particulier lorsque la méthode testée est nouvelle et fait appel à des techniques analytiques complexes. L'essai inter-laboratoires ne sera organisé que lorsque les participants auront été capables de retrouver la valeur attendue, dans un intervalle donné. Ainsi, sera respectée l'une des hypothèses de l'analyse de variance : la distribution sous-jacente des résultats est indépendante du temps, soit une absence de dérive.

Des instructions nettement plus détaillées que dans l'ISO 5725 sont fournies pour préparer le formulaire de renseignements sur le mode opératoire précis de chaque participant, ainsi que le rapport d'essai où chaque participant indiquera ses résultats.

Pour effectuer les répétitions, le protocole laisse le choix entre des doubles aveugles, ou des échantillons appariés de Youden, soit le plan à niveau fractionné introduit par l'ISO 5725-5 comme alternative au schéma classique. Nous avons déjà vu que cette deuxième solution présentait un net avantage (§ 1.1.5.1) et d'ailleurs elle est retenue ici comme l'option à préférer. En ce qui concerne le nombre de répétitions, l'AOAC justifie le choix de se limiter à des doubles : la répétition n'apporte une information que sur la variabilité intra-laboratoire qui est généralement la composante la moins importante de l'erreur. Il est préférable d'augmenter le nombre de niveaux étudiés. L'AOAC utilise le terme de « *material* » à la place du terme « niveau » employé par l'ISO 5725.

Les détails concernant le protocole d'organisation de l'essai inter-laboratoires ne seront pas repris ici, puisque AOAC a établi un protocole qui est spécifique aux méthodes microbiologiques et que nous aborderons en détail dans le § 1.1.10.

1.1.7.3 Analyse statistique

Certes, les tests d'identification par combinaison <matériau x niveau de concentration> des valeurs aberrantes sont bien ceux de l'ISO 5725-2, soit le test de Cochran pour les écarts-types intra-laboratoires et le test de Grubbs pour les moyennes de cellules (voir leur description détaillée en 2^{ème} partie, § 2.1.2).

Cependant, AOAC a retenu une valeur critique différente pour le rejet de valeur aberrante par les deux tests : 2,5 %, au lieu de 1 % dans l'ISO 5725-2. De plus, la séquence d'application des tests est différente :

- L'ISO 5725-2 prévoit une séquence relativement simple :
 - Si un écart-type intra-laboratoire est considéré comme aberrant par le test de Cochran, cet écart-type est éliminé des calculs. Le test est éventuellement répété sur les valeurs restantes, avec beaucoup de précaution pour éviter des rejets excessifs qui pourraient être dus à un défaut de normalité de la distribution sous-jacente.
 - Si une moyenne de cellule est considérée comme aberrante par le test simple de Grubbs parce que c'est une valeur extrême haute ou basse, cette moyenne est éliminée des calculs et le test est répété sur la moyenne extrême suivante. Si le test de Grubbs simple ne met pas en évidence de valeur aberrante, le test de Grubbs double est appliqué sur les deux moyennes extrêmes groupées, les deux plus grandes et les deux plus petites.
- L'AOAC prévoit une séquence plus complexe pour les moyennes de cellules :
 - Initialement, application unique du test de Cochran.
 - Si le test de Grubbs détecte comme aberrante(s) une ou deux moyennes extrêmes, ces valeurs sont non seulement éliminées des calculs, mais un nouveau cycle de tests (Cochran et Grubbs) est appliqué aux valeurs restantes. Les conditions d'application du test de Grubbs double sont identiques, mais là aussi la détection de valeurs aberrantes entraîne l'exécution d'un nouveau cycle de tests.
 - Une limite pratique est fixée à cette approche itérative d'élimination des valeurs aberrantes. Par niveau étudié, au plus 2/9^{èmes} des laboratoires participants peuvent être éliminés.

Ainsi, en dépit de cette limite en matière de données éliminées, il est probable que l'approche AOAC amène à exclure plus de moyennes de cellules que ne le ferait l'ISO 5725-2.

Le protocole reprend ici fidèlement l'ISO 5725-2 pour l'expression de la justesse, ainsi que pour les écarts-types/limites de répétabilité et de reproductibilité. Néanmoins, AOAC introduit la notion intéressante d'incertitude sur l'estimation des valeurs vraies d'écarts-types : pour environ 30 données par niveau, soit un ordre de grandeur habituel pour un essai en double avec 15 laboratoires, l'intervalle de confiance, à un niveau de 95 %, est de l'ordre de ± 25 %. Ceci devrait relativiser grandement la portée des essais inter-laboratoires tels qu'ils sont organisés !

1.1.8 Norme FIL 135 B

La Fédération Internationale Laitière (FIL) propose une norme qui découle directement de l'ISO 5725, mais qui constitue un guide nettement simplifié et dépourvu de considérations théoriques. Cette norme est destinée aux membres de groupes de normalisation, et elle fournit aux organisateurs d'essai inter-laboratoires des indications très pratiques sur le choix du protocole expérimental, le choix et la préparation des échantillons ou les relations avec les participants. Tout comme le protocole IUPAC/AOAC, le domaine d'application est limité aux méthodes chimiques quantitatives, mais la lecture du contenu de la norme révèle qu'une partie importante peut s'appliquer à priori à la microbiologie quantitative. Seuls certains aspects, comme le type de matrice à utiliser pour préparer les échantillons, sont spécifiques de l'analyse des composants des produits laitiers.

Le protocole expérimental ne retient qu'un des cas de l'ISO 5725 : l'analyse en double aveugle. Les exigences minimales sont précisées :

- au moins six matrices et trois niveaux de concentration, ou au moins trois matrices si un seul niveau est étudié ;
- et au moins 8 laboratoires participants.

Cette norme préconise l'organisation d'une étude pilote, impliquant au moins trois laboratoires. Afin d'éviter d'invalider un essai complet, cette étape préliminaire peut s'avérer très utile pour tester la faisabilité de l'essai, en particulier en ce qui concerne les matrices utilisées et leur contamination.

Les calculs et tests statistiques de valeurs aberrantes reprennent ceux de l'ISO 5725-2, même si la norme FIL a été publiée trois ans avant la dernière version de la norme ISO 5725. La norme FIL comporte cependant l'exigence complémentaire suivante. Si les valeurs considérées comme aberrantes représentent plus de 20 % de la totalité des résultats, aucune exclusion de valeurs aberrantes n'est effectuée. Un tel pourcentage est en effet considéré comme imputable à un nombre trop faible de participants, ou à un manque de normalisation de la méthode. Ce taux maximal d'exclusion est proche de celui de l'AOAC ($\approx 22\%$), mais l'utilisation qui en est faite est différente, puisque l'AOAC conserve l'élimination des valeurs aberrantes jusqu'à $2/9^{\text{èmes}}$ des valeurs. **Cette notion de taux maximal d'exclusion nous semble être un complément intéressant à l'ISO 5725-2, surtout dans le sens de la FIL.**

1.1.9 Norme EN ISO 16140

En ce qui concerne les référentiels spécifiques du domaine de la microbiologie des aliments, la Norme EN ISO 16140 et les lignes directrices de l'AOAC, leur présentation est ici aussi limitée aux aspects qui diffèrent de l'ISO 5725, ou qui sont spécifiques de l'analyse microbiologique des aliments. Nous allons les présenter cependant de façon plus détaillée que dans les deux paragraphes précédents, puisque le champ précis de notre travail est maintenant atteint.

1.1.9.1 Présentation

Une place particulière est accordée à cette norme, qui représente le premier référentiel normalisé à traiter de la validation de méthodes d'analyse microbiologique des aliments. Cette norme a été préparée au niveau européen par le Groupe de Travail 6 « Contaminants microbiens » du Comité Technique 275 « Analyse des denrées alimentaires – Méthodes horizontales » du CEN (CEN/TC 275/WG 6), sur la base d'un protocole développé par le projet MicroVal du programme européen Eurêka. Le projet MicroVal visait à établir un système européen de validation de méthodes alternatives en microbiologie des aliments [Rentenaar, 1996]. Il est intéressant de remarquer que si MicroVal a transmis au CEN un protocole complet dès 1996, la norme n'a pu être publiée que 7 ans plus tard. En effet, l'élargissement du consensus à l'ensemble des pays membres du CEN sur un sujet aussi vaste a été un processus difficile. Auparavant, ce sujet ne faisait pas l'objet d'un document de référence normalisé, et il a nécessité pour le traiter une norme de 70 pages.

Cette norme conçoit la validation d'une méthode dans son sens le plus complet, puisque deux étapes sont prévues :

1. Une étude préliminaire, effectuée par un seul laboratoire. Cette étude permet de caractériser la méthode par plusieurs critères de performance, soit la justesse, la limite de détection/quantification, la spécificité –inclusivité et exclusivité–, ainsi que la linéarité pour les méthodes quantitatives.
2. Un essai inter-laboratoires.

L'étude préliminaire est beaucoup plus complète que celle préconisée par AOAC (§ 1.1.7.2). Elle vise en effet à établir un plus grand nombre de critères de performance analytique, dont la justesse sur la base de l'étude d'échantillons naturellement contaminés.

Le protocole technique pour l'organisation et l'exploitation de l'étude préliminaire et de l'essai inter-laboratoires est défini séparément pour les méthodes qualitatives (présence/absence d'un microorganisme) et pour les méthodes quantitatives (dénombrement de microorganismes, ou méthodes indirectes de dénombrement).

La norme a pour objectif principal de valider une méthode alternative par rapport à une méthode de référence. Cependant, le protocole technique peut être utilisé, dans une large mesure, pour valider une méthode en elle-même. Notamment en ce qui concerne l'essai inter-laboratoires, le protocole expérimental et l'exploitation des résultats sont applicables à une seule méthode, en omettant les aspects liés à la comparaison de deux méthodes.

1.1.9.2 Organisation des essais et protocole expérimental

L'annexe H de la norme définit des lignes directrices pour l'organisation des essais inter-laboratoires, que ces essais portent sur des méthodes qualitatives ou quantitatives. La plupart de ces dispositions sont spécifiques de la microbiologie.

Choix des échantillons et types de matrices. Il est indiqué que les matrices liquides, comme le lait, permettent une meilleure homogénéité s'ils sont préparés et répartis de façon à ne pas modifier le contenu microbiologique. Dans ce cas, un ensemencement global de la matrice liquide peut être pratiqué, avant de préparer chaque échantillon. Outre la question de l'homogénéité, ce mode de contamination peut permettre de se rapprocher des conditions d'une contamination naturelle, le microorganisme cible pouvant établir des interactions avec la matrice et avec la flore annexe.

La quantité de matrice utilisée par échantillon est prévue de telle façon qu'une seule analyse puisse être effectuée. En d'autres termes, l'échantillon correspond à la prise d'essai. Ainsi, le laboratoire ne pourra pas répéter l'analyse d'un échantillon donné ; ce qui pourrait biaiser l'estimation de la justesse et de la répétabilité de la méthode, en sous-estimant la valeur vraie.

Type de contamination. La contamination artificielle est la règle générale – strictement pour les déterminations qualitatives –, à partir d'un matériau dont on a démontré, selon un protocole précisé, qu'il ne contient pas le microorganisme cible. La contamination artificielle doit être effectuée (annexe C de la norme) :

- soit par *mélange* du matériau liquide ou semi-solide avec un échantillon naturellement contaminé du même type ;
- soit par une contamination du matériau avec un matériau de référence ou une souche isolée du même type de produit et soumise à un stress ;
- toujours en présence d'une microflore de fond représentative dans sa nature et sa quantité de celle rencontrée naturellement dans le produit considéré.

Il peut être néanmoins noté que le texte de la norme manque de précision sur cet aspect, les modalités de contamination artificielle restant assez floues, tout comme le type de stress à appliquer, qui n'est pas décrit.

La norme préconise dans le cas général un ensemencement individuel de chaque échantillon, y compris pour chacune des répétitions. Ce type de contamination représente effectivement une façon satisfaisante de tendre vers l'homogénéité de contamination entre échantillons du même niveau, à condition de s'assurer de l'homogénéité de la solution d'ensemencement elle-même, ou de tout autre support d'ensemencement.

Homogénéité/stabilité. La norme est très lapidaire sur ce sujet, se contentant de :

- fixer les objectifs (*i*) d'homogénéité vis-à-vis du microorganisme cible et de la flore annexe, et (*ii*) de stabilité des échantillons pendant le transport jusqu'au lancement des analyses ;
- et demander au laboratoire organisateur de mener des études d'homogénéité et de stabilité avant de distribuer le matériau d'essai.

Il s'agit là certainement d'un des principaux **points faibles** de cette norme. Cet aspect pourtant essentiel dans la réussite d'un essai inter-laboratoires en microbiologie des aliments a été traité trop rapidement ; des indications sur la façon d'assurer l'homogénéité et la stabilité auraient dues être fournies, de même qu'une façon de les vérifier.

Transport. Des prescriptions adaptées au transport de denrées périssables sont données, concernant en particulier la température de réfrigération ou de congélation. Il est demandé d'inclure dans chaque paquet un échantillon témoin pour mesurer la température à réception. De plus, le laboratoire organisateur devrait vérifier l'effet des conditions de transport en faisant effectuer à une série d'échantillons un aller-retour vers les laboratoires participants, avant l'étude elle-même. Cette dernière demande paraît peu réaliste, compte tenu des coûts et du temps qu'elle requiert, et elle n'est pas appliquée, à notre connaissance.

Vérification de la qualité des échantillons. Chaque participant doit dénombrer la flore totale d'un des échantillons. Cette technique de vérification de la qualité microbiologique des échantillons apparaît facile à mettre en œuvre et il serait intéressant d'en généraliser la pratique.

De plus, le laboratoire organisateur doit analyser le jour du début de l'analyse pour l'ensemble des participants un certain nombre d'échantillons de chacune des combinaisons <matrices x niveau de contamination en microorganisme cible>, conservés dans des conditions équivalentes à celles des échantillons distribués.

Conditions opératoires de l'analyse. À propos de l'une des conditions opératoires critiques, à savoir la qualité et la composition des milieux de culture, la norme recommande deux options :

1. Soit limiter la variabilité des milieux de culture, en distribuant des milieux issus d'un même lot de fabrication à tous les participants.
2. Soit prendre en compte cette variabilité lors de l'interprétation des données.

La première option nous semble devoir être exclue, car l'objectif d'un essai inter-laboratoires de validation d'une méthode est d'obtenir une estimation des critères de fidélité qui caractérise la méthode lors de son application ultérieure, en routine. Or, à partir d'une composition-type décrite dans une méthode normalisée, on rencontre en pratique une variabilité dans la qualité des milieux de culture, en fonction des lots de fabrication et des fabricants. Il convient donc de prendre en compte cette variabilité lors de l'établissement des critères de fidélité de la méthode. Cette variabilité peut être très importante pour certains types de milieux, comme les essais inter-laboratoires d'aptitude le montrent. Ainsi, le RAEMA note, dans des essais récents sur le dénombrement de la flore totale mésophile, jusqu'à 30 % de laboratoires obtenant des résultats douteux ou non satisfaisants, alors qu'une méthode de référence normalisée a été utilisée par 85 % des participants [Augustin & Carlier, 2002 ; RAEMA, 2004].

Quant aux autres conditions opératoires (lancement de l'analyse, temps/température d'incubation, pesées...), la norme recommande à juste titre de les spécifier le plus précisément possible et d'indiquer le niveau de tolérance accepté.

Questionnaire accompagnant le rapport d'analyse. La norme précise le contenu du questionnaire que devra remplir chaque participant, en même temps que le rapport d'essai. Ce questionnaire doit recenser les points critiques pour la variabilité des résultats, en ce qui concerne :

- l'organisation, soit l'état des échantillons après transport, les conditions de leur conservation avant analyse ;
- le mode opératoire ;
- le contrôle qualité des milieux de culture et les éléments indiqués ci-dessus.

1.1.9.3 Cas des méthodes qualitatives

1.1.9.3.1 Protocole expérimental et examen des résultats

Aucun des référentiels que nous avons présentés jusqu'à maintenant a traité des méthodes qualitatives. Or ces analyses représentent une part importante des contrôles effectués en microbiologie des aliments, sinon la quasi-totalité des analyses pour les germes pathogènes, notamment dans le cadre des contrôles officiels. Ce sujet ne pouvait donc pas être éludé par une norme portant sur les essais inter-laboratoires en microbiologie des aliments.

Outre les aspects généraux déjà exposés du protocole expérimental, certaines exigences complémentaires sont précisées :

- au moins 10 laboratoires participants dont les résultats sont exploitables ;
- un seul matériau (une matrice) à choisir, en fonction du germe étudié ;
- 3 niveaux de contamination :
 - un témoin négatif (noté L_0),
 - un niveau juste supérieur à la limite de détection de la méthode (L_1),
 - un niveau à environ 10 fois cette limite de détection (L_2) ;
- 8 répétitions à chaque niveau de contamination ;
- analyse en aveugle.

Le choix d'un nombre élevé de répétitions, par rapport au nombre habituellement retenu de 2, est fondé sur des considérations statistiques. En effet, des doubles ne permettent pas d'affirmer avec un degré de confiance suffisant que deux laboratoires obtiennent les mêmes résultats. Ainsi deux résultats tous deux positifs, ou tous deux négatifs, ou un positif/un négatif ont une répétabilité équivalente, et de tels résultats peuvent n'être dûs qu'au hasard. Le nombre de 8 répétitions a été considéré comme suffisant pour déceler des différences de répétabilité entre laboratoires.

La norme proscrit l'élimination de valeurs pouvant être considérées comme aberrantes, à l'exception de celles qui peuvent être expliquées par de raisons techniques, telles que :

- mauvais état de l'échantillon (fuites), température et durée de transport non respectées ;
- lancement de l'analyse à une date autre que celle prévue ;
- valeurs du dénombrement témoin en dehors de la tolérance spécifiée ;
- écart significatif au mode opératoire de la méthode d'analyse, tel que l'utilisation d'un milieu de culture commercialisé dont la formule n'est pas conforme à celle de la méthode, temps/température d'incubation non respectés/hors tolérance,...

1.1.9.3.2 Critères de performance

Sont reprises ici les notions classiques de sensibilité, spécificité et exactitude (justesse), qui permettent de caractériser la **justesse** des méthodes qualitatives. Cependant, la définition et l'estimation de la justesse sont adaptées à l'objectif général de cette norme ; à savoir la comparaison d'une méthode alternative à une méthode de référence, qui donne la valeur de référence.

La définition et le mode de calcul de chacun de ces critères figurent dans le Tableau 3.

Tableau 3 : Critères de sensibilité, spécificité et exactitude pour les méthodes qualitatives, selon la Norme EN ISO 16140

Critère	Définition	Calcul
Spécificité	Capacité de la méthode à donner une réponse négative pour un échantillon dont on a démontré qu'il ne contient pas le germe cible, soit un échantillon témoin négatif.	<p>Pourcentage de spécificité SP :</p> $SP = \left(1 - \left(\frac{FP}{N-} \right) \right) \times 100\%$ <p>où :</p> <ul style="list-style-type: none"> - $N-$ est le nombre total de résultats au niveau L_0 ; - FP est le nombre de faux positifs.
Sensibilité	Capacité de la méthode à donner une réponse positive pour un échantillon dont on a dit qu'il contient le germe cible.	<p>Pourcentage de sensibilité SE pour le niveau L_1 (ou L_2) :</p> $SE = \frac{TP}{N+} \times 100\%$ <p>où :</p> <ul style="list-style-type: none"> - $N+$ est le nombre total des résultats à L_1 (ou L_2) ; - TP est le nombre de vrais positifs à ce même niveau.
Exactitude (justesse)	Niveau de correspondance, sur des échantillons identiques, entre la réponse de la méthode alternative et la réponse de la méthode de référence.	<p>Pourcentage d'exactitude AC :</p> $AC = \frac{PA + NA}{N} \times 100\%$ <p>où :</p> <ul style="list-style-type: none"> - PA (accords positifs) est le nombre de résultats positifs par les deux méthodes ; - NA (accords négatifs) est le nombre de résultats négatifs par les deux méthodes ; - N est le nombre total de résultats aux trois niveaux.

L'emploi du terme exactitude est ici surprenant. Une note sous la définition de ce terme justifie ce choix de la norme, se référant aux définitions des termes justesse et exactitude de l'ISO 5725-1. La justesse faisant intervenir, à la différence de l'exactitude, une « valeur moyenne obtenue à partir d'une large série de résultats d'essai », les auteurs de la norme ont estimé que le protocole expérimental ne faisait pas appel à un nombre suffisant d'échantillons pour pouvoir considérer que l'estimation de ce paramètre était fondée sur une large série d'essais. Cependant, ce choix apparaît contestable, car il introduit une confusion dans l'acceptation du terme exactitude. Ainsi, l'ISO 5725-1 emploie ce terme pour désigner une caractéristique globale de la méthode d'analyse, incluant justesse et fidélité, alors qu'ici il ne

recouvre, d'après le protocole expérimental, que la notion de justesse.

L'annexe E de la norme indique une façon de calculer un intervalle de confiance autour de chaque proportion, associé au nombre d'échantillons du protocole expérimental permettant d'estimer la proportion. Ainsi, l'incertitude associée à l'estimation de ces paramètres est connue, en fonction du nombre de résultats utilisés. Bien évidemment, cette incertitude diminue avec l'augmentation du nombre de résultats.

En ce qui concerne la justesse, si la définition de la norme est prévue pour une comparaison de deux méthodes, il est facile de l'adapter au cas de l'évaluation d'une seule méthode, en remplaçant la notion de méthode de référence par celle de valeur de référence. Les accords positifs deviennent alors des vrais positifs (on sait que l'échantillon contient le germe cible) et les accords négatifs deviennent des vrais négatifs (on sait que les échantillons ne contiennent pas le germe cible).

Les critères de sensibilité, spécificité et exactitude caractérisent, comme nous l'avons vu, la justesse d'une méthode qualitative. Or, l'objectif habituel d'un essai inter-laboratoires, du moins pour les méthodes quantitatives, est d'estimer la **fidélité** d'une méthode. Les critères de répétabilité et de reproductibilité, définis pour les méthodes quantitatives, ne peuvent a priori pas s'appliquer, puisqu'ils font appel à des écarts-types et les limites à des différences acceptables entre deux résultats.

C'est pourquoi la norme introduit, en annexe L, de nouveaux critères qui visent à caractériser la variabilité de la méthode au sein d'un laboratoire, le degré d'accord, et entre laboratoires, la concordance. La norme introduit avec prudence ces critères, en les qualifiant de « supplémentaires » et en les faisant figurer non dans le corps principal de la norme, comme les précédents, mais en annexe informative. En l'an 2000, ces critères venaient en effet d'être tout juste proposés par un projet européen du programme « Mesures et Essais », que nous présenterons plus loin (§ 1.1.11.5). Ils n'avaient pas eu encore le temps d'être diffusés et soumis à critique, démarche habituelle de la normalisation. Ces critères seront décrits et discutés dans la 2^{ème} partie (§ 2.1.3).

1.1.9.4 Cas des méthodes quantitatives

1.1.9.4.1 Protocole expérimental

Tout comme pour les méthodes qualitatives, les aspects spécifiques s'ajoutant aux lignes directrices communes sont recensés ci-après :

- Au moins 8 laboratoires participants dont les résultats sont exploitables.
- Mode de contamination. En dépit des avantages d'une contamination artificielle (homogénéité de la contamination), la possibilité d'avoir recours à une contamination naturelle est ouverte. En effet, certains dénombrements en microbiologie des aliments portent sur des cibles très larges, telles que la flore totale, les levures et moisissures, les coliformes,... Pour ces cibles, il est très difficile sinon impossible de reconstituer artificiellement un inoculum avec des souches d'espèces d'une diversité suffisante pour représenter la population ciblée.
- Un seul matériau (matrice) à choisir, en fonction du germe étudié.
- 4 niveaux de contamination : un témoin négatif, un niveau inférieur, moyen et supérieur couvrant l'ensemble du domaine d'application.

- Echantillons en double à chaque niveau de contamination, analysés en aveugle. Le schéma classique de doubles aveugles a donc été repris ici.

1.1.9.4.2 Modes de calcul des estimateurs de fidélité

Pour des raisons similaires à celles invoquées dans l'ISO 5725. Partie 5, il a été choisi de faire appel à des estimateurs robustes de répétabilité et de reproductibilité. En particulier, est invoquée la non normalité rencontrée fréquemment pour la distribution statistique des résultats de dénombrements microbiens. Une transformation logarithmique des résultats est néanmoins prescrite pour rapprocher leur distribution de la normalité. Conséquence de l'utilisation des statistiques robustes, les valeurs apparaissant comme aberrantes ne sont pas exclues, sauf pour des raisons techniques, les mêmes que pour les méthodes qualitatives.

Cependant, nous ne retrouverons pas les mêmes estimateurs robustes que ceux préconisés par l'ISO 5725-5. Ce manque d'harmonisation ne semble pas avoir de raison technique, liée à la spécificité de la microbiologie. Au moment de la rédaction de l'EN ISO 16140, la norme ISO 5725-5 n'était pas encore parue et, de ce fait, n'a pas pu être prise en compte par les auteurs. Nous verrons néanmoins que les estimateurs robustes de l'EN ISO 16140 présentent un avantage d'ordre statistique sur ceux de l'ISO 5725-5. Les algorithmes sont décrits dans la 2^{ème} partie, dans le § 2.1.2.

1.1.10 Lignes directrices AOAC

1.1.10.1 Présentation

Ces lignes directrices, publiées récemment [Feldsine *et al*, 2002], respectent les principes du protocole AOAC d'application générale, notamment dans la conception de la validation en deux étapes : étude préliminaire puis essai inter-laboratoires. Elles reposent elles aussi sur une grande expérience d'organisation d'essais inter-laboratoires. Expérience d'autant plus précieuse qu'unique pendant plusieurs années, dans le domaine de la microbiologie des aliments. Dès le début du document, les auteurs affichent leur intention d'harmoniser ce protocole avec la norme EN ISO 16140. De fait, les définitions sont pour la plupart tirées de la norme et nombre de spécifications du protocole expérimental sont communes aux deux protocoles. La préparation simultanée de ces deux référentiels a favorisé ce rapprochement et nous avons là un bel exemple d'harmonisation, où chaque partie a fait des concessions pour se rapprocher de l'autre. Les deux exemples suivants nous paraissent significatifs :

- Le groupe du CEN a accepté de faire passer de 8 à 10 le nombre minimal de laboratoires pour les essais portant sur des méthodes qualitatives, afin de trouver un accord avec AOAC. De son côté, AOAC, qui préconisait auparavant au moins 15 participants, a accepté de réduire ce minimum à 10.
- En cas de validation d'une méthode alternative, le CEN a accepté de mener l'essai inter-laboratoires simultanément sur les deux méthodes, à savoir la méthode alternative et la méthode de référence, alors qu'il était initialement prévu de mener cet essai sur la seule méthode alternative.

Nous verrons qu'il subsiste néanmoins des différences notables entre les deux référentiels.

AOAC prévoit que toute méthode qui n'est pas une méthode de référence soit comparée, lors de l'essai inter-laboratoires, à une méthode de référence par culture. Le schéma de l'EN ISO 16140 pour les méthodes alternatives est donc retrouvé ici, avec une différence importante, qui réside dans le choix des méthodes de référence. Pour ces dernières, la norme préconise logiquement les méthodes normalisées par ISO et CEN, alors que AOAC choisit soit les Méthodes Officielles AOAC, soit les méthodes officielles du BAM (*Bacteriological Analytical Manual*) de la FDA (*Food and Drug Administration*), soit celles de l'USDA (*United States Department of Agriculture*) pour l'analyse des viandes, volailles et leurs produits.

Par rapport au protocole général décrit au § 1.1.7.2, celui adapté à la microbiologie comprend une étude préliminaire nettement plus conséquente, en matière de caractérisation des performances analytiques. Trois critères sont étudiés :

1. **La robustesse.** Soit l'identification des facteurs internes à la méthode ou liés à l'environnement du laboratoire, qui sont critiques pour la variabilité de la méthode. Andrews décrit un exemple d'étude de robustesse en microbiologie des aliments, pour une méthode immuno-enzymatique [Andrews, 1996].
2. **La justesse.** Le protocole expérimental est très complet, proche de celui de l'EN ISO 16140, si ce n'est qu'en pratique une contamination artificielle est favorisée.
3. **La spécificité.** Soit l'étude de l'inclusivité/exclusivité avec des souches pures.

L'étude préliminaire vise donc principalement à définir le domaine d'application de la

méthode, tant en termes de matrices que de souches de microorganismes. Andrews discute des critères retenus par AOAC pour le choix des matrices et des souches, en fonction de l'objectif poursuivi [Andrews, 1996].

Comme la norme EN ISO 16140, les lignes directrices distinguent le cas des méthodes qualitatives de celui des méthodes quantitatives. Avant de les aborder séparément, des aspects communs aux deux types de méthode peuvent être discernés.

1.1.10.2 Protocole expérimental : aspects communs

Les aspects communs aux méthodes qualitatives et aux méthodes quantitatives sont les suivants :

- Sélection de six types de matrices, si on veut valider la méthode pour tous les aliments. Les types de matrices sont choisis parmi ceux connus pour être favorables à la croissance du germe cible.
- Cultures d'inoculation :
 - Les types sont choisis de façon à couvrir au mieux le domaine d'application de la méthode. En fonction de la cible de la méthode, un isolat/une souche/une souche productrice de toxines/un sérotype/une espèce différent(e) sera choisi(e) pour chaque type d'aliment. Les cultures mixtes sont déconseillées.
 - L'état physiologique est choisi de façon à se rapprocher au mieux des conditions de contamination naturelle, soit :
 - des microorganismes stressés pour les aliments transformés ;
 - des microorganismes non stressés pour les aliments crus ;
 - des inocula lyophilisés pour des aliments en poudre/granuleux ;
 - des aliments solides stockés après leur contamination, afin de stabiliser la population microbienne.

En matière de protocole expérimental, la différence essentielle entre le protocole AOAC et l'EN ISO 16140 réside dans le nombre de matrices utilisées pour valider une méthode applicable à l'ensemble des denrées alimentaires. Six matrices sont requises par l'AOAC et seulement une par l'EN ISO 16140. La norme a fait le choix d'alléger l'essai inter-laboratoires, compte-tenu de l'importance donnée à l'étude de justesse lors de l'essai préliminaire, qui a déjà couvert la diversité des matrices auxquelles la méthode s'applique.

1.1.10.3 Méthodes qualitatives

1.1.10.3.1 Protocole expérimental

Les points complémentaires et spécifiques aux méthodes qualitatives sont les suivants :

- Au moins 10 laboratoires participants dont les données sont exploitables.
- 3 niveaux d'inoculation :
 - un témoin négatif ;
 - un niveau bas, vers la limite de détection de la méthode, soit 1 à 5 ufc/25 g ;
 - un niveau élevé, soit 10 à 50 ufc/25 g.
- Le niveau bas est fixé de façon à obtenir une récupération partielle. Il s'agit d'obtenir, lors des répétitions de l'analyse, des résultats positifs dans une certaine proportion, idéalement 50 %, et des résultats négatifs. De fait, l'EN ISO 16140 définit la limite de détection d'une

façon similaire. Les niveaux de contamination sont vérifiés par le laboratoire organisateur, avec un dénombrement par NPP, le jour du lancement des analyses.

- 6 répétitions par combinaison <type d'aliment, niveau d'inoculation>, sous forme de 6 prises d'essai analysées en aveugle.

Deux différences mineures entre ce protocole et l'EN ISO 16140 peuvent être notées :

1. L'importance donnée par l'AOAC au niveau permettant une récupération partielle. Sa non réalisation est suffisante pour invalider une étude.
2. Le nombre de répétitions, légèrement supérieur dans la norme : 8 au lieu de 6.

1.1.10.3.2 Cohérence des résultats et test de comparaison

A la différence de l'EN ISO 16140, AOAC préconise l'utilisation de tests statistiques pour écarter les **valeurs aberrantes**. Cependant, le document ne rentre pas dans les détails, mais se contente de se référer aux travaux de McClure [McClure, 1990].

Dans cette publication, les valeurs aberrantes sont testées sur la base de la proportion de résultats corrects (vrais positifs ou vrais négatifs) obtenus par les laboratoires, par rapport à la valeur connue. On teste si les laboratoires ont des proportions de résultats corrects qui diffèrent significativement entre elles. Fondée sur un test statistique Q défini par Cochran, la statistique Q suivante est proposée par McClure :

$$Q = \frac{L(L-1) \sum (T_i - \bar{T})^2}{L \sum S_j - \sum S_j^2} \quad \text{Eq. 4}$$

où :

- L est le nombre de laboratoires ;
- T_i est le nombre de résultats corrects du laboratoire i ;
- \bar{T} est le nombre moyen de résultats corrects ;
- S_j est le nombre de résultats corrects pour l'échantillon j .

Q doit suivre une loi de χ^2 à $L-1$ degrés de liberté. Si le Q expérimental dépasse la valeur du χ^2 , la proportion de résultats corrects diffère significativement entre les laboratoires. L'étape suivante, détaillée dans la publication, est de déterminer le laboratoire ou le groupe de laboratoires qui diffère(nt) de l'ensemble des participants.

Pour chaque combinaison <type d'aliment x niveau de contamination>, une **comparaison** directe de la méthode alternative avec la méthode de référence est menée ici d'une façon similaire à l'EN ISO 16140. On vérifie par un test de χ^2 l'hypothèse d'une différence nulle entre les proportions de positifs confirmés par chacune des deux méthodes.

Le χ^2 à un degré de liberté, tel que défini par McNemar, est égal à :

$$\chi^2 = \frac{(|a-b|-1)^2}{a+b} \quad \text{Eq. 5}$$

où :

- a est le nombre d'échantillons confirmés positifs par la méthode alternative, mais négatifs par la méthode de référence ;
- b est le nombre d'échantillons négatifs par la méthode alternative, mais confirmés positifs par la méthode de référence.

On peut noter que l'EN ISO 16140 définit différemment la statistique de χ^2 : le dividende est égal à $(a-b)^2$. En effet, les auteurs de cette norme ont retenu la formule classique du χ^2 pour des valeurs de $a+b$ au-delà de 30, condition effectivement précisée dans la norme pour appliquer ce test. Alors que AOAC a opté pour la formule adaptée aux petits échantillons ($a+b < 30$).

L'existence d'une différence significative selon ce test entre les deux méthodes est une condition suffisante pour exclure du domaine de validation de la méthode le type d'aliment concerné.

1.1.10.3.3 Critères de performance

Quatre critères sont définis et calculés par combinaison <type d'aliment x niveau de contamination> :

Taux de sensibilité et taux de spécificité

Ces critères sont définis de la même façon que dans l'EN ISO 16140 (§ 1.1.9.3.2), mais leur mode de calcul, figurant dans le Tableau 4, est différent.

Tableau 4 : Critères de performance des méthodes qualitatives, selon AOAC

Critère	Définition	Calcul
Taux de sensibilité (<i>SE</i>)	Voir § 1.1.9.3.2	$SE = \frac{N_{+}}{N_{\Sigma+}}$ <p>où :</p> <ul style="list-style-type: none"> - N_{+} est le nombre de positifs de la méthode à valider ; - $N_{\Sigma+}$ est le nombre total de positifs confirmés par l'une des deux méthodes (alternative ou référence).
Taux de spécificité (<i>SP</i>)	Voir § 1.1.9.3.2	$SP = \frac{N_{-}}{N_{\Sigma-}}$ <p>où :</p> <ul style="list-style-type: none"> - N_{-} est le nombre de négatifs de la méthode à valider ; - $N_{\Sigma-}$ est le nombre total de négatifs confirmés par l'une des deux méthodes (alternative ou référence).
Taux de faux négatifs (<i>FN</i>)	Fréquence selon laquelle un échantillon positif connu a été classé négatif par la méthode à valider.	$FN = \frac{N_{FN}}{N_{\Sigma+}}$ <p>où :</p> <ul style="list-style-type: none"> - N_{FN} est le nombre d'échantillons connus pour être positifs, mais donnés négatifs par la méthode à valider ; - $N_{\Sigma+}$ est le nombre total de positifs confirmés par l'une des deux méthodes (alternative ou référence). <p>Soit : $FN = 100 - SE$</p>
Taux de faux positifs (<i>FP</i>)	Fréquence selon laquelle un échantillon négatif connu a été classé positif par la méthode.	$FP = \frac{N_{FP}}{N_{\Sigma-}}$ <p>où :</p> <ul style="list-style-type: none"> - N_{FP} est le nombre d'échantillons connus pour être négatifs, mais donnés positifs par la méthode à valider ; - $N_{\Sigma-}$ est le nombre total de négatifs confirmés par l'une des deux méthodes (alternative ou référence). <p>Soit : $FP = 100 - SP$</p>

La différence en matière de calcul des taux de sensibilité et de spécificité réside dans la confirmation des résultats, qui n'est pas prévue dans la norme. Cette confirmation concerne les méthodes dites de « dépistage », qui ne comprennent pas d'étape de confirmation biochimique et/ou sérologique. Pour la sensibilité, le fait de ne prendre en compte au dividende que les résultats confirmés positifs de la méthode à valider rend le critère plus exigeant : le taux sera inférieur ou égal à celui obtenu selon la norme. Pour les deux critères, la base de référence (le diviseur) ne prend en compte que les résultats confirmés par l'une des deux méthodes. Le diviseur selon AOAC est donc plus restrictif que celui de la norme. De plus, la méthode de référence n'est pas considérée comme référence absolue. Au final, AOAC prend le parti de ne pas évaluer strictement la méthode alternative telle qu'elle est conçue.

Taux de faux négatifs et taux de faux positifs

Ces deux critères, définis dans le Tableau 4, ne figurent pas dans l'EN ISO 16140, mais ils dérivent directement des deux premiers.

1.1.10.4 Méthodes quantitatives

Pour l'étude préliminaire, le critère d'inclusivité/exclusivité n'est pas applicable aux méthodes dénombrant une flore non spécifique. Contrairement à l'EN ISO 16140, la linéarité ne fait pas partie de cette étude préliminaire ; ce qui indique que AOAC ne s'intéresse pas prioritairement aux méthodes alternatives instrumentales, pour lesquelles la linéarité est une caractéristique indispensable.

En ce qui concerne le protocole expérimental, les exigences principales suivantes viennent s'ajouter à celles qui sont communes (§ 1.1.10.2) :

- Au moins 8 laboratoires participants dont les résultats sont exploitables, soit le minimum prévu par le protocole général AOAC. Dans des cas spéciaux, comme les méthodes faisant appel à un appareillage très onéreux ou à des laboratoires spécialisés, ce minimum peut être réduit à 5 laboratoires. La FIL a aussi repris cette variante.
- Contamination artificielle, telle que décrite pour les méthodes quantitatives, à utiliser pour les méthodes dénombrant des germes spécifiques, alors que la contamination naturelle est à préférer pour les flores non spécifiques, comme la flore totale.
- 4 niveaux d'inoculation :
 1. un témoin négatif ;
 2. un niveau bas, à environ la limite de détection de la méthode ;
 3. un niveau moyen, un logarithme plus haut ;
 4. un niveau élevé, deux logarithmes plus haut.
- Echantillons en double par combinaison <type d'aliment, niveau d'inoculation>, sous forme de 2 prises d'essai analysées en aveugle.

Toutes ces prescriptions sont identiques à celles de l'EN ISO 16140.

La cohérence des données est assurée, comme dans l'EN ISO 16140, par une transformation logarithmique initiale des comptages, afin de se rapprocher d'une distribution normale. Une représentation graphique des résultats est ensuite préconisée, pour identifier visuellement les incohérences majeures. Les données non valides pour des causes identifiées, similaires à celles de la norme, sont exclues des calculs. Les valeurs aberrantes sont également écartées des calculs en utilisant des tests statistiques qui ne sont pas décrits dans ces lignes directrices mais qu'on peut trouver dans les ouvrages statistiques [Youden & Steiner, 1987 ; Dixon &

Massey, 1969] et qui sont également référencés dans le protocole général AOAC.

Les critères de performance sont les suivants :

- **Comparaison méthode alternative/méthode de référence.** Elle est effectuée par la comparaison des moyennes, à l'aide d'une analyse de variance ou d'un test *t* apparié, pour chaque type d'aliment et chaque niveau de contamination. Ces statistiques ne sont pas détaillées dans le document, mais la référence du manuel de statistiques de l'AOAC est indiquée [Youden & Steiner, 1987].
- **Estimateurs de fidélité.** Les écart-types et limites de répétabilité/reproductibilité sont définis classiquement ; l'utilisation des écart-types relatifs est recommandée. Les calculs ne sont pas donnés, mais il est fait référence au même manuel de statistiques de l'AOAC, également repris dans le protocole général (§ 1.1.7.3).

1.1.11 Procédures de validation des méthodes de référence en microbiologie des aliments

Compte-tenu des difficultés particulières d'organisation des essais inter-laboratoires en microbiologie des aliments, évoquées en introduction, nous nous intéressons maintenant au degré de mise en œuvre des référentiels qui ont été présentés, tout d'abord pour l'estimation de la fidélité des méthodes d'analyse de référence.

1.1.11.1 Méthodes officielles AOAC

Exigeant depuis plusieurs années l'estimation par essai inter-laboratoires de critères de performance pour pouvoir accéder au statut de *AOAC Official Method* [AOAC, 2003], AOAC est sans nul doute l'organisation qui dispose de l'expérience la plus ancienne et la plus complète dans le domaine de la microbiologie des aliments, comme nous l'avons déjà évoqué d'une façon plus générale pour l'analyse des aliments. Ainsi, pas moins de 113 méthodes microbiologiques bénéficient à ce jour le statut de méthode officielle pour l'AOAC !

Par ailleurs, AOAC est l'organisme qui, dans le domaine de l'analyse des aliments, a développé le processus le plus formalisé et rigoureux en matière d'organisation et d'évaluation des résultats des essais inter-laboratoires. Une crédibilité très élevée est alors conférée aux données de fidélité qui caractérisent les Méthodes Officielles AOAC. Ainsi, l'obtention de ces données implique les intervenants suivants [Feldsine *et al*, 2002] :

- Un Directeur d'Etude, qui conduit le processus de validation, propose la méthode de référence, effectue l'étude préliminaire, développe un protocole pour l'essai inter-laboratoires, l'organise et en prépare le rapport.
- Un « médiateur » (*General Referee*), soit un microbiologiste senior, disposant d'une solide expérience en matière d'organisation d'essais inter-laboratoires AOAC, qui recommande le Directeur d'Etude, suit de près la réalisation de l'essai et recommande (ou non) au Comité de Méthodes la validation de la méthode.
- Un Comité de Méthodes spécifique de la microbiologie, qui approuve les protocoles d'essais et qui recommande au Bureau des Méthodes Officielles le statut de validation initiale ou finale à conférer à la méthode, en fonction du résultat de l'étude. Il nomme les médiateurs.
- Un statisticien pour le Comité de Méthodes, qui apporte un soutien statistique au déroulement de l'étude et donne un avis sur le protocole et les résultats.
- Un conseiller du Comité de Méthodes pour l'hygiène et la sécurité.
- Le Bureau des Méthodes Officielles, qui avalise toutes les décisions relatives aux méthodes officielles.

La validation initiale (*First Action*) correspond à un statut temporaire d'appropriation. A l'issue d'une période probatoire de 3 ans, durant laquelle les utilisateurs de la méthode sont invités à faire part de leurs commentaires à AOAC, le statut de validation finale (*Final Action*) est conféré à la méthode, à condition que les commentaires reçus aient été pris en compte de façon satisfaisante.

Andrews a illustré dans le domaine de la microbiologie des aliments la répartition des responsabilités entre les différents intervenants [Andrews, 1996].

AOAC définit de façon très précise la conséquence que peut avoir la modification, à différents degrés, d'une méthode validée sur la validation de cette méthode. La conséquence peut être, dans le cas de modifications substantielles, la nécessité de mener un nouvel essai inter-laboratoires [Andrews, 1996].

Un exemple de méthode officielle AOAC en microbiologie des aliments nous semble intéressant. Il s'agit de la méthode par culture de recherche de *Listeria monocytogenes* dans le lait et les produits laitiers. En effet, cette méthode a été la première, et la seule pendant plusieurs années, à avoir été préparée en collaboration par AOAC et par un groupe de normalisation ISO/FIL, avant d'être approuvée en 1993 par AOAC, sous le numéro 993.12, et publiée conjointement par la FIL (norme FIL 143) et par l'ISO (norme ISO 10560) [Hitchins, 1996]. L'essai inter-laboratoires, organisé par le groupe FIL/ISO, a associé 18 laboratoires européens et américains.

1.1.11.2 Normes NMKL

Le *Nordisk Metodikkomite for Næringsmidler* (NMKL, Comité Nordique sur l'Analyse des Aliments) est une association qui normalise des méthodes de référence pour les pays du Nord de l'Europe, notamment en microbiologie des aliments. Son Secrétariat Général est actuellement assuré par l'Institut National Vétérinaire de Norvège (Oslo).

Suivant l'exemple d'AOAC, NMKL a généralisé l'organisation, sous son égide, d'essais inter-laboratoires sur la plupart des méthodes microbiologiques qu'elle retient pour ses normes. Ces essais sont organisés suivant le Rapport NMKL No 20, établi à l'intention des « référents » (*referees*), pour les guider dans le développement de la méthode normalisée, ainsi que dans l'organisation des essais inter-laboratoires [NMKL, 2003]. En voici les principales caractéristiques :

- A la différence de AOAC et de l'EN ISO 16140, qui ont porté le nombre minimal de participants à 10 pour les méthodes qualitatives, ce nombre est maintenu à 8, selon l'usage général.
- Il est recommandé que le référent mène une pré-étude avec un nombre restreint de laboratoires.
- Les caractéristiques de performance sont celles retenues par l'ISO 5725 et AOAC. Le document ne les décrit pas mais fait une simple référence aux lignes directrices d'AOAC pour les méthodes microbiologiques, à l'EN ISO 16140 et au protocole NordVal (voir ci-dessous).
- Les différentes étapes de mise en œuvre d'un essai inter-laboratoires sont décrites de façon pratique.
- Pour le protocole expérimental, l'EN ISO 16140 est prise en référence quasi-systématique.
- En matière de traitement statistique, l'EN ISO 16140 est également recommandée, en particulier pour les méthodes quantitatives (statistiques robustes).
- Le format du rapport de l'essai inter-laboratoires est indiqué, de même que la façon de spécifier dans le texte des méthodes normalisées par le NMKL les caractéristiques de performance, ainsi que les résultats de l'essai inter-laboratoires.

1.1.11.3 Méthodes validées pour le contrôle officiel au Royaume Uni

L'exemple au Royaume-Uni de la validation par essais inter-laboratoires de méthodes microbiologiques nous est apparu intéressant, car il s'inscrit dans un cadre particulier, celui du contrôle officiel des aliments.

La législation européenne relative au contrôle des aliments, et notamment la Directive 85/397 concernant *l'introduction de méthodes communautaires d'échantillonnage et d'analyse pour la surveillance des aliments destinés à l'homme*, préconise l'utilisation de méthodes validées, dont les performances de fidélité sont établies. Avec une volonté de répondre à ces exigences, le Ministère anglais de l'Agriculture, de l'Alimentation et de la Pêche (MAFF) a lancé en 1980 un vaste programme d'essais inter-laboratoires sur l'ensemble des méthodes d'analyse déjà prescrites dans la législation nationale, ou devant être incorporées dans la législation anglaise ou européenne. Ce programme a porté sur les méthodes microbiologiques à partir de 1988 [Scotter & Wood, 1996]. En 1996, 13 méthodes microbiologiques relatives à différents germes avaient été validées. Il s'agit de méthodes normalisées au niveau national ou par l'ISO, applicables soit à l'analyse d'une famille de produits (lait et produits laitiers, ou œuf et ovoproduits), soit à l'analyse de l'ensemble des aliments.

Pour la conduite de ces essais et l'exploitation de leurs résultats, les lignes directrices IUPAC/AOAC ont été suivies (§ 1.1.7) – le protocole AOAC spécifique à la microbiologie n'existait pas encore-. Les caractéristiques principales de conception de ces essais sont les suivantes :

- laboratoires participants : au sein du réseau des laboratoires « agréés » pour le contrôle officiel, soit les « *public analysts* » et les laboratoires de santé publique ;
- nombre minimal de laboratoires participants : 8 pour les méthodes quantitatives, conformément aux référentiels actuels, et 15 pour les méthodes qualitatives, selon les pratiques d'alors d'AOAC ;
- nombre minimal d'échantillons : 6, soit des doubles à 3 niveaux de contamination, ce qui correspond au schéma actuel retenu pour les méthodes qualitatives ;
- échantillons analysés en doubles aveugles ;
- transformation en \log_{10} des résultats avant calcul des écarts-types de fidélité ;
- application de tests statistiques pour l'identification de valeurs aberrantes.

Les rapports de ces essais ont été publiés dans les Bulletins du MAFF, ainsi que de manière condensée dans des revues scientifiques, telles que le *Journal of the Association of Public Analysts* pour les méthodes d'analyse du lait et des produits laitiers [Scotter *et al.*, 1993].

1.1.11.4 Normes CEN/ISO/FIL

Aujourd'hui, la normalisation dans le domaine de la microbiologie des aliments porte principalement sur des méthodes dites « horizontales ». Il s'agit de méthodes applicables non seulement à l'ensemble des denrées alimentaires, mais aussi aux aliments pour animaux et aux prélèvements d'environnement de production agro-alimentaire. Ces normes sont établies :

- au niveau international par le Sous-Comité 9 « Microbiologie » du Comité Technique 34 « Produits alimentaires » d'ISO (ISO/TC 34/SC 9), structure sous responsabilité française [Lombard *et al.*, 1996 ; Lombard, 2004];
- et au niveau européen par une structure correspondante, le Groupe de Travail 6 « Contaminants microbiens » du Comité Technique 275 « Analyse des denrées alimentaires – Méthodes horizontales » du CEN (CEN/TC 275/WG 6), également sous responsabilité française [Lahellec, 1998].

En fonction des sujets, l'une de ces structures prend la responsabilité de préparer le projet de norme qui est souvent adopté à l'identique en parallèle par l'autre structure. A ce jour, outre les normes portant sur des aspects généraux de l'analyse microbiologique des aliments, il existe 20 normes ISO qui définissent des méthodes de référence pour la recherche ou le

dénombrement de germes pathogènes ou indicateurs d'hygiène². Onze d'entre elles, relatives à des germes pathogènes, ont été reprises en normes CEN³.

Par ailleurs, certaines méthodes normalisées sont spécifiques de l'analyse microbiologique de la viande et des produits carnés d'une part (4), ou du lait et des produits laitiers d'autre part (16). Ces dernières sont préparées en collaboration par la FIL et le Sous-Comité 5 « Lait et produits laitiers » de l'ISO/TC 34, puis publiées conjointement par la FIL et par l'ISO [Hitchins, 1996].

AOAC a établi des liaisons formelles avec ces différentes structures de normalisation ; ce qui permet de renforcer les échanges, et d'ouvrir la voie d'une harmonisation entre les différents documents de référence. Cette harmonisation a été tangible pour les protocoles de validation des méthodes microbiologiques, comme nous l'avons vu plus haut (§ 1.1.10.1).

Une norme ISO se doit d'abord de représenter le consensus le plus large possible au niveau mondial, et d'être applicable dans le plus grand nombre de laboratoires situés dans les différents continents [Gomy & Lombard, 1992 ; Lombard, 2001]. C'est pourquoi les méthodes de référence normalisées sont jusqu'à aujourd'hui essentiellement fondées sur les techniques de microbiologie pasteurienne, faisant appel à l'aptitude du germe cible à se développer dans des milieux de culture adaptés. Une évolution dans cette conception vient d'être actée et vise à introduire, dans certains cas, des technologies plus récentes, tel que la PCR, dans les méthodes normalisées de référence.

En matière de validation scientifique de ces méthodes normalisées par CEN, ISO et FIL, les choix méthodologiques sont actuellement fondés sur au moins des travaux publiés dans la littérature scientifique, ainsi que sur des essais de la méthode retenue pour la norme effectués dans différents laboratoires utilisant chacun leurs propres échantillons. Cependant, une critique que l'on peut formuler sur le mode de fonctionnement actuel de ces structures de normalisation, en comparaison d'autres structures telles que AOAC ou NMKL, est l'absence de validation systématique par essais inter-laboratoires des méthodes normalisées. Cette critique touche surtout l'ISO et la FIL, puisque le CEN préconise, pour l'analyse des denrées alimentaires, la normalisation des seules méthodes validées par essais inter-laboratoires, à l'exception près de la microbiologie, du moins à l'heure actuelle.

1.1.11.5 Projet SMT

Un projet de recherche européen du programme SMT (*Standards, Measurements & Testings*, Normes, Mesures & Essais) a voulu combler ce manque, tout en répondant aux exigences du CEN. Il s'agit du projet SMT4-CT 96 2098 « Évaluation de méthodes microbiologiques pour la détection et le dénombrement de contaminants microbiologiques dans les aliments », projet financé par la Commission Européenne dans le cadre du 4^{ème} Programme Commun de Recherche et Développement (PCRD). Ce projet a débuté en 1996 pour une durée de 5 ans [Lahellec, 1998]. Dans le reste du texte nous nous y référerons comme au projet SMT.

² Selon www.iso.ch

³ Selon www.cennorm.be

Ce projet, coordonné par Cécile Lahellec, comportait trois partenaires principaux, le Centre National d'Etudes Vétérinaires et Alimentaires (CNEVA), intégré à l'Agence Française de Sécurité Sanitaire des Aliments (AFSSA) en 1999, le *Central Science Laboratory* (CSL, York, RU) et le *Rijksinstituut voor Volksgezondheid en Milieu* (RIVM, Institut National de Santé Publique et d'Environnement, Bilthoven, Pays-Bas). Il avait pour objectif de valider par essai inter-laboratoires les sept principales méthodes normalisées en microbiologie des aliments :

- dénombrement de *Bacillus cereus* selon l'EN ISO 7932 ;
- détection et dénombrement de *Listeria monocytogenes* selon l'EN ISO 11290-1 & 2 ;
- dénombrement des staphylocoques à coagulase-positive selon l'EN ISO 6888-1 & 2 ;
- dénombrement de *Clostridium perfringens* selon l'EN ISO 7937 ;
- détection de *Salmonella* selon l'EN ISO 6579.

Nous présenterons en 2^{ème} partie la mise en œuvre de ce projet et les résultats obtenus (§ 2.1).

1.1.12 Procédures de validation des méthodes alternatives en microbiologie des aliments

Les méthodes dites « alternatives » par rapport aux méthodes de référence tiennent une place particulière dans notre domaine.

En effet, nous venons de rappeler que les méthodes de référence font à ce jour presque uniquement appel aux techniques classiques de microbiologie, fondées sur la croissance des microorganismes. Cette croissance, même dans des milieux favorables, est un long processus, surtout lorsque le nombre de cellules dans la prise d'essai est très faible, de l'ordre de quelques cellules pour les méthodes de détection. Et ces mêmes techniques requièrent une main d'œuvre importante et qualifiée pour la réalisation des différentes étapes : ensemencement, préparation des dilutions, étalement, repiquage, reconnaissance visuelle des colonies sur les boîtes de Petri, identification,....

Or les opérateurs de la chaîne agro-alimentaire (producteurs, transformateurs, distributeurs) et les pouvoirs publics ont besoin d'obtenir rapidement pour les auto-contrôles et pour les contrôles officiels les résultats d'analyse de denrées alimentaires périssables, notamment dans le contexte du commerce international. Ce besoin a favorisé le développement de méthodes plus rapides, faisant appel à des technologies récentes, telles que l'immuno-enzymatologie et les techniques de biologie moléculaire. Cependant, la multiplicité des méthodes mises sur le marché et les allégations des fabricants, pour certaines inexactes, ont incité la mise en place, tant en France qu'à l'étranger, de processus de validation par tierce-partie de ces méthodes.

1.1.12.1 Procédure AFNOR pour les méthodes alternatives

L'AFNOR, avec l'ensemble des partenaires français, a été un précurseur dans la mise en place, dès 1989, d'un processus de validation de méthodes alternatives en microbiologie des aliments, qui a fait la preuve de son efficacité [Gomy & Lombard, 1992 ; Lombard *et al*, 1996]. C'est toujours aujourd'hui le seul système de certification par tierce partie, et 50 méthodes commercialisées sont actuellement validées.

Outre des exigences d'assurance qualité pour le fabricant de la méthode alternative, le principe de cette validation est d'établir les performances de la méthode alternative par comparaison à la méthode de référence correspondante, au cours de deux étapes : une étude préliminaire et un essai inter-laboratoires.

Les deux études sont jusqu'ici conduites selon un référentiel propre à AFNOR [AFNOR Certification, 2002], mais très proche de l'EN ISO 16140. Rappelons que cette norme a été rédigée à partir d'un document développé par le projet MicroVal, qui s'est largement inspiré de l'expérience française [Rentennar, 1996]. AFNOR, de son côté, est en passe d'adopter la norme récemment publiée. Malgré la proximité des deux référentiels, ce passage ne se fait pas sans difficulté, en raison d'une différence notable en matière d'essai inter-laboratoires. Jusqu'ici, cet essai était réalisé sur la méthode alternative seulement, alors que la norme requiert la réalisation de cet essai sur les deux méthodes en parallèle, dans un effort d'harmonisation avec AOAC (voir § 1.1.10.1). Ce changement se traduit ainsi par un doublement ou plus, compte-tenu de la lourdeur des méthodes de référence, de la charge de travail incombant aux laboratoires participant à l'essai. Cette participation étant bénévole, les

demandeurs de la validation, les fabricants et les laboratoires organisateurs craignent de ne plus pouvoir recruter suffisamment de participants pour mener ces études collaboratives.

1.1.12.2 Procédures AOAC

AOAC International propose principalement deux possibilités de validation des méthodes alternatives :

1. la validation complète, donnant le statut de *Official Method* à la méthode alternative [AOAC, 2003] ;
2. une validation allégée, appelée *performance tested* et délivrée par une filiale, AOAC *Research Institute* [AOAC, 2000].

La première possibilité consiste en une comparaison de la méthode alternative à la méthode de référence correspondante, menée en deux étapes, étude préliminaire et essai inter-laboratoires, selon les lignes directrices déjà présentées en § 1.1.10. L'application du statut de *Official Method* à des méthodes brevetées correspond à un changement de politique de AOAC, opéré en 1991 : auparavant, seules des méthodes dont les réactifs étaient complètement décrits pouvaient être validées par AOAC [Andrews, 1996].

La seconde possibilité se réduit à l'étude préliminaire de la méthode alternative confiée à un laboratoire qualifié et indépendant du fabricant. Aujourd'hui, 47 kits microbiologiques ont été approuvés par AOAC *Research Institute*.

Seule la validation complète peut donc être comparée à la Validation AFNOR. Les protocoles techniques sont proches, surtout depuis l'établissement de l'EN ISO 16140, comme nous l'avons déjà étudié (§ 1.1.10.1 et articles suivants). Cependant, AOAC International n'est pas un organisme certificateur et n'apporte pas de garantie sur le système qualité du fabricant.

1.1.12.3 Procédure NordVal

Au sein des pays scandinaves, les autorités nationales chargées de l'alimentation ont créé récemment, en 1999, l'association NordVal, dont le secrétariat est assuré par le l'Institut Danois pour la Recherche Alimentaire et Vétérinaire. NordVal vise à valider pour les besoins de cette région des méthodes alternatives en microbiologie des aliments. Le principe de cette validation est le même que celui de AFNOR ou de AOAC (*Official Methods*) : une étude préliminaire et un essai inter-laboratoires, menés en comparaison à une méthode de référence. A ce jour, 24 méthodes microbiologiques ont été validées par NordVal.

Le protocole [NordVal, 2004] reprend le plan et certains éléments de l'EN ISO 16140, tout en étant un document plus concis et plus appliqué que la norme.

En ce qui concerne les essais inter-laboratoires :

- Méthodes qualitatives
 - Le *protocole expérimental* suit dans l'ensemble celui de l'EN ISO 16140 , y fait référence, mais en diffère sur le point important qui suit. L'essai inter-laboratoires ne porte que sur la méthode alternative et seul le laboratoire organisateur effectue la comparaison méthode alternative/méthode de référence. De plus, le nombre minimal de participants est 8, au lieu de 10, et le nombre minimal de répétitions est 2, au lieu de 8. Au total, un protocole allégé et proche de celui suivi actuellement par la Validation AFNOR.

- Dans *l'interprétation des résultats*, une différence non négligeable avec la norme apparaît dans la comparaison des deux méthodes. Comme AOAC, NordVal préconise une *confirmation* des résultats, mais de façon quelque peu différente. NordVal ne s'en tient pas à une stricte comparaison de la méthode alternative avec la méthode de référence, mais veut confirmer le résultat de la méthode de référence en cas de désaccord entre les deux méthodes.
- Ainsi,
 - ✓ une déviation positive (méthode alternative +/méthode de référence -) devient un faux positif si le vrai résultat peut être prouvé comme étant négatif ;
 - ✓ une déviation négative (méthode alternative -/méthode de référence +) devient un faux négatif si le vrai résultat peut être prouvé comme étant positif.
- On peut remarquer que le protocole ne précise pas comment prouver le résultat de la méthode de référence. Or la réalisation de cette confirmation ne semble pas évidente ; ce qui représente une nette limitation de l'approche de NordVal. Le cas de figure le plus réaliste semble être de pouvoir transformer une déviation positive en un vrai positif, à savoir confirmer que la méthode de référence n'a pas pu détecter le germe cible, à la différence de la méthode alternative.
- Méthodes quantitatives
Le texte est très succinct en ce qui concerne ce type de méthodes. Le protocole expérimental est strictement identique à celui de l'EN ISO 16140, y compris pour la réalisation de l'essai inter-laboratoires sur les deux méthodes. Ce manque de cohérence avec l'option retenue pour les méthodes qualitatives apparaît surprenant. Aucun détail n'est donné pour le calcul des écarts-types/limites de répétabilité et de reproductibilité. On peut en déduire que les estimateurs robustes de l'EN ISO 16140 ou non robustes de l'AOAC peuvent être utilisés. D'ailleurs, le protocole fait preuve de souplesse en indiquant qu'il est possible de prendre en compte des critères de fidélité fournis par le fabricant et estimés selon les protocoles acceptés au niveau international, tels que ceux d'AFNOR, d'AOAC ou d'ISO.

1.1.12.4 Projet FOOD-PCR

FOOD-PCR, l'un des projets du 5^{ème} Programme Cadre de Recherche & Développement (Programme Qualité de la Vie), a été lancé en 2000 avec pour objectif de faciliter la mise en œuvre, pour la détection de bactéries pathogènes dans les aliments, de la PCR, l'une des techniques de biologie moléculaire les plus intéressantes comme alternative aux méthodes de microbiologie traditionnelle.

Le projet a consisté à développer, harmoniser et valider des méthodes de diagnostic PCR pour la détection de cinq bactéries pathogènes majeures en hygiène alimentaire, à savoir *Salmonella*, *Campylobacter*, *E.coli* entérohémorragiques, *L.monocytogenes* et *Yersinia enterocolitica*. Ce projet, d'une durée de 3 ans, a associé 12 partenaires principaux, dont notre laboratoire, et 23 partenaires dits « utilisateurs finaux », issus de 21 pays européens différents.

D'une part, ont été effectuées la validation des thermocycleurs et la mise au point de modes de pré-traitement pour quatre matrices : lait liquide, ainsi que liquide de rinçage de carcasse de porc, de carcasse de volaille et de carcasse de bovin. D'autre part, pour chaque germe, le projet a comporté différentes phases :

- Construction d'un banque d'ADN lyophilisé de référence.
- Développement et mise au point d'une méthode de détection du germe considéré dans les aliments, incluant une étude entre plusieurs laboratoires pour évaluer différentes amorces possibles de PCR.

- Deux séries d'essais inter-laboratoires :
 1. Un essai inter-laboratoires (10-12 participants) sur des souches de référence pour établir la spécificité du test PCR.
 2. Un essai inter-laboratoires (10-12 participants) de validation de la méthode complète, sur une ou deux des quatre matrices ci-dessus. Dans l'intention de valider une méthode d'analyse de aliments, on peut s'étonner du choix de ces matrices, qui pour trois d'entre elles ne concernent pas les aliments eux-mêmes.

Cette dernière série d'essais inter-laboratoires a porté sur des échantillons artificiellement contaminés par le laboratoire organisateur, à quatre niveaux (témoin négatif, niveaux bas, moyen et haut), et répliqués en aveugle trois fois à chaque niveau. A noter que le laboratoire organisateur a effectué lui-même le pré-enrichissement de ces échantillons, dont le résultat a été envoyé aux participants. Ainsi, au sens strict, la méthode complète n'a pas été validée par essai inter-laboratoires. La performance de chaque méthode a été exprimée selon les critères de l'EN ISO 16140 : spécificité, sensibilité, degré d'accord, concordance et *odds ratio* de concordance (§ 1.1.9.3.2).

Les résultats de ces différentes études sont disponibles sur le site Internet du projet (www.foodpcr.dk) et ont donné lieu à des publications, en projet ou publiées, dont celles portant sur les essais inter-laboratoires : *L.monocytogenes* [D'Agostino et al], *Yersinia enterocolitica* [Josefsen et al, 2003], *E.coli* O157 [Abdulmawjood et al, 2004], *Salmonella* [Malorny et al, 2004] et *Campylobacter* [Lubeck et al, 2003].

Outre la diffusion de la technique de PCR pour l'analyse des aliments, sur la base de méthodes harmonisées et validées de façon très complète, ce projet aura permis la mise en place, par un sous-groupe du comité CEN/TC 275/WG 6, d'un ensemble de normes CEN/ISO portant sur l'utilisation de cette technologie pour l'analyse microbiologique des aliments. Dans le cadre de la normalisation, le passage de statut de méthode alternative vers celui de méthode de référence peut être prévu pour les années à venir, tout au moins dans certaines applications.

1.2 Vérifier la compétence d'un réseau de laboratoires

Résumé

Les essais inter-laboratoires visant ce deuxième objectif disposent, comme les essais de validation de méthodes, d'un document de référence de portée générale : le guide ISO 43-1. Sa présentation permet de définir les notions de base et les modalités générales d'organisation et d'exploitation des essais inter-laboratoires d'aptitude.

Ce dernier aspect, qui vise à évaluer la compétence des laboratoires, est complété par un protocole IUPAC/AOAC, ainsi que, d'une façon plus approfondie, par le projet de norme ISO/DIS 13528. Ces deux derniers référentiels traitent également d'autres points qui n'ont été abordés que très succinctement par le guide ISO 43-1, comme l'étude de la stabilité et de l'homogénéité des échantillons.

La mise en pratique des essais d'aptitude dans le domaine de la microbiologie des aliments est envisagée au travers de deux exemples de réseaux français, le RAEMA et CECALAIT, se distinguant par l'état de la matrice utilisée pour préparer les échantillons (lait liquide/fromage frais, ou lait en poudre), la fréquence d'organisation des essais, ainsi que par les statistiques de performance utilisées.

Ce chapitre traite du deuxième type d'essai inter-laboratoires, que nous désignerons de façon simplifiée par « essai inter-laboratoires d'aptitude », ou encore « essai d'aptitude ».

Les essais d'aptitude ont été, jusqu'à un passé récent, moins fréquemment mis en œuvre en microbiologie des aliments que le type d'essai objet du chapitre précédent. Effectivement, dans ce domaine analytique comme dans d'autres, l'attention a été portée historiquement sur l'utilisation de méthodes validées. Cependant, la pratique des essais d'aptitude se développe rapidement dans notre domaine depuis plusieurs années.

Les référentiels disponibles seront présentés avant d'envisager des exemples de mise en pratique.

Sans doute le reflet d'une pratique limitée jusqu'à un passé récent, le nombre et la nature des référentiels disponibles sont très restreints, en net contraste avec le type d'essai inter-laboratoires traité dans le chapitre précédent. Nous n'avons en effet recensé que trois référentiels, dont un est en projet et dont aucun ne porte spécifiquement sur l'analyse microbiologique des aliments :

1. le Guide ISO 43-1, de portée générale ;
2. le protocole IUPAC/AOAC [Thompson & Wood, 1993], développé pour l'analyse chimique des aliments, mais dont nous verrons que certains aspects sont applicables à notre domaine ;
3. et le projet de norme ISO/DIS 13528, qui complète le premier pour l'exploitation des résultats.

Conscient de cette situation, le comité ISO/TC 34/SC 9 a décidé de préparer une norme traitant des exigences en matière d'organisation et d'exploitation des essais d'aptitude, déclinées à l'analyse microbiologique des aliments. Par ailleurs, l'accréditation des organisateurs d'essais d'aptitude se met en place actuellement en France comme dans d'autres pays. Dans ce contexte, il est prévisible que les associations d'organismes d'accréditation, comme EA (*European Co-operation for Accreditation*) au niveau européen ou ILAC (*International Laboratory Accreditation Cooperation*) au niveau international, décideront de préparer un référentiel d'accréditation, spécifique à la microbiologie des aliments.

1.2.1 Guide ISO 43-1

1.2.1.1 Présentation

Le Guide ISO/CEI 43 est paru en 1997 dans une deuxième édition, nettement enrichie par rapport à la première édition de 1987. Il a été préparé par un groupe ad'hoc du CASCO (Comité pour l'évaluation de la conformité) de l'ISO et formellement approuvé tant par l'ISO que par la CEI (Commission Electrotechnique Internationale). La première partie de ce guide sera détaillée ici : elle fournit des lignes directrices sur le développement et la mise en œuvre d'essais inter-laboratoires d'aptitude. La deuxième partie de ce guide ne sera pas présentée ; elle contient en effet des recommandations à l'intention des organismes d'accréditation pour la sélection et l'utilisation de systèmes d'essais d'aptitude dans le cadre de l'accréditation, ainsi que les bases d'accréditation des organisateurs d'essais d'aptitude.

La première partie du guide représente un référentiel d'ordre général sur les essais inter-laboratoires, au même titre que la norme ISO 5725. Cependant, outre le fait que chacun traite d'essais à objectifs différents, la lecture de la première partie du guide dénote une orientation différente de la norme. En effet, le texte est pour une large part de nature organisationnelle, de façon à permettre la mise en place de système d'assurance qualité alors même que la partie 2 est strictement consacrée à ces aspects. De fait, cette orientation du guide n'est pas surprenante, sachant qu'il a été rédigé par un comité traitant de l'évaluation de conformité, qui a en particulier établi la norme EN ISO 17 025, référentiel pour l'accréditation des laboratoires.

1.2.1.2 Objectifs des essais d'aptitude

L'introduction du guide est à notre connaissance le seul référentiel portant sur les essais inter-laboratoires qui distingue de façon aussi claire les *différents types d'essais inter-laboratoires*, en fonction des objectifs poursuivis. Les essais d'aptitude sont ainsi replacés dans un contexte plus large.

En ce qui concerne les essais d'aptitude eux-mêmes, leurs *finalités* sont précisées, la principale étant de déterminer la performance de laboratoires et de surveiller l'évolution de cette performance dans le temps.

Parmi les utilisations possibles des essais d'aptitude, est mentionné en introduction leur rôle dans une démarche d'*assurance qualité*. En complément des procédures internes de maîtrise de la qualité du laboratoire, l'essai d'aptitude constitue une mesure externe supplémentaire de sa compétence et vient compléter également les audits mandatés par les organismes

d'accréditation. A ce stade, le guide insiste sur le fait que les **résultats de la participation d'un laboratoire à un essai d'aptitude ne constituent qu'un des éléments permettant d'apprécier sa compétence**. Cet avertissement est essentiel et n'est pas pris en compte de façon systématique. En France, le guide est suivi correctement, puisque tant le COFRAC (Comité Français d'Accréditation) que les autorités publiques en charge du contrôle officiel des aliments accordent leur juste place aux essais d'aptitude. Les laboratoires ont pour obligation de participer régulièrement aux essais d'aptitude, lorsqu'ils existent pour les domaines analytiques considérés. L'obtention de résultats non satisfaisants à un essai d'aptitude ponctuel n'a pas de conséquence pour le laboratoire sur son accréditation ou son agrément pour les contrôles officiels. Celui-ci doit en revanche identifier la cause du résultat non satisfaisant et définir une action corrective pour éviter qu'une telle anomalie ne se reproduise ; ce qui représente une mesure habituelle d'assurance qualité.

1.2.1.3 Définitions et conséquences

Comme dans tout document normatif de portée générale, une section importante est consacrée aux définitions. Nous ne retiendrons ici que les deux termes qui composent l'« essai d'aptitude par intercomparaison », dénomination la plus complète retenue par le guide pour le type d'essai inter-laboratoires objet du présent chapitre :

1. Essai d'aptitude (d'un laboratoire) :

« Évaluation des performances d'un laboratoire en matière d'essais, au moyen de comparaisons inter-laboratoires

NOTE – Pour les besoins de la présente partie du Guide ISO/CEI 43, le terme « essai d'aptitude d'un laboratoire » est pris au sens le plus large et inclut par exemple les aspects suivants :

- a) systèmes qualitatifs – par exemple lorsqu'il est exigé des laboratoires qu'ils identifient un composant d'un objet soumis à l'essai.
- b) exercices de transformation des données – par exemple lorsque les laboratoires sont dotés d'ensembles de données qu'ils sont priés de traiter pour obtenir de nouvelles informations.
- c) essai d'un seul objet soumis à l'essai – lorsqu'un objet soumis à l'essai est expédié successivement à plusieurs laboratoires, puis est retourné à l'organisateur à intervalles déterminés.
- d) exercices ponctuels – lorsque les laboratoires reçoivent une seule fois un objet soumis à l'essai.
- e) systèmes continus – lorsque les laboratoires reçoivent des objets soumis à l'essai à intervalles fixes, dans le cadre d'une activité continue.
- f) échantillonnage – par exemple lorsqu'il est demandé aux individus ou aux organisations de prélever des échantillons pour analyse ultérieure. Ceci n'inclut pas la recherche prénormative qui peut s'appuyer sur des intercomparaisons. »

2. Intercomparaisons

« Organisation, exécution et évaluation d'essais sur des objets soumis à l'essai identiques ou semblables par au moins deux laboratoires différents dans des conditions prédéterminées

NOTE – Dans certains cas, l'un des laboratoires participant à l'intercomparaison est le laboratoire qui a fourni la valeur assignée à l'objet soumis à l'essai. »

La première définition est certes longue, mais elle permet de couvrir une grande variété de types d'essais d'aptitude, donc d'en élargir la conception. Les déterminations qualitatives qui occupent une place essentielle en microbiologie des aliments correspondent au type a). En revanche, les types b) et c) ne semblent pas être applicables dans ce domaine, comme nous l'avons déjà signalé en Introduction de cette partie, en raison du manque de stabilité des matériaux au regard de l'analyte considéré. Les essais ponctuels du type d) sont pratiqués dans le cadre des contrôles officiels, comme les réseaux de laboratoires nationaux de référence au niveau européen et les réseaux de laboratoires agréés au niveau national. Les systèmes continus de type e) sont typiques des réseaux animés par des organisateurs « professionnels » d'essais d'aptitude dont c'est l'activité principale, sinon unique : dans ce cas, les participants souscrivent un abonnement au réseau. Enfin, l'objectif du type f) est d'évaluer l'aptitude du participant à effectuer un échantillonnage, opération effectivement essentielle dans la signification des contrôles. Il n'a pas été mis en œuvre dans notre domaine et l'envisager semble difficile, en raison du manque évident d'homogénéité entre échantillons qui en résulterait. Même si en théorie, il est possible de prendre en compte cette hétérogénéité dans l'interprétation des résultats, tel que l'ISO 5725-5 le prévoit (§ 1.1.5.2).

A ce stade, il apparaît nécessaire de préciser la terminologie relative au type d'essais d'aptitude le plus fréquemment rencontré dans notre domaine, soit les systèmes continus de type e). Pour maintenir une cohérence dans l'ensemble de notre texte, nous désignerons par « essai d'aptitude » un essai inter-laboratoires inclus dans un système continu d'essais d'aptitude, au sens du guide ISO. Alors que l'on trouve une terminologie différente, notamment dans le projet de norme ISO/DIS 13528 : l'essai d'aptitude désigne le système continu, qui est alors composé de « cycles » (« *rounds* » en anglais), ou de « campagnes » pour certains organisateurs français d'essais d'aptitude.

La deuxième définition correspond à la notion générale que nous avons convenu de désigner par « essai inter-laboratoires ». Nous avons d'ailleurs repris les éléments de cette définition, en les complétant et les précisant, dans la définition que nous avons retenue pour « essais inter-laboratoires » en Introduction de cette partie. Le terme « intercomparaisons » introduit la notion d'essai inter-laboratoires dans la dénomination complète avancée par le guide, à savoir « essai d'aptitude des laboratoires par intercomparaisons » ; nous pouvons cependant nous interroger sur sa réelle utilité, puisque la définition d'essai d'aptitude introduit déjà la référence à des comparaisons inter-laboratoires.

1.2.1.4 Les différents types d'essais d'aptitude

Le guide reprend en partie la typologie envisagée dans la définition d'un essai d'aptitude, en détaillant chaque cas. Nous retiendrons les cas suivants :

- Systèmes d'essais inter-laboratoires. Nous avons déjà vu que ce sont ceux qui font l'objet de nos travaux. Ils comprennent l'envoi simultané de sous-échantillons sélectionnés au hasard dans une même source de matériau. L'essai à niveaux fractionnés, traité dans

l'ISO 5725-5 apparaît particulièrement bien adapté au cas d'un essai d'aptitude. En effet, le laboratoire, dans un essai à niveau fractionné, ne s'attend pas à trouver deux échantillons identiques (cas de doubles aveugles), alors que dans un essai d'aptitude, le laboratoire est particulièrement attentif à trouver la valeur attendue, par rapport à un essai de validation de méthode.

- Systèmes de comparaison de mesures. Ils correspondent au type c) de la définition (§ 1.2.1.3). Nous avons vu qu'ils n'étaient pas applicables dans notre cas.
- Systèmes d'essais à échantillons partagés. Il s'agit d'une comparaison de données produites par un petit groupe de laboratoires qui se partagent un même matériau. Cette comparaison est souvent réduite à deux laboratoires seulement : le laboratoire du fabricant et le laboratoire du client, dans le cadre de transactions commerciales. Un échantillon du même matériau est habituellement conservé si une différence significative apparaît entre les deux laboratoires, nécessitant l'arbitrage d'un troisième laboratoire. Ce type d'essai n'est pas utilisé couramment dans notre domaine, à la différence de domaines proches tels que la biologie médicale, ou l'environnement.

1.2.1.5 Organisation et conception

Le guide fixe le cadre d'organisation des essais d'aptitude, dont la gestion nécessite un coordinateur qui pourra s'entourer des compétences d'un groupe consultatif, composé d'experts techniques et statisticiens. Une liste des différentes étapes dans le déroulement d'un essai d'aptitude est indiquée.

1.2.1.6 Plan statistique

Le guide fournit en annexe A des exemples de méthodes statistiques pour le traitement des données d'essais d'aptitude. Il est indiqué que ces méthodes seront décrites en détail par un document préparé par le comité ISO/TC 69SC 6 : il s'agit du projet de norme ISO/DIS 13 528, qui sera présenté ultérieurement.

L'annexe du guide distingue trois, sinon quatre étapes communes à toutes les évaluations de résultats d'essais d'aptitude :

1. la détermination de la valeur de référence acceptée ;
2. le calcul des statistiques de performance ;
3. l'évaluation de la performance ;
4. dans certains cas, la détermination préalable de l'homogénéité et de la stabilité de l'objet soumis à l'essai.

Pour des nouveaux systèmes d'intercomparaison, dont l'accord initial est souvent faible, le guide recommande l'usage de statistiques robustes pour la performance.

1.2.1.7 Détermination de la valeur de référence acceptée

Définition

La valeur de référence acceptée est définie par l'ISO 5725-1 dans l'objectif de caractériser la justesse d'une méthode d'analyse (voir § 1.1.2.1). Nous avons discuté dans ce même paragraphe l'application de cette définition au cas de la microbiologie des aliments.

Se fondant sur cette définition, le guide envisage cinq façons d'établir la valeur de référence acceptée, également appelée valeur assignée. Ces cinq possibilités sont énumérées dans un ordre qui produit généralement une incertitude croissante sur l'estimation de la valeur de référence :

a) *valeurs connues* : établies par une formulation spécifique de l'échantillon, fabrication ou dilution par exemple.

Selon notre interprétation de la définition de l'ISO 5725-1 pour la microbiologie, ce cas correspondrait au type b) de l'ISO 5725-1 (§ 1.1.2.1), soit une contamination artificielle maîtrisée au mieux par un laboratoire de référence.

b) *valeurs de référence certifiées* : déterminées par des méthodes définitives pour des essais quantitatifs.

Nous entendons par méthode « définitive » une méthode permettant de déterminer le nombre « vrai » de microorganismes dans un échantillon selon des principes scientifiques ; soit le type a) de l'ISO 5725-1. Nous avons vu que de telles méthodes ne sont pas disponibles à ce jour. Cette définition de la valeur de référence certifiée peut apparaître surprenante. En effet, il semblerait logique que la valeur de référence certifiée provienne d'un matériau de référence certifié ; ce cas aurait alors correspondu au type c) de l'ISO 5725-1. Peut-être était-ce là l'intention des auteurs du guide, mais la formulation manque alors de clarté.

c) *valeurs de référence*, déterminées par :

- l'analyse d'un matériau de référence ou d'un étalon susceptible d'être raccordé à un étalon national ou international ;
- ou par la comparaison du matériau pour essai à ce matériau de référence ou à cet étalon.

Ce cas correspond au type c) de l'ISO 5725-1. Comme nous l'avons déjà interprété pour la microbiologie, le matériau de référence peut être soit certifié, soit utilisé lors d'un essai inter-laboratoires, notamment un essai d'aptitude. Nous ne disposons pas à ce jour d'étalon national ou international en microbiologie. En effet, les souches des collections de référence ne peuvent pas jouer ce rôle, car elles ne représentent pas un étalon pour la méthode d'analyse dans sa totalité, qui comprend l'étape initiale de détection du germe dans l'aliment.

d) *valeurs consensuelles provenant de laboratoires spécialisés*

Les laboratoires, notamment des laboratoires de référence, doivent faire preuve de leur compétence pour effectuer l'analyse avec une méthode validée, fidèle et juste. Ce sont les conditions requises pour certifier un matériau de référence. A ce titre, ce cas correspond à une partie du type c) de l'ISO 5725-1.

Ce type d'essai correspond aux essais mis en œuvre par le Laboratoire Communautaire de Référence Lait pour les Laboratoires Nationaux de Référence Lait ; essais que nous détaillerons en 2^{ème} partie.

Par contre, les essais inter-laboratoires de validation de méthode, ainsi que la majorité des essais d'aptitude dans notre domaine, ne répondent généralement pas aux conditions fixées pour ce cas.

e) *valeurs consensuelles des laboratoires participants*

Il s'agit soit des essais inter-laboratoires visant à valider une méthode, soit de la plupart des essais d'aptitude dans notre domaine, à savoir l'autre partie du type c) de ISO 5725-1.

En définitive, la correspondance entre les deux référentiels, le guide ISO 43-1 et l'ISO 5725-1, n'est que partielle et le guide adapte la norme au cas des essais d'aptitude. Par exemple, le type d) de la norme n'est pas repris ; ce qui se comprend aisément, puisque d'après notre interprétation, il peut correspondre à des répétitions effectuées dans un seul laboratoire.

Calculs

Lorsque des techniques consensuelles sont suivies, la valeur de référence acceptée est calculée de la façon suivante :

- Pour une *détermination qualitative* : consensus d'un pourcentage majoritaire prédéterminé, exprimé sur une échelle nominale ou ordinale.
- Pour une *détermination quantitative* : moyenne pondérée ou transformée, telle que moyenne géométrique, médiane, mode ou autre estimateur de situation centrale robuste. Il est surprenant que le guide ne fasse pas mention de la moyenne arithmétique classique.

Le « Guide pour l'expression de l'incertitude de mesure » [ISO GUM, 1993] est cité pour estimer l'incertitude de mesure des valeurs de référence acceptées.

La façon de traiter les *valeurs extrêmes* est indiquée :

- Si les résultats des participants sont utilisés pour déterminer les valeurs de référence acceptées, l'influence des valeurs extrêmes doit être minimisée, soit par l'utilisation de statistiques robustes, soit par l'élimination des valeurs aberrantes. L'ISO 5725-2 est alors citée en référence.
- Pour la présentation des résultats d'un essai d'aptitude et l'évaluation des performances des laboratoires, aucun résultat ne doit être éliminé.

1.2.1.8 Statistiques de performance

1.2.1.8.1 Pour un essai unique

L'objectif est de mesurer l'écart à la valeur de référence acceptée, de manière à déterminer l'acceptabilité de cet écart. A notre avis, le guide traite trop rapidement le cas des déterminations qualitatives, en indiquant qu'elles ne nécessitent en général aucun calcul. Pour les déterminations quantitatives, le guide énumère les statistiques suivantes, dans l'ordre croissant de degré de transformation des résultats des participants :

1. Différence, ou biais de laboratoire ($x - X$), où x représente le résultat d'un participant et X la valeur de référence acceptée.

Cette différence correspond à l'estimation du biais du laboratoire, telle qu'introduite par l'ISO 5725-4 (§ 1.1.4.2). Elle est facilement comprise par les participants.

2. Différence en pourcentage :

$$\frac{(x - X)}{X} \times 100 \quad \text{Eq. 6}$$

L'effet du niveau de concentration est compensé. Cette notion est également aisément comprise par les participants.

3. Percentile ou rang. Comme on l'a déjà noté, ces statistiques sont utiles pour des résultats soit hautement dispersés, comme lors du lancement d'un nouveau cycle d'essais d'aptitude, soit biaisés, soit lorsque le nombre de réponses différentes est limité.
4. Scores z , tels que :

$$z = \frac{x - X}{s} \quad \text{Eq. 7}$$

où : s est une estimation appropriée de la variabilité, sélectionnée pour répondre aux exigences du système et devant être fiable, c'est-à-dire fondée sur un nombre suffisant d'observations.

Ce modèle est utilisable que X et s soient dérivés, ou non, des résultats des participants. Pour cette statistique de performance, le protocole IUPAC/AOAC est cité en référence. Les scores z sont à ce jour largement utilisés à l'étranger dans le domaine de l'analyse des aliments et notamment en microbiologie. On peut remarquer le manque de précision du guide quant à la façon d'estimer s .

5. Scores d'erreur normalisée E_n , tels que :

$$E_n = \frac{x - X}{\sqrt{U_{lab}^2 + U_{ref}^2}} \quad \text{Eq. 8}$$

où U_{lab} est l'incertitude du résultat d'un participant et U_{ref} est l'incertitude de la valeur de référence acceptée du laboratoire de référence.

De fait, le paramètre s du score z est ici estimé par une combinaison d'incertitudes, qui est en partie liée au participant dont le résultat est évalué. Les nombres E_n ne sont pas utilisés, à notre connaissance, en microbiologie des aliments. Avec la mise en œuvre de l'incertitude de mesure dans ce domaine, cette situation devrait changer et ainsi conférer une utilisation supplémentaire à l'incertitude de mesure.

En France, certains organisateurs d'essais d'aptitude, à l'inverse de leurs homologues anglo-saxons, conservent une attitude très réservée, sinon de rejet, vis-à-vis des différents scores de performance, en particulier les scores z . En effet, en accord avec la façon de prendre en compte les résultats d'essais d'aptitude (§ 1.2.1.2), ces scores risquent d'être pris comme une note unique et univoque des laboratoires, pouvant se transformer rapidement en moyen de sanction. Les essais d'aptitude perdraient alors leur dimension d'outil d'amélioration et d'aide aux laboratoires.

1.2.1.8.2 Scores combinés de performance

Une évaluation plus complète peut être effectuée à l'aide de plusieurs résultats obtenus au cours d'un même essai d'aptitude, avec plusieurs échantillons pour le même analyte, ou pour une famille d'analytes apparentés. Le diagramme h de Mandel, selon l'ISO 5725-2 (§ 1.1.3.3) est cité comme interprétation efficace de la performance dans ce cas.

Des exemples de combinaison de scores sont indiqués :

- nombre de résultats satisfaisants,
- score z moyen,
- différence moyenne absolue, en unité ou en pourcentage,
- différence cumulée absolue, ou carré de la différence.

Le guide précise que les scores peuvent être éventuellement transformés, de façon à ce qu'ils observent tous la même loi, telle que la loi Normale pour les scores z ou le χ^2 pour les différences au carré.

Une telle conception des essais d'aptitude est courante, sinon systématique, dans notre domaine : la combinaison consiste à associer plusieurs niveaux de contamination d'un même germe à une matrice donnée.

1.2.1.8.3 Évaluation initiale des performances

Il s'agit de déterminer l'acceptabilité des résultats de chaque participant, afin d'en évaluer la performance. Trois possibilités sont prévues :

1. Consensus des experts, par exemple le groupe consultatif : méthode type pour les résultats qualitatifs.
2. Détermination statistique relative aux scores :
 - pour les scores z : satisfaisant si $|z| \leq 2$, discutable si $2 < |z| < 3$, et non satisfaisant si $|z| \geq 3$;
 - pour les nombres E_n : satisfaisant si $|E_n| \leq 1$, et non satisfaisant si $|E_n| > 1$.
3. Consensus des participants : étendue des scores ou des résultats obtenus par un certain pourcentage de participants, ou par un groupe de référence, tel que :
 - un pourcentage central (80 %, 90 %, ou 95 %) satisfaisant ;
 - ou un pourcentage unilatéral (90 % minimal) satisfaisant.

À noter que cette dernière possibilité n'est pas utilisée, à notre connaissance, en microbiologie des aliments, alors que celle qui concerne les scores z est bien suivie.

1.2.1.8.4 Surveillance dans le temps des performances

Le suivi dans le temps des performances est rendu possible par les réseaux fonctionnant par abonnement et animés par des organisateurs « professionnels ». Les techniques graphiques, notamment les cartes de contrôle de Shewhart, selon l'ISO 8259, sont recommandées par le guide.

1.2.1.9 Détermination préalable de l'homogénéité des échantillons

Le paragraphe du guide est très court à ce sujet : il se contente de faire référence au protocole IUPAC/AOAC. Alors que le guide considère cette étape comme optionnelle, elle est sans conteste indispensable dans notre domaine, l'analyte étant un organisme vivant, comme nous l'avons souligné à plusieurs reprises.

1.2.1.10 Préparation et gestion des échantillons

Le guide énonce des prescriptions générales, relatives à la préparation des échantillons. Quant à leur gestion, sont indiquées des notions d'assurance qualité, qui recourent celles de l'ISO 5725-2. Outre l'étude de l'homogénéité évoquée ci-dessus, le coordinateur doit également s'assurer de la stabilité des échantillons et si nécessaire, préciser une date de lancement des essais par l'ensemble des participants.

1.2.1.11 Choix de la méthode d'analyse

Un des principes de base des essais d'aptitude est énoncé ici, les différenciant clairement des essais inter-laboratoires de validation de méthodes : les participants peuvent normalement utiliser la méthode de leur choix. Ce principe nous semble être tout à fait justifié, puisqu'il s'agit d'évaluer la performance des laboratoires. En effet, il est généralement admis qu'un laboratoire appliquera de façon plus satisfaisante une méthode qu'il a développée lui-même, ou qu'il a choisie parmi un ensemble de méthodes disponibles, plutôt qu'une méthode qui lui est imposée. Outre des facteurs objectifs, tels que l'adaptation de la méthode à la situation du laboratoire, son équipement, son personnel, son volume analytique, ..., l'aspect psychologique ne doit pas être sous-estimé. Tout « bon » microbiologiste considère que « sa » méthode est la meilleure !

Cependant, le guide prévoit que le coordinateur puisse, dans certaines circonstances, imposer la méthode à utiliser. Il doit alors s'agir de méthodes normalisées et/ou validées par essais inter-laboratoires. Nous verrons qu'il peut s'agir également du cas des essais d'aptitude organisés par des Laboratoires Communautaires de Référence.

1.2.1.12 Mise en œuvre et rapport d'essai

Dans un optique d'assurance qualité, le guide recommande de *documenter* toutes les pratiques et procédures dans un manuel qualité dont il indique les éléments en annexe B.

Le coordinateur doit fournir des *instructions* aux participants sur tous les aspects du système d'essai d'aptitude. En particulier, les participants devraient traiter les échantillons comme des échantillons de routine. Cette exigence se comprend aisément, l'objectif étant d'évaluer l'aptitude des laboratoires à rendre des résultats corrects dans le cadre d'une activité normale. Cependant, il est tentant de traiter avec un soin particulier les échantillons d'un essai d'aptitude ; ainsi l'évaluation de la compétence des laboratoires sera biaisée.

Emballage et mode de transport doivent être adaptés par le coordinateur au type d'échantillon transporté, afin d'en assurer la stabilité.

Outre les éléments de l'*évaluation des résultats* mentionnés plus haut, le guide recommande au coordinateur de définir des critères pour considérer un essai d'aptitude comme valide. En effet, il ne faut pas rendre un classement ou une notation si le matériau d'essai n'a pas été assez stable et/ou homogène.

Enfin, un plan-type des *rapports* d'essai d'aptitude est défini.

1.2.1.13 Évaluation de la performance

Il est indiqué que le coordinateur peut s'entourer de conseillers techniques pour traiter de certains aspects de la performance, tels que :

- la performance d'ensemble par rapport aux attentes, notamment en matière d'incertitude de mesure ;
- les variations au sein d'un laboratoire et entre laboratoires, appréciées par le biais d'une comparaison avec des essais antérieurs ou avec des données de fidélité publiées ;
- les variations entre méthodes d'analyse ;
- les sources possibles d'erreurs et suggestions d'amélioration ;
- les conclusions.

Le guide déconseille de classer les laboratoires selon leurs performances (§ 1.2.1.8).

1.2.1.14 Communication avec les participants et confidentialité

Une *communication* entre le coordinateur et les participants, tant écrite qu'orale, par le biais de réunions régulières, est recommandée, notamment en cas de mauvais résultats d'un participant, pour l'aider à identifier l'origine de son écart et à y remédier. Il est fait référence à la partie 2 du guide pour les actions correctives.

Notamment dans les rapports d'essais, la *confidentialité* de l'identité des participants doit être assurée, par le biais d'une codification des participants, cette identité n'étant connue que du coordinateur, sauf exception en accord avec les participants. Le guide recommande également au coordinateur de mettre en œuvre toute mesure permettant d'éviter une *collusion* entre participant, ou une *falsification* des résultats : par exemple un participant peut répéter une analyse et rendre une moyenne de répétitions, au lieu du résultat d'une analyse unique.

1.2.2 Protocole harmonisé IUPAC/AOAC

1.2.2.1 Contexte

Thomson et Wood ont publié, sous l'égide du groupe de travail « Harmonisation des schémas d'assurance qualité pour les laboratoires d'analyse » de IUPAC, un protocole international harmonisé pour les essais d'aptitude des laboratoires d'analyse (chimique) [Thompson & Wood, 1993]. L'AOAC a contribué aux travaux du groupe et approuvé le protocole qui en a résulté.

En introduction, le protocole situe les essais d'aptitude comme un des moyens, avec l'utilisation de méthodes validées, permettant au laboratoire de produire de façon constante des résultats fiables. L'usage que peut faire l'accréditation de ces essais est également mentionné.

En comparant ce protocole avec la version actuelle du guide ISO, nous pouvons constater que, d'une part, de nombreuses prescriptions sont communes aux deux référentiels, et d'autre part, le guide ISO se réfère explicitement au protocole IUPAC/AOAC pour deux applications des méthodes statistiques, le score z et la détermination de l'homogénéité. Il est d'ailleurs probable que la version actuelle du Guide ISO ait été fondée, du moins en partie, sur le protocole IUPAC/AOAC

Si le titre du protocole indique qu'il s'applique aux laboratoires d'analyse chimique, nous verrons que ses recommandations peuvent s'étendre telles quelles ou avec une légère adaptation à l'analyse microbiologique. D'ailleurs, les auteurs du protocole étaient conscients de cette généralisation éventuelle puisqu'ils ont mis le terme « chimique » entre parenthèses dans le titre.

Après une section habituelle de rappel des définitions, le protocole traite successivement de l'organisation, du traitement statistique des résultats, ainsi que d'un exemple de sélection des valeurs assignées et des valeurs cibles. Nous ne mentionnerons que les aspects complémentaires ou différents du guide ISO 43-1.

1.2.2.2 Organisation des essais

1.2.2.2.1 Échantillons

Le protocole demande que les matériaux utilisés soient proches de ceux analysés en routine, en matière de composition de la matrice et de niveau de concentration. Une des originalités réside, nous l'avons déjà mentionné, dans le fait de fournir en Annexe II un protocole pour établir l'*homogénéité* du matériau utilisé dans la préparation des échantillons distribués. Cette procédure s'applique lorsque l'analyte est présent dans un matériau en vrac, qui est homogénéisé avant d'en prélever les échantillons envoyés aux participants. La procédure d'étude de l'homogénéité comprend les étapes suivantes :

1. Sélection de n échantillons (au moins 10), qui sont homogénéisés et prélèvement de 2 prises d'essai par échantillon.
2. Analyse des $2n$ prises d'essai sous des conditions de répétabilité, avec une méthode d'analyse suffisamment précise.

3. Estimation de la variance d'échantillonnage (s_s^2) et de la variance analytique (s_a^2) par une analyse de variance, sans exclusion de valeurs aberrantes.
4. L'homogénéité est considérée comme satisfaisante si $\frac{s_s}{\sigma} < 0.3$, où σ est la valeur cible pour l'écart-type de l'essai d'aptitude.

Cette procédure est aujourd'hui généralement utilisée par les organisateurs d'essais d'aptitude dans le domaine de l'analyse des aliments.

En matière de *stabilité* du matériau, le protocole précise les modalités de l'étude requise par le Guide ISO 43-1. Le coordinateur doit s'assurer, avant le lancement de l'essai d'aptitude, de cette stabilité pendant l'intégralité du déroulement de l'essai. Ainsi, le matériau contenant l'analyte doit être conservé pendant la durée prévue entre l'envoi des échantillons et le lancement des analyses dans des conditions (température) représentatives des conditions attendues pendant le transport et dans les laboratoires avant le lancement des analyses.

La concentration en analyte ne doit pas varier significativement pendant cette durée, le caractère significatif étant défini par rapport à la variance d'échantillonnage de l'étude d'homogénéité. Le protocole indique que pour des raisons pratiques, 6 échantillons au plus par analyte sont distribués au cours d'un essai.

1.2.2.2 Fréquence de répétitions des essais

En fonction de certains facteurs énumérés, le document recommande une gamme de fréquence : entre un essai toutes les 2 semaines et tous les 4 mois. Au-delà de 2 semaines, la charge risque d'être trop lourde pour l'organisateur et la « pression » trop forte sur les participants, qui pourraient alors considérer que ces essais se substitueraient au contrôle qualité interne ; ce qu'il faut éviter. A moins de 4 mois, les délais d'identification et de correction des problèmes seraient trop longs et l'impact du système sur les participants affaibli.

1.2.2.3 Valeur assignée

Le protocole reprend 4 des 5 possibilités considérées par le guide ISO 43-1 (§ 1.2.1.7) pour établir la valeur de référence acceptée, appelée dans le protocole « valeur assignée ».

Le cas b) du guide est exclu ; nous l'avons nous-mêmes écarté comme ne pouvant pas s'appliquer à notre domaine, par manque de méthode « définitive », donnant une valeur « vraie » du nombre de microorganismes.

Si 4 des différentes possibilités du guide ISO 43-1 sont reprises, le protocole les recommande néanmoins dans un ordre de priorité différent. Il apporte de plus des détails précieux pour leur mise en œuvre dans notre domaine :

1. Valeur de consensus des laboratoires experts.

Elle est préférée dans la plupart des circonstances, alors qu'elle ne figure qu'en 4^{ème} position dans le guide ISO qui proposait une classification des différentes possibilités sur la base d'une incertitude croissante autour de l'estimation de la valeur assignée. Là encore, la médiane ou le mode sont préconisés pour l'estimation de cette valeur.

2. Contamination artificielle : correspond au cas a) du guide.

Elle s'applique de préférence lorsque l'analyte est ajouté individuellement à chaque échantillon. Ce mode de contamination présente deux avantages :

- il n'y a pas besoin de vérifier l'homogénéité du matériau avant la préparation des échantillons ;
- dans certains cas, l'homogénéisation de l'analyte ajouté dans le matériau en vrac peut s'avérer difficile à réaliser.

Le document mentionne les limites de cette option :

- l'analyte doit effectivement être absent initialement du matériau et la quantité ajoutée d'analyte doit être connue très précisément ;
- l'analyte ajouté peut être moins intimement lié à la matrice, ou sous une forme chimique différente, même s'il est ajouté à la matrice avant la préparation des échantillons.

Ce cas est particulièrement adapté à la microbiologie des aliments, comme nous avons pu déjà le noter pour les essais inter-laboratoires de validation de méthodes (EN ISO 16140, § 1.1.9.2). Les limites indiquées par le protocole sont également entièrement applicables à notre domaine, en adaptant le vocabulaire.

3. Comparaison directe avec des matériaux de référence certifiés (MRC) : correspond au cas c) du guide.

Le protocole en précise les modalités : le matériau est analysé en parallèle avec un MRC, sous des conditions de répétabilité. Le MRC doit avoir une matrice et une gamme de concentration proches de celles du matériau. Cette exigence limite beaucoup, dans le domaine des aliments et en particulier en microbiologie, l'utilisation de cette possibilité.

4. Valeur de consensus des participants : également dernière possibilité du guide.

Pour les méthodes empiriques – c'est-à-dire des méthodes qui définissent l'analyte – il est indiqué que le consensus d'un large nombre de participants peut être considéré comme acceptable.

Le document identifie les différentes limites de cette approche :

- Une traçabilité ou une incertitude est difficile à attribuer à une telle valeur de consensus, à moins que tous les participants n'utilisent la même méthode de référence.
- Il n'y a pas toujours un consensus entre les participants, en particulier si différentes méthodes empiriques sont utilisées.
- Le consensus peut être biaisé par l'utilisation généralisée d'une mauvaise méthode.
- En cas de co-existence entre plusieurs méthodes empiriques, la valeur assignée pourrait être obtenue par des laboratoires experts avec une certaine méthode, alors que les participants pourraient utiliser une autre méthode, laissant penser à tort à l'existence d'un biais.

Comme nous l'avons déjà mentionné à l'occasion de la définition de la justesse par l'ISO 5725-1, les méthodes de dénombrement microbiologique peuvent être considérées comme étant empiriques. Cette option est donc largement applicable dans notre domaine,

tout comme le sont les défauts recensés. Si le problème de multiplicité de méthodes empiriques se pose, le document propose plusieurs façons d'y remédier :

- une valeur assignée est attribuée à chaque méthode ;
- une seule méthode doit être utilisée ;
- et/ou les participants sont avertis qu'un biais peut résulter de l'utilisation d'une méthode différente de celle qui a servi à établir la valeur de référence.

1.2.2.2.4 Choix de la méthode

Sur ce point, le protocole IUPAC/AOAC est en accord avec le contenu du guide ISO, tout en précisant que la méthode choisie par le participant doit être la plus proche possible, sinon identique, à celle utilisée en routine. En particulier, une répétition des analyses, pour vérifier les conditions de répétabilité, ne doit être effectuée que si elle est pratiquée habituellement.

1.2.2.3 Traitement statistique des résultats

Le protocole IUPAC/AOAC introduit les différentes statistiques de performance que l'on trouve dans le guide ISO (§ 1.2.1.8), à l'exception du plus complexe, le nombre E_n , et de façon plus approfondie que le guide ISO.

1.2.2.3.1 Scores z

Le protocole apporte les précisions qui manquaient au guide sur la façon d'obtenir s , l'estimation de la valeur cible pour l'écart-type (§ 1.2.1.8).

Deux situations sont distinguées :

1. L'écart-type s est calculé à chaque essai et pour chaque combinaison <matrice x analyte>, sur la base des résultats des participants après élimination des valeurs aberrantes, ou en utilisant des statistiques robustes.
2. Une valeur fixe est attribuée à s . Cette option est préférée, car elle permet de comparer d'un essai à l'autre les scores z d'un laboratoire. Plusieurs possibilités sont offertes pour fixer cette valeur :
 - Par « perception » : de façon arbitraire, par accord entre le coordinateur et les participants.
 - Par prescription : en fonction de la fidélité requise pour l'interprétation des données et en fonction du niveau de concentration. Le protocole favorise cette possibilité, notamment dans le cadre de contrôles officiels.
 - Par référence à une méthode validée. Lorsqu'une méthode de référence doit être utilisée dans l'essai, l'écart-type de reproductibilité de la méthode, obtenu lors d'un essai inter-laboratoires, peut convenir.
 - Par référence à un modèle généralisé. La courbe de Horwitz, qui relie l'écart-type de reproductibilité au niveau de concentration, peut être utilisée à défaut d'autres possibilités. Cette courbe ayant été dérivée expérimentalement d'un grand nombre d'essais inter-laboratoires organisés sur des méthodes d'analyse physico-chimique, elle n'est pas applicable en microbiologie des aliments.

Dans les deux premiers cas, la valeur cible de l'écart-type est directement liée à une déclaration « d'aptitude à l'emploi » (*fit-for-purpose*) de la méthode d'analyse à employer.

1.2.2.3.2 Score Q

Le protocole introduit un score alternatif au score z , le score Q , fondé sur le biais relatif, que le guide ISO a également introduit sous le nom de différence en pourcentage (§ 1.2.1.8). Le protocole, bien que donnant en annexe V des détails sur ce score, ne le recommande pas, en affirmant, à l'inverse du guide, que la signification de ce score n'est pas évidente, sans apporter plus de précision.

1.2.2.3.3 Combinaison des scores z

Comme le guide ISO, le protocole mentionne cette possibilité, pour un même essai et détaille en annexe III différentes façons de combiner les scores z :

- la somme des scores (SZ) ;
- la somme des scores divisée par \sqrt{m} (RSZ), où m est le nombre de scores combinés ;
- la somme des carrés des scores (SSZ).

Les deux premières combinaisons conservent l'information du signe des biais. Ainsi, de faibles scores z individuels, acceptables individuellement, peuvent devenir non acceptables lorsqu'ils sont combinés ; ce qui est intéressant pour un même système analytique. La troisième (SSZ) ne prend pas en compte le signe des scores z ; ce qui semble en général approprié, puisque dans un essai d'aptitude, l'attention est surtout portée sur l'amplitude du biais, et non sur sa direction.

Toutefois, le protocole déconseille une utilisation généralisée des scores combinés, qui risquent d'être mal interprétés, surtout lorsque des analyses de nature très différente sont combinées. Il existe également le danger d'ignorer le biais d'un laboratoire pour une analyse, qui pourrait se répéter d'un essai à l'autre, mais qui serait masqué par ses autres scores z satisfaisants.

Les systèmes français d'essais d'aptitude, dans le domaine des aliments, ont bien tenu compte de cette mise en garde, et ne pratiquent généralement pas la combinaison des scores.

1.2.2.3.4 Scores glissants

En complément du guide qui a introduit l'intérêt d'un suivi de la performance dans le temps (§ 1.2.1.8), le protocole indique la possibilité d'adopter des scores glissants (*running*) pour quantifier ce suivi. Des détails sur la façon de les établir sont donnés en annexe IV. Cette dernière propose d'abord la simple moyenne de scores sur un nombre donné d'essais. Un tel score glissant présente certes l'avantage de pondérer la mauvaise performance d'un laboratoire limitée à un seul essai, mais il comporte l'inconvénient de l'« effet mémoire » d'une déviation importante isolée. Deux moyens d'atténuer cet effet mémoire sont fournis.

1.2.2.3.5 Évaluation des résultats

Le guide et le protocole coïncident pour l'interprétation des scores z . Le protocole fournit néanmoins la justification des valeurs limites. Si \hat{X} et s sont des estimateurs justes de la moyenne et de l'écart-type vrais et si la distribution sous-jacente est normale, z suit alors une loi Normale centrée réduite. Ainsi, 95 % des valeurs de z seront situées dans l'intervalle $[-2 ; +2]$, et 99,7 % dans l'intervalle $[-3 ; +3]$. Ces valeurs correspondent à un système analytique qui s'est bien « comporté », selon le terme du document.

Le protocole attire l'attention sur les précautions à prendre dans la classification des laboratoires : les valeurs limites de 2 et 3 sont fondées sur des hypothèses qu'il ne faut pas oublier de vérifier (voir paragraphe ci-dessus).

1.2.3 Le projet de norme ISO/DIS 13 528

Comme indiqué plus haut, le projet de norme ISO/DIS 13 528 complète le Guide ISO 43-1 en ce qui concerne les aspects statistiques des essais d'aptitude. Il a été préparé par le comité ISO/TC 69/SC 6, la même structure qui a développé la norme ISO 5725.

Ce document reprend en fait les principes et la démarche définie en annexe A du guide, consacrée aux méthodes statistiques (§ 1.2.1.6), et en détaille les différents aspects. Cet approfondissement est conséquent, puisqu'il résulte en un document de 70 pages ! Il est de plus fondé sur le protocole IUPAC/AOAC présenté ci-dessus, tout en l'approfondissant également. Cependant, le traitement statistique de mesures qualitatives n'est toujours pas traité ; ce qui est un manque important pour notre domaine, comme nous l'avons déjà souligné.

En introduction, la notion de *performance* de laboratoires est traduite en termes statistiques. Trois propriétés peuvent caractériser cette performance : le biais de laboratoire, la stabilité et la répétabilité. Pour définir et estimer ces propriétés, un lien est établi avec les différentes parties de l'ISO 5725, en indiquant que les essais d'aptitude représentent une alternative aux moyens définis par l'ISO 5725 pour caractériser les trois critères de performance :

1. Biais de laboratoire. Nous avons vu que ce biais est normalement estimé par l'utilisation de matériaux de référence (§ 1.1.4.2). En l'absence de matériaux de référence appropriés, le biais de laboratoire peut être estimé en participant à un essai d'aptitude. Cependant, il est indiqué que cette estimation sera affectée par la stabilité et la répétabilité. Ainsi le biais important d'un laboratoire lors d'un essai peut être attribuable en réalité à une stabilité ou à une répétabilité insuffisantes.
2. Stabilité. Elle peut être évaluée par des analyses répétées régulièrement sur un même matériau de référence, conformément à l'ISO 5725-3, norme que nous n'avons pas présentée dans le chapitre précédent, car elle traite de la fidélité intermédiaire, sans mettre en jeu un essai inter-laboratoires. Alternativement, cette stabilité peut être estimée en construisant une carte de contrôle avec les biais de laboratoire obtenus au cours d'essais d'aptitude successifs.
3. Répétabilité. Elle est habituellement estimée au sein d'un laboratoire, au cours de l'activité normale ou lors d'essais dédiés à cette estimation. Cependant, il est possible que le laboratoire mette à profit sa participation à des essais d'aptitude pour estimer et surveiller sa répétabilité, en construisant une carte de contrôle avec les étendues d'analyse en double, conformément à l'ISO 5725-6 (§ 1.1.6.4).

Certains aspects de ce document sont maintenant présentés.

1.2.3.1 Conception des essais d'aptitude

Incertitude de la valeur assignée. Le document précise une exigence sur l'incertitude u_X de la valeur assignée X , alors que l'annexe A du guide se contente de demander au coordinateur d'estimer cette incertitude et d'en fixer un critère d'acceptabilité. En effet, si u_X est trop grande par rapport à l'écart-type cible de l'essais d'aptitude s , défini pour les scores z (§ 1.2.1.8), les résultats de certains laboratoires risquent d'être considérés comme suspects ou non acceptables, non pas en raison de leur propre performance, mais à cause d'une inexactitude dans l'estimation de la valeur assignée. La règle d'acceptabilité de u_X est définie ainsi : $u_X \leq 0,3 s$. Si cette règle n'est pas respectée, il est préconisé de choisir une meilleure

méthode d'estimation de X , ou de prendre en compte u_X dans l'interprétation des résultats de l'essai d'aptitude.

Répétitions des mesures. A l'inverse du protocole IUPAC/AOAC, le document recommande dans certains cas de répéter les mesures dans chaque laboratoire. En effet, une variation de la répétabilité entraîne une variation entre les biais de laboratoire dans un essai d'aptitude. S'il est souhaitable de limiter l'influence de cette variation, le nombre n de répétitions effectuées par chaque laboratoire sera choisi de la façon suivante : $\frac{s_r}{\sqrt{n}} \leq 0,3s$, où s_r est l'écart-type de répétabilité établi lors d'un essai précédent.

Homogénéité et stabilité des échantillons. L'annexe B de ce document est consacrée à la vérification de l'homogénéité et de la stabilité des échantillons. En matière d'*homogénéité*, la procédure définie par le protocole IUPAC/AOAC (§ 1.2.2.3) est reprise, avec les précisions suivantes :

- Les échantillons utilisés doivent provenir du même lot que celui dont sont extraits les échantillons de l'essai d'aptitude.
- Il est possible, pour les essais portant sur plusieurs analytes à la fois, de ne vérifier l'homogénéité qu'au regard d'un seul analyte, à condition que celui-ci soit le plus sensible à l'hétérogénéité.

Par ailleurs, si le critère d'homogénéité n'est pas respecté, il est recommandé :

- soit d'améliorer la procédure de préparation des échantillons ;
- soit d'évaluer avec les laboratoires l'écart-type inter-échantillons s_s . Ce dernier est combiné de la façon suivante à l'écart-type de répétabilité s_r :

$$s_{r1} = \sqrt{s_r^2 + s_s^2} \quad \text{Eq. 9}$$

s_{r1} est utilisé à la place de s_r pour déterminer le nombre de répétitions.

- soit d'inclure s_s de la façon suivante dans l'écart-type cible de l'essai d'aptitude, s :

$$s = \sqrt{s_1^2 + s_s^2} \quad \text{Eq. 10}$$

où s_1 est l'écart-type cible n'incluant aucune tolérance pour l'hétérogénéité des échantillons.

Une procédure de vérification de la *stabilité* est introduite, alors qu'elle n'existe pas dans le protocole IUPAC/AOAC :

1. Sélection au hasard de g (au moins 3) échantillons d'un même lot de matériau.
2. A l'issue du délai séparant la préparation des échantillons de leur mise en analyse par les participants, préparation de 2 prises d'essais par échantillon, analysées chacune une fois, dans des conditions de répétabilité.
3. Comparaison de la moyenne générale des résultats d'analyse de l'étude d'homogénéité à celle des résultats obtenus lors de l'étude de stabilité. Les échantillons peuvent être considérés comme suffisamment stables si la différence absolue des deux moyennes générales est inférieure ou égale à $0,3 s$.

1.2.3.2 Détermination de la valeur assignée et de son incertitude-type

Les cinq possibilités avancées en annexe A de l'ISO 43-1 pour estimer la valeur assignée sont reprises intégralement. Outre des détails sur les différentes possibilités, le guide est complété par la façon de calculer l'incertitude-type de ces estimations :

1. Valeurs connues obtenues par contamination artificielle
L'incertitude-type est estimée par combinaison des incertitudes selon le GUM.
2. Valeurs de référence certifiées
L'incertitude-type est déduite des informations sur l'incertitude fournies par le certificat du matériau de référence certifié (MRC). On peut remarquer que le projet de norme confirme la signification que nous avons accordée à la valeur de référence certifiée, à l'encontre de ce que le texte de l'ISO 43-1 pouvait laisser entendre par « méthode définitive » (voir 1.2.1.7).
3. Valeurs de référence
Lorsque la valeur assignée d'un matériau est établie par comparaison avec un MRC, l'incertitude-type est déduite de celle du MRC.
4. Valeurs consensuelles provenant de laboratoires spécialisés
En accord avec le protocole IUPAC/AOAC, la médiane des résultats constitue la valeur assignée. De plus, le projet décrit en annexe C la méthode robuste à employer pour les deux types de valeurs consensuelles, soit l'algorithme A de l'ISO 5725-5. Dans ce cas, une formule permet de déduire l'incertitude-type sur la valeur assignée à partir des incertitudes-types de chaque laboratoire.
5. Valeurs consensuelles des laboratoires participants
Une formule permet de déduire l'incertitude-type de l'écart-type robuste s^* calculé selon l'algorithme A de l'ISO 5725-5 sur les moyennes de résultats pour un matériau. Cette incertitude-type n'est pas affectée par l'utilisation générale d'une méthode d'analyse erronée, qui biaiserait alors la valeur de consensus. Cette limite, déjà notée par le protocole IUPAC/AOAC, peut être rencontrée en particulier lors de l'analyse de traces.

1.2.3.3 Détermination de l'écart-type pour l'évaluation de l'aptitude

Le document adopte l'approche du protocole IUPAC/AOAC pour estimer cet écart-type (noté s), que le protocole appelle écart-type cible. En particulier, les 5 possibilités de l'estimer sont retrouvées, avec des compléments, comme les formules de l'écart-type selon les différentes possibilités.

1.2.3.4 Calcul des statistiques de performance

Le projet de norme détaille les différentes statistiques de performance introduites dans l'annexe A du guide ISO. Pour chacune d'entre elles, il fournit la façon de les interpréter. Ici aussi, il reprend le contenu du protocole IUPAC/AOAC, en apportant quelques détails. A l'exception, bien évidemment, de ceux qui ne sont pas traités dans le protocole, comme les scores E_n , ainsi que les deux statistiques supplémentaires suivantes, qui ne figurent pas dans le guide ISO :

1. Score z' , défini ainsi :

$$z' = \frac{x - X}{\sqrt{s^2 + u_X^2}} \quad \text{Eq. 11}$$

où u_X est l'incertitude-type de la valeur assignée X .

Les scores z' sont toujours inférieurs aux scores z . Si la règle d'acceptabilité d' u_X (§ 1.2.3.1) est respectée, cette différence est négligeable. Sinon, certains scores z peuvent dépasser les valeurs limites de 2 ou 3, donnant ainsi des signaux de surveillance ou d'action, alors que les scores z' peuvent ne pas les dépasser. Les scores z' sont alors préférables, à condition que u_X soit connue.

2. Score $z\hat{\eta}$, défini ainsi :

$$z\hat{\eta} = \frac{x - X}{\sqrt{u_x^2 + u_X^2}} \quad \text{Eq. 12}$$

où :

- u_x est l'incertitude-type du résultat x du laboratoire, estimée par le laboratoire lui-même ;
- u_X est l'incertitude-type de la valeur assignée X .

Ce score est très proche du nombre E_n , à l'exception de l'estimation de l'incertitude sur la valeur assignée. Comme pour E_n , la mise en œuvre de cette option est conditionnée par la disponibilité de l'incertitude de mesure de chaque laboratoire pour la combinaison <matrice, analyte> considérée. Comme nous l'avons déjà noté, l'incertitude de mesure est de plus en plus fréquemment estimée par les laboratoires, et son utilisation pour les essais d'aptitude représenterait une valorisation supplémentaire d'une caractéristique dont l'obtention requiert un investissement lourd en microbiologie des aliments.

1.2.3.5 Méthodes graphiques de combinaison de scores de performance

1.2.3.5.1 Sur un seul essai d'aptitude

Afin de combiner les scores de performance de plusieurs analytes pour un même essai d'aptitude, six types de graphiques sont proposés. Ils sont décrits dans le Tableau 5.

Tableau 5 : Graphiques de combinaison de scores de performance pour un seul essai d'aptitude, selon l'ISO/DIS 13528

Type de graphique	Description	Applicabilité
Histogramme de scores de performance	Un histogramme par analyte.	Adapté à un faible nombre d'analytes, ou à des analytes différents.
Diagramme en bâtons du biais de laboratoire normalisé	Identique au diagramme <i>h</i> de Mandel de l'ISO 5725-2 (§ 1.1.3.3). Les scores pour tous les analytes sont rassemblés sur le même diagramme.	Adapté à des analytes semblables, étudiés simultanément.
Diagramme en bâtons de répétabilité normalisée	Identique au diagramme <i>k</i> de Mandel de l'ISO 5725-2 (§ 1.1.3.3).	Applicable lorsque des répétitions on été effectuées.
Graphique de Youden	Graphique construit avec les deux scores appariés pour chaque participant. L'interprétation de ce graphique est facilitée par la construction d'une ellipse de confiance, dont le calcul est fourni.	Applicable lorsque l'essai a inclus des paires de Youden, soit un essai à niveau fractionné selon l'ISO 5725-5 (§ 1.1.5.1).
Graphique des écarts-types de répétabilité	Représentation de l'écart-type intra-laboratoire de chaque participant en fonction de sa moyenne.	Applicable lorsque des répétitions on été effectuées.
Graphique à échantillons partagés (§ 1.2.1.4)	Représentation sur un axe de la variation entre analyses répétées dans deux laboratoires (ou plus), et sur l'autre axe de la différence entre les moyennes de ces laboratoires pour chaque échantillon.	

Ces techniques graphiques de combinaison de scores présentent l'avantage de ne pas masquer une valeur élevée de score individuel par plusieurs autres scores satisfaisants.

Le graphique de Youden est particulièrement intéressant pour l'évaluation de la performance de laboratoires par essais d'aptitude.

1.2.3.5.2 Sur plusieurs essais d'un même système

Il s'agit ici de combiner les scores de performance sur plusieurs essais d'un même système d'essais d'aptitude. Le projet de norme, se démarquant ici nettement du protocole IUPAC/AOAC (§ 1.2.2.3), déconseille d'utiliser les scores « glissants » en l'absence de représentation graphique, compte-tenu des risques de masquer une mauvaise performance ponctuelle. Le Tableau 6 décrit les quatre méthodes graphiques proposées.

Tableau 6 : Graphiques de combinaison de scores de performance pour plusieurs essais d'aptitude d'un même système, selon l'ISO/DIS 13528

Type de graphique	Description	Applicabilité
Carte de contrôle de Shewhart	Construite selon l'ISO 8258.	Efficace pour identifier les problèmes entraînant une dispersion importante des valeurs de scores z .
Carte de contrôle à somme cumulée	Construite selon le fascicule de documentation AFNOR FD V 03-115.	Appropriée pour l'identification de problèmes entraînant un biais persistant sur plusieurs essais successifs.
Diagramme du biais de laboratoire normalisé par rapport à la moyenne des laboratoires		Adapté aux cas où le niveau de l'analyte varie sur une large étendue, d'un essai à l'autre du même système.
Diagramme en points	Semblable à la carte de contrôle de Shewhart, mais avec plusieurs points pour un même essai et avec des droites reliant la moyenne des scores à chaque essai.	Adapté à la représentation de systèmes qui prévoient plusieurs échantillons pour le même analyte durant le même essai. Situation souvent rencontrée dans notre domaine.

1.2.4 Mise en œuvre des référentiels en microbiologie des aliments

L'organisation d'essais d'aptitude se généralise depuis plusieurs années tant en France qu'à l'étranger, dans le domaine de l'analyse des aliments et en particulier en microbiologie, couvrant une variété toujours plus grande de combinaisons <matrices x analyte>. Cette évolution est sans doute largement attribuable à l'importance donnée à cette évaluation par les organismes d'accréditation et par les autorités en charge du contrôle officiel des aliments. Nous nous limiterons cependant à présenter l'activité de deux organisateurs d'essais d'aptitude, le RAEMA et CECALAIT, les deux seuls réseaux français opérant à ce jour en microbiologie des aliments, et disposant d'une expérience déjà assez ancienne.

1.2.4.1 RAEMA

Le champ d'activité du « Réseau d'Analyses et d'Echanges en Microbiologie des Aliments » coïncide exactement avec notre domaine d'intérêt. Le RAEMA a été créé en 1988, à l'initiative de deux associations, l'AVHA (Association Vétérinaire d'Hygiène Alimentaire) et l'ADILVA (Association des Directeurs de Laboratoires Vétérinaires d'Analyse). Le nombre de participants a considérablement augmenté : de 76 pour le premier essai, il est actuellement d'environ 450. Au départ presque uniquement formé des laboratoires publics français, le réseau comprend aujourd'hui des laboratoires au statut diversifié : 70 % des laboratoires sont privés (prestataires ou d'entreprise) et environ 60 laboratoires sont étrangers (surtout belges) [Augustin & Carlier, 2002 ; RAEMA, 2004].

La plupart des germes recherchés ou dénombrés en hygiène alimentaire sont couverts. Il s'agit de la détection de *Salmonella* et de *Listeria monocytogenes*, ainsi que du dénombrement de la flore aérobie mésophile, des *Enterobacteriaceae*, des coliformes totaux et thermotolérants, des *Escherichia coli* β -glucuronidase positive, des bactéries anaérobies sulfito-réductrices, de *Clostridium perfringens*, des staphylocoques à coagulase positive et de *L.monocytogenes*.

Deux essais sont organisés par an.

Echantillons, méthodes d'analyse et rapport

Pour chaque essai, chaque laboratoire reçoit 5 échantillons. Une seule matrice est utilisée : du lait en poudre. Un lot de lait en poudre est contaminé artificiellement par un mélange de germes cibles, sous forme déshydratée, avant de prélever les échantillons distribués aux laboratoires. La quantité d'échantillon est prévue de façon à permettre au laboratoire d'effectuer plusieurs prises d'essai pour la détermination de chacun des germes. Le nombre d'échantillons contaminés par les différents microorganismes ainsi que les niveaux de contamination sont variables d'un essai à l'autre. Habituellement, les 5 échantillons sont contaminés de façon identique par les germes à dénombrer. Ainsi, ce système a fait le choix d'un nombre limité de répétitions, inférieur à celui recommandé par IUPAC/AOAC, mais en contre partie il propose des essais très complets, en termes d'analytes. Par ailleurs, la matrice choisie est stable à température ambiante au regard des microorganismes qu'elle contient. Ce choix facilite les conditions d'organisation des essais inter-laboratoires mais il ne peut pas représenter les conditions d'analyse des denrées périssables, fraîches.

Conformément au projet ISO/DIS 13528, la *stabilité* est étudiée à chaque essai par analyse en début et fin d'essai de toutes les flores à déterminer. Quant à l'*homogénéité*, elle est vérifiée à chaque essai par dénombrement en double de la flore aérobique mésophile sur 10 échantillons de chacune des 5 préparations. Ainsi, à condition de considérer que la flore aérobique mésophile n'est pas moins sensible que les autres germes cibles à l'hétérogénéité de contamination, les prescriptions du guide ISO 43-1 complétées par le projet ISO/DIS 13528 (§ 1.2.3.1) sont bien respectées.

Le choix de la méthode d'analyse est laissé aux participants. Dans le modèle de rapport adressé aux participants, doivent être en particulier renseignés les méthodes utilisées, les milieux de culture (nature, contrôles de performance, conditions d'incubation) ainsi que les modalités de confirmation.

Évaluation de la performance pour les dénombrements

Selon le projet ISO/DIS 13528, la performance est évaluée tant en termes de justesse (biais de laboratoire) que de fidélité (répétabilité). Pour déterminer les valeurs assignées, le RAEMA a choisi les valeurs consensuelles de l'ensemble des participants, obtenues selon les méthodes de calculs robustes préconisées par le projet de norme (§ 1.2.3.2). Comme l'indique le protocole IUPAC/AOAC (§ 1.2.2.2), il s'agit effectivement de l'option la plus appropriée pour estimer les valeurs consensuelles, en cas de méthodes empiriques et ce d'autant que le nombre de participants aux essais est très élevé. Conformément au protocole et au projet ISO, une analyse statistique est effectuée afin d'évaluer l'impact des méthodes utilisées sur les résultats obtenus. Si un biais significatif est identifié, plusieurs valeurs assignées seront calculées pour chacune des méthodes.

Tous les calculs sont effectués après transformation des résultats de dénombrements en logarithmes décimaux.

Pour le biais de laboratoire, le score z est employé avec les limites de surveillance et d'actions communes aux trois référentiels présentés plus haut. Le RAEMA se démarque ainsi des réticences françaises.

La répétabilité est estimée pour deux flores, l'une non spécifique à fort taux de contamination, à savoir la flore aérobique mésophile, et l'autre spécifique à taux plus faible, soit *E.coli* ou les staphylocoques. Les cinq échantillons sont contaminés à des taux identiques par ces deux flores. L'estimation de la répétabilité est ainsi fondée sur des quintuples répétitions.

Un indice de performance r_i est défini de la façon suivante :

$$r_i = \frac{(k-1)s_i^2}{s^{*2}} \quad \text{Eq. 13}$$

où :

- k est le nombre de répétitions (en général 5) ;
- s_i est l'écart-type des logarithmes décimaux des k résultats obtenus par le $i^{\text{ème}}$ laboratoire ;
- s^* est la valeur cible pour l'écart-type de l'essai d'aptitude, soit l'estimation robuste à partir des écarts-types s_i de l'ensemble des participants.

Cet indice est dérivé de la statistique de cohérence intra-laboratoire k de Mandel, comme le recommande le projet ISO/DIS 13528 (§ 1.1.3.3). En l'absence de valeurs limites fixées par

ce projet ISO, le RAEMA fixe des limites pour r_i : un indice r_i inférieur à 0,45 ou supérieur à 11,5 donne un signal de surveillance et un indice r_i inférieur à 0,1 ou supérieur à 18 donne un signal d'action. Ces valeurs limites ont été calculées en supposant que le logarithme de la contamination suit une distribution Normale ; ce qui implique que l'indice r_i suit une distribution du χ^2 à $k-1$ degrés de liberté. Il peut paraître surprenant que les valeurs faibles de répétabilité donnent lieu à des signaux de surveillance ou d'action, indiquant alors normalement la nécessité d'une action corrective ! De fait, les organisateurs souhaitent ainsi attirer l'attention du laboratoire qui a obtenu une répétabilité significativement meilleure que les autres laboratoires. C'est également une façon de décourager les falsifications. En effet, l'adoption de ces limites inférieures a fait nettement baisser le nombre de résultats identiques ou très proches, au sein d'un même laboratoire !

Évaluation de la performance pour les détections

Deux critères sont utilisés pour évaluer la performance des laboratoires pour la détection de *Salmonella* et de *L.monocytogenes* : la sensibilité, soit la capacité du laboratoire à détecter les échantillons contaminés, et la spécificité, soit la capacité du laboratoire à ne pas détecter de contamination dans les échantillons témoins négatifs. Les taux de résultats faux positifs et de résultats faux négatifs sont calculés.

Nous pouvons remarquer que cette interprétation n'est pas plus approfondie que celle des trois référentiels disponibles.

1.2.4.2 CECALAIT

Le « Centre d'Expertise et de Contrôle des Analyses LAITières » (CECALAIT) est une association fondée en 1990 par l'INRA, le CNIEL (Centre National Interprofessionnel de l'Economie Laitière) et l'Institut de l'Élevage [Rollier *et al*, 1996]. CECALAIT a pour mission principale d'organiser des essais inter-laboratoires d'aptitude dans le secteur laitier, tant en physico-chimie qu'en microbiologie. Des essais dans ce dernier domaine sont organisés depuis 1992 avec du lait cru et depuis 1994 avec du fromage frais. CECALAIT dispose donc d'une expérience relativement plus récente que le RAEMA.

Les essais d'aptitude suivants sont organisés par campagnes. Une campagne correspond aux envois répétés d'échantillons d'une même matrice – lait cru ou fromage frais – pour un même groupe de germes :

- Avec du lait cru :
 - flore aérobie mésophile et coliformes : 4 fois par an ;
 - dénombrement de spores butyriques : 3 fois par an.
- Avec du lait cru et du fromage frais : dénombrement des staphylocoques et des *E.coli*, détection de *L.monocytogenes* et de *Salmonella*, 3 fois par an.

Le nombre de participants par campagne varie de 10 à 119. La fréquence des essais est donc supérieure à celle du RAEMA, respectant le minimum recommandé par le protocole IUPAC/AOAC, mais le nombre de germes étudiés est plus restreint.

Echantillons

Comme le RAEMA, CECALAIT contamine le lot avant de préparer les échantillons distribués. Les échantillons de « fromages frais » (plus exactement un caillé frais) sont préparés de la façon suivante :

- Un lot de lait cru est contaminé artificiellement avec des souches cultivées des germes cibles et d'espèces spécifiques des fromages, telles que *Lactococcus lactis* sbsp. *lactis*, *Enterococcus faecalis*, *Lactobacillus paracasei* et *L.plantarum*.
- Un agent bactériostatique non divulgué est ajouté pour éviter la croissance de la flore lactique et pour stabiliser la flore cible. L'effet stabilisant de cet agent disparaît avec les dilutions effectuées en début d'analyse.
- Le lait est ensuite distribué dans des flacons où la coagulation s'effectuera par ajout de présure.

L'homogénéité de ces échantillons est considérée comme aussi satisfaisante que celle du lait liquide. Leur durée de vie a été testée avec succès jusqu'à 4 jours à 4°C. Les échantillons, soit 5 ou 10 par laboratoire selon la campagne, sont distribués par courrier rapide, dans des emballages isolés et comportant des éléments réfrigérants.

Par rapport au RAEMA, la nature de la matrice utilisée entraîne beaucoup plus de contraintes. En effet, cette matrice nécessite un transport rapide dans un emballage réfrigéré et a une durée de vie courte. En revanche, la représentativité de ce type d'échantillons par rapport aux analyses effectuées en routine nous semble meilleure, puisqu'une telle matrice, contaminée par une flore à l'état frais est proche des types de nombreux échantillons frais, analysés habituellement en microbiologie des aliments.

Cependant, des études d'homogénéité et de stabilité ne sont pas réalisées systématiquement à chaque essai, selon les référentiels cités précédemment, alors même que la stabilité et l'homogénéité de sa contamination sont plus difficiles à maîtriser. Il s'agit là, à notre avis, d'une déficience d'organisation de ce réseau.

Évaluation de la performance pour les dénombrements

Comme le RAEMA, CECALAIT transforme les résultats en logarithmes décimaux pour rendre la distribution des résultats proche de la Normale. La performance des laboratoires est appréciée tant en répétabilité qu'en biais.

Pour estimer la *répétabilité*, chaque laboratoire analyse deux fois chacun des 5 (ou 10) échantillons, soit une répétition en doubles non aveugles. Le paramètre habituel est d'abord calculé, à savoir l'écart-type de répétabilité en unité logarithmique ($s_r \log$). Cependant, ce paramètre, compte-tenu de son unité, n'est pas considéré par CECALAIT comme étant d'une utilisation facile par les laboratoires pour l'interprétation de leurs résultats, en termes d'écarts acceptables entre répétitions. C'est pourquoi CECALAIT utilise un paramètre introduit par Piton et Grappin [Piton & Grappin, 1991], *l'écart-type relatif géométrique (GRSDr %)*.

Ce dernier se fonde sur la propriété suivante. La moyenne arithmétique de valeurs en log est égale au log de la moyenne géométrique en unité d'origine, soit :

$$\bar{x}^*_{A} = \frac{\sum x^*}{N} = \log \bar{x}_G \quad \text{Eq. 14}$$

où :

- $\bar{x}_G = \prod x^{1/N}$ (moyenne géométrique) ;
- * indique que la valeur est en unité logarithmique.

La façon de définir l'écart-type relatif géométrique est présentée dans le Tableau 7.

Tableau 7 : Calcul de l'écart-type relatif géométrique

Etape	Calcul
Intervalle de confiance	$\bar{x}^* \pm s_r^*$ où \bar{x}^* et s_r^* sont la moyenne et l'écart-type de répétabilité de la population, tous deux exprimés en log ufc/ml.
Limite haute de l'intervalle de confiance, dans la population log-normale d'origine	$10^{\bar{x}^* + s_r^*} = \bar{x}_G \times 10^{s_r^*}$
Ecart-type relatif géométrique <i>GRSDr</i> (%)	$GRSDr = \frac{\bar{x}_G \times 10^{s_r^*} - \bar{x}_G}{\bar{x}_G} \times 100 = (10^{s_r^*} - 1) \times 100$

Dans le Tableau 7, on peut noter la façon inhabituelle de définir l'intervalle de confiance : l'écart-type de répétabilité (ou de reproductibilité) est en général multiplié par 2, soit $\bar{x} \pm 2s_r$, tel que dans l'exemple d'application au § 2.5.4.

L'écart-type relatif géométrique caractérise l'étendue de la partie supérieure de l'intervalle de confiance, ou la variabilité des résultats au-dessus de la moyenne géométrique.

Ainsi, pour chaque laboratoire, sont calculés s_r et *GRSDr*. CECALAIT a défini, sur la base de données publiées, issues d'essais inter-laboratoires de validation des méthodes de référence, les valeurs limites d'acceptabilité reportées dans le Tableau 8.

Tableau 8 : Valeurs limites d'acceptabilité de répétabilité pour les essais d'aptitude CECALAIT

Germe	s_r (log)	<i>GRSDr</i>
<i>E.coli</i>	0,14	40 %
Staphylocoques à coagulase positive	0,25	80 %

En ce qui concerne le *biais de laboratoire*, il est estimé pour chaque laboratoire par deux paramètres :

1. la moyenne \bar{d} des différences entre les résultats des 5 (ou 10) échantillons et les valeurs de référence acceptées pour chacun des échantillons, respectivement. Les valeurs de référence acceptées sont égales à la moyenne de tous les participants, après exclusion des valeurs aberrantes.
2. l'écart-type s_d des 5 (ou 10) différences individuelles.

Ces deux paramètres sont également exprimés sous forme relative, à l'aide de la moyenne géométrique, comme pour la répétabilité. Soit :

- la moyenne relative géométrique des différences (*GRd*) ;
- l'écart-type relatif géométrique des différences (*GRSd*).

Enfin, les laboratoires sont classés par la distance euclidienne D du laboratoire :

$$D = \sqrt{\bar{d}^2 + s_d^2} \quad \text{Eq. 15}$$

Un graphique est construit, où \bar{d} figure en abscisse et s_d en ordonnée.

A nouveau sur la base de données publiées, CECALAIT a défini des valeurs de référence figurant dans le Tableau 9.

Tableau 9 : Valeurs limites d'acceptabilité de biais pour les essais d'aptitude CECALAIT

Germe	\bar{d} (log)	<i>GRd</i>	s_d (log)	<i>GRSd</i>
<i>E.coli</i>	± 0,3	+ 100 % -50 %	0,3	100 %
Staphylocoques à coagulase positive	± 0,4	+ 150 % - 60 %	0,4	150 %

Au total, l'exploitation statistique a certes des points communs avec le RAEMA, notamment en ce qui concerne les premières étapes. Cependant, CECALAIT se démarque du RAEMA en introduisant des paramètres adaptés à la transformation logarithmique des résultats, transformation habituelle pour les dénombrements microbiens et qui sera discutée en 2^{ème} partie. Par ailleurs, CECALAIT n'a pas hésité à introduire un classement des participants en fonction de leur biais, malgré les réserves des référentiels sur ce sujet.

Évaluation de la performance pour les détections

Comme le RAEMA, CECALAIT calcule pour chaque laboratoire la fréquence de résultats corrects par rapport aux résultats vrais définis par l'organisateur (contamination artificielle ou pas des échantillons). Par contre, comme pour les dénombrements, une classification des participants est opérée. Si différentes cibles sont incluses dans l'essai, un poids différent peut être attribué à chacun des rangs, en fonction de l'importance de la cible, pour en déduire une classification unique des participants pour l'essai.

1.3 Caractériser un matériau de référence

Résumé

Un ensemble de guides ISO porte sur les matériaux de référence. Parmi ceux-ci, le guide ISO 30 introduit la définition d'un matériau de référence, dont les implications en microbiologie des aliments sont envisagées, ainsi qu'une définition des matériaux de référence certifiés, dont les caractéristiques sont soulignées, en prenant l'exemple des matériaux BCR.

Une typologie des matériaux de référence est proposée par le Fascicule de Documentation AFNOR FD V 03-115 pour l'analyse des aliments, ajoutant aux deux types de matériaux ci-dessus les matériaux de référence externes, dont les valeurs de référence sont établies lors d'essais d'aptitude.

Les seuls matériaux de référence certifiés existant à ce jour en microbiologie des aliments sont décrits : les matériaux BCR, soit des capsules de poudre de lait artificiellement contaminée. La disponibilité de matériaux de référence externes est potentiellement nettement plus importante.

1.3.1 Généralités

Nous ne consacrerons qu'une place limitée à ce troisième type d'essais inter-laboratoires. En effet, très peu de matériaux de référence, notamment certifiés, existent à ce jour en microbiologie des aliments. Par ailleurs, nous n'avons pas mis en œuvre d'essais inter-laboratoires dans cet objectif.

D'une façon générale, les matériaux de référence bénéficient d'un ensemble de publications de référence, principalement sous forme de guides ISO, préparés par une structure de l'ISO qui s'y consacre exclusivement : le « Comité sur les Matériaux de Référence » (REMCO). On peut retenir en particulier le guide ISO 30 qui fixe la terminologie et le guide ISO 35 qui définit les principes généraux et statistiques pour la certification de matériaux de référence. Un autre guide complète ce dernier pour des aspects liés à la qualité. Il s'agit du guide ISO 34 qui prescrit des exigences générales relatives à la compétence des organismes producteurs de matériaux de référence certifiés.

1.3.2 Définitions et typologie

1.3.2.1 Matériaux de référence (MR)

Le guide ISO 30 propose la **définition** suivante d'un matériau de référence :

« Matériau ou substance dont une (ou plusieurs) valeur(s) de la (des) propriété(s) est (sont) suffisamment stable(s), homogène(s) et bien définie(s) pour être utilisée dans l'étalonnage d'un appareil, l'évaluation d'une méthode de mesure, ou pour assigner des valeurs à des matériaux. »

La valeur de la propriété est établie en soumettant ce matériau à un essai inter-laboratoires, dans lequel au moins deux laboratoires indépendants déterminent la valeur de la propriété en utilisant leur propre méthode de mesure. La valeur de référence est la valeur de consensus de l'essai inter-laboratoires et elle est exprimée comme la valeur moyenne, assortie de son incertitude. Lorsque plusieurs méthodes de mesure sont utilisées, les valeurs obtenues doivent se trouver à l'intérieur de l'intervalle d'incertitude assigné à chaque méthode. Lorsque les différences entre méthodes sont trop importantes, la valeur de référence peut être exprimée en fonction de la méthode d'analyse.

Nous pouvons donc tout de suite noter le rôle central tenu par l'essai inter-laboratoires dans le processus d'obtention d'un matériau de référence. Les essais inter-laboratoires de caractérisation des matériaux de référence se distinguent nettement des deux autres types d'essais inter-laboratoires, abordés dans les chapitres précédents, par la nature des participants. En effet, il doit s'agir de laboratoires experts ou spécialisés, selon la terminologie du guide ISO 43-1, dans la détermination de la propriété concernée pour le matériau candidat à devenir matériau de référence. L'organisateur de l'essai doit donc opérer une sélection des laboratoires participants.

De cette description peuvent être déduites les trois principales **propriétés** d'un matériau de référence :

1. Homogénéité. L'homogénéité, qui est la variation de la valeur de la propriété entre les différentes parties du matériau, doit rester à l'intérieur de limites spécifiées. Dans une situation idéale, cette variation est négligeable par rapport à la variation due à la méthode de mesure (fidélité) ;
2. Stabilité. La valeur de la propriété doit rester stable dans des limites spécifiées sur une période spécifiée de temps pendant la conservation et le transport.
3. Représentativité. Le matériau de référence doit être représentatif de son utilisation prévue. En ce qui nous concerne, il s'agit principalement du type de matrice et du niveau de contamination.

Nous pouvons remarquer que ces propriétés ne sont pas exclusives des matériaux de référence et qu'elles sont aussi requises pour les matériaux utilisés dans les deux autres types d'essais inter-laboratoires traités dans les chapitres précédents. La différence réside principalement dans le degré d'exigence pour ces propriétés, notamment en ce qui concerne la stabilité. Dans les deux premiers types d'essais, la stabilité est nécessaire uniquement pour la durée de l'essai inter-laboratoires, alors qu'une stabilité se comptant en mois sinon en années est ici exigée.

1.3.2.2 Matériaux de référence certifiés (MRC)

Considérons maintenant un type particulier de matériau de référence, les matériaux de référence certifiés. Le guide ISO 30 les définit ainsi :

« Matériau de référence, accompagné par un certificat, dont une ou plusieurs valeur(s) de propriété(s) a (ont) été certifiée(s) par une procédure qui établit la traçabilité à une réalisation exacte de l'unité dans laquelle les valeurs de la propriété sont exprimées et pour lequel chaque valeur certifiée est accompagnée par une incertitude à un niveau de confiance donnée. »

De cette définition, un peu compliquée dans sa formulation, ressortent trois caractéristiques s'ajoutant à celles d'un matériau de référence : l'existence d'un *certificat*, qui s'engage sur l'existence d'une *traçabilité* à une valeur exacte de l'unité de la valeur certifiée et sur une *incertitude* associée à cette valeur certifiée.

1.3.3 Matériaux de référence en agroalimentaire

Se rapprochant de notre domaine d'intérêt, le fascicule de documentation AFNOR FD V 03-115 traite de l'utilisation des matériaux de référence dans le domaine de l'analyse des aliments. Trois types de matériaux de référence sont distingués par le FD V 03-115 :

1. Matériau de référence certifié : défini en accord avec le guide ISO 30, mais de façon simplifiée.
L'importance est donnée au recours à un essai inter-laboratoires, à l'incertitude attachée à la valeur de référence et à l'organisme reconnu certifiant ces matériaux, dont une liste est donnée en annexe.
2. Matériau de référence externe (MRE): matériau de référence dont la (les) valeur(s) de consensus ont été déterminée(s) à la suite d'essais inter-laboratoires.
3. Matériau de référence interne : matériau de référence dont la (les) valeur(s) de référence est attribuée par l'utilisateur :
 - par comparaison aux valeurs certifiées ou aux valeurs de consensus des matériaux de référence (MRC ou MRE) ;
 - par ajout d'une quantité connue de l'analyte à la matrice exempte de cet analyte.

Ce document a donc l'intérêt d'élargir la conception de matériau de référence, dont l'établissement requiert systématiquement, selon les guides ISO cités ci-dessus, un essai inter-laboratoires. Par ailleurs, il est clairement indiqué qu'un matériau de référence, distingué par sa dénomination (MRE) des MRC, peut être également caractérisé soit par les essais inter-laboratoires de validation de méthode, soit par les essais d'aptitude.

1.3.3.1 Matériaux de référence certifiés BCR

Au niveau mondial, les deux principaux organismes certificateurs de matériaux de référence sont le Centre Commun de Recherches (*Joint Research Centre, JRC*) de la Commission Européenne (CE) et le NIST (*National Institute for Science & Technology*) aux USA. Nous présenterons brièvement le premier [JRC/IRMM, 2004].

L'Institut pour les Matériaux de Référence et pour les Mesures (IRMM) du JRC, situé à Geel (Belgique), certifie des matériaux de référence appelés « BCR », en référence à l'ancien organisme certificateur, le Bureau Communautaire de Référence. Ces MRC sont soit le produit de projets de recherche financés par la CE, soit directement développés par l'IRMM.

Les cinq étapes majeures du processus de certification, conçu en accord avec les guides ISO, notamment les Guides 34 et 35, sont les suivantes :

1. étude de faisabilité ;
2. préparation du matériau de référence envisagé ;
3. essais d'homogénéité ;
4. essais de stabilité ;
5. mesures de certification, normalement par essai inter-laboratoires impliquant en général de 6 à 15 laboratoires.

L'IRMM se charge au minimum de la certification au sens strict, soit la soumission des documents de certification à un groupe d'experts indépendants des participants à l'exercice de certification.

1.3.3.2 Matériaux de référence en microbiologie

Comme nous l'avons déjà mentionné en début de ce chapitre, il n'existe que peu de matériaux de référence, et encore moins de MRC, dans notre domaine d'étude. Cette situation est clairement imputable à ce qui singularise l'analyse microbiologique, à savoir le caractère vivant de l'analyte, caractéristique que nous avons déjà soulignée à plusieurs reprises. Ainsi, deux des propriétés d'un matériau de référence sont difficiles à respecter :

- l'homogénéité de la contamination entre échantillons,
- la stabilité, qui, pour un MRC, doit se compter en années.

Une autre raison vient s'ajouter à la nature de l'analyte. Nous avons vu que les méthodes conventionnelles d'analyse microbiologique, et notamment les méthodes de dénombrement, sont de nature « empirique » : le germe cible est défini par la méthode employée. Ainsi, surtout dans le cas de matériaux de référence dits quantitatifs, pour lesquels la valeur de référence est quantifiée, cette dernière est dépendante d'une méthode donnée (§ 1.3.2.1). L'universalité de tels matériaux et l'étendue de leur utilisation potentielle en sont singulièrement limitées.

1.3.3.3 Les matériaux de référence certifiés préparés par le RIVM

Malgré ces difficultés, la CE a financé un projet de recherche monté par le RIVM ayant pour objectif de développer des MRC en microbiologie des aliments à travers le BCR puis le programme « Normes, Mesures et Essais » (SMT, *Standards, Measurements & Testing*). Ce projet a été une réussite, puisque des MRC pour différents microorganismes ont pu être certifiés conformément aux prescriptions générales du Guide ISO 35, soit les premiers MRC à être disponibles au niveau mondial dans notre domaine d'intérêt [In't Veld, 1998].

Ces matériaux de référence ont été **préparés** de la façon suivante :

- Une suspension bactérienne dans du lait a été séchée par atomisation afin d'obtenir une poudre de lait fortement contaminée. L'influence de ce traitement sur la survie des microorganismes ciblés a été caractérisée.
- La poudre de lait a été conservée un certain temps pour stabiliser la concentration en microorganismes.
- Cette poudre de lait a été utilisée comme poudre étalon pour préparer les matériaux de référence, par mélange avec de la poudre de lait stérile. Ce mélange est l'étape critique pour l'homogénéité du matériau. La poudre étalon est extrêmement stable (un lot préparé en 1978 était toujours utilisé en 1998) et le niveau de contamination du MR peut être ajusté en modifiant le ratio du mélange poudre étalon/poudre de lait stérile.
- La poudre mélangée a été finalement introduite par quantité d'environ 0,3 g dans des capsules de gélatine.

Deux **types** de MR peuvent être distingués en fonction de leur utilisation :

1. les MR quantitatifs, de fort niveau de contamination (10^3 à 10^4 ufc par capsule), pour évaluer les déterminations quantitatives ;
2. les MR qualitatifs, à bas niveau de contamination (environ 5 ufc par capsule), pour évaluer les déterminations qualitatives.

La procédure de vérification de l'**homogénéité des MR quantitatifs** se distingue des procédures retenues pour les essais d'aptitude par le protocole IUPAC/AOAC (§ 1.2.2.2) ou par l'ISO/DIS 13528 (§ 1.2.3.1). Deux tests ont été introduits, fondés sur la statistique de dispersion de Cochran :

1. Le premier test, appelé T_1 , caractérise la variation au sein d'une même capsule, soit la variation entre échantillons issus de la même capsule reconstituée. La statistique du test T_1 est la suivante :

$$T_1 = \sum_{i=1}^I \sum_{j=1}^J \frac{(z_{ij} - z_{i+} / J)^2}{z_{i+} / J} \quad \text{Eq. 16}$$

où :

- - z_{ij} est le comptage de l'échantillon j sur la capsule i ;
- - z_{i+} est la somme des comptages de tous les réplicats de la capsule i ;
- - J est le nombre de réplicats par capsule ;
- - I est le nombre total de capsules examinées.

T_1 suit approximativement une loi de χ^2 à $I(J-1)$ d.d.l (degrés de liberté), avec l'hypothèse que la variation entre les comptages issus de la même capsule suit une loi de Poisson.

2. Le second test, appelé T_2 , caractérise la variation entre les capsules, soit la variation entre les sommes de comptages d'échantillons répliqués par capsule. La statistique du test T_2 est la suivante :

$$T_2 = \sum_{i=1}^I \frac{(z_{i+} - z_{++} / I)^2}{z_{++} / I} \quad \text{Eq. 17}$$

où : z_{++} est la somme des comptages de tous les réplicats pour les I capsules.

T_2 est également distribué approximativement selon une loi de χ^2 à $(I-1)$ d.d.l. Quand la variation des comptages est conforme à la distribution de Poisson, $T_2/(I-1) = 1$. On considère qu'il y a une surdispersion lorsque $T_2/(I-1) > 2$.

L'homogénéité d'un lot de MR a été testée en même temps que la **stabilité**. Une valeur globale d'homogénéité T_{hom}/N du lot de MR a été calculée : $T_{\text{hom}}/N = \Sigma T_2 / (I \cdot J)$, où I est le nombre total de capsules examinées pour l'étude de stabilité et J le nombre de jours d'examen. On considère que l'homogénéité du lot est satisfaisante si la variation entre comptages est conforme à la distribution de Poisson, soit T_{hom} ne diffère pas significativement d'une distribution de χ^2 à N d.d.l.

Lorsqu'il s'est agi de tester l'**homogénéité de MR qualitatifs**, sachant que la capsule est entièrement utilisée pour l'analyse, seul le test T_2 a été utilisé, avec une formule adaptée de la statistique de test :

$$T_2 = \sum_{i=1}^I \frac{(z_i - z_+ / I)^2}{z_+ / I} \quad \text{Eq. 18}$$

où : z_i est le comptage de la capsule i et z_+ la somme des comptages des I capsules.

L'étude de **stabilité** pour les deux types de MR a consisté en deux étapes : la stabilité du MR conservé à -20°C a été étudiée pendant deux ans et celle du MR conservé à différentes températures a été étudiée sur un mois. Après transformation en logarithmes décimaux, une régression linéaire est effectuée et le test t est utilisé pour déterminer si les coefficients de régression diffèrent significativement de 0.

Les MR ont été certifiés avec au moins un essai inter-laboratoires. Pour *L.monocytogenes*, trois essais inter-laboratoires ont été conduits, en présence ou non de flore compétitrice.

Les trois MRC aujourd'hui disponibles sont les suivants [JRC/IRMM, 2004] :

1. *Salmonella typhimurium* (BCR-507R) : la valeur certifiée de ce MR qualitatif a été obtenue lors d'un essai qui a impliqué 11 laboratoires, ayant chacun analysé 50 capsules par la méthode normalisée de détection (ISO 6579) et par une technique de dénombrement.
Le certificat indique la valeur certifiée, ainsi que la fraction de capsules dans lesquelles *Salmonella* n'a pas pu être détectée (voir Tableau 10).
2. *Bacillus cereus* (BCR-528) : l'essai inter-laboratoires de certification de ce MR quantitatif a impliqué 12 participants, ayant chacun analysé 16 capsules.
Quatre valeurs de référence ont été certifiées, chacune attribuée à une méthode de dénombrement différente, en fonction de la gélose de dénombrement et du temps d'incubation (voir Tableau 10).
3. *L.monocytogenes* (BCR-595) : la valeur certifiée de ce MR qualitatif a été obtenue comme pour *Salmonella* par 11 laboratoires ayant chacun analysé en détection et en dénombrement 50 capsules, mais dans ce cas en l'absence de méthode normalisée de référence.
Ici aussi, le certificat indique la valeur certifiée et la fraction de capsules dans lesquelles *L.monocytogenes* n'a pas pu être détectée (voir Tableau 10).

Tableau 10 : Valeurs certifiées des matériaux de référence BCR en microbiologie des aliments

MRC	Propriété	Valeur(s) certifiée(s)	Intervalle de confiance à 95 %
BCR-507R (<i>Salmonella typhimurium</i>)	Nombre d'ufc/capsule	5,0	[4,5 ; 5,4]
	Fraction (%) de capsules dans lesquelles <i>Salmonella</i> n'a pas pu être détectée	1,1 (dénombrement) 1,6 (détection)	[0 ; 2,1] [0 ; 2,8]
BCR-528 (<i>B.cereus</i>)	Nombre d'ufc/capsule	53,4 (MYEP + 24 h incubation)	[51,7 ; 55,1]
		53,7 (MYEP + 48 h incubation)	[52,1 ; 55,4]
		55,0 (PEMBA + 24 h incubation)	[52,9 ; 57,3]
		55,8 (PEMBA + 48 h incubation)	[53,6 ; 58,0]
BCR-595 (<i>L.monocytogenes</i>)	Nombre d'ufc/capsule	7,2	[6,8 ; 7,6]
	Fraction (%) de capsules dans lesquelles <i>L.monocytogenes</i> n'a pas pu être détectée	0,075 (dénombrement) 1,2 (détection)	[0,050 ; 0,11] [0 ; 2,3]

Conclusion de la première partie

A l'issue de la revue comparative menée dans cette première partie, nous pouvons tenter de dégager des éléments de réponse aux différentes questions que nous nous étions posées initialement :

- **Le degré d'applicabilité des référentiels de portée générale au domaine de la microbiologie des aliments.** Pour les trois objectifs qui peuvent être attribués aux essais inter-laboratoires, il apparaît que les concepts et les méthodologies de ces référentiels sont, dans l'ensemble, applicables à notre domaine d'intérêt. Cependant, ils ne sont pas suffisants et nécessitent des compléments, principalement en raison de la nature vivante de l'objet de l'analyse, à savoir le microorganisme.
- **La compatibilité des référentiels spécifiques avec les référentiels généraux et leur justification technique.** Dans l'ensemble, les référentiels qui sont spécifiques soit à l'analyse des aliments, soit à l'analyse microbiologique des aliments, respectent les notions définies par les référentiels généraux, ainsi que les méthodologies prescrites. Quelques exceptions peuvent être néanmoins notées, comme l'algorithme robuste de l'EN ISO 16140 qui diffère de celui choisi pour l'ISO 5725-5. La plupart des exigences et pratiques spécifiques peuvent se justifier par la spécificité soit de l'analyse d'aliments, soit de l'analyse microbiologique (instabilité de l'analyte). Cependant, quelques dispositions introduites dans des documents sectoriels sont d'application générale et pourraient très bien figurer dans un référentiel de portée plus large. On peut citer à titre d'exemple les essais d'homogénéité introduits par le protocole IUPAC/AOAC pour les essais d'aptitude.
- **Le degré de correspondance entre référentiels.** Ici, une disparité entre référentiels de même domaine d'application peut être constatée. Cette disparité apparaît surtout pour les essais inter-laboratoires de validation de méthode, là où le nombre de référentiels est le plus élevé. On peut citer par exemple les différences entre protocoles IUPAC/AOAC et normes ISO :
 - différence entre le protocole IUPAC/AOAC général et l'ISO 5725, en matière de test des valeurs aberrantes ;
 - pour les méthodes microbiologiques, différences entre le protocole AOAC et l'EN ISO 16140, en termes de nombre de matrices utilisées, de mode de calcul des critères de fidélité, de choix et de calcul des critères de performance des méthodes de détection, et ce malgré les efforts indéniables de rapprochement.

2. Deuxième partie.
Mise en œuvre d'essais
inter-laboratoires en
microbiologie des
aliments

Introduction

Nous avons été amenés à organiser des essais inter-laboratoires en microbiologie des aliments, selon deux des trois objectifs pour lesquels on les applique classiquement :

1. **Estimer les performances métrologiques de méthodes qualitatives et quantitatives.** Cette étude a été conduite dans le cadre du projet SMT. Elle a permis la mise en œuvre pratique de la norme EN ISO 16140 avant même sa parution, alors qu'elle n'avait pas encore eu d'application expérimentale, notamment en ce qui concerne les estimateurs robustes de fidélité. Le même projet SMT a également introduit des critères de fidélité relatifs aux méthodes qualitatives ; or de tels critères n'étaient pas disponibles avant le lancement de ce projet européen. Ce dernier a par ailleurs permis d'obtenir pour la première fois des données de fidélité caractérisant les normes CEN et ISO relatives aux méthodes de référence pour l'analyse microbiologique des aliments.
2. **Évaluer la compétence d'un réseau de laboratoires.** Dans ce cas, des essais inter-laboratoires d'aptitude ont été conduits dans le cadre particulier de l'activité de notre laboratoire en tant que Laboratoire Communautaire de Référence ; cadre officiel se distinguant à plusieurs titres de celui des organisateurs privés d'essais d'aptitude.

Pour chacun de ces deux types d'essais, une fois décrite la méthodologie adoptée en termes de protocoles expérimentaux et d'exploitation des résultats, les données obtenues seront présentées. A la suite des résultats obtenus pour les essais de validation de méthodes, nous discuterons d'abord certains aspects relatifs à l'estimation de la performance de méthodes microbiologiques qualitatives et quantitatives. Sur la base conjointe des résultats obtenus et de l'analyse critique de l'existant en termes de référentiels et pratiques, nous envisagerons enfin les avantages et les limites des différents types d'essais inter-laboratoires en microbiologie.

2.1 Études européennes de validation de méthodes de référence

Résumé

Les essais inter-laboratoires pour la validation de méthodes ont été mis en œuvre dans le cadre d'un projet européen qui visait à valider les principales méthodes d'analyse microbiologique des aliments, normalisées par le CEN et l'ISO.

Les protocoles expérimentaux des cinq essais auxquels nous avons été associés sont présentés. De plus, la méthodologie d'exploitation des résultats des essais est décrite : celle portant sur les méthodes de dénombrement, pour lesquelles les statistiques robustes de la norme EN ISO 16140 ont été appliquées, et celle concernant les méthodes de détection, pour lesquelles le projet a introduit de nouveaux critères de fidélité.

Les résultats de chacun des essais sont résumés, ainsi que les recommandations qui ont été adressées au CEN et à l'ISO. La publication des critères de fidélité pour les méthodes qualitatives est introduite.

2.1.1 Protocoles expérimentaux

2.1.1.1 Matériel et méthodes

Les cinq essais inter-laboratoires, présentés ici, ont été organisés dans le cadre du projet européen que nous avons convenu de désigner par « projet SMT », décrit au § 1.1.11.5. Les essais de ce projet avaient pour objectif d'estimer les performances des méthodes normalisées suivantes :

- Cinq méthodes quantitatives :
 - dénombrement de *Bacillus cereus* (norme EN ISO 7932) ;
 - dénombrement de *Listeria monocytogenes* (norme EN ISO 11290-2) ;
 - deux méthodes de dénombrement de staphylocoques à coagulase positive (parties 1 & 2 de la norme EN ISO 6888) ;
 - dénombrement de *Clostridium perfringens* (normes ISO 7937) ;
- Deux méthodes qualitatives :
 - détection de *L.monocytogenes* (norme EN ISO 11290-1) ;
 - détection de *Salmonella* (norme EN ISO 6579).

L'organisation de chaque essai était placée sous la responsabilité de l'un des partenaires du projet :

- le CSL pour les deux essais sur *L.monocytogenes* ;
- le RIVM pour les deux essais sur *B.cereus* et sur *C.perfringens* ;
- l'AFSSA pour les deux essais sur les staphylocoques (deux méthodes) et sur *Salmonella*.

Tous ces essais inter-laboratoires ont été organisés et exploités selon les protocoles internationaux alors disponibles et applicables à la microbiologie, soit les lignes directrices IUPAC/AOAC (§ 1.1.7) et la norme EN ISO 16140 (§ 1.1.9), à l'état de projet pendant la conduite des essais. L'utilisation de protocoles harmonisés et acceptés au niveau international permet une large reconnaissance des résultats des essais. De plus, la comparaison entre résultats obtenus par différents essais inter-laboratoires est rendue possible, car quelque soit l'organisateur des essais, le même protocole a été utilisé.

L'organisation ainsi que le protocole expérimental de quatre de ces essais sont décrits en détail dans les publications et rapports suivants, figurant en annexe :

1. Annexe 3.1 : dénombrement de *L.monocytogenes* [Scotter *et al*, 2001b] ;
2. Annexe 3.2 : dénombrement de staphylocoques à coagulase positive [De Buyser *et al*, 2003] ;
3. Annexe 3.3 : détection de *L.monocytogenes* [Scotter *et al*, 2001a] ;
4. Annexe 3.4 : détection de *Salmonella* [Bohnert *et al*, 2001]. Une autre publication a résulté de cet essai, mais elle a été rédigée dans un objectif particulier ; à savoir la validation par AOAC de la méthode normalisée EN ISO 6579 [Feldsine *et al*, 2003].

Nous n'avons pas annexé la description de l'essai concernant le dénombrement de *C.perfringens*, car cet essai n'a pas donné lieu à une publication scientifique, et nous n'avons contribué qu'à l'exploitation statistique de cet essai.

Par ailleurs, nous n'avons pas été associés à l'essai sur le dénombrement de *B.cereus*, réalisé au début du projet SMT [Schulten *et al*, 2000].

2.1.1.2 Pré-essai

Selon les recommandations de la norme FIL 135B et du NMKL, un pré-essai a été effectué par les trois partenaires principaux, CSL, RIVM et AFSSA, pour vérifier si :

- le mode opératoire ne comportait pas d'ambiguïté,
- le rapport d'essai était complet,
- les différentes conditions opératoires avaient été prévues,
- les matériaux d'essai étaient satisfaisants, en termes d'homogénéité et de stabilité pendant le transport.

2.1.1.3 Mode opératoire et rapport d'essai

Bien que l'objectif du projet ait été de valider les méthodes déjà complètement normalisées par CEN et ISO, il s'est avéré cependant nécessaire de rédiger des modes opératoires conformes aux normes, mais :

1. résolvant les ambiguïtés,
2. précisant en particulier la façon de traiter les échantillons reçus par les participants, avant l'obtention de la suspension mère pour les dénombrements, ou du bouillon d'enrichissement pour les détections.

Par ailleurs, un rapport d'essai très détaillé a été préparé. Il prévoit :

1. toutes les conditions de mise en œuvre de l'analyse qui sont susceptibles d'influencer le résultat, conformément aux recommandations de EN ISO 16140 ;
2. les différents résultats « bruts » de l'analyse avant l'obtention du résultat final, tel que le nombre de colonies caractéristiques par boîte de Pétri, le nombre de colonies confirmées et le résultat de chaque test d'identification.

En particulier, les participants ont utilisé leurs propres milieux de culture et réactifs, certes conformes à la norme, mais issus de fabricants différents et se présentant sous des formes différentes : milieux reconstitués à partir des ingrédients de base, milieux déshydratés ou milieux prêts à l'emploi. Cette latitude correspond aux conditions réelles d'utilisation des méthodes normalisées et a permis d'estimer de façon réaliste la performance de ces méthodes.

2.1.1.4 Types de matrices

Suivant les pratiques du NMKL, on a décidé de valider des méthodes dites horizontales, c'est-à-dire applicables à l'ensemble des denrées alimentaires, aliments pour animaux et environnement, en incluant trois types de matrices :

- un produit laitier (fromage frais) ;
- un produit carné (viande hachée) ;
- un produit déshydraté (poudre d'œuf ou poudre de lait).

Ce choix permet, avec trois familles d'aliments essentielles en hygiène alimentaire, de couvrir une bonne partie de la portée de ces méthodes d'analyse microbiologique. De plus, une matrice fraîche (fromage frais) et deux matrices stables à température ambiante (viande hachée et poudre d'œuf/de lait) ont été choisies.

Par ailleurs, un matériau de référence a été utilisé pour identifier d'éventuels problèmes de performance des participants. Il s'agissait d'un matériau de référence certifié (MRC) BCR, préparé par le RIVM, tel que décrit au § 1.3.3.3. Ce matériau est une poudre de lait artificiellement contaminée par une suspension bactérienne lyophilisée. Comme nous l'avons souligné, c'est le seul matériau de référence à être actuellement certifié en microbiologie des aliments.

2.1.1.5 Contamination

Conformément à l'ensemble des référentiels, les matrices étaient contaminées artificiellement, y compris pour les méthodes de dénombrement portant sur des germes spécifiques. Le Tableau 11 présente les niveaux de contamination choisis.

Tableau 11 : Niveaux de contamination choisis pour le projet SMT

Type de méthode	Nombre de niveaux	Valeurs ciblées pour les niveaux
Qualitative	3	5-10 ufc/g, 50-100 ufc/g + témoin négatif
Quantitative	4	10 ² ufc/g, 10 ³ ufc/g, 10 ⁴ ufc/g + un témoin négatif

Une flore annexe était également présente dans les échantillons, de même que des germes pouvant mettre en évidence un défaut de spécificité de la méthode, tel que *L.innocua* pour *L.monocytogenes*.

2.1.1.6 Préparation des échantillons pour essai

Les échantillons ont été préparés par un laboratoire différent pour chaque type de matrice, tel que le Tableau 12 le présente.

Tableau 12 : Préparation des échantillons pour le projet SMT

Type de matrice	Laboratoire préparateur	Mode de préparation
Fromage frais	CECALAIT	Identique à celle adoptée pour les essais d'aptitude (§ 1.2.4.2)
Viande hachée	CSL	<ol style="list-style-type: none">1. Lyophilisation d'un lot de viande hachée, répartie ensuite dans des flacons et stérilisation par irradiation.2. Contamination individuelle des échantillons avec une suspension bactérienne du germe cible et d'une flore annexe reconstituée.3. 2^{ème} lyophilisation des flacons, afin de stabiliser la flore.
Poudre d'œuf	RIVM	Préparation extemporanée de chaque échantillon par le participant, par mélange de deux flacons : <ul style="list-style-type: none">- l'un contenant la poudre d'œuf entier avec une flore annexe,- l'autre contenant deux capsules du germe cible suspendu dans de la poudre de lait (capsules préparées de façon similaire aux MRC du BCR).

Les modes de préparation sont nettement différents d'un matériau à l'autre car chaque préparateur a suivi ses propres pratiques, en particulier pour le mode de contamination qui était soit global pour tout le lot, soit individuel et effectué par le préparateur ou par les participants. Cette diversité ne semble pas préjudiciable car, d'une part, elle est adaptée à la nature de chacune des matrices, et d'autre part, elle garantit une meilleure qualité d'échantillons préparés selon une technique bien maîtrisée et qui a fait ses preuves.

2.1.1.7 Stabilité et homogénéité des matériaux d'essai

Là aussi, les modalités ont varié en fonction des pratiques de chaque préparateur, comme le montre le Tableau 13.

Tableau 13. Procédures de contrôle de la qualité des matériaux soumis à l'essai dans le projet SMT

Matrice	Critère	Procédure
Fromage frais	Homogénéité	Analyse en double de 10 échantillons d'un même lot, 3 jours après sa préparation.
	Stabilité	Analyse en double et à 5 jours différents de 3 échantillons d'un lot de matériau conservé à 4°C pendant 7 jours.
	Interprétation	Calcul du χ^2 sur le dénombrement total par échantillon, en unité logarithmique, pour comparer la dispersion de la contamination à celle d'une série de Poisson.
Viande hachée	Homogénéité	Analyse en double de 10 échantillons de chaque lot, après production et le jour du début des analyses par les participants.
	Stabilité	Analyse en double et à 5 jours différents de 10 échantillons d'un lot de viande hachée conservé à 4°C pendant un mois.
	Interprétation	Test F sur les comptages transformés en log : stabilité et homogénéité considérées comme satisfaisantes si le test F ($\alpha=0,05$ %) n'est pas significatif.
Poudre d'œuf/ Capsules	Homogénéité	Analyse en double de 25 capsules de chaque lot.
	Stabilité	Analyse en double de 10 capsules par semaine de chacun des lots conservés à -20 °C pendant 10 semaines.
	Interprétation	Caractérisation de la variation entre doubles, à l'intérieur d'une capsule (test T_1) et entre capsules (test T_2) : tests statistiques identiques à ceux utilisés pour les MRC (§ 1.3.3.3).

Ainsi, les protocoles IUPAC/AOAC et ISO/DIS 13528 (§ 1.2.2.2, § 1.2.3.1) ont-ils été respectés pour le plan d'expérience des essais d'homogénéité. Le CSL a également appliqué le traitement statistique préconisé, alors que CECALAIT a adapté le traitement statistique au cas de dénombrements bactériens et le RIVM a utilisé ses propres tests statistiques.

2.1.1.8 Nombre d'échantillons et envoi

Le Tableau 14 précise le nombre d'échantillons et de répétitions en fonction du type de méthodes.

Tableau 14. Nombre d'échantillons et de répétitions en fonction du type de méthodes, dans le projet SMT

Méthodes	Répétitions par niveau	Nombre de niveaux	Échantillons par type de matrice
Qualitatives	5	3	15
Quantitatives	2	4	8

Par ailleurs, 5 matériaux de référence ont été envoyés à chaque participant.

Pour les méthodes qualitatives, ce protocole est donc allégé par rapport à l'EN ISO 16140 qui requiert huit répétitions par niveau, mais il est proche de AOAC qui en demande six.

Les échantillons ont été codés de telle façon que les participants ne puissent pas reconnaître les répétitions et échanger leurs résultats. Les échantillons de viande hachée et de poudre d'œuf ont été distribués ensemble, alors que ceux de fromage frais ont été directement distribués par CECALAIT en raison de leur stabilité limitée. Ils ont tous été transportés à température de réfrigération (4-8 °C), maintenue par des emballages isolants et des éléments réfrigérants. Chaque paquet comportait un dispositif de suivi de la température. Les paquets ont été acheminés par transporteur rapide, généralement en 24 à 48 h, en Europe.

La même date de début des analyses a été fixée pour l'ensemble des participants. Pour le fromage frais, elle a été fixée dès le 3^{ème} jour suivant l'envoi des échantillons, en raison de la courte durée de vie de cette matrice.

2.1.1.9 Participants

Entre 17 et 24 laboratoires, provenant de 16 pays européens, ont participé aux différents essais inter-laboratoires. La liste des participants à chaque essai figure dans les annexes 0 à 3.4. Dans le cas de *Salmonella*, l'essai a de plus inclus 10 laboratoires américains.

Les coordonnateurs ont sélectionné les laboratoires participants, sur la base de leur compétence à effectuer des analyses microbiologiques des aliments. Ils devaient ainsi tous être accrédités, ou en voie de l'être. La plupart d'entre eux avaient déjà participé aux essais inter-laboratoires de certification des matériaux de référence, organisés par le RIVM (voir § 1.3.3.3). Nous sommes donc ici dans un cas un peu particulier d'essai inter-laboratoires de validation de méthode. Il correspond exactement à la conception de AOAC telle que décrite au § 1.1.7.2. Il est quelque peu plus restrictif que la conception de l'ISO 5725, qui recommande de choisir les participants au hasard parmi la population des laboratoires susceptibles d'utiliser la méthode. Les partenaires principaux ont ainsi voulu réduire le risque d'invalider un essai inter-laboratoires à cause d'un nombre trop important de participants qui auraient rendu des résultats inexploitable.

2.1.2 Méthodologie d'exploitation des résultats pour les méthodes de dénombrement

La norme ISO 5725 caractérise la fidélité d'une méthode par deux critères, la répétabilité et la reproductibilité, correspondant respectivement aux conditions minimales et maximales de variation. Selon l'ISO 5725 Partie 1, la méthode statistique d'estimation de ces deux critères s'applique à des méthodes quantitatives, produisant des mesures sur une échelle continue, mais elle peut s'appliquer également à des méthodes produisant des mesures discrètes. La méthodologie développée dans l'ISO 5725 Partie 2 est donc bien applicable aux méthodes de dénombrement bactérien.

A l'occasion de la présentation de l'ISO 5725 Partie 5, nous avons déjà souligné l'intérêt des statistiques robustes pour estimer les écart-types de répétabilité et de reproductibilité, à la place des statistiques fondées sur la moyenne et utilisées dans le modèle de base de ISO 5725 Partie 2. Cet intérêt réside principalement dans la moindre sensibilité de ces statistiques aux valeurs dites aberrantes, permettant ainsi d'éviter d'avoir recours à l'exclusion de résultats par des tests statistiques.

Lors de la conception de l'essai inter-laboratoires sur le dénombrement de *L. monocytogenes*, en 1998, nous avons suggéré aux partenaires principaux de mettre en œuvre les estimateurs robustes tout récemment introduits dans le projet de norme EN ISO 16140, qui ne sera publiée que cinq ans plus tard. Ces estimateurs ont ainsi été utilisés, en parallèle de la démarche classique, pour cet essai et pour les deux autres essais de dénombrement qui suivirent.

Transformation initiale en logarithmes décimaux

Quelle que soit la méthode suivie, les résultats de dénombrement obtenus par les participants ont d'abord été transformés en logarithmes décimaux, selon la procédure généralement admise pour les dénombrements bactériens [Jarvis, 1989]. Cette transformation est justifiée par la normalité qu'on considère qu'elle introduit pour la distribution des dénombrements. Or cette normalité est une condition nécessaire pour pouvoir utiliser en particulier les tests statistiques d'identification des valeurs aberrantes. Cette transformation est préconisée par les référentiels s'appliquant à la microbiologie, soit l'EN ISO 16140 et le protocole AOAC. Elle est utilisée également pour exploiter les essais d'aptitude, entre autres par le RAEMA et CECALAIT.

Approche classique

Le protocole de l'ISO 5725-2 a été suivi pour deux de ses étapes : l'élimination de valeurs aberrantes selon des tests statistiques, puis le calcul des écart-types de répétabilité et de reproductibilité. Les calculs ont été effectués avec le logiciel AIL 5725 (sous DOS) et la macro Excel EasyVal™. Le calcul des *écarts-types de répétabilité et de reproductibilité* est également détaillé en annexe 3.5, paragraphe 3.5.1.

Approche robuste de l'EN ISO 16140

L'approche suivie par le projet SMT a été celle du projet de norme EN ISO 16140, laquelle approche n'a pas subi de modification dans la norme publiée. Nous avons effectué les calculs en développant des feuilles de calcul sous Excel. Un estimateur robuste de *l'écart-type de répétabilité* est déduit de la médiane des écarts-types intra-laboratoires. Pour déduire *l'écart-type de reproductibilité*, un estimateur robuste de l'écart-type des moyennes de répétitions est

déduit de l'estimateur d'échelle de Rousseeuw S_n , qui consiste en deux médianes successives. Les calculs sont détaillés en annexe 3.5, paragraphe 3.5.2.

2.1.3 Méthodologie d'exploitation des résultats pour les méthodes de détection

Nous avons vu au § 1.1.9.3.1, à propos de la norme EN ISO 16140, que les critères habituellement retenus pour exploiter un essai inter-laboratoires se fondent sur la justesse de la méthode ou son erreur systématique, et non pas sur sa fidélité ou son erreur aléatoire. Or il s'agit de deux aspects complémentaires de la variabilité totale d'une méthode d'analyse et l'estimation de l'un ne peut pas remplacer l'estimation de l'autre.

Selon l'ISO 5725-1, l'estimation de la fidélité par les écarts-types de répétabilité et de reproductibilité ne peut pas être appliquée à des méthodes qualitatives. Aucun critère n'était alors disponible pour caractériser la fidélité de méthodes qualitatives par essais inter-laboratoires. C'est pourquoi un des objectifs importants que s'était fixé le projet SMT était de développer de tels critères, qui puissent être considérés comme des équivalents de la répétabilité et de la reproductibilité. Les trois biostatisticiens réunis au sein de ce projet ont pu atteindre cet objectif de façon à pouvoir appliquer les nouveaux paramètres dès l'exploitation du premier essai qualitatif, organisé en 1998 sur la détection de *L. monocytogenes*, soit deux ans après le lancement du projet. Pour les deux essais qualitatifs, les critères habituels de justesse ont été calculés ainsi que les nouveaux critères de fidélité.

Justesse. Notamment selon l'EN ISO 16140, deux critères sont calculés par combinaison <matrice x niveau de contamination> pour caractériser la justesse de la méthode :

- pour les échantillons négatifs ou témoins, c'est la spécificité qui représente le pourcentage d'échantillons qui sont correctement analysés comme négatifs par les participants ;
- pour les échantillons à bas et haut niveau de contamination, c'est la sensibilité qui est le pourcentage d'échantillons qui sont correctement analysés comme positifs par les participants.

Fidélité. Les deux critères ont été appelés degré d'accord (*accordance* en anglais) et concordance. Globalement, l'estimateur statistique de dispersion qu'est l'écart-type a été remplacé par une probabilité d'obtenir des résultats répétés identiques. La définition et le calcul de ces critères ont été repris en Annexe L de l'EN ISO 16140.

Degré d'accord. Le degré d'accord est la probabilité de trouver le même résultat (c'est-à-dire tous les deux positifs ou tous les deux négatifs) pour deux prises d'essai identiques analysées par le même laboratoire, dans des conditions de répétabilité. Il s'agit donc d'un équivalent de la répétabilité. Pour déduire le degré d'accord des résultats d'un essai inter-laboratoires, la probabilité que deux échantillons identiques donnent le même résultat est calculée pour chaque participant à son tour, et cette probabilité est ensuite moyennée sur l'ensemble des laboratoires.

Le Tableau 15 indique le calcul du degré d'accord.

Tableau 15 : Calcul du degré d'accord selon l'EN ISO 16140 (avec p le nombre de laboratoires participants, n le nombre de répétitions, et k_i le nombre de résultats positifs trouvés par le laboratoire i)

Etape de calcul	Formule
Probabilité pour le laboratoire i de trouver le résultat d'une répétition positif	k_i/n
Probabilité pour le laboratoire i de trouver le résultat de deux répétitions positives	$(k_i/n)^2$
Probabilité pour le laboratoire i de trouver deux répétitions négatives	$((n - k_i)/n)^2$
Probabilité que le laboratoire i obtienne deux résultats identiques ; soit le degré d'accord pour le laboratoire i , ACC_i	$ACC_i = (k_i/n)^2 + ((n - k_i)/n)^2$
Degré d'accord ACC	$ACC = \frac{1}{n} \sum_{i=1}^p ACC_i$

Concordance. La concordance est la probabilité de trouver le même résultat pour deux échantillons identiques, analysés dans deux laboratoires différents. C'est un équivalent de la reproductibilité. Pour calculer la concordance à partir d'un essai inter-laboratoires, le résultat de chaque répétition dans chaque laboratoire est pris l'un après l'autre et est apparié avec les résultats identiques de tous les autres laboratoires. La concordance est la proportion du nombre de paires donnant les mêmes résultats sur le nombre total de paires possibles.

Le calcul de la concordance est indiqué dans le Tableau 16.

Tableau 16 : Calcul de la concordance selon l'EN ISO 16140 (avec les notations utilisées pour le degré d'accord)

Etape de calcul	Formule
Nombre total de résultats positifs (N_+)	$N_+ = \sum_{i=1}^p k_i$
Nombre total de résultats négatifs (N_-)	$N_- = \sum_{i=1}^p (n - k_i)$
Nombre de possibilités d'apparier un résultat positif du laboratoire i avec un résultat positif d'un autre laboratoire	$k_i(N_+ - k_i)$
Concordance (CON)	$CON = \frac{1}{p} \sum_i \frac{(n - k_i)(N_- - (n - k_i)) + k_i(N_+ - k_i)}{n(N_+ + N_- - n)}$

2.1.4 Résultats des études menées

2.1.4.1 Organisation générale

Pour chaque essai inter-laboratoires, les résultats bruts des participants, exprimés sous forme de présence/absence du germe cible ou sous forme de nombre d'ufc/g ou ml, ont été reportés dans un rapport final, rédigé par le partenaire « leader » pour cet essai, pour le compte de la Commission Européenne (l'extrait d'un tel rapport figure en annexe 3.4). Ces rapports relatent également les conditions de réalisation de ces essais, les estimations de performance de méthode et les conclusions qui ont pu être tirées.

A l'issue de chaque essai inter-laboratoires, une réunion a été organisée avec l'ensemble des participants à l'essai, pour discuter avec eux des conditions de réalisation de l'essai et des résultats obtenus. En particulier ont été discutés l'élimination des calculs de certains résultats, ainsi que les recommandations qu'il convenait d'adresser au CEN et à ISO sur la méthode normalisée qui avait été étudiée et sur les critères de performance qui pouvait la caractériser. Les publications qui ont résulté de quatre de ces essais reprennent ces éléments de façon synthétique, et notamment les valeurs de fidélité et de justesse établies pour ces méthodes. Trois publications figurent dans les annexes 3.1, 3.2, 3.3.

Nous allons présenter brièvement quelques éléments généraux sur les conditions de réalisation de ces essais, avant d'évoquer, pour chacun d'eux, les conclusions qui ont pu être tirées.

Globalement pour les 5 essais, les études sur les matériaux d'essai ont montré une homogénéité et une stabilité suffisantes puisque les critères de performance de l'essai inter-laboratoires n'en n'ont pas été affectés. A titre d'exemple, pour le dénombrement de *L. monocytogenes* dans le lot de viande hachée au niveau moyen de contamination, bien qu'on ait conclu à une hétérogénéité, les valeurs de fidélité obtenues n'ont pas été statistiquement plus élevées qu'aux autres niveaux. Les échantillons de fromage frais, les moins stables, ont été généralement reçus dans les temps prévus, entre 24 et 48 h, dans de bonnes conditions de température. Un nombre limité de laboratoires les a reçus plus tardivement, mais l'état des échantillons a permis de prendre en compte les résultats, dans la mesure où la stabilité du matériau avait été vérifiée jusqu'à 7 jours.

En ce qui concerne l'élimination de résultats du calcul des critères de performance, il s'agit là d'un point particulièrement délicat. Pour les essais sur les méthodes tant qualitatives que quantitatives, l'élimination des résultats s'est faite sur la base de raisons techniques dûment identifiées avec les participants qui avaient fourni ces résultats. Même dans le cadre de l'approche « classique » pour l'estimation de la fidélité de méthodes de dénombrement, des valeurs identifiées comme aberrantes par des tests statistiques n'ont pas été exclues en l'absence de raison connue. En effet, il a été considéré que de telles valeurs pouvaient refléter la (grande) variation naturelle des dénombrements bactériens, et éliminer ces valeurs aurait pu conduire à sous-estimer la valeur vraie de la fidélité de la méthode. La prise en compte des résultats de tests statistiques n'est donc pas entièrement conforme à celle recommandée par l'ISO 5725-2. Cette dernière recommande en effet d'éliminer une valeur aberrante même si aucune explication n'a été trouvée.

Voici les principales raisons retenues pour éliminer des résultats :

- les échantillons ont subi une température excessive pendant le transport, ou ils ont atteint le laboratoire en mauvais état (exemple : fuite) ;
- le mode opératoire n'a clairement pas été respecté ;
- la performance du laboratoire, notamment son degré de maîtrise de la méthode, est nettement mise en cause par :
 - un nombre élevé et non aléatoire de faux positifs ou de faux négatifs pour un essai qualitatif. Ainsi pour *Salmonella*, des règles spécifiques ont été définies, notamment en ce qui concerne la poudre d'œuf utilisée pour préparer les échantillons qui, n'étant pas stérile, aurait pu contenir *Salmonella* (voir Annexe 4).
 - des résultats de comptages beaucoup plus bas que ceux attendus, ou des doubles hautement dispersés.

Dans certains cas, des études complémentaires ont été menées pour établir si tel écart pouvait avoir une influence sur le résultat de l'analyse. En l'absence d'influence observable, le résultat a été conservé dans les calculs.

En définitive, le nombre de résultats éliminés par combinaison <matrice x niveau de contamination> est toujours resté faible et le nombre de résultats pris en compte a été suffisant dans tous les cas pour estimer de façon satisfaisante les critères de fidélité ou de justesse. Les minima fixés par les référentiels (EN ISO 16140 et AOAC) sont nettement respectés. Ainsi, pour les méthodes qualitatives, les résultats d'au moins 17 laboratoires ont été pris en compte (minimum exigé 10) et pour les méthodes quantitatives, les résultats d'au moins 17 laboratoires, sauf 13 pour *C. perfringens*, ont été pris en compte (minimum exigé 8).

2.1.4.2 Dénombrement de *L. monocytogenes*

Les estimations robustes et non robustes des écarts-types de répétabilité et de reproductibilité ont conduit à des valeurs très similaires pour chacune des combinaisons. Il a donc été décidé de ne retenir que les valeurs issues de l'estimation robuste, conformément aux recommandations du projet de norme EN ISO 16140 d'alors.

Le projet SMT a décidé d'exprimer des valeurs générales de répétabilité et de reproductibilité, soit des moyennes des valeurs par matrice, pour caractériser globalement la méthode. Il s'agit là certainement d'une vision synthétique, mais réductrice, car ces valeurs sont alors considérées comme valables pour l'analyse de l'ensemble des denrées alimentaires et pour des niveaux de contamination très différents. Les valeurs générales figurent dans le Tableau 17.

Tableau 17 : Valeurs générales de fidélité exprimées en \log_{10} ufc/g, pour le dénombrement de *L.monocytogenes*

Critère de fidélité	Valeurs
Répétabilité <i>r</i>	0,58
Reproductibilité <i>R</i>	0,81

Ces valeurs sont particulièrement élevées et elles ont pu être expliquées, comme pour l'essai sur la détection de *L. monocytogenes*, par une faiblesse de la méthode normalisée. La gélose de dénombrement (PALCAM) n'étant pas spécifique de *L. monocytogenes*, l'étape de

confirmation de 5 colonies choisies au hasard parmi celles formées sur la boîte de Pétri introduit une grande variabilité dans la méthode. Les conditions expérimentales avaient été choisies de façon à mettre en avant cette faiblesse, puisque les échantillons ont été ensemencés avec des niveaux équivalents de *L. monocytogenes* et de *L. innocua*.

Le projet SMT a recommandé au comité ISO/TC 34/SC 9 :

- de caractériser la fidélité de la méthode normalisée par les valeurs obtenues lors de l'essai inter-laboratoires ;
- d'engager en parallèle une révision de la norme pour améliorer la spécificité de la gélose de dénombrement.

L'ISO/TC 34/SC 9 a suivi ces recommandations en décidant d'inclure les valeurs de fidélité dans un amendement à la norme EN ISO 11290-2 : même mauvaises, ces valeurs renseignent l'utilisateur sur les limites de la norme. Les valeurs générales de fidélité ont été incluses dans le corps principal du texte de la norme, alors que les valeurs par combinaison <matrice x niveau de contamination> ont été introduites en annexe. Ces dernières figurent dans le Tableau 1 de l'annexe 3.1. De plus, des travaux expérimentaux ont été lancés par le comité ISO sur la gélose, et ils ont aujourd'hui aboutis.

Enfin, notre laboratoire a mené des travaux visant à améliorer la sensibilité de la méthode de référence normalisée, en introduisant une étape de concentration de la suspension mère par filtration [Gnanou Besse *et al*, 2004].

2.1.4.3 Dénombrement des staphylocoques à coagulase positive

Comme pour l'essai précédent, les estimations robustes et non robustes des écarts-types de fidélité sont similaires, ainsi que le montrent les Tableaux 1 et 2 de l'annexe 3.2.

La norme EN ISO 6888, relative au dénombrement des staphylocoques, comporte deux parties : la première utilise la gélose Baird-Parker (BP), et la deuxième utilise le milieu gélosé au plasma de lapin et au fibrinogène (RPFA). De fait, l'essai inter-laboratoires a porté sur les deux méthodes, qui ont pu être comparées. Une différence statistiquement significative en termes de justesse (comparaison des moyennes des deux méthodes par le test *F*) a été trouvée pour 5 combinaisons <matrice x niveau> sur 9, mais l'amplitude de ces différences était limitée pour des dénombrements bactériens. Aucune différence significative entre les deux méthodes n'est apparue en termes de fidélité, sauf pour les deux combinaisons <fromage frais x niveau bas> et <viande x niveau moyen>. Pour ces dernières, comme pour 5 autres combinaisons, la méthode RPFA a une meilleure fidélité que la méthode BP : cette tendance se comprend parfaitement car l'étape de confirmation des colonies, propre à la méthode BP, est une source importante de variabilité.

Le Tableau 18 présente les valeurs générales de répétabilité et de reproductibilité pour chacune des deux parties de l'EN ISO 6888.

Tableau 18 : Valeurs générales de fidélité exprimées en \log_{10} ufc/g, pour le dénombrement des staphylocoques à coagulase positive

Méthode	Répétabilité <i>r</i>	Reproductibilité <i>R</i>
Méthode BP	0,28	0,43
Méthode RPFA	0,22	0,33

Ces valeurs sont acceptables pour des méthodes de dénombrement bactérien, en particulier pour la méthode BP qui comporte une étape de confirmation.

Ainsi, le projet SMT a pu formuler un certain nombre de recommandations au comité ISO/TC 34/SC 9 :

- caractériser la fidélité des deux méthodes normalisées par les valeurs issues de l'essai inter-laboratoires ;
- utiliser les statistiques robustes du projet de norme EN ISO 16140 pour estimer les critères de fidélité de méthodes de dénombrement ;
- modifier la norme pour les aspects suivants :
 - sur la base de la comparaison expérimentale, donner un statut équivalent aux deux parties de la norme, alors que jusqu'alors, une priorité d'utilisation était donnée à la 1^{ère} partie ;
 - améliorer la description de la méthode normalisée, soit en 1^{ère} partie, la description des colonies typiques et atypiques.

L'ISO/TC 34/SC 9 a suivi toutes ces recommandations et aujourd'hui les amendements aux parties 1 & 2 de l'EN ISO 6888 sont publiés.

2.1.4.4 Dénombrement de *C. perfringens*

Le rapport final de cet essai [Schulten *et al*, 2001] détaille les conditions de réalisation de l'essai, ainsi que les résultats obtenus. A nouveau, les calculs robustes et non robustes des écarts-types de répétabilité et de reproductibilité ont donné des valeurs similaires, conduisant encore à retenir les estimations robustes de la fidélité.

Cet essai inter-laboratoires a été l'occasion de comparer les deux normes portant sur ce microorganisme : la norme internationale ISO 7937 (version de 1997) et la norme européenne EN 13401 (version de 1999). Ces deux normes différaient uniquement par l'étape de confirmation, comme l'indique le Tableau 19.

Tableau 19 : Comparaison des normes ISO 7937 : 1987 et EN 13401 : 1999 pour le dénombrement de *C.perfringens*

Norme	Étape de confirmation des colonies
ISO 7937	Milieu Lactose Sulfite (LS)
EN 13401	Milieu LS ou combinaison des milieux Nitrate-Mobilité (MN) et Lactose-Gélatine (LG)

Les deux techniques de confirmation n'ont pas montré de différence significative en terme de justesse (test *F* sur le nombre de colonies confirmées par les deux techniques), et les valeurs de fidélité étaient également très proches.

A la différence des deux essais précédents, le projet SMT n'a pas jugé possible de définir des valeurs de répétabilité et de reproductibilité caractérisant la méthode dans son ensemble, mais s'en est tenu à exprimer les valeurs obtenues par type de matrice. Les valeurs du Tableau 20 montrent que la méthode est caractérisée par une bonne répétabilité, mais par une reproductibilité moins satisfaisante. Une reproductibilité particulièrement médiocre a été obtenue pour la poudre d'œuf à faible taux de contamination, dont l'étude d'homogénéité avait été pourtant satisfaisante, bien qu'à la limite de l'acceptabilité.

Tableau 20. Valeurs de fidélité exprimées en log₁₀ ufc/g, pour le dénombrement de *C. perfringens* avec deux techniques de confirmation (LS / MN+LG)

Type de matrice	Niveau moyen	Répétabilité <i>r</i>	Reproductibilité <i>R</i>
Fromage	3,53	0,24 / 0,25	0,26 / 0,31
Viande	3,62	0,26 / 0,29	0,55 / 0,52
Poudre d'œuf	3,74	0,19 / 0,14	0,65 / 0,72

Ainsi, le projet SMT a pu recommander au comité ISO/TC 34/SC 9 :

- de caractériser la fidélité de la méthode normalisée par les valeurs obtenues lors de l'essai inter-laboratoires ;
- de laisser le choix dans l'ISO 7937 entre les deux techniques de confirmation qui ont montré des performances équivalentes, soit d'harmoniser cette norme avec la norme européenne équivalente ;
- d'améliorer la description de la méthode normalisée en différents points :
 - définir le temps maximal de conservation d'une des géloses (TSC),
 - incuber le milieu LS seulement dans des tubes et non dans des bouteilles,
 - préciser les conditions d'incubation, en aérobiose ou en anaérobiose, du milieu LS.

L'ISO/TC 34/SC 9 a suivi ces recommandations et la norme ISO 7937 est en phase finale de révision dans ce sens, devenant une norme commune avec le CEN.

2.1.4.5 Détection de *L. monocytogenes*

Cet essai inter-laboratoires a été l'occasion de mettre en œuvre pour la première fois les critères de fidélité d'une méthode qualitative.

Outre l'estimation de paramètres de performance globale de la méthode, cet essai a également permis d'investiguer de façon quantifiée l'efficacité des différentes étapes de cette méthode, qui est complexe comme la plupart des méthodes de détection. Elle comporte effectivement deux phases d'enrichissement et, à l'issue de chacun de ces deux enrichissements, une étape d'isolement sur deux géloses, suivie d'une confirmation biochimique. Les éléments suivants ont pu ainsi être mis en évidence :

- 95,5 % des échantillons à fort taux de contamination ont été trouvés positifs à l'issue du premier enrichissement, mais seulement 30 % pour les échantillons à faible niveau de contamination. L'intérêt d'un double enrichissement a donc été confirmé, notamment pour détecter *L. monocytogenes* en faibles nombres. Cet intérêt était jusqu'alors contesté par certains, notamment pour des raisons de lourdeur de l'analyse.
- Pour ce même pourcentage de positifs à l'issue du premier enrichissement, une différence assez nette, compte tenu de l'interaction forte entre laboratoires et matrices, a été observée entre les trois types de matrice : le pourcentage est plus élevé pour le fromage frais, dont la flore était effectivement moins stressée que dans les autres matrices.
- En ce qui concerne l'étape d'isolement, aucune différence significative n'a été trouvée sur la plupart des données entre les deux géloses d'isolement.
- Les résultats de confirmation ont montré que l'enrichissement en deux étapes tel que normalisé permet à *L. innocua*, présent initialement en quantité équivalente à *L. monocytogenes*, de devenir prédominant. Ainsi, sur des géloses telles que Oxford et PALCAM ne permettant pas de distinguer ces deux espèces, le risque est grand, et réel

dans cet essai, de ne pas conclure à la présence de *L. monocytogenes* en confirmant seulement 5 colonies par boîte.

Les valeurs caractérisant la performance de la méthode, en termes de justesse et de fidélité, sont présentées dans le Tableau 21.

Tableau 21 : Valeurs de justesse et de fidélité, exprimées en pourcentages, pour la détection de *L. monocytogenes*

Valeurs	Niveau	Justesse		Fidélité	
		Spécificité	Sensibilité	Degré d'accord	Concordance
Valeurs globales		97,4	85,6	88,2	87,0
Fromage	Négatif	96,5	-	92,9	93,1
	Bas	-	83,3	80,0	75,2
	Haut	-	100	100	100
Viande	Négatif	97,8	-	96,7	95,6
	Bas	-	85,9	76,7	71,7
	Haut	-	100	100	100
Poudre d'œuf	Négatif	97,9	-	95,8	95,8
	Bas	-	53,7	48,4	49,8
	Haut	-	88,4	83,2	79,1

Nous avons introduit dans ce tableau les valeurs par combinaison <matrice x niveau de contamination>, pour illustrer le fait que dans le cas de cette étude, les valeurs globales masquent la variation importante des critères de performance (justesse comme fidélité) entre différentes matrices, et entre différents niveaux de contamination. En particulier, la matrice poudre d'œuf se distingue des deux autres matrices par des valeurs particulièrement basses, en termes de justesse et de fidélité, hors témoins négatifs. Aucune explication claire n'a pu être trouvée, l'homogénéité et la stabilité de ce matériau ayant été satisfaisantes. Par ailleurs, une relation directe entre la sensibilité et les paramètres de fidélité peut être observée ici.

Le projet SMT a pu recommander au comité ISO/TC 34/SC 9 :

- de caractériser la justesse et la fidélité de la méthode normalisée par les valeurs issues de l'essai inter-laboratoires ;
- d'apporter les précisions et améliorations suivantes à la norme :
 - corriger la formulation du test à l'hémolyse en puits ;
 - laisser la possibilité d'effectuer la confirmation par des méthodes moléculaires, en alternative aux essais biochimiques ;
 - comme pour la méthode de dénombrement, lancer une révision de la norme pour améliorer l'étape d'isolement.

La structure ISO a suivi le projet SMT et très prochainement, un amendement à la norme sera publié incluant les données de performance de la méthode, en précisant que ces données caractérisent la méthode non modifiée. Cet amendement prévoira pour l'étape d'isolement deux géloses : une gélose spécifique (ALOA) et une 2^{ème} gélose au choix.

2.1.4.6 Détection de *Salmonella*

Cet essai a revêtu une importance particulière, tant en matière de charge de travail que de résultats, car il a été organisé de façon à pouvoir permettre une reconnaissance des résultats par AOAC, soit une validation par cet organisme américain de la méthode normalisée EN ISO 6579. Ainsi, 10 laboratoires américains ont pris part à l'essai et il a fallu pour cela réussir à acheminer dans des délais courts des échantillons contaminés vers les USA : les échantillons de fromage frais ont pu atteindre les laboratoires américains en 5 jours. Cet acheminement a représenté un défi, et un tel essai inter-laboratoires associant laboratoires américains et européens a constitué une première expérience de ce type dans le domaine de la microbiologie des aliments. Par ailleurs, le protocole de l'essai a dû être complété, puisqu'une méthode ne peut être validée par AOAC que si l'essai inter-laboratoires porte en parallèle sur cette méthode et la méthode « officielle » AOAC.

Les valeurs caractérisant la performance (justesse et fidélité) de la méthode normalisée sont rassemblées dans le Tableau 22.

Tableau 22 : Valeurs de justesse et de fidélité, exprimées en pourcentages, pour la détection de *Salmonella*

Valeurs	Niveau	Justesse		Fidélité	
		Spécificité	Sensibilité	Degré d'accord	Concordance
Valeurs globales		100	92	96,7	91,4
Fromage	Négatif	100	-	100	100
	Bas	-	74,3	83,8	60,5
	Haut	-	83,8	95,2	71,7
Viande	Négatif	100	-	100	100
	Bas	-	98,0	96,9	96,0
	Haut	-	100	100	100
Poudre d'œuf	Négatif	100	-	100	100
	Bas	-	98,1	96,2	96,2
	Haut	-	99	98,1	98,1

Comme pour l'essai précédent, les valeurs globales pour la méthode, satisfaisantes, masquent une dispersion des valeurs par matrice et par niveau. Ici, hormis les témoins négatifs, la matrice fromage frais se distingue clairement des deux autres par ses faibles valeurs de justesse (jusqu'à plus de 20 % d'écart pour la sensibilité au niveau bas) et de fidélité (jusqu'à 36 % d'écart pour la concordance au niveau bas). Ce manque de sensibilité et cette mauvaise concordance pour la matrice fromage frais n'ont pas pu être expliqués, les études de stabilité et d'homogénéité du matériau ayant été satisfaisantes. Il est probable que ces résultats sont dûs à la nature de la matrice, qui est à l'état frais et qui comporte une flore annexe « naturelle » importante. L'analyse microbiologique est alors rendue particulièrement « difficile », par rapport aux deux autres types de matrice.

A la suite de cet essai, le projet SMT a pu formuler les recommandations suivantes à l'ISO/TC 34/SC 9 :

- caractériser la justesse et la fidélité de la méthode normalisée, en phase finale de révision ;
- envisager les précisions et améliorations suivantes de la norme :
 - reconsidérer le choix du 2^{ème} bouillon d'enrichissement, difficile à préparer et à utiliser ;
 - prévoir l'utilisation éventuelle d'une deuxième boîte pour l'isolement, dans le cas d'échantillons fortement contaminés ;
 - préciser l'étape de confirmation et restreindre la portée de la confirmation sérologique ;
- harmoniser avec AOAC la méthode de référence pour la détection de *Salmonella*.

Dans ce cas, la structure ISO n'a pu suivre que partiellement ces recommandations. Si les données de performance ont été introduites dans la norme publiée à ce jour, il n'a pas été possible de modifier le bouillon d'enrichissement – ce qui aurait invalidé les valeurs de performance inter-laboratoires – ni de choisir une méthode commune avec l'AOAC. Par ailleurs, la réussite de cet essai a constitué un argument fort pour éviter que certains membres de l'ISO ne remettent en cause le principe de la norme en phase finale de révision.

Quant à la reconnaissance par l'AOAC de la méthode normalisée, elle n'était possible que s'il n'y avait pas de différence significative (test de χ^2) en terme de justesse, entre la méthode AOAC et l'EN ISO 6579. Les résultats de l'essai ont conclu à une telle équivalence pour la poudre d'œuf. On a montré que la méthode normalisée est significativement meilleure que la méthode AOAC pour la viande hachée et le contraire pour le fromage frais, avec une différence à la limite de la signification. Il a fallu donc organiser un nouvel essai inter-laboratoires, à échelle plus réduite, dont le résultat a été satisfaisant. L'EN ISO 6579 a donc pu être finalement validée par AOAC [Feldsine *et al*, 2003] ; ce qui représente la première reconnaissance formelle par l'Amérique du Nord d'une méthode microbiologique normalisée par le CEN et l'ISO !

2.1.4.7 Critères de fidélité des méthodes qualitatives

Suite à la mise au point des critères de fidélité pour les méthodes qualitatives et à leur mise en œuvre pour exploiter les deux essais inter-laboratoires présentés ci-dessus, les partenaires du projet SMT ont souhaité que cette démarche fasse l'objet d'une publication spécifique, qui a été effectivement réalisée (voir Annexe 3.6). Par ailleurs, l'auteur principal a développé une feuille de calculs sous Excel, disponible sur le site Internet <http://www.slangton.co.uk/>.

2.1.4.7.1 Degré d'accord et concordance

Le degré d'accord et son mode de calcul qui ont déjà été définis, sont différents de ceux employés pour exploiter les deux essais inter-laboratoires. En effet, lors de la préparation de la publication, nous nous sommes aperçus que le mode de calcul retenu initialement introduisait un biais dans l'estimation. Un estimateur non biaisé a été introduit de la façon suivante.

Pour le degré d'accord du laboratoire i (ACC_i) la nouvelle formule de calcul est la suivante :

$$ACC_i = \frac{k_i(k_i - 1) + (n - k_i)(n - k_i - 1)}{n(n - 1)} \quad \text{Eq. 19}$$

Où : n est le nombre de répétitions et k_i le nombre de résultats positifs trouvés par le laboratoire i .

Le degré d'accord de la méthode (ACC) est la moyenne des degrés d'accord de chaque laboratoire :

$$ACC = \frac{1}{p} \sum_{i=1}^p ACC_i \quad \text{Eq. 20}$$

La *concordance* est définie et calculée de la même façon qu'au § 2.1.3.

La façon de calculer un écart-type et un intervalle de confiance pour le degré d'accord et pour la concordance est abordée en annexe de la publication. Elle est fondée sur la méthode du *bootstrap* et s'appuie sur une technique de ré-échantillonnage, qui dépend de la façon dont les participants à l'essai ont été sélectionnés.

2.1.4.7.2 Amplitude de la variation entre laboratoires

Nous avons voulu apprécier l'amplitude de la variation entre laboratoires, en comparant le degré d'accord avec la concordance. Un pourcentage de degré d'accord plus élevé que le pourcentage de concordance devrait indiquer que la variabilité inter-laboratoires est dominante. Cependant, ces deux paramètres sont fortement dépendants de la sensibilité de la méthode. Le *odds ratio* de concordance (COR) a été alors introduit :

$$COR = \frac{\text{degré d'accord} \times (100 - \text{concordance})}{\text{concordance} \times (100 - \text{degré d'accord})} \quad \text{Eq. 21}$$

Ce ratio est moins dépendant du niveau de sensibilité que ne le sont les deux critères pris individuellement. Il peut être interprété comme la probabilité relative d'obtenir le même résultat lorsque deux échantillons sont analysés par le même laboratoire, par comparaison à la probabilité d'obtenir le même résultat s'ils sont envoyés à deux laboratoires différents. Un ratio supérieur à 1,0 indique que la variation inter-laboratoires l'emporte sur la variation intra-laboratoire.

Cependant, des valeurs de COR supérieures à 1 pourraient être dues à des variations aléatoires. Un test statistique de différence significative est alors recommandé, tel que le test exact ou à défaut un test de χ^2 , pour confirmer la prédominance de la variation inter-laboratoires. L'interprétation de ces tests doit néanmoins prendre en compte leur puissance, liée au nombre de laboratoires et au nombre de répétitions. En effet, alors qu'une différence entre laboratoires existe, le test pourrait conclure à une différence non significative, en raison d'un nombre insuffisant de données.

2.2 Fidélité des méthodes de dénombrement

Résumé

Dans le prolongement des cinq essais du projet européen SMT, et prenant comme exemple l'un de ces essais, celui qui a porté sur le dénombrement de *L.monocytogenes*, sont abordés certains aspects relatifs à l'exploitation des résultats d'un essai inter-laboratoires sur une méthode de dénombrement.

La représentation graphique des données (les résultats d'un essai inter-laboratoires) s'avère être un outil efficace de vision synthétique de ces résultats d'une part, et d'identification simple des valeurs anormales d'autre part. Cette représentation est particulièrement utile lorsque des statistiques robustes sont utilisées.

Des considérations théoriques rejoignent l'exemple choisi pour remettre en question le bien-fondé de la pratique courante de la transformation initiale des données en logarithmes décimaux.

L'intérêt des statistiques robustes pour estimer les critères de répétabilité et de reproductibilité est souligné, et est illustré avec l'exemple choisi. Cependant, la diversité actuelle des approches normalisées est notée, et des voies d'amélioration sont envisagées.

A l'issue du projet SMT qui vient d'être présenté, nous avons approfondi et discuté un certain nombre d'aspects du volet « exploitation des résultats » d'un essai inter-laboratoires sur une méthode de dénombrement :

- la représentation graphique des données ;
- la transformation initiale des données en logarithmes décimaux ;
- l'utilisation de statistiques robustes pour l'estimation de la fidélité des méthodes de dénombrement.

Ces trois aspects sont illustrés par les données de l'essai du projet SMT sur le dénombrement de *L.monocytogenes*, et les deux derniers aspects ont fait l'objet d'un projet de publication [Lombard *et al*] inclus en annexe 3.7. L'essai sur le dénombrement de *L.monocytogenes* a été choisi en raison des valeurs de fidélité particulièrement mauvaises qui en ont résulté. Ainsi, la comparaison des estimateurs robustes et non robustes de fidélité a pu être menée dans un cas de dispersion élevée des données.

2.2.1 Représentation graphique des données

Alors que cette étape est vivement recommandée par l'ISO 5725-2 pour l'examen initial des données (voir § 1.1.3.3), elle n'est que rarement, sinon jamais mise en œuvre dans notre domaine. Nous avons souhaité envisager son application pour les dénombrements bactériens.

2.2.1.1 Construction des graphiques

Conformément à l'ISO 5725-2, deux types de graphiques sont construits sur la base des statistiques h et k de Mandel.

La statistique de cohérence inter-laboratoires h est calculée selon le Tableau 23.

Tableau 23 : Statistique de cohérence inter-laboratoires h de Mandel (laboratoire i , niveau j)

Etape de calcul	Formule
Moyenne par cellule, \bar{y}_{ij}	$\bar{y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$
Moyenne générale pour le niveau j \bar{y}_j	$\bar{y}_j = \frac{\sum_{i=1}^p n_{ij} \bar{y}_{ij}}{\sum_{i=1}^p n_{ij}}$
Ratio h_{ij} de l'écart de la cellule sur l'écart-type des moyennes de cellules à ce niveau	$h_{ij} = \frac{\bar{y}_{ij} - \bar{y}_j}{\sqrt{\frac{1}{(p_j - 1)} \sum_{i=1}^{p_j} (\bar{y}_{ij} - \bar{y}_j)^2}}$

Les valeurs de h_{ij} sont ensuite reportées sur un graphe par niveau, ainsi que par laboratoire.

La statistique de cohérence intra-laboratoire k est calculée selon le Tableau 24.

Tableau 24 : Statistique de cohérence intra-laboratoires k de Mandel (laboratoire i , niveau j , p_j laboratoires au niveau j)

Etape de calcul	Formule
Variance intra-cellule, s_{ij}^2	$s_{ij}^2 = \frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2$
Ecart-type intra-cellule commun pour chaque niveau j	$\sqrt{\frac{\sum s_{ij}^2}{p_j}}$
Ratio k_{ij} de l'écart-type intra-cellule s_{ij} sur l'écart-type intra-cellule commun	$k_{ij} = \frac{s_{ij} \sqrt{p_j}}{\sqrt{\sum s_{ij}^2}}$

Les valeurs de k_{ij} sont également reportées sur un graphe par niveau, ainsi que par laboratoire.

Nous avons effectué les calculs sur les estimateurs non robustes des différentes grandeurs des deux tableaux ci-dessus, à l'aide du logiciel EasyVal™ (macro pour Excel), sur la base des données de l'essai inter-laboratoires sur le dénombrement de *L.mobocytogenes*, dont les résultats figurent dans le Tableau 1 en annexe 3.7.

2.2.1.2 Application au dénombrement de *L.monocytogenes*

Les statistiques h et k de Mandel, groupées par laboratoire, sont représentées sous forme d'histogrammes dans la

Figure 1 et la

Figure 2. A noter que les statistiques h n'ont pas été calculées par le logiciel pour la matrice poudre d'œuf.

Les graphiques peuvent indiquer qu'un ou plusieurs laboratoires donnent des ensembles de résultats différant nettement de ceux des autres laboratoires.

Pour la statistique h de justesse, on n'observe pas globalement dans la

Figure 1 de situation anormale. En effet, aucun laboratoire ne se distingue par des valeurs de h d'un signe, par rapport aux autres laboratoires dont les valeurs de h seraient du signe contraire. Le nombre de laboratoires donnant des valeurs négatives de h équilibre approximativement le nombre de laboratoires donnant des valeurs positives de h . La situation idéale est un équilibre au sein même du laboratoire entre les valeurs positives et négatives de h , pour les différentes combinaisons <matrice x niveau de contamination> (tel que le laboratoire 2). Par contre, certaines valeurs individuelles (par combinaison) apparaissent

clairement anormales. Retenant la valeur critique de Mandel au niveau de signification de 1 %, soit 2,35 pour 17 laboratoires et 2 répétitions, la valeur h du laboratoire n°16 est très largement anormale (-24) pour la combinaison <viande x niveau haut>, ainsi que dans une moindre mesure celle des laboratoires n°4, 11, 13, 14 et 17. Le nombre de laboratoires ayant obtenu des valeurs de justesse anormale pour cette combinaison suggère qu'il ne s'agit pas d'un problème lié au laboratoire, mais très certainement lié à ce matériau d'essai. Les autres valeurs de justesse respectent ou dépassent légèrement les valeurs critiques.

En ce qui concerne la statistique k de fidélité, le diagramme de la

Figure 2 indique que la moitié des laboratoires participants (8) a obtenu une bonne à très bonne répétabilité, soit valeurs de k respectant approximativement la valeur de 1. A l'opposé, la valeur critique de Mandel, au niveau de signification de 1 %, soit 2,44 pour 17 laboratoires et 2 répétitions, a été dépassée par 6 laboratoires pour chaque fois une seule combinaison <matrice x niveau de contamination>, différente d'un laboratoire à l'autre. Ce cas de figure reflète un défaut de répétabilité des laboratoires concernés.

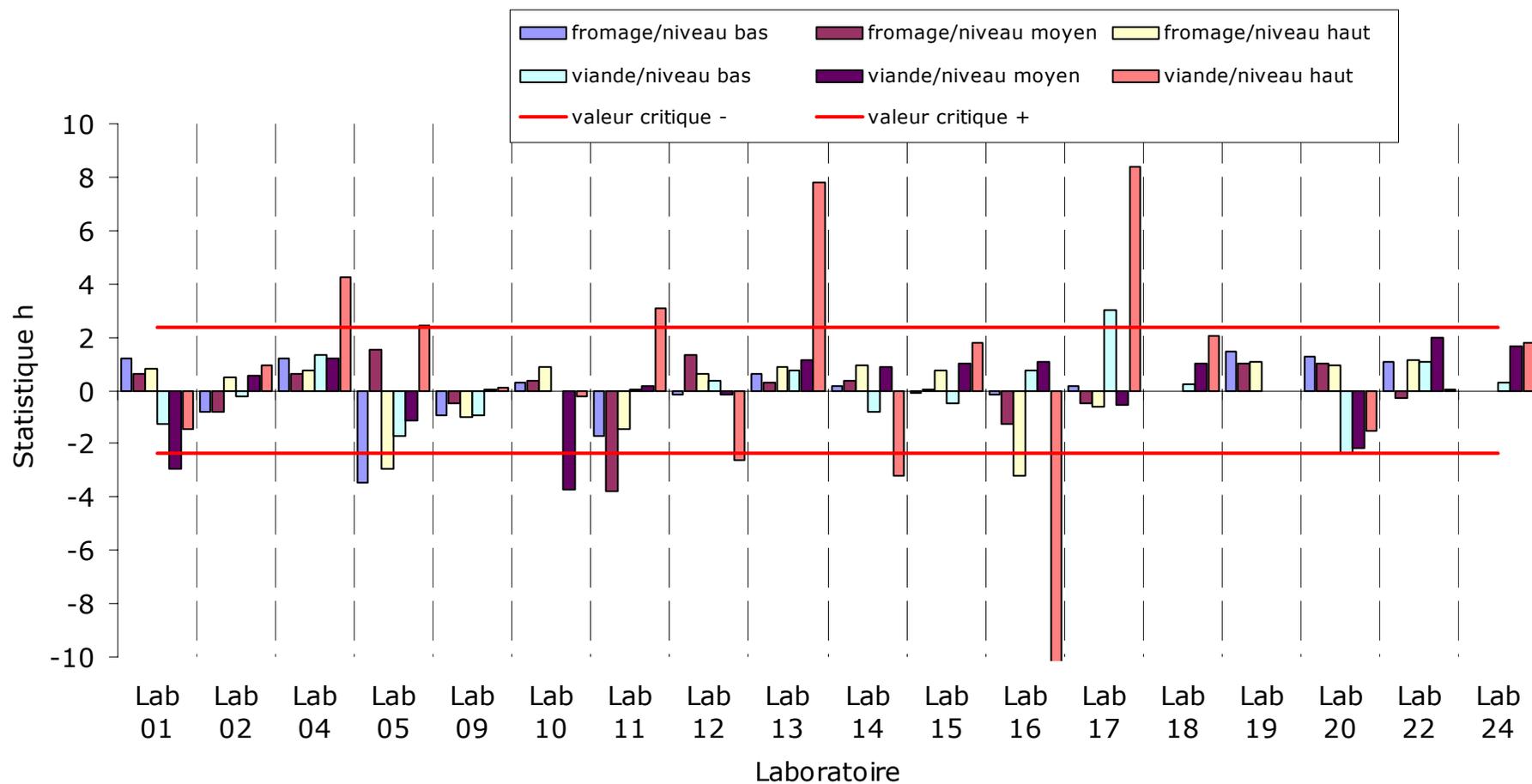


Figure 1 : Histogramme de justesse, selon la statistique h de Mandel – Exemple des résultats de l’essai inter-laboratoires sur le dénombrement de *L.monocytogenes*

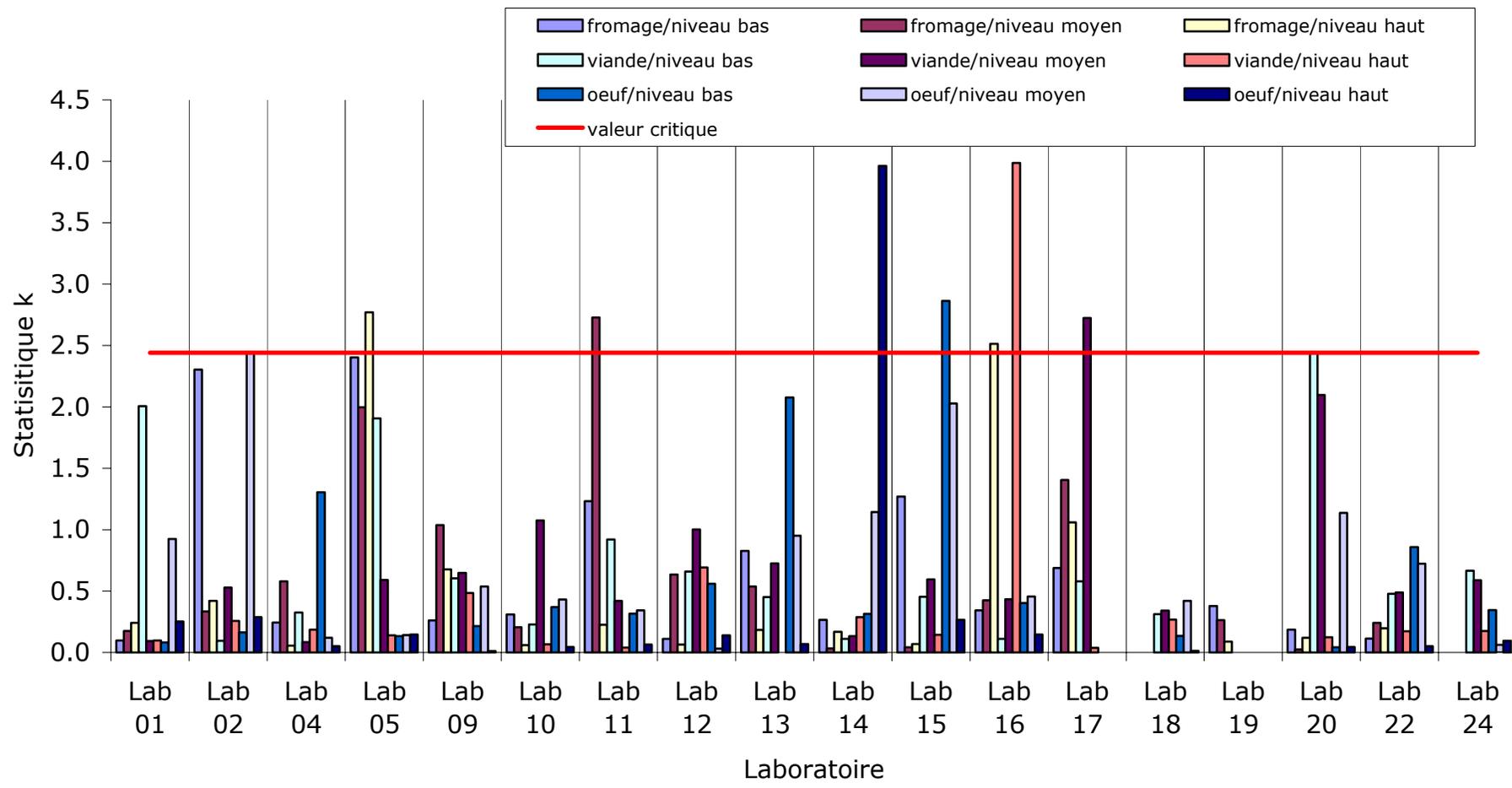


Figure 2 : Histogramme de fidélité, selon la statistique k de Mandel – Exemple des résultats de l’essai inter-laboratoires sur le dénombrement de *L.monocytogenes*

2.2.1.3 Intérêt de la représentation graphique

La représentation graphique qui vient d'être illustrée permet donc d'avoir une vision synthétique des résultats d'un essai inter-laboratoires, plus lisible que l'identification individuelle des valeurs aberrantes par des tests statistiques. Elle constitue un moyen d'identification facile des résultats anormaux, permettant d'instaurer un dialogue avec les participants pour tenter d'identifier les causes de déviation.

Comme nous le verrons plus loin, un telle représentation graphique est d'autant plus utile que des statistiques robustes sont utilisées pour estimer les critères de fidélité. En effet, dans ce cas, l'emploi de tests statistiques pour identifier les valeurs aberrantes n'est plus nécessaire.

2.2.2 Transformation des données en logarithmes décimaux

On peut s'interroger sur la validité de la transformation des données brutes en logarithmes décimaux. En effet, cette transformation logarithmique est largement pratiquée en microbiologie quantitative, aussi bien pour la validation de méthodes (AOAC, EN 16140, projet SMT) que dans le cadre d'essais d'aptitude (RAEMA, CECALAIT, essais du LCR Lait). Deux raisons principales sont avancées pour justifier cette pratique :

1. Elle permet de normaliser la distribution des données, en supposant que les dénombrements bactériens suivent une distribution log-normale [Jarvis, 1989].
2. Elle permet d'atteindre l'homoscédasticité, soit la stabilité des variances de répétabilité ou de reproductibilité, quel que soit le niveau de contamination. En effet, nous avons montré en Annexe 3.7 que l'écart-type d'un dénombrement est proportionnel, à un facteur de dilution près, à l'écart-type des comptages sur boîte de Pétri ; il varie donc de façon proportionnelle avec le niveau de contamination. Une transformation en \log_{10} , selon une propriété des logarithmes décimaux, permettra alors d'obtenir des valeurs « relatives » d'écart-type, qui ne seront plus proportionnelles aux valeurs des niveaux de contamination.

2.2.2.1 Log-normalité de la distribution des données

Nous avons voulu vérifier que la première hypothèse s'appliquait à l'essai inter-laboratoires sur le dénombrement de *L. monocytogenes* :

- Sans élimination des valeurs aberrantes, l'hypothèse de normalité des données log-transformées est rejetée par le test de Shapiro-Wilks ($P < 5\%$) pour 8/9 des combinaisons <matrice x niveau de contamination> et pour 5/9 combinaisons des données non transformées.
- Après élimination des données aberrantes, la proportion de rejet de la log-normalité est encore de 4/9 et de 1/9 pour la normalité.

Il en ressort que l'hypothèse de log-normalité est rarement vérifiée sur cet exemple.

Cette remise en cause de la log-normalité est confortée par des considérations théoriques et plusieurs travaux expérimentaux :

- Certains travaux [Hedges, 1967 et Jarvis, 1987] ont porté sur les sources d'erreur associées à la technique de comptage des colonies :
 - erreurs de nature technique, telle que celle liée au pipetage ;
 - erreurs liées à la distribution des microorganismes, telles que celle provenant de la prise d'essai dans l'échantillon d'aliment, ou celles découlant de la réalisation des dilutions successives, et de la prise d'essai pour l'étalement sur boîte.

Une modélisation des différentes sources d'erreur (à l'exception de l'hétérogénéité de contamination entre échantillons et de la sélection de 5 colonies pour confirmation) a été effectuée sur la base de ces travaux, et elle a abouti à une distribution normale des comptages [Hedges, 2002].

- L'analyse à grande échelle d'échantillons naturellement contaminés (tels que produits laitiers avec des bactéries anaérobies, concentrés congelés de jus de pomme avec levures), dans le cadre d'autocontrôles industriels, a fourni de nombreuses séries de résultats qui ont toutes pu être correctement décrites par une distribution normale [Corradini *et al*, 2001].

2.2.2.2 Stabilisation des variances

Outre la normalité de la distribution des données, il reste toujours nécessaire, pour pouvoir comparer les valeurs de fidélité entre différents niveaux, de stabiliser des variances.

La Figure 4 du § 2.2.3.2 illustre clairement la stabilité de l'écart-type de reproductibilité, après transformation en \log_{10} des données, sur les différents niveaux de contamination, à l'exception de la combinaison <fromage x niveau haut>. Cependant, cette dernière se distingue par une dispersion plus élevée des résultats par rapport aux résultats des autres niveaux de la même matrice : cette exception à la stabilité semble donc imputable à une dispersion anormale des résultats, et n'apparaît donc pas constituer une exception à l'homoscédasticité de la variance de reproductibilité.

On peut noter dans la

Figure 5 du § 2.2.3.2 que l'utilisation des coefficients de variation obtenus sans transformation en \log_{10} permet aussi de visualiser une certaine stabilité. Ce qui n'est pas surprenant puisqu'on les obtient en divisant les écarts-types par les moyennes : si l'écart-type est proportionnel à la moyenne, le coefficient de variation est alors constant.

2.2.2.3 Comparaison des écarts-types sur des données transformées ou non en \log_{10}

La justification de calculer les écarts-types sur des données log-transformées a déjà été discutée, ainsi que l'effet de cette transformation sur la comparaison entre estimateurs robustes et non robustes de l'écart-type de reproductibilité. Les écarts-types, exprimés en unité log et non log, sont comparés d'une façon supplémentaire au travers d'une utilisation possible de l'écart-type de reproductibilité, à savoir l'estimation de l'incertitude de mesure (IM). L'applicabilité de cette notion à la microbiologie des aliments sera envisagée plus loin, dans le § 2.5.4.

Selon les principes du *Guide pour l'expression de l'incertitude de mesure* [GUM, ISO, 1995], document de référence en la matière, un laboratoire peut associer un intervalle d'incertitude de mesure à un résultat de mesure individuel z . Conformément à la spécification technique ISO/TS 21748, lorsque le résultat a été obtenu avec une méthode de dénombrement caractérisée par un écart-type de reproductibilité s_R , l'intervalle peut être exprimé, sous les conditions abordées dans le § 2.5.4.2, et avec un niveau de confiance de 95 %, comme $z \pm 2 s_R$. Cet intervalle correspond à l'IM étendue, selon la terminologie du GUM.

En fonction de la transformation logarithmique ou non, l'intervalle d'IM est exprimé de deux façons :

1. avec des données brutes, soit x le résultat individuel exprimé en nombre d'ufc/g ou ml :

$$\text{Intervalle IM} = [x - 2 s_R ; x + 2 s_R] \quad \text{Eq. 22}$$

2. avec des données log-transformées, en notant $y = \log_{10} x$, l'intervalle d'IM pour y est $[y - 2 s_R ; y + 2 s_R]$ et pour x :

$$\text{Intervalle IM}_x = [10^{y-2s_R} ; 10^{y+2s_R}] = \left[\frac{x}{10^{2s_R}} ; x \cdot 10^{2s_R} \right] \quad \text{Eq. 23}$$

Nous avons reporté dans le Tableau 25 et représenté dans la Figure 3 les intervalles calculés selon les équations 27 et 28, à partir des données du laboratoire 1.

Tableau 25 : Comparaison des intervalles d'incertitude de mesure (IM) du laboratoire 1, avec et sans transformation en \log_{10} des données, lors de l'essai sur le dénombrement de *L. monocytogenes*

Matrice	Niveau	Transformation en log (valeurs en antilog)		Données brutes	
		Médiane	IM (intervalle)	Médiane	IM (intervalle)
Fromage frais	Bas	89	[11 ; 708]	123	[-39 ; 285]
	Moyen	1349	[427 ; 4266]	1497	[-195 ; 3189]
	Haut	7943	[1905 ; 33113]	16811	[-10425 ; 44047]
Viande hachée	Bas	457	[132 ; 1585]	524	[-144 ; 1192]
	Moyen	3162	[1047 ; 9550]	3624	[-14 ; 7262]
	Haut	21878	[4365 ; 109648]	25709	[-4711 ; 56129]
Poudre d'œuf	Bas	209	[52 ; 832]	261	[-113 ; 635]
	Moyen	2042	[676 ; 6166]	2320	[-268 ; 4908]
	Haut	24547	[11220 ; 53703]	25646	[5098 ; 46194]

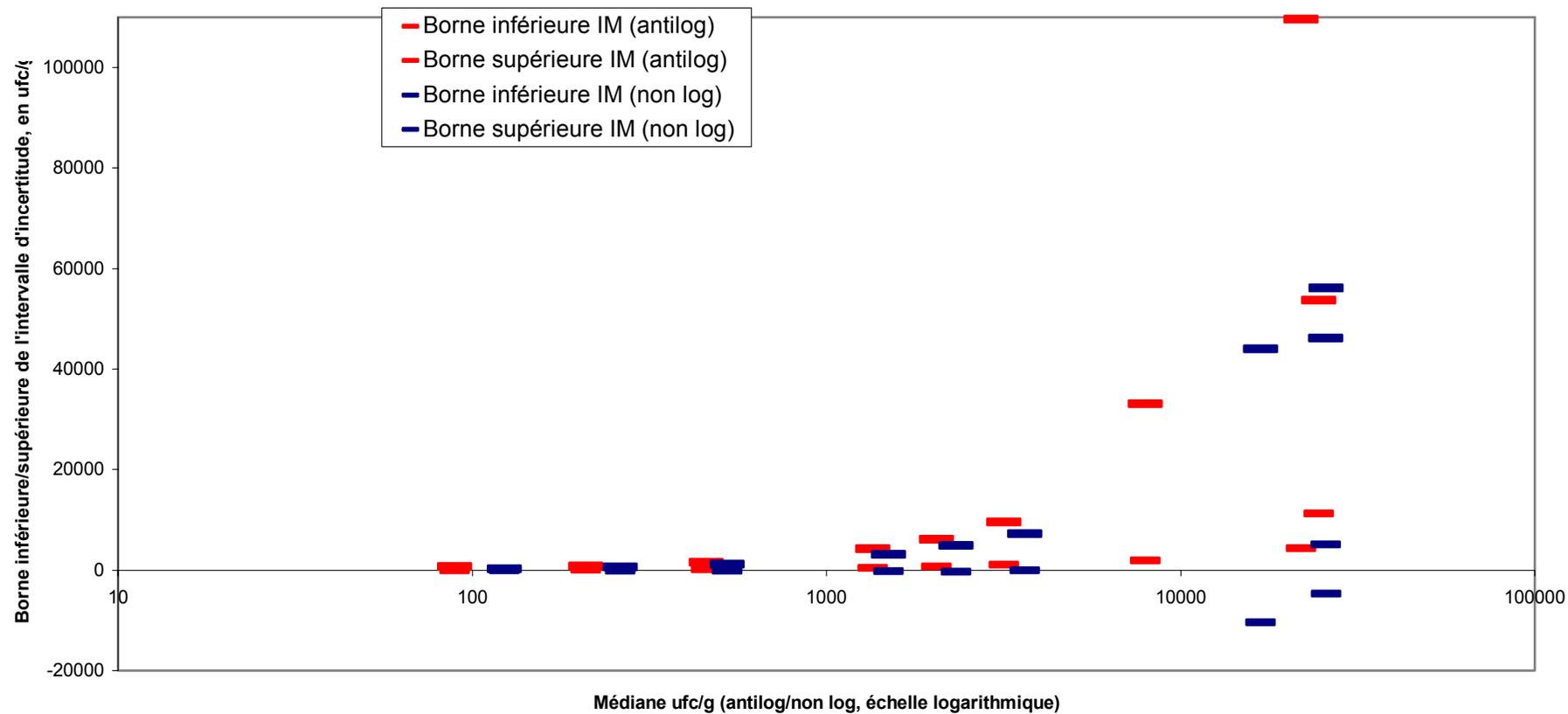


Figure 3 : Représentation des intervalles d'incertitude de mesure (IM), avec ou sans transformation en \log_{10} des données. Exemple du laboratoire 1, essai sur le dénombrement de *L.monocytogenes*.

On peut voir dans la Figure 3 que les valeurs obtenues sont cohérentes, ce qui confirmerait que les deux approches ne sont pas contradictoires. Cependant, les bornes supérieures des intervalles anti-log sont plus hautes que pour les données brutes, et presque toutes les bornes basses des intervalles non-log sont négatives, sauf deux. Ces valeurs négatives sont bien sûr embarrassantes si on veut leur donner un sens microbiologique. Un avantage de la transformation logarithmique semble donc d'obtenir directement des bornes basses réalistes. Cependant, il serait également possible de remplacer, par convention, les bornes basses négatives par 0.

2.2.2.4 Conclusion sur la transformation des données en logarithmes décimaux

Il ressort de notre travail que la log-normalité des comptages bactériens ne devrait pas être tenue pour acquise, et qu'il est possible d'estimer des écarts-types de fidélité sur la base de données non transformées en \log_{10} , à deux conditions :

1. Exprimer ces écarts-types de façon relative.
2. Remplacer par convention les limites basses des intervalles d'incertitude de mesure par 0.

Cependant, il ne faudrait pas généraliser trop rapidement cette conclusion. Si la transformation en \log_{10} ne semble pas nécessaire pour des résultats obtenus sur un échantillon homogénéisé, elle semble devoir être toujours préférable lorsque l'échantillon est hétérogène. Ces travaux montrent également qu'il sera nécessaire à l'avenir de proposer des modèles sans doute plus élaborés que la loi normale ou la loi log-normale pour valoriser au mieux les comptages microbiologiques.

2.2.3 Comparaison des estimateurs robustes et non robustes

Si l'EN ISO 16140 a été logiquement choisie par le projet SMT comme référentiel propre à la microbiologie des aliments, les estimateurs robustes qui y figurent sont néanmoins différents de ceux introduits par l'ISO 5725-5. A l'issue du projet SMT, nous avons donc mis en œuvre également les estimateurs de cette deuxième norme avec les données de l'essai sur le dénombrement de *L. monocytogenes*. Une analyse comparative a été menée sur ces deux types d'estimateurs robustes, par rapport aux estimateurs non robustes de l'ISO 5725-2.

2.2.3.1 Méthodologie

Les estimateurs de fidélité robustes de l'EN ISO 16140, et ceux non robustes de l'ISO 5725-2, ont été décrits dans le § 2.1.2.

En ce qui concerne les estimateurs robustes de l'ISO 5725-5, deux algorithmes, appelés A et S, sont utilisés dans une procédure itérative. Ils ont été retenus à l'issue de travaux du Comité des Méthodes d'Analyse de la *Royal Society of Chemistry* [Analytical Methods Committee, 1989 a et b].

L'annexe 3.5 (paragraphe 3.5.3) décrit l'algorithme S qui permet d'estimer l'écart-type de répétabilité par combinaison robuste des écarts-types intra-laboratoires de tous les laboratoires. Quant à l'algorithme A, il comprend deux étapes principales :

1. l'estimation de l'écart-type inter-laboratoires à partir des moyennes de doubles de chaque laboratoire ;
2. l'estimation par itérations de l'écart-type de reproductibilité, à partir d'une valeur initiale s^* qui est la Médiane des Déviations Absolues (MAD).

Nous avons développé des feuilles de calculs sous Excel pour effectuer ces calculs, et nous avons par ailleurs utilisé la macro EasyVal™ pour Excel.

Lors de la présentation des résultats des trois essais du projet SMT sur les méthodes quantitatives (§ 2.1.4), l'utilisation des estimateurs robustes de l'EN ISO 16140 a été recommandée pour exprimer la fidélité des méthodes, sur la base d'une similarité globale entre ces estimateurs robustes et les estimateurs non robustes de l'ISO 5725-2. Nous avons voulu approfondir cette comparaison, en l'élargissant aux estimateurs robustes de ISO 5725-5. Nous avons également étudié l'influence de la transformation logarithmique des données sur cette comparaison.

Les trois estimations de fidélité ont été comparées de deux façons différentes :

1. Graphiquement.
2. Par un critère pragmatique de cohérence. En effet, les tests statistiques usuels ne pouvaient pas être utilisés, car ils ne s'appliquent pas à la comparaison de variances estimées sur le même échantillon de la même population. Le critère pragmatique a été choisi de la façon suivante : la différence entre deux écarts-types est considérée comme acceptable jusqu'à $0,1 \log_{10}$ (sur la base de l'expression habituelle des écarts-types à un chiffre significatif), ou jusqu'à 10 % pour les écarts-types relatifs.

2.2.3.2 Application au dénombrement de *L.monocytogenes*

La comparaison présentée porte uniquement sur les écart-types de reproductibilité, celle des écarts-types de répétabilité conduisant aux mêmes conclusions.

Les trois estimations d'écart-type de reproductibilité, après transformation des données en \log_{10} , sont représentées dans la Figure 4. Ce graphique indique une très bonne cohérence entre les trois estimations, à l'exception de la combinaison <fromage x niveau haut>.

Le recours au critère pragmatique de cohérence que nous avons retenu conforte cette indication : il est vérifié pour 7 des 9 combinaisons <matrice x niveau de contamination>. Il s'agit là d'un argument en faveur des statistiques robustes, puisqu'il leur a été objecté qu'elles pouvaient cacher la variabilité véritable d'une méthode, en sous-évaluant le poids des valeurs extrêmes. Pour la seule combinaison <fromage frais x niveau haut> distinguée graphiquement, l'approche non robuste a donné une valeur (1,15) nettement plus élevée que les deux approches robustes (0,31 et 0,81). Ici, les approches robustes ont effectivement sous-estimé la variabilité des résultats. Cependant, comme nous l'avons noté au § 2.2.2.2, les résultats pour cette combinaison étaient extrêmement dispersés par rapport aux autres niveaux de cette même matrice. De ce fait, la qualité de l'expérience, notamment en matière d'échantillons, peut être dans ce cas raisonnablement remise en cause et la fidélité de la méthode n'aurait sans doute pas dû être estimée pour cette combinaison.

La

Figure 5 présente la comparaison des trois estimateurs, cette fois-ci sans transformation en \log_{10} des données, et avec l'expression de l'écart-type de reproductibilité sous forme relative. La cohérence entre les trois estimateurs apparaît toujours graphiquement ; elle semble toutefois moins forte, même si les échelles différentes empêchent toute comparaison véritable entre les deux graphiques. Si sur le graphe, n'apparaît plus l'exception de la combinaison <fromage x niveau haut> en matière de cohérence entre estimateurs, notre critère pragmatique de cohérence, soit une différence entre écarts-types relatifs de 10 % au plus, n'est pas respecté pour quatre combinaisons sur neuf. La cohérence des estimateurs était donc meilleure sur des données transformées en log ; ce qui pourrait être imputé à la réduction, par cette transformation, de la dispersion des données.

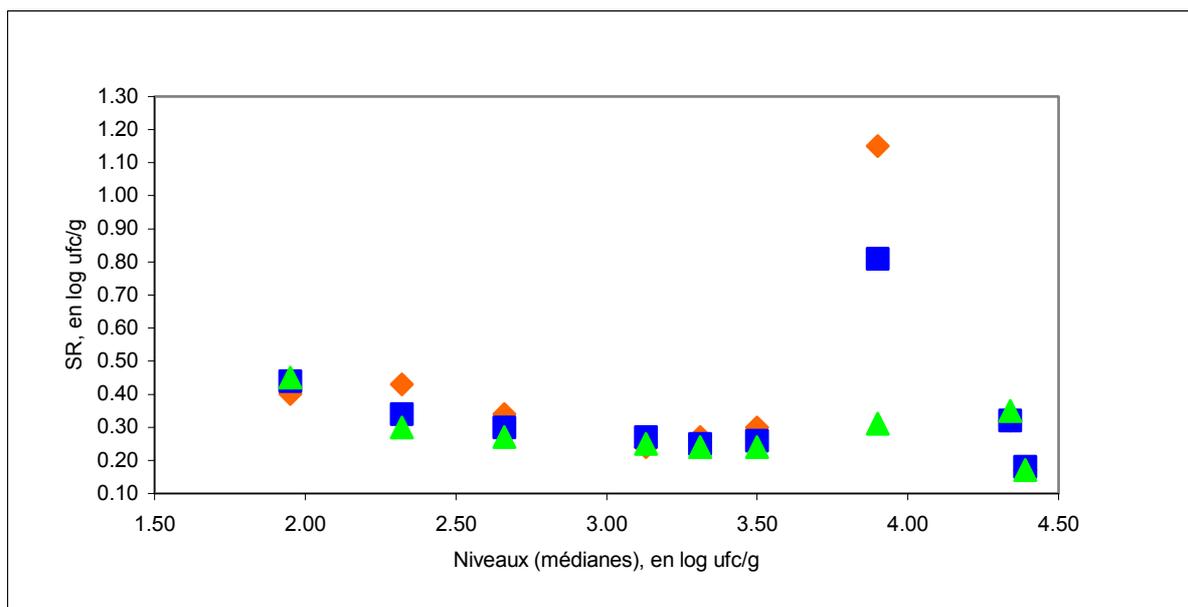


Figure 4 : Ecart-types de reproductibilité s_R pour l'essai inter-laboratoires sur le dénombrement de *L. monocytogenes*, calculés selon ISO 5725-2 (\diamond), ISO 5725-5 (\square) et EN ISO 16140 (\triangle), sur des données transformées en \log_{10} .

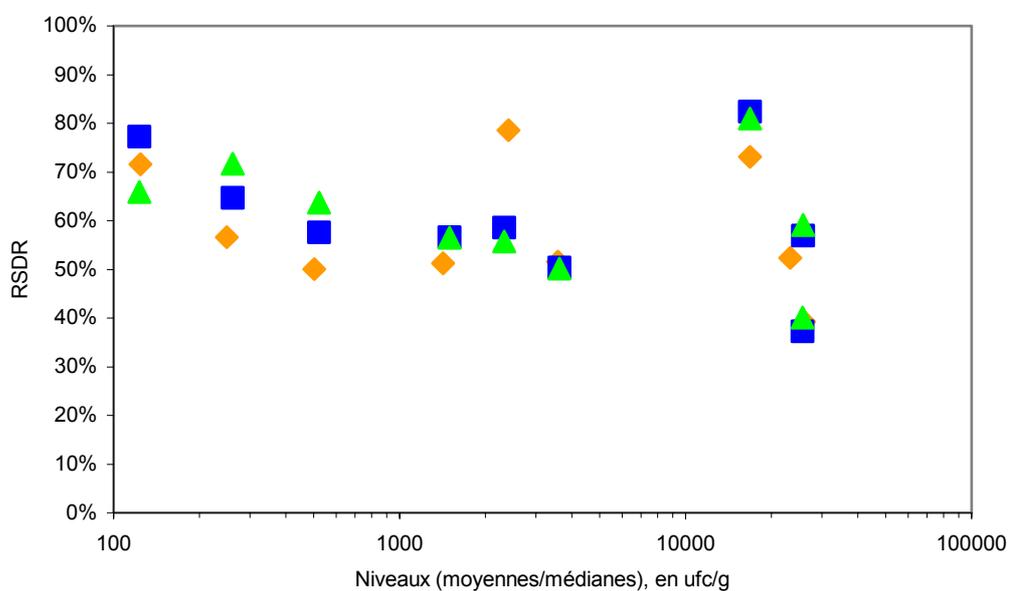


Figure 5 : Coefficients de variation de reproductibilité RSD_R pour l'essai inter-laboratoires sur le dénombrement de *L. monocytogenes*, calculés selon ISO 5725-2 (\diamond), ISO 5725-5 (\square) et EN ISO 16140 (\triangle), sur des données non transformées en \log_{10} .

2.2.3.3 Intérêt des statistiques robustes

L'intérêt comparatif des statistiques robustes peut d'abord être envisagé d'un point de vue théorique. Les estimateurs non robustes de l'ISO 5725-2 utilisent la moyenne, et supposent que la distribution des erreurs est Normale. Cette hypothèse est souvent justifiée mais nécessite l'élimination des calculs des valeurs aberrantes [Analytical Methods Committee, 1989 a]. Or, ce rejet des valeurs aberrantes présente plusieurs inconvénients :

- Ainsi que l'a souligné l'ISO 5725-5 (voir § 1.1.5.3.1), la décision d'exclure des valeurs peut comporter un certain degré de subjectivité.
- De plus, dans le domaine de l'analyse des aliments, nous avons remarqué qu'il n'existait pas d'accord général sur la procédure d'identification des valeurs aberrantes. Ainsi le protocole IUPAC/AOAC se distingue de l'ISO 5725-2 par le niveau de risque choisi, ainsi que par la séquence d'application de ces tests (voir § 1.1.7.3).

Par ailleurs, nous pouvons distinguer d'autres avantages comparatifs des statistiques robustes, qui s'accommodent des valeurs aberrantes, au lieu de les rejeter :

- Comme l'a noté ISO 5725-5 :
 - La méthode « classique » requiert une stabilité des variances intra-laboratoires ; ce qui n'est pas toujours vérifié en pratique. Les statistiques robustes, ne reposant pas sur cette hypothèse, représentent alors une alternative satisfaisante.
 - Les représentations graphiques de Mandel reposent sur des statistiques dont les dénominateurs sont des écarts-types calculés avec des données brutes. La présence de valeurs aberrantes produira alors un effet déformant sur les graphiques.
- L'approche classique est très sensible aux valeurs réelles des données. Par exemple, l'oubli d'une virgule dans un résultat aura un impact important sur la moyenne et beaucoup plus limité sur la médiane.
- Une exclusion trop importante de valeurs aberrantes pourrait nettement sous-estimer la variabilité véritable de la méthode en pratique ; ce qui a d'ailleurs été reconnu par FIL ou AOAC, qui ont défini des valeurs maximales pour les taux d'exclusion.
- Enfin, un avantage comparatif des statistiques robustes réside dans l'absence d'influence sur la médiane de la convention choisie pour indiquer les résultats nuls, qui sont en fait inférieurs à la limite de dénombrement de la méthode. D'autres conventions que 0 sont possibles, et peuvent influencer la moyenne.

Compte tenu de ces différents avantages, ainsi que de leur illustration dans l'exemple choisi, nous pouvons donc confirmer l'intérêt d'utiliser des statistiques robustes pour estimer les critères de fidélité d'une méthode de dénombrement bactérien.

2.2.3.4 Statistiques robustes : limites et perspectives

Si l'intérêt des statistiques robustes pour estimer la fidélité des méthodes de comptage bactérien est indéniable, leur utilisation comporte deux risques principaux :

- **Sous-estimer la variabilité réelle de la méthode.** Pour éviter ce risque, il convient de s'assurer que la proportion de valeurs aberrantes reste faible, après l'identification des valeurs aberrantes par les tests statistiques –ce que nous avons effectivement vérifié dans notre exemple, puisqu'aucune valeur aberrante n'a été identifiée pour 5 des 9 combinaisons, et une seule combinaison a été identifiée comme aberrante pour chacune des 4 autres combinaisons.
- **Supprimer l'étape initiale d'examen critique des données.** Ainsi, l'étape de représentation graphique, dont nous avons souligné l'utilité plus haut, devrait être effectuée systématiquement. Les valeurs aberrantes devraient être investiguées avec les participants concernés, afin d'une part d'identifier les raisons de ces déviations et d'autre part de corriger si possible les valeurs aberrantes lorsqu'il s'agit, par exemple, d'erreurs de transcription ou d'inversion de résultats.

En ce qui concerne les propriétés statistiques des estimateurs robustes que nous avons utilisés, à savoir la Médiane des Déviations Absolues (ou MAD) et le S_n de Rousseuw, il faut être tout d'abord conscient qu'ils sont quand même très dépendants de la normalité de la distribution, contrairement à ce que l'on pouvait présumer. En effet, il convient de distinguer deux types de robustesse : les estimateurs ci-dessus sont certes robustes envers les valeurs aberrantes et ils ont une bonne *résistance* contre les données dites polluantes, mais ils manquent de robustesse ou d'*efficacité* contre les distributions non normales ou dissymétriques. Ces propriétés ont été mises en évidence par des simulations effectuées dans le cadre de travaux du groupe de travail ISO/TC 34/SC 9/WG 2 qui traite des aspects statistiques liés à l'analyse microbiologique des aliments [Wilrich, 2004a].

Par ailleurs, nous pouvons nous interroger sur l'existence de deux approches robustes différentes pour un même objectif. Comme mentionné lors de la présentation de l'EN ISO 16140, les rédacteurs de cette norme n'ont pas pris en compte l'ISO 5725-5 non encore publiée. Cependant, les deux auteurs de l'estimateur S_n [Rousseuw et Croux, 1993] ont critiqué la MAD, retenue par l'ISO 5725-5, pour deux raisons :

1. sa faible efficacité asymptotique envers les distributions normales, qui est de 37 % au lieu de 64 % pour la médiane ;
2. la nécessité d'appliquer la MAD à une distribution symétrique, car la MAD estime d'abord une valeur centrale, la médiane, et donne ensuite un même poids aux déviations positives et négatives à la médiane.

Cette critique justifie *à posteriori* le choix de l'EN ISO 16140 de différer de l'ISO 5725-5.

L'EN ISO 16140 a adopté la première alternative de Rousseeuw à la MAD, S_n . D'un point de vue théorique, ce choix est justifié de la façon suivante :

- S_n est fondé sur une distance typique entre données, toujours valide pour les distributions asymétriques. Cette propriété est particulièrement importante pour les comptages bactériens, souvent distribués de façon dissymétrique.
- S_n a le point de rupture asymptotique⁴ le plus élevé possible pour un estimateur d'échelle (50 %) et une efficacité asymptotique meilleure que la MAD (58 %).

En revanche, la fonction d'influence de S_n présente des discontinuités. C'est ainsi que Rousseeuw, dans la même publication, a proposé une seconde alternative plus intéressante que S_n : Q_n . Le Tableau 26 indique la façon de calculer cet estimateur.

Tableau 26 : Calcul de l'estimateur d'échelle Q_n , selon [Wilrich, 2004a] (p : nombre de laboratoires, \bar{x}_i : moyennes des répétitions du laboratoire i)

Etape	Formule
Déviations absolues de chacune des \bar{x}_i avec les $(p-1)$ autres valeurs \bar{x}_j	$ \bar{x}_i - \bar{x}_j $
Classement en ordre croissant des $ \bar{x}_i - \bar{x}_j $ (en supprimant les doublons)	
Q_n : $k^{\text{ième}}$ ordre statistique des $ \bar{x}_i - \bar{x}_j $ (soit la $k^{\text{ième}}$ valeur des $ \bar{x}_i - \bar{x}_j $ classées en ordre croissant)	$Q_n = \left(\bar{x}_i - \bar{x}_j _{i < j} \right)_{[k]}$ <p>avec $k = C_h^2$ et $h = [n/2] + 1$</p> <p>où la notation $[x]$ indique le nombre entier immédiatement inférieur ou égal à x</p>

$c_n Q_n$ est un estimateur non biaisé de l'écart-type, avec l'hypothèse de normalité, où :

$$c_n = 2,2219 \frac{n}{n + 1,4} \text{ si } n \text{ est impair}$$

$$c_n = 2,2219 \frac{n}{n + 3,8} \text{ si } n \text{ est pair}$$

Eq. 24

Par rapport à S_n , $c_n Q_n$ est autant adapté aux distributions asymétriques, a également un point de rupture de 50 %, mais possède une fonction d'influence lissée et une efficacité asymptotique très élevée (82 %). Il apparaît donc préférable de retenir Q_n à la place de S_n comme estimateur robuste à la base de l'écart-type de reproductibilité.

⁴ Le point de rupture d'un estimateur est la plus grande fraction de valeurs très aberrantes (mathématiquement qui tendent vers l'infini) qui n'entraîne pas l'estimateur à tendre vers l'infini, c'est-à-dire à une rupture de l'estimateur. Lorsque $n \rightarrow \infty$, il s'agit du point de rupture asymptotique.

2.2.4 Perspectives

L'EN ISO 16140 et le protocole AOAC étant d'abord conçus pour la validation d'une méthode alternative par rapport à une méthode de référence, il n'existe pas à ce jour de document de référence applicable strictement à l'estimation de la fidélité d'une méthode de dénombrement de référence.

Dans ce contexte, un groupe de travail mixte ISO/FIL sur les statistiques des données analytiques a décidé de développer, en rapport avec ISO/TC34/SC9, un tel référentiel, qui couvrirait aussi bien l'organisation des essais inter-laboratoires, que l'exploitation de leurs résultats [FIL, 2003]. Sa rédaction nous a été confiée. La version initiale est fondée sur l'EN ISO 16140, mais nous comptons y incorporer les éléments qui faisaient défaut dans cette norme, comme les études de stabilité et d'homogénéité, ainsi que les améliorations décrites ci-dessus, telles que la représentation graphique initiale des données et l'utilisation de Q_n à la place de S_n pour l'estimation robuste de l'écart-type de reproductibilité. Nous soumettrons également à discussion la question de la transformation logarithmique des données.

2.3 Performance des méthodes de détection

Résumé

Toujours dans le prolongement des essais de validation du projet européen, est abordée dans ce chapitre la façon d'évaluer la performance d'une méthode de détection. En ce qui concerne la fidélité, les limites des critères de degré d'accord et de concordance sont discutées, et une approche alternative est introduite, se calquant directement sur celle adoptée pour les méthodes quantitatives. Quant à la justesse, un protocole expérimental et un mode de calcul de la sensibilité et de la spécificité sont proposés, afin de pouvoir d'une part s'affranchir de la comparaison à une méthode de référence, et d'autre part prendre en compte la distribution des microorganismes à faibles taux.

2.3.1 Fidélité des méthodes de détection

2.3.1.1 Limites des critères de degré d'accord et de concordance

En matière d'exploitation des résultats des essais inter-laboratoires portant sur les méthodes de détection, la façon de caractériser leur fidélité est d'abord traitée. Alors que l'intérêt des critères de degré d'accord et de concordance, introduits par le projet SMT, a déjà été mis en évidence, leurs limites sont maintenant abordées, avant d'envisager une voie alternative, tout récemment proposée, pour caractériser la fidélité des méthodes qualitatives.

Le projet SMT a eu le mérite de se « risquer » à proposer une approche de la fidélité inter-laboratoires de méthodes qualitatives ; ce qui n'avait pas été encore fait à notre connaissance. Cette démarche a attiré les critiques, comme les auteurs de ces travaux s'y attendaient. Ces critiques portent, d'une part, sur le mode d'estimation de ces critères, et d'autre part, sur leur interprétation.

2.3.1.1.1 Mode d'estimation

Dans un article publié récemment [van der Voet et van Raamsdonk, 2004], les auteurs notent que les critères de degré d'accord et de concordance, tels que décrits dans l'article présenté dans le § 2.1.4.7 [Langton *et al*, 2002], sont des probabilités estimées sur la base d'un modèle à effets fixes. En effet, l'estimation de ces critères repose sur la division du nombre de paires de résultats identiques par le nombre total de paires de l'ensemble des données. Ce qui signifie que l'ensemble des laboratoires ayant participé à l'essai inter-laboratoires serait exhaustif, et que les critères de performance déduits ne seraient applicables qu'à ces laboratoires et pas aux autres laboratoires qui utilisent la même méthode. A l'évidence, ce modèle n'est pas adapté à l'objectif des essais inter-laboratoires visant à estimer la fidélité d'une méthode.

Comme nous l'avons indiqué à plusieurs reprises, il s'agit au contraire de produire des valeurs de fidélité applicables à l'utilisation ultérieure et correcte de la méthode par un ensemble bien plus large de laboratoires que ceux qui ont participé à l'essai. Dans cet objectif, un modèle à effets aléatoires apparaît plus justifié. Ce modèle considère l'ensemble des données de l'essai

inter-laboratoires comme un échantillon représentatif de la population d'intérêt. Un tel postulat implique que les laboratoires participant à l'essai soient *représentatifs* de la population des laboratoires compétents pour utiliser cette méthode.

Pour le degré d'accord, il s'avère que le calcul dans un cadre aléatoire correspond exactement à l'approche initiale adoptée par le projet SMT, telle qu'elle a été mise en œuvre pour les deux essais qualitatifs (§ 2.1.3) !

Pour la concordance, le Tableau 27 indique le mode de calcul dans un cadre aléatoire.

Tableau 27 : Mode de calcul de la concordance dans un cadre aléatoire (n : nombre de répétitions, k_i : nombre de résultats positifs)

Etape	Formule
Probabilité pour un laboratoire i d'obtenir un résultat positif, $p_{+,i}$	$p_{+,i} = k_i/n$
Probabilité pour un laboratoire i d'obtenir un résultat négatif, $p_{-,i}$	$p_{-,i} = (n - k_i)/n$
Probabilité moyenne d'obtenir un résultat positif, P_+	$P_+ = \frac{1}{p} \sum_{i=1}^p p_{+,i}$
Probabilité moyenne d'obtenir un résultat négatif, P_-	$P_- = \frac{1}{p} \sum_{i=1}^p p_{-,i}$
Probabilité que deux <i>nouveaux</i> résultats dans une paire soient tous deux positifs ou négatifs, soit la concordance CON_a	$CON_a = P_+^2 + P_-^2$

Les auteurs ont appliqué les modèles fixe et aléatoire à un exemple et ils ont constaté que le modèle aléatoire donnait des estimations supérieures ou égales à celles du modèle fixe. Par ailleurs, comme il est logique de s'y attendre, les valeurs du degré d'accord ont toujours été supérieures à celles de la concordance dans le modèle aléatoire, à la différence du modèle fixe pour lequel cette propriété n'a pas été vérifiée pour une combinaison.

Un lien entre la formule du degré d'accord dans un cadre fixe (§ 2.1.4.7) et celle dans un cadre aléatoire (§ 2.1.3, Tableau 15) a été établi. Ainsi, la formule du degré d'accord dans un cadre fixe (ACC_f) peut être réécrite de la façon suivante :

$$ACC_f = \frac{1}{p} \sum_i \frac{p_{+,i}^2 + p_{-,i}^2 - 1/n}{1 - 1/n} \quad \text{Eq. 25}$$

Cette formule diffère de celle dans un cadre aléatoire seulement par la fraction $1/n$ soustraite au numérateur et au dénominateur.

Nous nous trouvons devant un dilemme pour le calcul du degré d'accord, puisque nous avons vu au § 2.1.4.7 que la formule retenue initialement, correspondant au modèle aléatoire, avait été abandonnée, en raison du biais introduit dans l'estimation. Rappelons que ce biais est égal à $\frac{2p_{+,i}(1-p_{+,i})}{n}$ pour un laboratoire. Comme la formule du degré d'accord dans un modèle

fixe diffère de la formule dans un modèle aléatoire par la fraction $1/n$, il serait intéressant d'étudier si, pour une valeur de n suffisamment grande – rappelons que l'EN ISO 16140

recommande $n = 8$ –, le cadre aléatoire pourrait être retenu, donnant une estimation pas trop différente de celle dans un cadre fixe. Ainsi, un biais significatif ne serait pas introduit par l'utilisation d'un cadre aléatoire.

Un argument supplémentaire qui conforte le choix du modèle aléatoire pour calculer le degré d'accord a été apporté dans le cadre des travaux de ISO/TC34/SC9/WG2 [van der Voet, 2004]. Ces travaux montrent que l'estimateur aléatoire du degré d'accord a presque toujours un plus petit carré moyen des erreurs que l'estimateur fixe : le premier estimateur est donc plus fidèle que le second.

En définitive, nous pouvons donc recommander l'utilisation du modèle aléatoire pour l'estimation de la concordance, comme du degré d'accord. Pour ce dernier, les arguments favorables au modèle aléatoire semblent l'emporter sur le modèle fixe, et il s'impose de toute façon en raison de l'objectif visé par les essais inter-laboratoires.

2.3.1.1.2 Interprétation du degré d'accord et de la concordance

Tout d'abord, il est ressorti, au sein du groupe ISO/TC34/SC 9/WG2, que le degré d'accord combine la répétabilité et la sensibilité d'une méthode. En effet, le degré d'accord, selon son équation dans le modèle aléatoire pour un laboratoire (Eq. 14, § 2.1.3), est une fonction non linéaire de la sensibilité p de la méthode ($p^2+(1-p)^2$). Dans ces conditions, le degré d'accord n'aurait pas de valeur ajoutée par rapport à la sensibilité, laquelle est de plus considérée comme la caractéristique essentielle d'une méthode qualitative.

Par ailleurs, l'absence d'interprétation pratique de ces critères a été souligné à juste titre. Comment prendre en compte les valeurs lorsqu'un laboratoire utilise la méthode d'analyse qui a été validée et, en particulier, comment juger de l'acceptabilité de répétitions effectuées au sein d'un laboratoire et entre laboratoires ? Aucune valeur critique n'a été définie. Ces questions méritent très certainement des réponses, qui ne peuvent pas être apportées à ce jour.

2.3.1.2 Approche alternative de la fidélité des méthodes qualitatives

Le degré d'accord et la concordance sont certes dans leur conception des équivalents, pour les méthodes qualitatives, des notions de répétabilité et de reproductibilité pour les méthodes quantitatives, mais leur estimation ne correspond pas aux écarts-types de répétabilité/reproductibilité.

Toujours dans le cadre des travaux du groupe ISO/TC34/SC9/WG2, des mesures de fidélité équivalentes aux écarts-types de répétabilité et reproductibilité ont été récemment proposées [Wilrich, 2004b]. Wilrich reconstitue la démarche de l'ISO 5725-2, tel que le Tableau 28 le résume.

Tableau 28 : Application aux méthodes qualitatives du modèle de l'ISO 5725-2 pour la fidélité des méthodes quantitatives (résultats y_{ij} du laboratoire i , n répétitions ; résultats notés '1' pour présence et '0' pour absence)

Étape	Formule
Moyenne des résultats (\bar{y}_i) : fraction des résultats positifs, soit la sensibilité estimée \hat{p}_i pour le laboratoire i	$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}$
Moyenne générale des résultats ($\bar{\bar{y}}$) : fraction de tous les résultats positifs, soit la sensibilité moyenne estimée pour tous les laboratoires, \hat{p}	$\bar{\bar{y}} = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n y_{ij} = \frac{1}{k} \sum_{i=1}^k \bar{y}_i$
Réécriture du modèle de base de l'ISO 5725-2 (§ 1.1.2.2, Eq. 1)	$y_{ij} = p + (p_i - p) + e_{ij}$ <p>où :</p> <p>p_i est la probabilité d'obtenir un résultat $y_{ij} = 1$ dans le laboratoire i (sa sensibilité) ;</p> <p>p la sensibilité moyenne dans la population des laboratoires à partir de laquelle les laboratoires participant à l'essai ont été sélectionnés ;</p> <p>e_{ij} l'erreur aléatoire du résultat y_{ij}, sous des conditions de répétabilité.</p>

Les différentes variances théoriques sont alors définies de la façon suivante.

Sous l'hypothèse d'indépendance des répétitions, les résultats y_{ij} du laboratoire i correspondent à un schéma de Bernoulli et leur somme $x_i = \sum_{j=1}^n y_{ij}$ (=nombre de résultats positifs) suit une loi binomiale de paramètres p_i et n . Les résultats y_{ij} ont donc pour moyenne p_i et pour variance $p_i(1 - p_i)$. Les variances de répétabilité et de reproductibilité en sont déduites, comme le présente le Tableau 29.

Tableau 29 : Variances théoriques de répétabilité et de reproductibilité pour une méthode qualitative (K laboratoires)

Etape	Formule
Variance intra-laboratoire du laboratoire i , σ_{ri}^2	$\sigma_{ri}^2 = p_i(1 - p_i)$
Variance de répétabilité σ_r^2 (malgré la non égalité des σ_{ri}^2)	$\sigma_r^2 = \frac{1}{K} \sum_{i=1}^K \sigma_{ri}^2 = \frac{1}{K} \sum_{i=1}^K p_i(1 - p_i)$
Variance inter-laboratoires σ_L^2 (une mesure de la variation entre les sensibilités p_i des laboratoires)	$\sigma_L^2 = \frac{1}{K-1} \sum_{i=1}^K (p_i - p)^2$
Variance de reproductibilité σ_R^2	$\sigma_R^2 = \sigma_r^2 + \sigma_L^2$

Une analyse de variance à effet aléatoire permet ensuite de calculer, avec les résultats d'un essai inter-laboratoires, les estimations de ces variances, comme dans l'ISO 5725-2.

L'auteur illustre ensuite cette approche par une simulation. Il remarque en particulier que la variance inter-laboratoires est nettement inférieure à la variance de répétabilité, la variance de reproductibilité n'étant alors que légèrement supérieure à la variance de répétabilité. Il s'agit effectivement d'une propriété des méthodes qualitatives, les différenciant des méthodes quantitatives : la variation entre 0 et 1 dans un laboratoire est seulement légèrement inférieure à cette variation entre laboratoires, même en cas de différence de sensibilité entre laboratoires. Il ne faudrait malgré tout pas en déduire que les méthodes qualitatives sont caractérisées par une variation entre laboratoires plus limitée que celle des méthodes quantitatives. Pour corriger cet effet d'une réponse binaire, l'auteur propose de définir la variance de répétabilité non pas comme celle des mesures 0 ou 1 individuelles, mais celle des résultats \hat{p} des n répétitions dans chaque laboratoire. La variance de répétabilité serait alors divisée par n .

Par ailleurs, Wilrich propose dans le même document un test d'égalité des sensibilités des laboratoires participants, le test du χ^2 d'indépendance, pour lequel il fixe cependant une condition d'application, soit un nombre de répétitions $n > 10$, de sorte que le test ait une puissance suffisante pour pouvoir détecter une différence de sensibilité entre laboratoires. Cette condition nous semble irréaliste en pratique !

En définitive, cette approche de l'estimation de la fidélité de méthodes qualitatives, alternative à celle du degré d'accord et de la concordance, apparaît séduisante à plusieurs titres :

- Elle représente un transfert aux méthodes qualitatives de l'approche de l'ISO 5725-2, conservant l'estimation de la répétabilité et de la reproductibilité par leur écart-type. Une cohérence entre l'estimation de la fidélité des méthodes quantitatives et celle des méthodes qualitatives est ainsi rétablie.

- L'interprétation et l'utilisation des critères de fidélité sont faciles, puisqu'il est possible de se calquer sur l'ISO 5725 et tous les protocoles qui en sont dérivés. Ainsi, lors de l'utilisation d'une méthode validée,
 - un laboratoire mettant en œuvre correctement la méthode devra respecter la sensibilité moyenne \hat{p} et l'écart-type de répétabilité s_r ,
 - les différences de sensibilité entre les laboratoires devront respecter l'écart-type inter-laboratoires s_L .

2.3.1.3 Perspectives

A l'issue de cette discussion, il apparaît de toute façon souhaitable d'engager une révision de l'annexe L de l'EN ISO 16140. En effet, si les critères de degré d'accord et de concordance étaient maintenus, il conviendrait d'introduire le calcul de la concordance selon un modèle aléatoire, et de maintenir le modèle aléatoire pour le degré d'accord, tel qu'initialement prévu par le projet SMT. Dans tous les cas de figure, il serait surprenant au plan statistique de mélanger les deux types de modèle.

L'approche proposée par Wilrich, qui introduit des écart-types pour l'estimation de la répétabilité et de la reproductibilité, apparaît une alternative tout à fait intéressante qui pourrait être retenue dans l'annexe de l'EN ISO 16140.

2.3.2 Justesse des méthodes de détection

Nous avons déjà eu l'occasion de mentionner que la caractéristique souvent considérée comme essentielle pour une méthode qualitative est la justesse. Ceci explique peut-être pourquoi la quasi-totalité des travaux sur la performance inter-laboratoires pour ces méthodes ne traitent que de la justesse. A l'inverse, les essais inter-laboratoires sur les méthodes quantitatives caractérisent presque uniquement leur fidélité. A l'évidence, la performance des méthodes qualitatives ne pourra pas être réduite à l'estimation de la fidélité, et la justesse devra être systématiquement estimée.

Lors de la revue des référentiels disponibles en 1^{ère} partie, un large consensus s'est dégagé pour caractériser la justesse d'une méthode qualitative par deux critères, sensibilité et spécificité, définis de façon similaire. AOAC complète certes ces critères par le taux de faux positifs et le taux de faux négatifs, mais ces deux derniers critères découlent directement des deux premiers.

Cependant, une divergence a été notée dans le calcul des critères de sensibilité et de spécificité. En effet, dans le cas particulier de la comparaison d'une méthode alternative à une méthode de référence, l'EN ISO 16140 s'en tient à une stricte comparaison des deux méthodes, alors que AOAC et NordVal prévoient chacun différemment une confirmation des résultats en cas de désaccord entre les deux méthodes.

Compte tenu de cette difficulté d'harmonisation des critères de sensibilité et spécificité dans le cadre de la validation d'une méthode alternative, une autre approche a été proposée, qui ne passe pas par la comparaison de la méthode alternative à la méthode de référence [Blodgett & Hitchins, 2000]. Une raison supplémentaire de s'affranchir de cette comparaison tient au protocole expérimental, qui requiert l'analyse d'échantillons appariés à faible concentration. Or, à ce niveau, la probabilité est forte qu'un des échantillons appariés soit contaminé et

l'autre pas. En effet, selon la distribution de Poisson généralement applicable aux faibles taux, la probabilité qu'une paire soit différente (un échantillon contenant le germe et l'autre pas) est :

- de 50 % au niveau 0,026 ufc/g,
- supérieure à 40 % entre 0,013 et 0,051 ufc/g, donc en particulier pour 0,04 ufc/g, soit 1 ufc/25 g. C'est-à-dire la limite de détection habituellement requise pour les méthodes de détection, qui est généralement incluse dans le protocole expérimental de détermination de la justesse.

Deux modèles sont introduits, portant seulement sur la méthode à valider : un modèle constant et un modèle logistique. Chacun utilise une procédure de maximum de vraisemblance pour estimer les critères, soit une estimation des critères qui maximise la vraisemblance des données observées.

Le **modèle constant** repose sur l'hypothèse que la probabilité pour une méthode de donner un résultat positif, ne dépend que de la présence ou de l'absence du micro-organisme dans l'échantillon, et non pas de sa concentration. Ce modèle comprend deux étapes :

1. Le calcul de la probabilité que la méthode donne un résultat positif en fonction :
 - des taux de faux positifs et de faux négatifs,
 - et des probabilités que l'échantillon contienne ou pas le microorganisme cible, selon la loi de Poisson.
2. En prenant un ensemble de combinaisons <matrice x niveaux de contamination>, les taux de faux positifs et de faux négatifs sont déduits en maximisant les probabilités d'observation du polynôme formé par le nombre de résultats positifs par combinaison.

Ce modèle constant s'applique bien aux méthodes conventionnelles, telles que les méthodes de référence normalisées, qui comportent une phase d'enrichissement augmentant la concentration du germe ciblé. On peut alors considérer que la capacité de la méthode à donner un résultat positif ne dépend pas de la concentration initiale du microorganisme recherché, à condition toutefois de se trouver au-dessus de la limite de détection. Ce modèle a été mis en œuvre à plusieurs reprises, à l'occasion d'essais inter-laboratoires réalisés pour AOAC sur des méthodes de détection de *L. monocytogenes*.

Dans le **modèle logistique**, une fonction relie la probabilité que la méthode donne un résultat positif au nombre de microorganismes présents dans l'échantillon. Ce modèle s'applique donc aux méthodes ne comportant pas de phase de croissance ; ce qui n'est pas le cas des méthodes de référence et est même rarement rencontré pour les méthodes alternatives.

En définitive, cette caractérisation de la justesse d'une méthode qualitative, qu'elle soit une méthode de référence ou une méthode alternative, apparaît intéressante à plusieurs titres. En effet, elle permet de :

- s'affranchir du besoin, pour les méthodes alternatives, de les comparer à une méthode de référence, ce qui allègerait grandement les essais inter-laboratoires ;
- prendre en compte la distribution des microorganismes à faible taux dans les échantillons, et éviter de considérer comme faux négatif un résultat issu d'un échantillon en fait non contaminé ; cas de figure fréquent à des taux voisins de la limite de détection, soit 1 à 5 ufc/g.

Il conviendrait donc de prendre en compte le modèle constant lors de la révision de l'EN ISO 16140.

2.4 Essais d'aptitude des Laboratoires Nationaux de Référence « Lait »

Résumé

Le deuxième type d'essais inter-laboratoires a été mis en œuvre par le biais d'essais d'aptitude que nous avons organisés pour les Laboratoires Nationaux de Référence « Lait ». L'organisation, le protocole expérimental, ainsi que le traitement des résultats retenus pour les déterminations quantitatives ou qualitatives, sont présentés. Deux exemples sont choisis pour illustrer cette présentation : le dénombrement des coliformes et la détection de *Salmonella*. Les résultats de ces deux essais sont résumés, ainsi que les enseignements qui ont pu en être tirés.

2.4.1 Contexte

La Directive 92/46/CEE du 16 juin 1992 *arrétant les règles sanitaires pour la production et la mise sur le marché de lait cru, de lait traité thermiquement et de produits à base de lait* (Journal Officiel de la Commission Européenne du 14/09/92) porte sur l'hygiène de la filière laitière, comprenant le lait cru à la sortie de la ferme jusqu'au produit laitier transformé en sortie usine. Par cette directive, notre laboratoire a été désigné Laboratoire Communautaire de Référence pour l'analyse du lait et des produits laitiers (noté en abrégé LCR Lait). Le LCR Lait est chargé de la mise en œuvre analytique de cette directive, soit la responsabilité des aspects analytiques des contrôles officiels effectués conformément à la directive. Dans ce but, il coordonne un réseau de Laboratoires Nationaux de Référence (en abrégé « LNR Lait »), nommés par les autorités nationales chargées de l'hygiène des produits laitiers. Chaque LNR Lait est chargé à son tour de coordonner un réseau des laboratoires agréés pour effectuer les analyses requises par les contrôles officiels. La coordination des LNR Lait comprend l'évaluation de leur compétence et l'amélioration de cette compétence par différents moyens, tels que formations, séminaires,...

Pour évaluer la compétence des LNR Lait, le LCR Lait organise des essais inter-laboratoires d'aptitude, à raison d'un essai par an en microbiologie, consacré chaque fois à un germe différent. Dans ces conditions, nous considérons qu'il s'agit d'essais d'aptitude ponctuels, selon la typologie introduite par le Guide ISO 43-1 (§ 1.2.1.3). L'objectif n'est d'ailleurs pas de se substituer aux essais d'aptitude continus, tel que ceux organisés par CECALAIT en France. La participation à ce dernier type d'essais est de toute façon une exigence de l'accréditation, et tous les LNR sont accrédités. La spécificité des essais d'aptitude organisés par le LCR Lait réside dans l'évaluation de la capacité des LNR à utiliser de façon satisfaisante les méthodes de référence prescrites pour le contrôle officiel. En d'autres termes, il s'agit du cas particulier d'essai d'aptitude où la méthode d'analyse est imposée.

Nous avons coordonné et exploité les résultats de six essais d'aptitude :

- le dénombrement des coliformes [Lafarge *et al*, 2000] ;
- le comptage des cellules somatiques [Voitoux et Lombard, 2002] ;
- deux essais sur la détection de *L.monocytogenes* [Lafarge & Lombard, 1999 ; Le Barillec *et al*, 2003] ;
- la détection de *E.coli* O 157 [Voitoux *et al*, 2001] ;
- et la détection de *Salmonella* [Voitoux & Lombard, 2002].

Notre laboratoire a complété l'essai d'aptitude sur la détection de *E.coli* O157 par une étude de l'applicabilité aux produits laitiers de la méthode de référence correspondante [Voitoux *et al*, 2002].

Nous avons choisi de présenter deux de ces six essais, car les autres sont très proches de ceux-ci, dans leur conception et leur exploitation :

1. l'un porte sur une détermination quantitative, soit le dénombrement de coliformes, dont le rapport figure en Annexe 3.8 ;
2. l'autre porte sur une détermination qualitative, soit la détection de *Salmonella*, dont le rapport figure en Annexe 3.9.

2.4.2 Matériel et méthodes

2.4.2.1 Mode opératoire et rapport d'essai

Aucun mode opératoire spécifique n'a été préparé, la méthode d'analyse normalisée devant être utilisée telle quelle :

- Détection de *Salmonella*. Les laboratoires ont eu le choix entre deux méthodes de référence, la norme EN ISO 6579, méthode dite « horizontale », et la norme FIL 93B/ISO 6785, méthode spécifique des produits laitiers. En effet, ces deux normes peuvent être utilisées pour le contrôle officiel.
- Dénombrement des coliformes. Seule la norme FIL 73/ISO 5541-1, méthode prescrite pour le contrôle officiel, pouvait être utilisée.

Un rapport d'essai pour chacun des deux essais a été préparé, selon un format moins détaillé que celui employé pour le projet SMT.

2.4.2.2 Échantillons

On a choisi le lait pasteurisé pour l'essai *Salmonella* et le lait cru pour l'essai sur les coliformes. Cette matrice liquide se prête bien à l'homogénéisation, notamment lors de la contamination artificielle. On a procédé à une contamination mixte pour l'essai sur les coliformes : mélange de flore naturellement présente dans le lait cru et de flore ajoutée. Ce mode de contamination présente le net avantage de se rapprocher des conditions de contamination naturelle, en analyse de routine.

Les niveaux de contamination étaient les suivants :

- pour la détection de *Salmonella* : trois niveaux ciblés à 5-15 ufc/25 ml, 50-70 ufc/25 ml et un témoin négatif ;
- pour le dénombrement des coliformes : quatre niveaux ciblés à $5 \cdot 10^2 - 10^3$ ufc/ml, $5 \cdot 10^3 - 10^4$ ufc/ml, $5 \cdot 10^4 - 10^5$ ufc/ml et un témoin négatif avec du lait pasteurisé. La flore de coliformes naturellement présente dans le lait cru a été prise en compte pour atteindre ces niveaux.

Les souches utilisées étaient :

- un mélange de souches de différents genres, pour les coliformes ;
- toutes isolées de lait ou de fromage, et non pas des souches de collection qui sont souvent isolées chez l'Homme.

Deux modes de préparation des échantillons différents ont été utilisés :

- Pour l'essai sur *Salmonella*, notre laboratoire a préparé les échantillons par contamination individuelle de chacun des échantillons, c'est-à-dire directement dans les flacons dans lesquels le lait pasteurisé avait été distribué. Le niveau de contamination de l'inoculum a été déterminé précisément par densité optique, la pente de la courbe d'étalonnage de la souche utilisée ayant été auparavant déterminée. La contamination individuelle de chaque échantillon a été choisie pour mieux maîtriser l'homogénéité de la contamination.
- La préparation et l'envoi des échantillons pour l'essai sur les coliformes ont été sous-traités à CECALAIT qui a suivi son mode habituel de préparation de contamination globale pour du lait liquide (voir § 1.2.4.2).

L'homogénéité et la stabilité des matériaux pour essai ont été contrôlées de façon suivante :

- Pour l'essai sur *Salmonella*, l'étude d'homogénéité des échantillons n'a pas été nécessaire, en raison de la contamination individuelle. En revanche, nous avons vérifié l'homogénéité et le niveau exact de la suspension bactérienne utilisée pour contaminer les échantillons. La stabilité d'un lait contaminé a été étudiée sur 4 jours et à plus haut niveau de contamination que pour l'essai, afin de pouvoir dénombrer *Salmonella*, avec un dénombrement en double chaque jour. Une ANOVA a été réalisée.
- Pour l'essai sur les coliformes, l'homogénéité a été étudiée de la même façon que pour le projet SMT sur les trois lots aux différents taux de contamination. La stabilité a été étudiée avec 5 échantillons d'un seul matériau à un niveau de contamination, analysés en double sur 4 jours différents pendant 6 jours.

Le schéma classique de doubles aveugles a été suivi, soit 6 échantillons par laboratoire pour *Salmonella* et 8 échantillons par laboratoire pour les coliformes. Les modalités de transport étaient les mêmes que pour le projet SMT, si ce n'est que le suivi de température est plus simple : à la place du dispositif enregistreur, chaque participant prend la température d'un flacon d'eau ajouté dans le paquet.

Au total, 21 LNR Lait ont pris part à l'essai sur *Salmonella* et 20 à l'essai sur les coliformes. Tous les LNR nommés n'ont pas participé à chaque essai puisqu'ils étaient 29 en 2000 et 24 en 2002.

2.4.3 Méthodes de traitement des données

2.4.3.1 Dénombrement des coliformes

Après avoir transformé en logarithmes décimaux les dénombrements réalisés par les laboratoires, la médiane des résultats a été retenue comme valeur de consensus assignée. Nous avons vu en qu'il s'agissait d'une des statistiques définies par le Guide ISO 43-1 et le protocole IUPAC/AOAC, et qu'elle est recommandée lorsqu'on détermine la valeur de référence d'un essai d'aptitude portant sur une analyse empirique, comme un dénombrement bactérien (voir § 1.2.2.2). Nous pouvons même considérer ici que la valeur de consensus a été déterminée par des laboratoires « experts », puisque qu'il s'agit de laboratoires nationaux de référence. Dans ce cas, il s'agit du cas de figure que le protocole IUPAC/AOAC a placé en 1^{ère} position, selon un ordre de préférence.

Pour chacun des niveaux de contamination, les résultats de chaque participant (répétitions et moyenne) ont été représentés graphiquement. Comme la méthode d'analyse était imposée, les écart-types de répétabilité et de reproductibilité ont été calculés selon les estimateurs robustes présentés précédemment (§ 2.1.2), afin d'apprécier la compétence globale du réseau des LNR Lait à mettre en œuvre la méthode considérée. Pour évaluer la performance de chacun des participants, nous avons retenu, comme le RAEMA, les scores z individuels, tels que définis par le Guide ISO 43-1 et le protocole IUPAC/AOAC :

$$z = (m_i - X) / s \quad \text{Eq. 26}$$

où : m_i est la moyenne des doubles du laboratoire i , X est la valeur assignée et s l'écart-type cible pour l'essai d'aptitude.

Nous avons choisi ici pour s l'écart-type de reproductibilité obtenu lors de l'essai, avec l'ensemble des participants.

Pour évaluer la compétence de chaque laboratoire, ont été appliquées les valeurs limites de 2 et 3, indiquées dans les référentiels mentionnés ci-dessus, notamment le protocole IUPAC/AOAC qui précise leur origine et l'hypothèse de normalité sous-jacente des scores z (§ 1.2.2.3.5).

2.4.3.2 Détection de *Salmonella*

La performance globale du réseau de laboratoires a été appréciée par les critères de spécificité pour le témoin négatif, et de sensibilité pour les niveaux de contamination bas et élevé. Ces critères ont été calculés de la même manière que dans le cadre du projet SMT (§ 2.1.3).

2.4.4 Dénombrement des coliformes dans le lait

La stabilité d'un des lots de lait contaminé a été satisfaisante sur 3 trois jours, de même que l'homogénéité des 3 lots de lait contaminés, à l'exception du lot à fort taux de contamination. Le nombre élevé de dilutions pour le dénombrement peut expliquer cette dispersion des résultats, qui serait ainsi plutôt imputable à une variabilité analytique qu'à une hétérogénéité du matériau.

La plupart des laboratoires ont reçu les échantillons sous 48 h et dans un bon état, à température inférieure à 8°C. 16 des 18 participants ont utilisé complètement la méthode imposée. La *performance globale* du réseau des LNR Lait dans l'application de la méthode de référence pour les contrôles officiels a été satisfaisante, puisque caractérisée par une répétabilité de 0,1 \log_{10} et une reproductibilité de 0,16-0,24 \log_{10} , en fonction des niveaux.

La Figure 6 donne un exemple de représentation graphique illustrant la *performance des participants*, parmi les graphes figurant en annexe 3.8. Un tel graphe permet d'avoir une vision rapide et claire de cette performance. Ainsi, dans ce cas, la majorité des laboratoires est caractérisée par une justesse satisfaisante, soit une moyenne des doubles proche de la valeur assignée, ainsi que par une bonne fidélité, en termes d'écart entre doubles. Cependant, le laboratoire n°13 et surtout le laboratoire n°20 se distinguent par leurs performances moins satisfaisantes, tant en termes de justesse que de fidélité.

Quant à la *performance individuelle des participants en terme de justesse*, appréciée par le score z , on peut noter, à partir du tableau des scores z par laboratoire et par niveau (Tableau C de l'annexe 3.8) que :

- 4 laboratoires ont obtenu un score non satisfaisant, soit $|z| > 3$;
- dont 3 ont eu également un autre score douteux, soit $2 < |z| \leq 3$.

Au moins pour ces trois derniers laboratoires, dont 2 sont apparus anormaux visuellement dans la Figure 6, une déficience de performance a pu être établie.

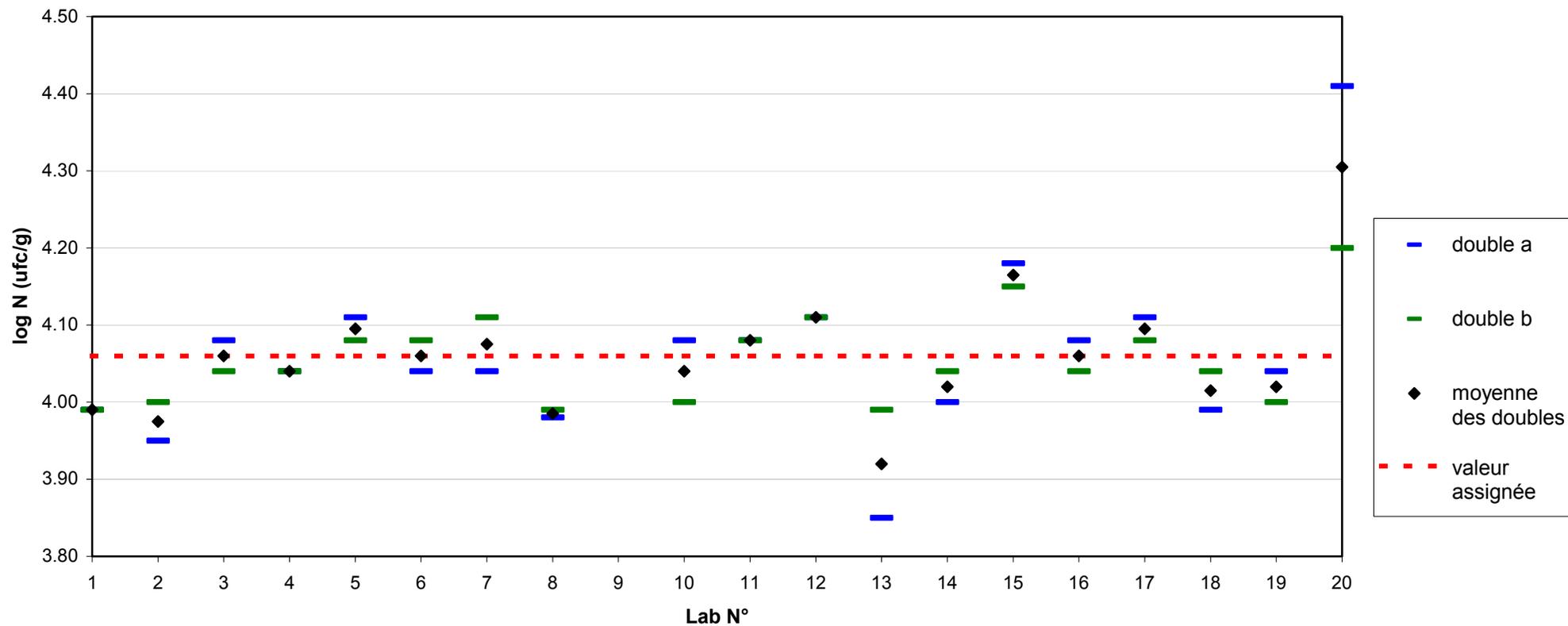


Figure 6 : Représentation graphique de la performance des laboratoires – Essai d’aptitude des LNR Lait sur le dénombrement des coliformes dans le lait – Exemple du lot à taux moyen de contamination (4,06 log ufc/ml)

2.4.5 Détection des *Salmonella*

L'étude d'homogénéité de la suspension d'inoculation a été satisfaisante. Quant à l'étude de stabilité, une décroissance nette du nombre de germes a été constatée de J0 à J+1 (différence significative, $p < 0,001$), mais ensuite la contamination s'est stabilisée. Cette diminution initiale, peut-être due à un stress des microorganismes lors du stockage du lait à température réfrigérée, n'a pas eu de conséquence négative pour deux raisons :

1. la stabilisation ultérieure de la contamination ;
2. la valeur assignée est la valeur de consensus des participants, et non la valeur initiale de contamination.

21 des 22 participants ont reçu les échantillons en 48 h, en bon état et à température inférieure à 8°C, sauf les échantillons ayant atteint le laboratoire en 3 jours. 14 laboratoires ont utilisé la méthode horizontale, l'EN ISO 6579, 11 la méthode sectorielle, l'ISO 6785 et 3 les deux méthodes en parallèle.

La performance globale du réseau de LNR Lait dans l'application d'une des méthodes de référence a été caractérisée par sa spécificité (93,2 %) et par sa sensibilité (97,7 % au bas niveau de contamination et 100 % au niveau élevé). Soit une performance du réseau très satisfaisante.

Nous avons cherché à identifier avec les participants concernés les raisons de leurs résultats incorrects. Un laboratoire a trouvé les 2 témoins positifs à l'issue de 48 h d'enrichissement et aucune raison n'a pu être trouvée pour expliquer ces résultats. Un autre laboratoire a trouvé positif un des témoins et négatif un des échantillons à bas niveau. N'ayant pas rapporté d'écart au protocole, ce laboratoire a sans doute inversé les résultats des deux échantillons, ou il a inversé leur identification lors de l'analyse.

Par ailleurs, cet essai a indiqué une performance équivalente et satisfaisante des deux méthodes, soit une justesse de 97 % pour chacune des deux méthodes.

2.4.6 Conclusion sur les deux essais d'aptitude

Ces deux essais ont permis d'apprécier la compétence globale du réseau des LNR Lait à mettre en œuvre les méthodes sélectionnées pour le contrôle officiel du lait et produits laitiers, ainsi que d'évaluer la compétence individuelle de chacun de ces laboratoires.

La tentation a été grande pour la Commission Européenne, prescripteur de ces essais, d'utiliser les résultats, en particulier surtout les scores z des dénombrements, pour « noter » les LNR Lait et proposer une échelle de « sanctions » en cas de résultats non satisfaisants, tel qu'inciter les autorités nationales à suspendre la nomination du laboratoire concernés. Nous nous sommes vivement élevés contre cette intention qui aurait enlevé leur caractère pédagogique à ces essais et aurait sans doute amené à prendre des décisions non justifiées, compte tenu des limites expérimentales des essais en microbiologie des aliments. En effet, les valeurs limites du score z reposent sur l'hypothèse de normalité de distribution de ce score ; ce qui n'est pas forcément vérifié. Par ailleurs, en microbiologie des aliments, il est toujours

difficile de s'assurer de l'homogénéité et de la stabilité de la totalité des échantillons envoyés aux laboratoires. Aussi un score non satisfaisant peut être dû à un défaut de l'échantillon envoyé au laboratoire, et non à une déficience du laboratoire lui-même.

En revanche, nous nous sommes engagés à donner une suite aux essais d'aptitude que nous organisons. Cette action, menée pour l'essai sur *Salmonella*, passe par le biais d'échanges individuels avec les laboratoires ayant obtenus des résultats non satisfaisants, cherchant à identifier la ou les causes des résultats incorrects et à envisager les actions correctives à mettre en œuvre. Il est prévu qu'une visite de ces laboratoires puisse avoir lieu, si besoin est.

2.5 Valorisation des résultats d'essais inter-laboratoires en microbiologie des aliments

Résumé

L'organisation d'essais inter-laboratoires en microbiologie des aliments présente un intérêt à plusieurs titres.

Les essais visant à établir les performances d'une méthode constituent un élément incontournable de validation complète de la méthode. Ils permettent de plus la comparaison entre méthodes, ils apportent une crédibilité scientifique aux méthodes de référence, et ils permettent de les améliorer.

Les essais d'aptitude constituent un élément essentiel du système d'assurance qualité externe d'un laboratoire, ainsi qu'un moyen d'évaluation externe pour l'accréditation et pour l'organisation du contrôle officiel des aliments.

Les essais de caractérisation des matériaux de référence permettent d'obtenir des outils utiles à l'assurance qualité du laboratoire.

Ces trois types d'essais permettent finalement au laboratoire, sous certaines conditions, d'estimer l'incertitude de mesure attachée à ses résultats.

Nous avons discuté plus haut de l'intérêt et des limites des différents critères utilisés pour caractériser les méthodes microbiologiques, et des perspectives d'amélioration ont été envisagées. Nous allons maintenant tenter de recenser les différents avantages des essais inter-laboratoires, en fonction de chacun des trois objectifs qui peuvent leur être assignés, à savoir la caractérisation de la performance d'une méthode d'analyse, l'évaluation de la compétence de laboratoires et la caractérisation de matériaux de référence. Nous terminerons sur une finalité commune, l'estimation de l'incertitude de mesure.

Ayant concentré jusqu'ici notre travail sur l'organisation et l'exploitation des essais inter-laboratoires, nous nous intéressons donc maintenant à l'*utilisation* que l'on peut faire des résultats d'un essai inter-laboratoires. Cependant, bon nombre d'intérêts de ces essais ont déjà été esquissés en première partie ou lors de la présentation des résultats de la phase expérimentale.

2.5.1 Essais inter-laboratoires visant à établir les performances d'une méthode

2.5.1.1 Validation des méthodes

Lorsque les essais inter-laboratoires caractérisent les performances d'une méthode d'analyse, leur principal intérêt nous semble être de disposer de méthodes dont la fiabilité est établie de façon objective, c'est-à-dire validée. La validation de méthodes constitue notamment un outil d'assurance qualité pour le laboratoire. En effet, les limites de répétabilité et de reproductibilité de la méthode, définies dans le § 1.1.6.1, peuvent être utilisées par le laboratoire pour établir qu'il maîtrise la méthode d'analyse, en montrant qu'il respecte ces limites [Scotter & Wood, 1996].

Pour les méthodes qui nous intéressent, la validation ne saurait se réduire à une étude des performances de la méthode dans un seul laboratoire. Lorsqu'une méthode d'analyse est appelée à être mise en œuvre dans plus d'un laboratoire, un essai inter-laboratoires est fortement souhaitable. Et une validation complète ne peut pas se concevoir sans que la méthode ait fait l'objet d'un tel essai. C'est le principe de base de la reconnaissance des méthodes officielles AOAC ; il ne connaît pas d'exception. Nous avons d'ailleurs souvent désigné ce type d'essais inter-laboratoires par « essais inter-laboratoires de validation de méthodes », indiquant par là même que la validation complète d'une méthode était indissociable de la réalisation d'un tel essai. A ce stade de la discussion, il semble nécessaire de clarifier le concept de validation, utilisé à de nombreuses reprises. Le fascicule de documentation AFNOR FD V 01-000 [AFNOR, 1999] définit ainsi la validation :

« Action de confirmer par examen et apport de preuves tangibles obtenues par des études statistiques intra-laboratoires et/ou inter-laboratoires que les exigences particulières pour un usage spécifique prévu de la méthode d'analyse sont satisfaites. Dans le cadre du présent document, les exigences particulières peuvent porter sur le domaine d'application, la linéarité, la spécificité, la fidélité et la justesse. »

De cette définition, nous retiendrons deux aspects :

1. La validation ne passe pas obligatoirement par la réalisation d'un essai inter-laboratoires ; une étude interne peut suffire pour prétendre au statut de méthode validée. C'est pour cette raison que nous avons employé ci-dessus le terme de validation complète (*full validation*) pour désigner une validation incluant un essai inter-laboratoires. Le terme validation employé seul peut donc prêter à confusion. De même, lorsqu'une méthode n'a fait l'objet que d'une étude interne, le terme de validation intra-laboratoire devrait être employé.
2. La validation d'une méthode implique que des objectifs de performance prédéfinis soient satisfaits. En d'autres termes, il ne suffit pas d'avoir réalisé un essai inter-laboratoires sur une méthode pour qu'elle soit validée ; il faut que les valeurs de performance obtenues soient jugées satisfaisantes au regard de l'utilisation prévue de la méthode. Il s'agit là de l'interprétation des résultats d'analyse, sur laquelle nous reviendrons plus loin à l'occasion de l'incertitude de mesure.

2.5.1.2 Autres acquis

Outre la caractérisation de la performance d'une méthode donnée, les essais inter-laboratoires peuvent également permettre de comparer de façon objective plusieurs méthodes de référence, établies pour une même détermination par des organisations ou des pays différents, sans avoir à faire appel à des avis d'experts ou à des consensus qui comportent toujours une part de subjectivité.

Si l'obtention de valeurs inter-laboratoires de fidélité ou de justesse est une condition *sine qua non* de reconnaissance d'une méthode par AOAC, nous avons vu en 1^{ère} partie que ce n'était pas le cas des méthodes microbiologiques de référence CEN ou ISO (§ 1.1.11.4). Le projet SMT a voulu combler ce manque pour les six principales méthodes normalisées. Il a ainsi nettement renforcé la crédibilité de ces normes d'un point de vue scientifique, en matière de degré de validation. De plus, comme ces essais ont été organisés selon un référentiel reconnu au niveau international, à savoir les lignes directrices AOAC et la norme EN ISO 16140, la fiabilité de la méthode a pu acquérir une **reconnaissance internationale**. Ainsi, nous avons vu que AOAC a reconnu l'une d'elles comme méthode « officielle ». Le projet SMT a donc concouru à élargir le champ de reconnaissance de ces normes à un pays qui est incontournable, ainsi qu'à l'ensemble des pays sous sa sphère d'influence. Enfin, cette validation ne peut que renforcer le degré de crédibilité des normes auprès du Codex Alimentarius. En effet, le Codex définit des critères microbiologiques, en se référant à des méthodes d'analyse qui sont développées par des organisations telles que ISO ou AOAC, et qui doivent être de préférence validées [Lahellec, 1998].

Toujours au chapitre des bénéfices pour la normalisation des méthodes d'analyse, nous avons noté [Langton *et al*, annexe 3.6] que les valeurs de fidélité des méthodes normalisées peuvent, lorsqu'elles sont trop élevées, mettre en évidence :

- un défaut de normalisation de la méthode, au sens d'un mode opératoire comportant des ambiguïtés ou des choix trop larges,
- ou une faiblesse de la méthode en matière de fidélité. Ce fut effectivement le cas de l'essai sur la méthode de détection de *L.monocytogenes*.

Au-delà des indications apportées par les valeurs de fidélité, nous avons vu que chacun des essais du projet SMT avait permis d'améliorer les normes, en mettant le doigt sur tel ou tel aspect de la méthode normalisée, qui n'avait pas été compris de façon homogène par les participants, ou qui présentait des difficultés de mise en œuvre. Ainsi, les six méthodes de référence ont été révisées ou amendées pour bénéficier de ces recommandations.

En ce qui concerne les méthodes commerciales, utilisées en alternative des méthodes de référence, la validation selon un référentiel et par un système reconnu internationalement confère à ces méthodes un « passeport » leur permettant d'être utilisées dans différents pays, sans avoir à être revalidées dans chaque pays [Archer, 1996]. De telles méthodes validées peuvent alors devenir partie intégrante d'une part des auto-contrôles effectués par les opérateurs, et d'autre part des contrôles officiels visant à assurer la sécurité microbiologique des aliments.

Enfin, il s'avère préférable d'utiliser pour les contrôles officiels des méthodes qui ont été validées par essais inter-laboratoires, que ce soient des méthodes de référence ou des méthodes alternatives. Cette validation est non seulement garante de la fiabilité des méthodes utilisées, mais elle renforce également le statut de méthodes dont les résultats peuvent être

utilisés dans le cadre de procédures judiciaires. Cette préférence pour des méthodes validées par essais inter-laboratoires est affichée par plusieurs pays, dont le Royaume-Uni ([Scotter & Wood, 1996], voir § 1.1.11.3), l'Australie [Fleet, 1996] et les USA [Jackson, 1996]. Aux Etats-Unis, les méthodes validées notamment par AOAC sont introduites dans le BAM (*Bacteriological Analytical Manual*), qui regroupe toutes les méthodes utilisées par la US FDA (*US Food and Drug Administration*). Néanmoins, ce manuel ne comprend pas uniquement des méthodes validées par essais inter-laboratoires. Des méthodes n'ayant pas fait l'objet d'un essai inter-laboratoires sont retenues, lorsque de telles méthodes ne sont pas disponibles pour une détermination donnée, tel que pour les microorganismes émergents, ou lorsqu'elles sont jugées pas praticables.

2.5.2 Essais d'aptitude des laboratoires

Nous pouvons tenter de résumer ainsi le bénéfice des essais d'aptitude des laboratoires : concourir à évaluer de façon objective et quantifiée l'aptitude des laboratoires à effectuer correctement une analyse microbiologique. L'évaluation est *objective et quantifiée*, car elle repose sur l'obtention de résultats d'analyse, obtenus en « aveugle », c'est-à-dire sans connaître la nature ou le nombre de germes cibles à trouver, par opposition à une évaluation de compétence fondée sur un jugement humain, l'auditeur.

Nous avons vu, notamment lors de la présentation en 1^{ère} partie du Guide ISO 43-1 (§ 1.2.1.2), que les essais d'aptitude représentent maintenant un point essentiel de l'assurance qualité externe des laboratoires, ainsi que, dans le cadre de l'accréditation, un des deux éléments d'évaluation externe des laboratoires, avec l'audit. La suite donnée aux résultats d'essais d'aptitude peut favoriser la mise en œuvre d'un des aspects souvent délaissés du processus d'assurance qualité, à savoir les actions correctrices, à la suite de résultats non satisfaisants. En fonction de la conception des essais et de leur mode d'exploitation mais surtout compte tenu de l'utilisation qui en est faite (absence de sanction), la participation à des essais d'aptitude constitue un facteur de progrès indéniable pour le laboratoire, notamment par le biais des actions correctrices.

Dans le cadre plus particulier du contrôle officiel des produits laitiers, nous avons pu montrer que les essais d'aptitude représentent un moyen important à la disposition du Laboratoire Communautaire de Référence pour :

- coordonner le réseau des Laboratoires Nationaux de Référence ;
- apprécier la capacité du réseau à mettre en œuvre les méthodes de référence retenues pour les contrôles officiels ;
- identifier les laboratoires pouvant présenter des déficiences, en complément évidemment de leur participation à des systèmes continus d'essai d'aptitude.

Aujourd'hui, la communauté des laboratoires, en France comme en Europe, peut bénéficier d'une offre variée de circuits continus d'essais d'aptitude couvrant une bonne partie du domaine de la microbiologie des aliments, tant en termes de matrices que de germes. La plupart de ces systèmes ont pu atteindre un niveau de maîtrise suffisant, notamment en ce qui concerne l'homogénéité et la stabilité des échantillons, pour pouvoir proposer des systèmes performants dont la qualité peut être désormais reconnue par l'accréditation.

2.5.3 Essais inter-laboratoires de caractérisation des matériaux de référence

Ici aussi, tentons de résumer l'intérêt des matériaux de référence : disposer de valeurs de référence pour l'assurance qualité interne et externe du laboratoire.

Tout comme les essais d'aptitude, les matériaux de référence s'inscrivent donc également et même de façon plus exclusive, dans la démarche d'assurance qualité des laboratoires. Ils permettent au laboratoire, lorsque les analyses sont effectuées sur des produits dont la nature, l'état et les germes sont proches des matériaux de référence, de disposer de valeurs indépendantes de la méthode utilisée et/ou du laboratoire qui la met en œuvre. En fait, l'indépendance du matériau de référence vis-à-vis de la méthode d'analyse utilisée pour le caractériser est, nous l'avons vu, très relative, surtout pour les matériaux de référence quantitatifs, caractérisés par un nombre de germes cibles. On peut en revanche considérer que cette indépendance est acceptable pour les matériaux de référence qualitatifs, caractérisant la présence du germe cible. En effet, les différentes techniques de confirmation, tant conventionnelles (biochimiques ou sérologiques) que moléculaires, comme les sondes à ADN, la PCR, le ribotypage, ou la PFGE..., permettent de confirmer avec certitude la présence d'un germe dans un échantillon, lorsqu'il en a été isolé. Concernant l'isolement, il est tout de même nécessaire de s'assurer de la sensibilité de la méthode, c'est-à-dire sa capacité à retrouver le germe cible présent dans la prise d'essai. La sensibilité est en général satisfaisante pour les méthodes de référence, fondées sur une ou plusieurs étapes d'enrichissement.

Le faible nombre de matériaux de référence certifiés à ce jour en microbiologie des aliments peut être compensé par le recours aux matériaux de référence externes, issus d'essais d'aptitude. Pour constituer un matériau de référence externe, il semble raisonnable que le matériau possède une stabilité au regard du germe considéré se comptant en mois. Ainsi, les matrices fraîches, comme le lait liquide ou le fromage frais de CECALAIT, dont la flore est en état d'activité même ralentie, et dont la stabilité est inférieure au mois, ne peuvent donc pas être retenues. En revanche, les matrices à l'état sec en poudre ou congelées, comme les poudres de lait déshydratées du RAEMA dont la flore est en état de latence, peuvent être utilisées à cette fin. Le nombre de réseaux d'essais d'aptitude en microbiologie des aliments augmentant rapidement depuis quelques années, la disponibilité en matériaux de référence devrait donc aussi augmenter, tant en termes de matrices que de germes.

2.5.4 Estimation de l'incertitude de mesure

2.5.4.1 Contexte

Nous allons maintenant envisager une autre retombée, et non des moindres, des essais inter-laboratoires pour l'interprétation des résultats d'analyse : il s'agit de l'incertitude de mesure. Nous verrons en quoi cette dernière peut être estimée à l'aide des différents types d'essais inter-laboratoires.

Une fois un résultat d'analyse obtenu, il peut être utilisé à différentes fins : connaissance d'un produit, de sa contamination, épidémiologie, investigation de toxi-infections alimentaires, conformité du produit à des critères microbiologiques,.... Quelle que soit l'utilisation finale du résultat, il est important d'en connaître la fiabilité. En particulier, dans le cadre des contrôles officiels, l'interprétation d'un résultat d'analyse consiste à décider de la conformité de l'échantillon par rapport à une limite qualitative ou quantitative, fixée par un critère microbiologique. Compte tenu de la variabilité des dénombrements microbiens, que nous eu l'occasion d'illustrer, il semble nécessaire, d'un point de vue scientifique, de prendre en compte cette variabilité dans le jugement de conformité à une limite. Sinon, un résultat proche de la limite pourrait être considéré à tort conforme, ou non conforme, selon les cas.

Un consensus s'est petit à petit formé au niveau mondial, quel que soit le domaine analytique considéré, pour caractériser cette fiabilité par la notion d'incertitude de mesure, qui caractérise la dispersion des valeurs autour du résultat d'analyse. Alors que nous avons jusqu'à présent étudié la variabilité attachée à une méthode d'analyse, l'incertitude de mesure est attachée quant à elle :

- au sens strict, à un résultat unique obtenu dans des conditions données de laboratoire, personnel, matériel,...
- et par extension à un laboratoire, si les conditions analytiques restent constantes.

Le large consensus autour de cette notion et de la façon de l'estimer s'est concrétisé par la publication du *Guide pour l'expression de l'incertitude de mesure (GUM)* [ISO, 1995]. Cependant, l'incertitude de mesure est restée longtemps étrangère au champ de l'analyse des aliments et en particulier de l'analyse microbiologique, en raison des retards de l'approche métrologique dans ce domaine. L'introduction en 1999 dans la norme EN ISO 17025 d'une exigence d'estimation de l'incertitude des mesures a fortement incité, sinon obligé les laboratoires accrédités à se pencher sur cette question. Aujourd'hui, de nombreux laboratoires agro-alimentaires se sont engagés dans une démarche de caractérisation de l'incertitude attachée à leurs résultats d'analyse microbiologique.

2.5.4.2 Estimation en microbiologie des aliments

Dans ce contexte, le comité ISO/TC 34/SC 9 a décidé d'établir un guide pour l'estimation de l'incertitude de mesure des dénombrements bactériens (ISO/TS 19036). Il a été décidé de traiter du cas des déterminations qualitatives dans un deuxième temps, compte tenu de la complexité du sujet et du manque d'approches disponibles à ce jour. La première version de ce guide, datant de 2003 et dont la rédaction nous a été confiée, pose les principes de la démarche. L'approche recommandée par le GUM, décomposant l'incertitude selon ses différentes sources dans le processus de mesure, est apparue difficile à appliquer au cas de l'analyse microbiologique des aliments. Pour cette dernière, il est en effet difficile de bâtir un modèle réellement exhaustif du processus de mesure et une source d'incertitude peut être facilement omise, d'où un risque de sous-estimer l'incertitude. Une approche plus globale a été préférée, faisant appel à l'écart-type de reproductibilité sur le résultat final du processus de mesure. Cette approche est en cohérence avec l'ISO/TS 21748, document de portée générale, et avec les recommandations du Codex Alimentarius.

Trois possibilités ont été retenues pour estimer l'incertitude-type, classées par ordre décroissant de préférence, dont deux font directement appel aux essais inter-laboratoires objets de notre travail :

1. L'écart-type de reproductibilité intra-laboratoire ;
2. L'écart-type de reproductibilité inter-laboratoires estimé par un essai de validation de méthode ;
3. L'écart-type de reproductibilité inter-laboratoires estimé par un essai d'aptitude.

Des conditions sont définies pour qu'un laboratoire donné puisse utiliser les résultats des essais inter-laboratoires pour estimer sa propre incertitude de mesure :

- Dans le cas d'essais de validation de méthode :
 - Le laboratoire doit strictement utiliser en routine le protocole opératoire de la méthode qui a été soumise à l'essai inter-laboratoires.
 - Le laboratoire doit démontrer que son biais et sa fidélité sont compatibles avec ceux attendus sur la base des estimations de répétabilité et de reproductibilité de la méthode, dérivées de l'essai inter-laboratoires.
- Dans le cas d'essais d'aptitude :
 - Le laboratoire doit avoir utilisé pendant l'essai d'aptitude la méthode qu'il pratique en routine.
 - L'écart-type de reproductibilité doit avoir été calculé sur la méthode d'analyse utilisée par le laboratoire.
- Dans tous les cas :
 - Les échantillons qui ont été utilisés dans l'essai doivent être comparables, en termes de matrice, de germe cible, de niveau de contamination, d'état physiologique et de flore annexe, à ceux analysés en routine.
 - L'essai inter-laboratoires doit avoir couvert correctement toutes les sources d'incertitude, en particulier l'influence de la prise d'essai. Dans le cas de facteurs additionnels non couverts par l'essai, une fois ces derniers estimés, il est possible de les combiner avec l'estimation de l'incertitude issue de l'essai.

Par ailleurs, il est possible d'utiliser des matériaux de référence, notamment les matériaux de référence externes ou certifiés à la suite des essais inter-laboratoires que nous avons présentés en 1^{ère} partie, pour estimer l'écart-type de fidélité intermédiaire. Cette estimation est l'option préférée dans le projet de guide ISO/DTS 19036 pour estimer l'incertitude de mesure en microbiologie. L'emploi de matériaux de référence est même la seule possibilité d'estimer cet écart-type sur plusieurs jours différents.

2.6 Limites expérimentales des essais inter-laboratoires en microbiologie des aliments

Résumé

Les limites des essais inter-laboratoires, au regard de leur utilisation, tiennent au protocole expérimental suivi, et pour l'essentiel à des questions de représentativité.

D'une façon générale, la représentativité des valeurs de performance de méthodes, lorsque ces valeurs sont issues d'essais inter-laboratoires, peut être affectée par le mode de sélection des participants, ainsi que par les échantillons choisis. Par ailleurs, les essais d'aptitude ne sauraient à eux seuls permettre d'apprécier en totalité la compétence des laboratoires.

D'autres limites, liées à la nature vivante de l'objet de l'analyse, sont spécifiques de la microbiologie. Elles sont dues à la nature des échantillons utilisés, à la difficulté de caractériser une valeur vraie, et au degré de validation d'une méthode par le seul biais d'un essai inter-laboratoires.

Après avoir envisagé aux § 2.2 et 2.3 les limites et perspectives en matière d'exploitation des essais inter-laboratoires de validation de méthode, nous allons maintenant aborder de façon plus générale les limites des différents essais inter-laboratoires en microbiologie. Comme pour leur intérêt, nous nous attacherons à l'utilisation qui peut être faite de ces essais, dont les limites résident essentiellement dans le volet « protocole expérimental ». Ces limites sont liées pour la plupart à la spécificité de la microbiologie des aliments et tiennent surtout à des questions de représentativité.

2.6.1 Limites inhérentes aux essais inter-laboratoires en général

Avant d'envisager les contraintes spécifiques de la microbiologie des aliments, nous commencerons par des considérations plus générales sur deux des trois types d'essais inter-laboratoires, à savoir les essais de validation de méthodes et les essais d'aptitude.

2.6.1.1 Essais de validation de méthodes

La question principale qui se pose quant à l'utilisation des essais de validation de méthode semble être la suivante. Les estimations de répétabilité, reproductibilité et justesse, issues d'un essai inter-laboratoires, sont-elles représentatives des écarts acceptables et attendus dans l'utilisation en routine de la méthode ?

Comme éléments intervenant dans cette représentativité, on peut citer la méthode de sélection des laboratoires participant à l'essai. Nous avons déjà vu, lors de la présentation de l'ISO 5725-1, qu'il ne fallait pas choisir des laboratoires experts. En effet, les valeurs de fidélité obtenues seraient trop exigeantes lors de leur utilisation ultérieure par d'autres

laboratoires, par exemple pour juger de l'acceptabilité de résultats obtenus sous des conditions de répétabilité ou de reproductibilité. A ce titre, on peut considérer que le projet SMT a été peut-être un peu trop exigeant dans la sélection des participants, dans un souci de réussite des essais, mais au détriment de la représentativité. L'ISO 5725-1 requiert de constituer un échantillon représentatif de la population des utilisateurs, au sein de laquelle les laboratoires sont choisis au hasard. Cette prescription apparaît difficile à réaliser, compte tenu de l'effort requis pour participer à ces essais, notamment dans le domaine de la microbiologie, caractérisée par des analyses particulièrement lourdes. De fait, il serait difficile de trouver de cette manière un nombre suffisant de participants, sans déployer des efforts démesurés.

Interviennent également dans cette représentativité la nature des échantillons utilisés dans l'essai, ainsi que leur traitement dans le cadre de l'essai. Cette question sera abordée plus loin, d'une façon spécifique à la microbiologie.

2.6.1.2 Essais d'aptitude

D'une part, la question de la représentativité des résultats d'un essai inter-laboratoires se pose ici aussi, mais d'une façon différente : les résultats obtenus dans le cadre d'un essai d'aptitude ne sont-ils pas optimisés par rapport à ceux obtenus en routine ?

En effet, outre les aspects liés à la nature et au traitement des échantillons, qui seront abordés ci-dessous, la tentation est grande pour un laboratoire, se sachant évalué, d'accorder un soin tout particulier à la réalisation des analyses dans le cadre d'un essai d'aptitude. L'idéal serait que ces échantillons soient analysés de façon « banalisée », ainsi que le Guide ISO 43-1 le demande (§ 1.2.1.12). Cet objectif apparaît difficile à mettre en œuvre, l'échantillon devant toujours être identifié à son arrivée au laboratoire, même s'il peut être ensuite codé pour le lancement des analyses. Le laboratoire peut être également tenté de répéter l'analyse. Une façon de limiter cette possibilité est de réduire l'échantillon à une seule prise d'essai : c'est ce qui a été fait lors de la mise en œuvre des essais, tant dans le cadre du projet SMT que pour les Laboratoires Nationaux de Référence. Cependant, ce risque ne peut pas être totalement éliminé : en microbiologie, il est possible de dupliquer des sous-analyses à partir de la même suspension-mère. Le risque de collusion entre laboratoires a été également évoqué ; il peut être néanmoins facilement évité en codant différemment les échantillons envoyés à chaque laboratoire. C'est ce qui a été effectué dans les essais mis en œuvre.

D'autre part, le Guide ISO 43-1 insiste certes sur le fait que l'essai d'aptitude ne constitue qu'un des éléments permettant d'apprécier la compétence des laboratoires (§ 1.2.1.2). Il est néanmoins tentant de réduire l'évaluation de la compétence à ce seul moyen, pour les raisons suivantes. Une première raison est de considérer ce moyen d'évaluation comme « objectif » car quantifié. Cependant, le terme « objectif » a été mis entre guillemets, car nous avons vu dans le paragraphe précédent que la performance d'un laboratoire peut être surévaluée intentionnellement. Son évaluation peut être également faussée par le manque de représentativité des échantillons (voir § 2.6.2.1). Une autre raison est d'ordre pratique. Une évaluation des laboratoires se limitant aux seuls essais d'aptitude est plus facile à mettre en œuvre qu'un processus complet d'accréditation de laboratoires, passant par la réalisation d'audits et l'évaluation des rapports d'audits.

Si ce dernier risque semble écarté à ce jour en France par les organisateurs d'essais d'aptitude, le COFRAC et les services publics chargés de l'agrément des laboratoires pour le contrôle officiel des aliments, d'autres pays, notamment anglo-saxons, sont moins prudents. Nous avons également noté que les services de la Commission Européenne en charge de l'hygiène des aliments auraient volontiers accordé ce rôle aux essais d'aptitude que nous organisons en tant que Laboratoire Communautaire de Référence, si nous n'avions pas insisté pour éviter cette dérive.

Enfin, une diversité sinon une disparité des choix en matière de critères d'évaluation des laboratoires a été notée en 1^{ère} partie lors du recensement des référentiels et des pratiques actuelles. Aucun critère ne s'est actuellement imposé, car chacun présente ses propres limites. Des nouvelles pistes ont été récemment investiguées, telles qu'une représentation graphique appelée « Naji Plot » [Robouch P. *et al*, non publié]. Dans ce graphe, le score z est porté en abscisse, et le ratio $(u_{\text{lab}}/u_{\text{ref}})^2$ en ordonnée, où : u_{lab} est l'incertitude de mesure de chaque laboratoire participant, et u_{ref} est l'incertitude de mesure de la valeur de consensus. Ce graphe est dérivé du score de performance E_n , introduit par le guide ISO 43-1 (voir § 1.2.1.8.1). Un tel graphe permet de comparer simultanément les valeurs des participants à la valeur de consensus, ainsi que les incertitudes des participants à l'incertitude de référence.

2.6.2 Limites liées aux spécificités de la microbiologie des aliments

Comme nous l'avons déjà annoncé, les limites liées spécifiquement à notre domaine de travail concernent le protocole expérimental des différents types d'essais inter-laboratoires présentés et mis en œuvre. Ces limites découlent principalement de la caractéristique essentielle de l'analyse microbiologique, à savoir la nature vivante de l'analyte. Nous traiterons d'abord des aspects liés aux échantillons utilisés dans les essais, avant d'aborder la question centrale de l'existence d'une valeur vraie, pour enfin tenter d'apprécier le degré de validation d'une méthode d'analyse microbiologique au travers d'un essai inter-laboratoires.

2.6.2.1 Échantillons utilisés dans les essais

Les aspects liés aux échantillons occupent une place importante dans la conception des protocoles expérimentaux des essais inter-laboratoires dans notre domaine. Nous avons pu le constater lors de la présentation des référentiels et des protocoles des différents essais inter-laboratoires mis en œuvre. La nature vivante de l'analyte rend particulièrement critique le choix de la matrice utilisée pour préparer les échantillons, le mode de contamination, les conditions de transport des échantillons et les délais d'analyse. Tous ces paramètres, mal maîtrisés, peuvent facilement invalider les résultats d'un essai inter-laboratoires. Ce sont ces difficultés qui ont sans doute empêché pendant longtemps la pratique des essais inter-laboratoires dans notre domaine et qui continuent à la freiner, surtout lorsqu'il s'agit de valider des méthodes.

2.6.2.1.1 Représentativité des échantillons vis-à-vis de l'analyse microbiologique

Nous pouvons résumer par la question suivante les limites des essais inter-laboratoires liées aux échantillons : est-ce que les conditions expérimentales de préparation et de traitement des échantillons utilisés ne s'écartent-elles pas trop des *conditions réelles d'analyse microbiologique des aliments* ?

Pour tenter de répondre à cette question, les aspects suivants peuvent être distingués :

- **Contamination artificielle par rapport à la contamination naturelle.** La contamination artificielle est apparue comme la règle générale car elle permet de mieux maîtriser, d'une part, l'homogénéité des échantillons au regard du germe cible, et d'autre part, le niveau de contamination des échantillons. La contrepartie réside justement dans son caractère artificiel. Une flore ajoutée dans une matrice risque en effet d'avoir un comportement différent d'une flore naturellement présente dans cette même matrice, qui aura eu la possibilité d'établir des interactions avec les constituants de cette matrice, ainsi qu'avec la flore annexe. La flore annexe peut inhiber le germe cible, ou produire des métabolites qui réduisent le pouvoir inhibiteur des milieux de culture sélectifs, favorisant ainsi le développement de cette même flore annexe sur les milieux de culture, au détriment du germe ciblé [Struijk, 1996]. Selon les recommandations des différents référentiels, une flore annexe a souvent été ajoutée dans les échantillons utilisés pour les essais que nous avons présentés, mais un ajout au moment de la préparation des échantillons ne saurait reconstituer toute la complexité des interactions établies dans le temps entre le germe cible, la flore annexe et la matrice.
- **Etat physiologique des bactéries.** Cet aspect découle directement du précédent, car généralement l'état physiologique des bactéries est différent en contamination artificielle et en contamination naturelle. En effet, la bactérie est souvent stressée dans le second cas, surtout dans les produits alimentaires transformés. Bon nombre de procédés de traitement, tels que la réfrigération, le salage, la pasteurisation, ... visent justement à détruire, sinon à freiner le développement des bactéries, en provoquant des lésions subléthales structurelles et métaboliques [Leclercq *et al*, 2000 ; Struijk, 1996]. Certains protocoles, tel que la norme EN ISO 16140, prévoient de stresser les bactéries inoculées. Nous avons appliqué un stress pour contaminer les échantillons utilisés dans les essais pour les Laboratoires Nationaux de Référence, en conservant au froid la suspension bactérienne, et un stress par lyophilisation a été appliqué aux inocula de deux matrices utilisées pour le projet SMT. Cependant, ces modes de stress ne sauraient être identiques à celui appliqué aux bactéries dans l'environnement de la matrice.

La conséquence de ces deux aspects pour la question posée est que la détection ou le dénombrement du germe cible seront différents en contamination artificielle ou en contamination naturelle. Généralement, compte tenu des éléments avancés ci-dessus, la détection ou le dénombrement seront plus faciles en contamination artificielle. En particulier, la détection de germes pathogènes sera facilitée en l'absence de flore annexe, qui souvent domine les pathogènes [Leclercq *et al*, 2000]. Les dénombrements en contamination artificielle donneront souvent des valeurs plus élevées qu'en contamination naturelle.

Ainsi, en ce qui concerne la validation des méthodes, un essai inter-laboratoires pourra donner une vision trop optimiste de la performance d'une méthode, par rapport à sa performance lors d'une utilisation en routine, avec des échantillons naturellement contaminés. De même, un essai d'aptitude pourra surestimer la performance d'un laboratoire, par rapport à sa compétence à analyser des échantillons naturellement contaminés.

Cette discussion rejoint directement la condition principale d'utilisation des matériaux de référence : la *représentativité du matériau de référence* pour son utilisation prévue, soit une des trois propriétés découlant de la définition d'un matériau de référence selon le Guide ISO 30 (§ 1.3.2.1). Les limites relatives à la contamination artificielle et au stress artificiel ne peuvent pas être évitées lors de l'utilisation des matériaux de référence. Un stress, au moins par dessiccation, est en effet systématiquement appliqué pour un matériau de référence en microbiologie. En revanche, il convient de choisir le matériau de référence de telle façon que sa matrice et sa (ses) souche(s) se rapprochent le plus possible de celles analysées en routine. Ce choix n'est pas possible pour les matériaux de référence certifiés, puisqu'une seule série du BCR, utilisant du lait en poudre, est actuellement disponible au niveau international. Les matériaux de référence externes, eux, laissent un choix nettement plus large, comme nous l'avons noté plus haut.

2.6.2.1.2 Représentativité du mode de préparation des échantillons

Deux aspects relatifs au mode de préparation des échantillons peuvent être envisagés :

1. **Le mode opératoire de traitement des échantillons.** En présentant les essais de validation de méthodes, dans le cadre du projet SMT, nous avons noté au § 2.1.1.3 que ceux-ci reprenaient les modes opératoires des méthodes normalisées. Toutefois, ces modes opératoires normalisés ont été complétés par les modalités de traitement de chacun des trois types d'échantillons en vue de préparer soit la suspension mère dans le cas de méthodes de dénombrement, soit le bouillon d'enrichissement, dans le cas de méthodes de détection. Or l'objectif de ces essais était de caractériser la performance de méthodes déjà normalisées, qui ne décrivent pas cette étape initiale de l'analyse (même si une série de normes récentes, les parties 2 à 4 de l'ISO 6887, définissent en partie ces modalités par grande famille d'aliments). Dans ces conditions, nous pouvons nous demander si ces essais n'ont pas réduit la variabilité réelle des méthodes normalisées.
2. **L'homogénéisation des échantillons.** Un autre aspect plus large, concernant les essais inter-laboratoires de validation de méthode et d'aptitude, est la conséquence de l'homogénéisation des échantillons sur l'interprétation des résultats. L'homogénéisation la plus minutieuse est requise pour tendre vers une comparabilité optimale, bien que jamais parfaite, des échantillons soumis à l'essai. Dans ce cas, l'estimation de la fidélité de la méthode ou des laboratoires qui en résulte ne reflètera pas la variabilité liée à l'hétérogénéité de contamination des échantillons. Au sens strict, la fidélité de la méthode sera correctement estimée, puisque la méthode d'analyse débute une fois la prise d'essai constituée à partir de l'échantillon reçu par le laboratoire. Cependant, l'hétérogénéité de contamination des échantillons pour laboratoire étant souvent importante, cet aspect ne peut pas être éludé dans l'interprétation des résultats d'une analyse, ainsi que le projet ISO/DTS 19036 sur l'incertitude de mesure en microbiologie l'a acté. Il faut bien être conscient que la fidélité d'une méthode estimée par un essai inter-laboratoires n'inclut pas cette source de variabilité.

2.6.2.1.3 Représentativité des échantillons vis-à-vis du domaine d'application de la méthode d'analyse

Nous venons d'aborder les aspects liés au mode de préparation de l'échantillon utilisé dans l'essai : contamination artificielle, stress absent ou artificiel et mode de traitement des échantillons.

Nous nous proposons maintenant d'envisager plus largement la possibilité de caractériser, par le biais d'un essai inter-laboratoires ayant porté sur un nombre limité de matrices et de souches du microorganisme cible, les performances d'une méthode d'analyse dite « horizontale », dont le domaine d'application est très vaste, en termes de matrices et de germes :

- Types de matrice. Il s'agit de l'ensemble des denrées alimentaires qui nécessitent un contrôle microbiologique, ainsi que des aliments pour animaux et des échantillons d'environnement de production agro-alimentaire.
- Variété de microorganismes. En fonction de la cible de la méthode, il peut s'agir de l'ensemble des genres d'une même famille, ou l'ensemble des espèces d'un même genre, ou l'ensemble des sérotypes/génotypes d'une même espèce.

En ce qui concerne le nombre de matrices, nous avons remarqué que AOAC veut privilégier cet aspect, en préconisant d'utiliser dans un essai inter-laboratoires visant à valider une méthode horizontale au moins six types de matrice, alors que le projet SMT en a retenu trois, pour limiter la lourdeur du protocole expérimental.

Selon une lecture au sens strict de la norme ISO 5725-1, les données de fidélité et de justesse déterminées au cours de l'essai inter-laboratoires ne sont applicables qu'aux combinaisons <matrice x germe x niveau de contamination> étudiées, puisque l'hypothèse d'indépendance ou de dépendance prévisible de la fidélité de la méthode vis-à-vis de ces combinaisons n'est pas vérifiée dans notre domaine. Cette dépendance non prévisible est, comme décrit plus haut, attribuable aux interactions s'établissant entre les différentes composantes de la combinaison <type de matrice x souche du microorganisme cible (nature et état physiologique) x niveau de contamination>, ainsi qu'avec la flore annexe.

Nous avons vu qu'en pratique il était néanmoins possible de caractériser, sur la base des résultats d'un essai inter-laboratoires, la fidélité de certaines méthodes par des valeurs qui pouvaient être considérées comme générales, au moins par niveau de contamination. Ce fut le cas, pour le projet SMT, des méthodes de dénombrement de *L. monocytogenes* et de dénombrement des staphylocoques. Cependant, cette généralisation porte seulement sur le domaine d'application de la méthode, les types de matrices et les souches cibles, et non pas sur l'applicabilité des valeurs de fidélité en routine, avec des échantillons naturellement contaminés.

Le comité ISO/TC 34/SC 9 a été conscient de ces limites. S'il a certes décidé d'inclure, sur les recommandations du projet SMT, des valeurs générales de répétabilité et de reproductibilité dans le texte principal des normes, ont été incluses en annexe des normes les valeurs pour chaque combinaison <matrice x niveau de contamination> testée lors de l'essai. Un avertissement mentionne que ces dernières valeurs peuvent ne pas être applicables à d'autres combinaisons que celles testées.

2.6.2.2 Question de la valeur vraie

Certes, la détermination de la valeur vraie représente un défi pour de nombreux domaines analytiques, dont l'analyse physico-chimique des aliments, et on ne parvient souvent, au mieux, qu'à l'approcher. Mais en microbiologie des aliments, les obstacles sont particulièrement importants, et les valeurs déterminées sont généralement éloignées de la valeur vraie.

Dès la présentation des concepts de base de l'ISO 5725-1 (§ 1.1.2.1), nous avons mis en évidence l'impossibilité, à ce jour, de définir une valeur quantitative vraie dans notre domaine. En particulier, les techniques actuelles d'analyse microbiologique ne permettent pas de connaître le nombre vrai de microorganismes présents dans l'échantillon. En effet, a été évoqué le lien intime existant entre le mode opératoire des méthodes conventionnelles de la microbiologie pasteurienne, et le résultat obtenu, soit le développement de la bactérie cible sur/dans une gélose, ou le nombre de bactéries cibles se développant sur/dans une gélose. Ce lien, attribuable à la nature vivante de l'analyte, est avéré lorsqu'il s'agit d'un dénombrement. Il est encore plus étroit lorsque la cible du dénombrement n'est pas définie taxonomiquement. Dans ce dernier cas, la cible de l'analyse est uniquement définie par la méthode employée, laquelle est alors dite empirique.

Outre la méthode d'analyse, nous pouvons également évoquer d'autres aspects liés à la nature vivante de l'analyte.

Ainsi, il convient de prendre en compte l'état physiologique de la bactérie, puisque le résultat du dénombrement dépend aussi de l'état initial du germe cible, en phase de croissance, ou altéré, ou stressé [Leclercq *et al*, 2000].

Par ailleurs, les bactéries sont soumises à des variations génétiques, qui peuvent être particulièrement importantes et rapides, compte-tenu de la vitesse de multiplication de ces microorganismes. Or les techniques classiques de microbiologies font appel, pour mettre en évidence un germe donné, à des propriétés qui sont déterminées génétiquement. On peut citer la capacité du microorganisme cible à se développer dans un milieu donné, à une température donnée, ou les réactions biochimiques/sérologiques caractéristiques qui sont à la base des techniques d'identification/confirmation des méthodes normalisées. De ce fait, une méthode d'analyse qui a été établie de façon à cibler au mieux un germe à un moment donné peut rapidement ne plus être adaptée à ce germe, qui aurait subi des variations génétiques notables. En d'autres termes, l'écart à la valeur vraie se serait amplifié.

L'impossibilité actuelle de définir une valeur vraie représente certes un argument de poids en faveur de la démarche d'harmonisation d'une seule méthode de référence au niveau international, par le biais de la normalisation [Lombard *et al*, 1996], qui permet de définir une valeur de référence acceptée. Néanmoins, cette impossibilité entraîne plusieurs inconvénients notables :

- La justesse d'une méthode d'analyse est difficile à caractériser, notamment par le biais d'un essai inter-laboratoires.
- En matière de validation de méthodes alternatives, en résulte une disparité des approches ISO, AOAC et NordVal pour la comparaison à la méthode de référence, avec ou sans confirmation des résultats de cette dernière (voir § 1.1.10.3.3 et 1.1.12.3).

- Dans le cadre des essais d'aptitude, l'estimation de la valeur de consensus est rendue délicate lorsque plusieurs méthodes empiriques sont utilisées. Ces difficultés ont été soulignées à l'occasion de la présentation du protocole IUPAC/AOAC (§ 1.2.2.2).
- Lorsqu'ils sont de nature quantitative, les matériaux de référence, notamment certifiés, voient leur portée restreinte. En effet, la valeur quantifiée est dépendante de la méthode d'analyse utilisée lors de l'essai inter-laboratoires de certification. La valeur certifiée n'est alors valable au sens strict que si le matériau de référence est analysé avec cette même méthode : la portée des matériaux de référence s'en trouve nettement limitée !

Des perspectives de dénombrement par des techniques nouvelles de biotechnologie, telles que la PCR quantitative, laissent entrevoir la possibilité d'atteindre la connaissance du nombre vrai de microorganismes, par la mise en évidence d'une caractéristique propre à chaque individu bactérien, tel qu'un fragment d'ADN de son noyau. Ces techniques, bien qu'en plein essor, ne sont pas encore utilisables pour ce type d'application.

2.6.2.3 Degré de validation d'une méthode

Nous pouvons finalement aborder la question plus large du degré de validation d'une méthode d'analyse microbiologique par un essai inter-laboratoires. Au chapitre de l'intérêt de ces essais, nous avons souligné que l'obtention, à l'issue d'un essai inter-laboratoires, de valeurs satisfaisantes de justesse et de fidélité est nécessaire à la validation complète d'une méthode d'analyse. Cependant, la caractérisation inter-laboratoires des performances d'une méthode d'analyse n'est pas suffisante pour sa validation complète.

En effet, un essai inter-laboratoires permet d'estimer la justesse et la fidélité d'une méthode d'analyse ; critères qui ne sauraient résumer toutes les facettes de la performance qu'on peut attendre d'une méthode d'analyse. Il convient donc qu'une validation complète repose également sur une étude intra-laboratoire qui permette de caractériser d'autres critères de performance, tels que la limite de détection, la limite de quantification pour les méthodes de dénombrement et la spécificité (inclusivité/exclusivité sur souches pures) [Lombard, 2001]. L'EN ISO 16140 décrit ces différents critères, et la façon de les estimer. De plus, cette étude intra-laboratoire permet d'estimer la justesse de la méthode sur :

- un nombre bien plus élevé de combinaisons <matrice x souche x niveau de contamination> que lors de l'essai inter-laboratoires ;
- et surtout sur des échantillons naturellement contaminés.

Cette caractérisation peut être également obtenue par le biais des essais réalisés par plusieurs laboratoires sur les méthodes retenues pour être normalisées par le CEN, l'ISO ou la FIL (voir § 1.1.11.4). Dans ce cas, chaque laboratoire utilise des produits alimentaires caractéristiques de son pays, naturellement contaminés par des souches locales du microorganisme cible.

De ce fait, de telles études intra-laboratoires ou « multi-laboratoires » viennent compléter l'essai inter-laboratoires, en suppléant à une partie des limites de ce dernier, surtout en matière de représentativité des échantillons utilisés, dont la déficience est liée à la contamination artificielle et au nombre limité de matrices et souches. La valeur ajoutée d'une validation intra-laboratoire existe pour toute validation de méthode, mais elle nous semble revêtir une importance toute particulière dans le cas de l'analyse microbiologique, compte-tenu de la nécessité de valider une méthode d'analyse sur des matrices alimentaires naturellement contaminées.

2.6.2.4 Conclusion

Les limites propres à la microbiologie des aliments, envisagées ci-dessus, peuvent être résumées de la façon suivante. Il s'avère particulièrement difficile de reproduire artificiellement une analyse qui doit répondre à des exigences très élevées. A savoir rechercher ou dénombrer un analyte *vivant*, souvent présent en minorité parmi une *flore annexe* compétitrice, dans un *milieu* (l'aliment) qui complique l'essai car pouvant posséder des propriétés antimicrobiennes, ou des substances interférant avec la procédure analytique. Un tel défi pourrait être comparé à la recherche d'une aiguille dans une botte de foin souvent hostile [Archer, 1996] !

Conclusion de la deuxième partie

La mise en œuvre de deux types d'essais inter-laboratoires en microbiologie des aliments a permis d'illustrer et de discuter différents aspects des deux volets d'un essai inter-laboratoires :

1. **Organisation et conception.** Nous avons mis en application dans le domaine de la microbiologie des aliments une façon de répondre aux règles générales établies quel que soit le secteur analytique. Les difficultés rencontrées et la manière de les surmonter ont été analysées, tant pour organiser des essais de validation de méthodes dans le cadre du projet SMT, que pour mener des essais d'aptitude pour les Laboratoires Nationaux de Référence. Ont été en particulier abordées les questions liées aux échantillons utilisés (nature, contamination) et à leur transport, de façon à respecter les exigences d'homogénéité et de stabilité.
2. **Exploitation des résultats.**
 - Les résultats des essais inter-laboratoires ont été traités selon la norme EN ISO 16140 et les critères introduits par le projet SMT.
 - Pour les **méthodes quantitatives**, nous avons mené une comparaison de trois approches normalisées d'estimation des critères de fidélité, soit deux approches robustes et l'approche classique non robuste. Cette comparaison a été illustrée par l'exemple du dénombrement de *L. monocytogenes*. Les conclusions suivantes ont pu en être tirées :
 - Compte tenu d'une part de la forte cohérence constatée entre les estimations issues de ces trois approches, et d'autre part d'une discussion sur les avantages comparatifs des statistiques robustes, nous avons conclu à l'intérêt de retenir une estimation robuste des critères de répétabilité et de reproductibilité.
 - L'approche robuste qui s'est avérée préférable pour des raisons statistiques est celle définie par l'EN ISO 16140. Nous avons cependant recommandé (i) d'adopter un meilleur estimateur d'échelle à la base de l'écart-type de reproductibilité, et (ii) d'inclure une représentation graphique initiale de données.
 - Nous avons discuté du bien-fondé de la transformation initiale des données en logarithmes décimaux ; laquelle ne se justifiait pas dans notre exemple. Nous avons recommandé de ne pas tenir pour acquis l'hypothèse de log-normalité de distribution des données.
 - Pour les **méthodes qualitatives**, les deux dimensions de performance inter-laboratoires ont été abordées :
 - En matière de *fidélité*, nous avons discuté des limites des critères de degré d'accord et de concordance, récemment introduits pour caractériser la fidélité de ces méthodes, et nous avons recommandé d'utiliser un cadre aléatoire pour leur calcul. Une approche très récemment introduite nous a semblé représenter une alternative intéressante, puisqu'il s'agit d'un transfert aux méthodes qualitatives de la méthodologie habituelle des écarts-types de répétabilité/reproductibilité pour les méthodes quantitatives.

- Quant à la *justesse*, nous avons mis en évidence l'intérêt d'une estimation qui ne passe pas par la comparaison à une valeur de référence, et qui prend en compte la distribution des microorganismes à faibles taux.

Les **avantages** que nous avons distingués pour les différents types d'essais inter-laboratoires résident dans l'utilisation que l'on peut faire de leurs résultats :

- Les essais visant à établir les *performances d'une méthode* permettent :
 - d'établir de façon objective la fiabilité de la méthode, et ainsi de constituer un élément incontournable de sa validation complète ;
 - d'apporter une crédibilité scientifique aux méthodes d'analyse microbiologiques normalisées, dont les défauts de « normalisation » peuvent être en outre mis en évidence.
- Les *essais d'aptitude* constituent un moyen d'évaluer de façon objective l'aptitude des laboratoires à effectuer correctement une analyse microbiologique. Ils représentent par là même un aspect essentiel de l'assurance qualité externe, ainsi qu'un des éléments de coordination d'un réseau de Laboratoires Nationaux de Référence.
- Les *essais de caractérisation de matériaux de référence* permettent de définir une valeur de référence qui ne soit pas biaisée par le laboratoire, ni par la méthode d'analyse dans certains cas. Les matériaux de référence, soit certifiés, soit externes, sont comme les essais d'aptitude des moyens de mettre en œuvre l'assurance qualité interne et externe dans un laboratoire.
- Les résultats de ces trois types d'essais peuvent être utilisés pour estimer l'*incertitude de mesure* en microbiologie des aliments.

Quant aux **limites** de ces essais inter-laboratoires, elles sont inhérentes au protocole expérimental utilisé, et tiennent principalement à des questions de représentativité :

- Pour les essais inter-laboratoires en général, le mode de sélection des laboratoires peut influencer la représentativité des estimations de fidélité et de justesse d'une méthode pour son utilisation en routine. Dans le cadre d'un essai d'aptitude, la performance d'un laboratoire peut être artificiellement optimisée, et il serait dangereux de réduire l'évaluation d'un laboratoire à ce seul moyen.
- Les limites spécifiques au domaine de la microbiologie des aliments se rapportent essentiellement à la nature vivante de l'analyte, et tiennent à deux aspects :
 - Les échantillons utilisés dans les essais. Il s'agit de la représentativité de la contamination et du mode de préparation des échantillons, ainsi que de la représentativité des échantillons vis-à-vis du domaine d'application de la méthode d'analyse.
 - L'impossibilité actuelle de définir une valeur vraie, surtout en matière de nombre de microorganismes.

Enfin, si la validation complète d'une méthode d'analyse ne peut pas se concevoir sans qu'elle ait été soumise avec succès à un essai inter-laboratoires, ce dernier ne saurait suffire à la valider, et d'autres critères de performance doivent être étudiés.

Conclusion générale

A l'issue de ce travail, trois caractéristiques ressortent de la revue des documents de référence et des pratiques, et permettent d'en tirer un enseignement :

1. Les référentiels de portée générale se sont avérés poser les concepts de façon satisfaisante pour la microbiologie des aliments. Ils nécessitent cependant d'être complétés par des référentiels particuliers, en raison des spécificités de la microbiologie des aliments, soit principalement la nature vivante de l'objet de l'analyse.
2. Les référentiels spécifiques soit à l'analyse des aliments, soit à l'analyse microbiologique des aliments, respectent généralement les concepts des référentiels de portée générale. De plus, les particularités des référentiels et pratiques se justifient la plupart du temps par les spécificités techniques de notre domaine.
3. En revanche, une disparité apparaît entre référentiels et pratiques d'un même domaine, surtout pour la validation de méthodes.

En ce qui concerne ce dernier point, ce manque de cohérence apparaît surprenant, et amène à se poser des questions sur le mode de fonctionnement des instances de normalisation, dont la mission est d'harmoniser les pratiques, et d'amener les différents acteurs, au niveau international, à se mettre d'accord sur une méthodologie unique et commune. Pour ce sujet, le processus d'harmonisation n'est pas allé jusqu'à son terme. Par ailleurs, on peut se demander pourquoi ne pas accepter en l'état ce qui a déjà été établi ailleurs, et qui a fait ses preuves. Ici, peuvent apparaître des raisons d'ordre psychologique, telle qu'une réticence à accepter ce que d'autres ont fait. Cette irruption de la subjectivité dans ce qui devrait être un sujet rationnel peut donner à réfléchir !

Le travail expérimental que nous avons mené a permis d'illustrer les contraintes propres à la microbiologie des aliments, tant en matière d'organisation d'essais inter-laboratoires, que d'exploitation des résultats de ces essais.

Quant aux perspectives, outre les voies d'amélioration que nous avons envisagées pour l'exploitation des résultats des essais inter-laboratoires portant sur des méthodes quantitatives ou qualitatives, nous pourrions citer :

- l'harmonisation au niveau mondial des référentiels dans notre domaine ;
- la validation complète des méthodes de référence normalisées, soit la réalisation d'essais inter-laboratoires complétés par l'étude d'autres critères de performance que la justesse et la fidélité ;
- l'établissement d'une valeur « vraie » en microbiologie quantitative, avec l'apport attendu des nouvelles technologies.

Les limites des essais inter-laboratoires, que nous avons attribuées essentiellement aux protocoles expérimentaux de ces essais en microbiologie des aliments, sont certes nombreuses, et l'organisation d'essais inter-laboratoires dans notre domaine est particulièrement lourde. Cependant, les intérêts qui résultent de l'utilisation des résultats de ces essais nous semblent correspondre à des enjeux suffisamment importants pour accepter de consacrer les moyens et le temps nécessaires à effectuer ces essais inter-laboratoires qui, bien qu'imparfaits, apportent un degré d'objectivité indéniable dans une discipline souvent considérée comme empirique.

3. Annexes

3.1 Annexe 1. Publication de l'essai de validation sur le dénombrement de *L.monocytogenes*

Validation of ISO method 11290 Part 2 - Enumeration of *Listeria monocytogenes* in foods.

Scotter S. *et al.* *International Journal of Food Microbiology*, 2001, No 70, p. 121-129

Validation of ISO method 11290 Part 2. Enumeration of *Listeria monocytogenes* in foods

S.L. Scotter^{a,*}, S. Langton^b, B. Lombard^c, C. Lahellec^d, S. Schulten^e,
N. Nagelkerke^e, P.H. in't Veld^f, P. Rollier^g

^a Food Microbiology Consultancy Service, Blackthorn House, Farlington, York, YO61 1NW, UK

^b Ministry of Agriculture, Fisheries and Food, Central Science Laboratory, Sand Hutton, York, YO41 1LZ, UK

^c Agence Française de Sécurité Sanitaire des Aliments, Lab. d'Etudes et de Recherches sur l'Hygiène et la Qualité des Aliments,
41 Rue du 11 Novembre 1918, 94700 Maisons-Alfort cedex, France

^d Agence Française de Sécurité Sanitaire des Aliments, Direction Hygiène des Aliments, 23 Avenue du General de Gaulle, BP 19,
94701 Maisons-Alfort cedex, France

^e Microbiological Laboratory for Health Protection, National Institute of Public Health and the Environment, P.O. Box 1,
3720 BA Bilthoven, Netherlands

^f Inspectorate for Health Protection, Commodities and Veterinary Public Health, P.O. Box 2280, 5202 CG Den Bosch, Netherlands

^g CECALAIT INRA-SRTAL BP 89, 39801 Poligny cedex, France

Received 13 November 2000; received in revised form 1 March 2001; accepted 17 April 2001

Abstract

The European and International Standard method for the enumeration of *Listeria monocytogenes*, described in EN ISO 11290 Part 2: 1998 [EN ISO 11290-2 Microbiology of Food and Animal Feedingstuffs-Horizontal Method for the Detection and Enumeration of *L. monocytogenes*: Part 2. Enumeration; International Organisation for Standardisation, Geneva.] was validated by order of the European Commission (Standards, Measurement and Testing Fourth Framework Programme Project SMT4-CT96-2098). The objective was to determine the precision of the method in terms of repeatability (r) and reproducibility (R) using three different food types inoculated with various levels of *L. monocytogenes* and a typical background flora. The results are intended for publication in the associated standards. Cheese, meat, dried egg powder and reference materials were examined by 21 laboratories in 16 countries in Europe. Each participant received eight test materials per food type: blind duplicates at four inoculum levels (0, 10^2 , 10^3 , 10^4 cfu/g). In addition, two reference materials containing *L. monocytogenes* were included in the study. All test materials were subjected to stringent homogeneity and stability testing before being used in the collaborative trial. Participants were required to use only PALCAM agar for enumeration of *L. monocytogenes*, as prescribed by the reference method. Statistical analyses has been performed using a newly introduced approach for food microbiology (draft standard prEN ISO 16140 [prEN ISO 16140 Microbiology of Food and Animal Feedingstuffs-Protocol for the Validation of Alternative Methods, International Organisation for Standardisation, Geneva.], the precision data being calculated using robust estimates.

Overall values for repeatability (r) of EN ISO 11290-2 when used with food test materials were $r = \log_{10} 0.58$ (expressed as an absolute difference between \log_{10} -transformed test results) or $r = 3.8$ (expressed as an absolute ratio

* Corresponding author. Fax: +44-1347-811636.

E-mail address: sue.scotter@tesco.net (S.L. Scotter).

between test results on the normal scale). For the reference materials (capsules containing approximately 5000 cfu), $r = \log_{10} 0.34$ (expressed as an absolute difference between \log_{10} -transformed test results) or $r = 2.2$ (expressed as an absolute ratio between test results on the normal scale).

Overall values for reproducibility (R) of EN ISO 11290-2 when used with food test materials were $R = \log_{10} 0.81$ (expressed as a difference between \log_{10} -transformed test results) or $R = 6.5$ (expressed as an absolute ratio between test results on the normal scale). For the reference materials, $R = \log_{10} 0.51$ (expressed as a difference between \log_{10} -transformed test results) or $R = 3.2$ (expressed as an absolute ratio between test results on the normal scale).

Further studies have been initiated by ISO TC34/SC9 to try to enhance the isolation of *L. monocytogenes* from foods and improve the confirmation procedures. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: *Listeria monocytogenes*; EN ISO 11290-2; Validation; Performance characteristics; Food

1. Introduction

It is widely accepted that the use of validated methods is an essential part of any sound laboratory quality assurance programme. The need for methods that have documented performance characteristics and the associated policies of the major international standardisation bodies, e.g. International Organisation for Standardisation (ISO) and European Committee for Standardisation (CEN), have been previously discussed by Scotter et al (2000). Project SMT4-CT96-2098, financed by the European Commission under the fourth Framework Standards, Measurements and Testing Program (SM&T), was elaborated to determine the precision data in terms of repeatability (r) and reproducibility (R) or performance characteristics (qualitative methods) of the six main ISO methods. These were prioritised for acceptance by CEN in the order of *Bacillus cereus* (enumeration), *Listeria monocytogenes* (detection and enumeration), *Staphylococcus aureus* (enumeration), *Clostridium perfringens* (enumeration) and *Salmonella* (detection). The determined performance characteristics are intended for publication by ISO in the corresponding ISO methods and subsequently by CEN via the Vienna Agreement.

The objectives of this EU-funded project have been described previously by Schulten et al. (2000). This is the third publication in a series describing the validation of the six ISO methods for the detection/enumeration of microorganisms in foods.

In this paper, the results of the collaborative study to validate EN ISO 11290 Part 2 (Anon., 1998) for the enumeration of *L. monocytogenes* in foods are reported. The trial was carried out in May 1998.

2. Materials and methods

2.1. Design of the trial

Twenty-one laboratories from 16 countries in Europe participated in the collaborative trial. All laboratories were accredited or in the process of being accredited according to EN 45001 standard (Anon., 1989a). The design of the trial was described to participants in a detailed trial chronology, standard operating procedure (SOP) and a test report on which to record the results and return to the trial leader for analysis. A SOP was written, based on EN ISO 11290-2, because additional information needed to be provided to ensure that all participants in the trial treated the test materials the same way. In the test report, participants were required to detail all supplementary information that could have influenced their results. Prior to each collaborative trial, a 'pre-trial' was organised between the three contractors to verify that the SOP, as written, contained no potential ambiguities and to establish that the test materials were fit for purpose.

The method was challenged with three food types (a dairy product, a meat product and a dried product) and each food type was inoculated with *Listeria* spp. at four different inoculum levels:

- negative sample (containing only *L. innocua* at 2.5×10^2 cfu/g and autochthonous flora);
- low level (2.5×10^2 cfu *L. monocytogenes*/g plus equal numbers of *L. innocua*/g and autochthonous flora);
- medium level (2.5×10^3 cfu *L. monocytogenes*/g plus equal numbers of *L. innocua*/g and autochthonous flora);

- high level (2.5×10^4 cfu *L. monocytogenes*/g plus equal numbers of *L. innocua*/g and autochthonous flora).

Acceptance of all batches for use in the trials was made on the basis of achieving satisfactory homogeneity and stability. All test materials were tested extensively prior to the pre-trial and collaborative trial to ensure that they could be maintained in a stable and homogeneous state for the duration of the distribution and trial period.

Each participant received eight test materials for each food type comprising blind duplicates at each inoculum level. In addition, two reference materials (RM) were included in the trial to determine the highest achievable level of precision of the method under test. The test materials were shipped to the participants by courier in polystyrene insulated boxes containing temperature monitoring devices and several ice packs to prevent exposure of the test materials to high temperatures during transport. Participants provided all media and reagents needed for the collaborative trial. Test material codes were randomised for each participant to prevent collusion among laboratories. The examination of the test materials was carried out within a specified time period as prescribed in the trial chronology.

2.2. Method under collaborative trial

The enumeration of *L. monocytogenes* according to EN ISO 11290-2 required the following successive stages:

- (i) preparation of the initial suspension in buffered peptone water (BPW, one of two dilution fluids permitted in the standard, used when enumeration is not carried out in parallel with detection of *L. monocytogenes*);
- (ii) resuscitation for 1 h at 20°C;
- (iii) surface plating of the initial suspension and further decimal dilutions in duplicate, onto PALCAM medium;
- (iv) incubation of the plates at 37°C and examination after 24 and 48 h;
- (v) confirmation of presumptive *L. monocytogenes* colonies;
- (vi) calculation of the final number of *L. monocytogenes*/g.

2.3. Preparation of test materials

Cheese test materials were prepared by CECALAIT, Poligny, France. All batches were prepared according to the following procedure: Liquid cheese was inoculated with the test strains comprising *L. monocytogenes* (LM2-serovar 1/2a) and *L. innocua* (LI2) both of which had been isolated from milk by CECALAIT. In addition, a background flora comprising a mixture of *Lactococcus lactis* subsp. *lactis*, *Enterococcus faecalis*, *Lactobacillus paracasei* and *Lac. plantarum*, all isolated from cheese, were also added at a total level of approximately 10^4 cfu/g. *Listeria* strains were isolated on Oxford (Oxoid) and PALCAM agars (Merck), and then on Trypticase–Soya–Yeast Extract agar (Merck), *Lactococcus* and *Enterococcus* on M17 (Biokar Diagnostics) and *Lactobacillus* on de Man, Rogosa and Sharpe (MRS) (Biokar Diagnostics) according to usual laboratory procedures. Prior to inoculation, the *Listeria* strains were cultivated in Brain–Heart–Infusion broth (BHI) 18 h at 37°C, and lactic bacteria in MRS or M17 broth for 18 h at 30°C. The cultures were then mixed: *Lactococcus* and *Enterococcus* together, and the *Lactobacilli* together. The cultures were diluted in 1/4 Ringer's solution and inoculated into the liquid cheese. Subsequently, the inoculated cheese was clotted after dispensing aliquots of 100 g into individual vials by the addition of rennet. The microflora within the cheese was stabilised by the addition of an undisclosed bacteriostatic mixture. The bacteriostatic effect is negated when the sample is diluted during examination. The inoculum levels for the target organisms were verified using a spiral plating system at a single dilution. Batches of cheese test materials were prepared for the collaborative trial immediately prior to dispatch to participants due to their relatively short stability.

Meat test materials were prepared by MAFF-CSL, UK. The test strains used to inoculate the meat test materials were *L. monocytogenes* NCTC 5214 and *L. innocua* NCTC 11288. All test materials were additionally inoculated with a simulated autochthonous meat flora comprising 10^4 cfu/g *B. subtilis*, 10^6 cfu/g *Lactobacillus* sp., 10^6 cfu/g *Micrococcus* sp. and 10^4 cfu/g *Pseudomonas* sp. Batches of test materials were prepared using retail fresh minced beef which was freeze-dried in bulk,

sub sampled into vials (portions of 10 g equivalents) and irradiated at 25 kGy. The test strains and background flora were cultivated into Nutrient broth (Oxoid), diluted to the appropriate level and added as a 'cocktail' to each vial. Finally, the vials were freeze-dried a second time to stabilise the microflora and stored at 4°C until use.

Egg powder test materials were prepared by the RIVM-MGB, The Netherlands. Each sample, as presented to participants, consisted of a vial containing whole egg powder with an autochthonous flora at approximately 10^2 cfu/g and also a vial containing one unmarked orange/white capsule and one unmarked green/white capsule containing *L. monocytogenes* and *L. innocua*, respectively. Capsules were used as it was not possible to achieve a sufficiently homogeneous matrix by contaminating the egg powder directly with both strains. Participants were instructed to prepare the test material by mixing the vials immediately before use according to a Standard Operating Procedure. A batch of 25 kg of egg powder was obtained from an egg powder factory (Enthoven, Bennebroek, NL). Two series of 10 test materials of egg powder were tested for aerobic plate count (Anon., 1989b), *Enterobacteriaceae* count (Anon., 1993), *Listeria* enumeration (Anon., 1998, Oxford medium) and *Listeria* detection (Anon., 1997). The test strains used in the egg test materials for the collaborative trial were *L. monocytogenes* Scott A, ALM 92 and *L. innocua* ALM 105. Both strains were cultivated in BHI for 24 h (shaken at 100 rpm) at 37°C. Each culture was centrifuged and the pellet suspended into 100 ml of milk supplemented with 2 M saccharose. This suspension was dried onto sterile milk powder (Melkproduct 17, Nestlé) by means of a fluid bed spray granulation (STREA-1, Niro-Aeromatic). This yielded two highly contaminated milk powders (ca. 10^8 cfu/g). The required contamination levels were prepared by further mixing of these powders with sterile milk powder in several steps (to optimise homogeneity of the test material) until the required contamination level was reached. For each prepared batch, the mixed powder was filled into the gelatin capsules (ca. 0.3 g/capsule) in a laminar flow cabinet using an aluminium filling apparatus.

For the R.M., *L. monocytogenes* capsules were used which contained ca. 5000 cfu/capsule. The

capsules were obtained from the Foundation for the Advancement of Public Health and Environmental Protection (SIVM, Bilthoven, The Netherlands).

2.4. Homogeneity and stability of the test materials

The homogeneity and stability of the test materials were verified according to the usual procedures of the laboratory expert in their preparation. Acceptance of all production batches for use in the trials was made on the basis of achieving satisfactory homogeneity and stability according to the laboratories' own criteria.

For the cheese test materials, stability at 4°C was measured by analysing, over several days, three test materials in duplicate. For each analysis, 10 g of cheese was diluted in 90 ml K_2HPO_4 solution (Anon., 1992) and blended for 3 min. This suspension was inoculated onto two plates of Oxford agar using a spiral plating system. Only the batch containing 10^4 cfu *L. monocytogenes* was tested. For homogeneity, 10 test materials of the same batch were examined in duplicate on Oxford medium. The CE-CALAIT computer program was used for the statistical treatments. χ^2 (chi square) was calculated on the total count (\log_{10}) of colonies per sample to compare the dispersion of the contamination to the dispersion obtained in a Poisson series.

For the meat test materials at each level of contamination, the stability of *L. monocytogenes* in minced beef stored at 4°C was verified over a 30-day period. Ten replicate vials were enumerated on PAL-CAM agar at each interval at days 0, 11, 17, 25 and 30. The homogeneity of the batch was verified by enumerating 10 replicate vials immediately after production and again at the same time that participants examined the test materials. An *F*-test ($\alpha = 0.05$) was applied using \log_{10} transformed counts and the stability and homogeneity considered satisfactory if the *F*-test was not significant.

For the egg test materials, the homogeneity of both the egg powder and the capsules containing the *Listeria* spp. was determined. For the low, medium and high contamination levels, 10 capsules were examined to determine the variation between duplicate counts, i.e. within sample variation (T_1 test)

also the variation between test materials (T_2 test). A value for $T_2/(I-1) < 2$ (where I is the number of capsules), is regarded as good homogeneity. Two lots of 10 replicates of the egg powder were also tested for aerobic count, Enterobacteriaceae and *Listeria* spp. The stability of the capsules at -20°C was assessed for two batches (*L. monocytogenes* high level and *L. innocua* medium level). Ten test materials per week were examined over a period of 10 weeks by plating on PALCAM agar and incubating for 48 h. Additionally, stability tests were carried out at elevated temperatures. The batches were tested at day 0, 2, 4 and 8. At each interval, 10 capsules were analysed as described above. Stability of the background flora at -20°C was monitored over a period of 10 weeks. Ten replicates per week were examined by plating on PCA and incubating for 3 days at 30°C .

2.5. Statistical analysis of the data

Analysis of data was firstly carried out using standard statistical procedures for identification of outlying results and the determination of the precision characteristics repeatability and reproducibility, according to ISO 5725-2 (Anon., 1994). However, based on data from another study in this SMT project on the enumeration of coagulase-positive staphylococci, and given the progress of standardisation work in the field of validation of microbiological methods (official draft Standard prEN ISO 16140 (Anon., 1999), it was considered that the method of calculation used in ISO 5725-2, based on the mean, was not the most appropriate for quantitative microbiological data. Moreover, such data do not always display a normal (Gaussian) distribution and the rejection of outlying results based only on statistical grounds is often controversial, e.g. are the values truly abnormal or do they reflect the reality of method variability? Such elimination also reduces the number of useful values for the statistical analyses.

Therefore, it was decided to use the draft Standard prEN ISO 16140, which was developed specifically for microbiological analyses by CEN (in the framework of the European Eurêka "MicroVal" project; Rentenaar, 1996). This draft introduces robust

estimators for calculating repeatability and reproducibility. Three main estimators are used:

- the median;
- the standard-deviation of repeatability derived from the median of the duplicate standard-deviations;
- the spread between the duplicate means, including the reproducibility, based on the recursive median (Rousseuw and Bassett, 1990; Rousseuw and Croux, 1993).

These robust statistics have the advantage of being less sensitive to extreme values, thus permitting outliers to be retained. They also do not depend heavily on the distribution of the data. As is usual for calculations based on microbiological counts, in order to get a more symmetric distribution, the raw data were transformed into logarithms before performing any calculations.

3. Results and discussion

3.1. Stability and homogeneity of the test materials

The majority of the test materials were considered sufficiently stable and homogeneous for the trial. Cheese test materials stored at 4°C were stable and homogeneous for up to 7 days. Meat test materials were generally stable and homogeneous when stored at 4°C for up to 1 month. However, statistical analysis revealed some intra-batch heterogeneity between the meat test materials inoculated at the medium level when they were re-tested by the production laboratory at the time of the trial. This heterogeneity was not evident when the vials were verified immediately after production and may be an anomalous result. The precision data for this test material must therefore be treated with some caution. Separate examinations were carried out on the egg powder and capsules containing *L. innocua* and *L. monocytogenes*. Both were stable and homogenous over a 10-week period when stored at -20°C .

3.2. General results of the trial

A total of 21 laboratories from 16 countries in Europe participated in the collaborative trial. Labora-

Table 1

(A) Statistical analysis of dried egg powder test materials

Trial parameters	Level of contamination		
	Low level	Medium level	High level
Year of inter-laboratory test	1998	1998	1998
Number of laboratories having returned results	18	18	18
Number of samples per laboratory	2	2	2
Number of excluded laboratories	2	2	2
Number of laboratories retained after exclusion	16	16	16
Number of accepted samples	32	32	32
Median value (in log ₁₀ cfu/g)	2.36	3.27	4.40
Repeatability standard deviation s_r (in log ₁₀ cfu/g)	0.29	0.2	0.14
Repeatability relative standard deviation RSD _r (in %)	12.3	6.14	3.19
Repeatability limit, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	0.81	0.56	0.39
Reproducibility standard deviation s_R (in log ₁₀ cfu/g)	0.30	0.24	0.17
Reproducibility relative standard deviation RSD _R (in %)	12.66	7.26	3.83
Reproducibility limit, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	0.84	0.67	0.47

(B) Statistical analysis of minced meat test materials

Trial parameters	Level of contamination		
	Low level	Medium level	High level
Year of inter-laboratory test	1998	1998	1998
Number of laboratories having returned results	18	18	18
Number of samples per laboratory	2	2	2
Number of excluded laboratories	1	1	1
Number of laboratories retained after exclusion	17	17	17
Number of accepted samples	34	34	34
Median value (in log ₁₀ cfu/g)	2.68	3.56	4.35
Repeatability standard deviation s_r (in log ₁₀ cfu/g)	0.18	0.23 ^a	0.21
Repeatability relative standard deviation RSD _r (in %)	6.59	6.43 ^a	4.76
Repeatability, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	0.49	0.64 ^a	0.58
Reproducibility standard deviation s_R (in log ₁₀ cfu/g)	0.27	0.24 ^a	0.35
Reproducibility relative standard deviation RSD _R (in %)	10.16	6.82 ^a	8.09
Reproducibility, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	0.76	0.68 ^a	0.99

(C) Statistical analysis of fresh cheese test materials

Trial parameters	Level of contamination		
	Low level	Medium level	High level
Year of inter-laboratory test	1998	1998	1998
Number of laboratories having returned results	19	19	19
Number of samples per laboratory	2	2	2
Number of excluded laboratories	2	2	2
Number of laboratories retained after exclusion	17	17	17
Number of accepted samples	34	34	34
Median value (in log ₁₀ cfu/g)	1.96	3.17	4.29
Repeatability standard deviation s_r (in log ₁₀ cfu/g)	0.19	0.15	0.27
Repeatability relative standard deviation RSD _r (in %)	9.88	4.81	6.22
Repeatability, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	0.54	0.43	0.75
Reproducibility standard deviation s_R (in log ₁₀ cfu/g)	0.45	0.25	0.31
Reproducibility relative standard deviation RSD _R (in %)	23.11	8.03	7.21
Reproducibility, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	1.27	0.71	0.87

Table 1 (continued)

(D) Statistical analysis of reference materials	
Trial parameters	RM (5000 cfu/g)
Year of inter-laboratory test	1998
Number of laboratories having returned results	19
Number of samples per laboratory	2
Number of excluded laboratories	1
Number of laboratories retained after exclusion	18
Number of accepted samples	36
Median value (in log ₁₀ cfu/g)	3.63
Repeatability standard deviation s_r (in log ₁₀ cfu/g)	0.12
Repeatability relative standard deviation RSD _r (in %)	3.33
Repeatability, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	0.34
Reproducibility standard deviation s_R (in log ₁₀ cfu/g)	0.18
Reproducibility relative standard deviation RSD _R (in %)	4.98
Reproducibility, as difference on log ₁₀ scale (in log ₁₀ cfu/g)	0.51

^aThese data must be treated with caution as some intra-batch heterogeneity was observed at the time of the trial.

tory 7 elected not to return their results. Laboratory 19 examined only cheese test materials and Laboratory 21 was able to examine the reference materials only.

The cheese test materials were dispatched directly by CECALAIT separately from the other food types and RM due to their shorter shelf life. All participants except one commenced examination of the cheese on the prescribed day. Laboratory 5 commenced the examinations late due to the delayed arrival of their test materials, however, as they had not been subjected to any temperature abuse (as indicated by the temperature monitoring device), they were instructed to continue with the trial. For the meat, egg powder test materials and RMs, all but three laboratories commenced their examinations on the prescribed day. Of these, only one of the laboratories test materials had received excessive temperature abuse.

3.3. Results excluded from further analysis

For technical reasons only, the following results were excluded from further statistical analysis: cheese data submitted by Laboratory 24 as they failed to confirm any colonies from any cheese sample; all data submitted by Laboratory 8 as they failed to discriminate between *L. monocytogenes* and *L. innocua* throughout the trial; egg powder data submitted by Laboratory 17 as their data was highly variable. The r and R values were calculated both with

and without data from those laboratories whose test materials had received some temperature abuse. The calculations indicated that the results from these test materials had little or no influence on the r and R values and thus the data were not excluded from the final calculation of precision.

3.4. Analysis of negative samples

A total of six negative samples were examined by each laboratory (two per food type). No false positive results were returned by any laboratory.

3.5. Determination of r and R values

The repeatability r of a method is defined as the value below which the absolute difference between two single test results obtained with the same method on identical test materials under the same conditions, may be expected to lie with a probability of 95% (Anon., 1999).

The reproducibility R of a method is defined as the value below which the absolute difference between two single test results obtained with the same method on identical test materials under different conditions, may be expected to lie with a probability of 95% (Anon., 1999).

The precision parameters for each food type and inoculum level are shown in Table 1a–c and for the RM in Table 1d.

4. Conclusion

4.1. Repeatability

For the EN ISO method 11290-2, a general value (mean of all r values for foods) of $r = \log_{10} 0.58$ (expressed as an absolute difference between \log_{10} -transformed test results) can be used when examining food samples or $r = 3.8$ (expressed as an absolute ratio between test results on the normal scale). When reference materials are used, $r = \log_{10} 0.34$ or $r = 2.2$ (ratio), respectively.

4.2. Reproducibility

For the EN ISO method 11290-2, a general value (mean of all R values for foods) of $R = \log_{10} 0.81$ (expressed as an absolute difference between \log_{10} -transformed test results) can be used when examining food samples or $R = 6.5$ (expressed as an absolute ratio between test results on the normal scale). When reference materials are used, $R = \log_{10} 0.51$ or $R = 3.2$ (ratio), respectively.

4.3. General conclusions from the trial

These values indicate that there are no major problems with the precision of ISO 11290-2. However, as previously reported by Scotter et al (2000), further work is required to improve the confirmation stages in order to enhance the isolation of *L. monocytogenes* when present in foods containing other *Listeria* spp., in particular *L. innocua*.

The outcome of this study has been presented at the annual meeting of ISO TC34 SC9 (April 1999, Paris). The committee has decided to include the precision data generated by this study in the standard EN ISO 11290-2 as an amendment. At the same time, the committee has launched experimental studies to improve the agar media used for enumeration with a view to revising this standard in the future.

Acknowledgements

The validation of International Standard EN ISO 11290-1 has been carried out under the framework of

a Standards, Measurement and Testing Project No. SMT4-CT96-2098. The authors wish to thank the following laboratories for their participation and cooperation in this collaborative trial: Mr. ir H. Stegeman, RIKILT-DLO, P.O. Box 230, 6700 AE Wageningen, The Netherlands; Mr. P.H. in't Veld, National Institute of Public Health and the Environment, Microbiological Laboratory for Health Protection, Antonie van Leeuwenhoeklaan 9, 3720 BA Bilthoven, The Netherlands; Dr. Boleslaw Wojton, National Veterinary Institute, Al Partyzantow 57, 24-100-Pulawy, Poland; G. Vlaemynck, Dairy Research Station, Microbiology Laboratory, Brusselssteenweg 370, 9090 Melle, Belgium; I.M. Vicente da Cruz, INETI-IBQTA-DTIA, Estrada do Paco do Lumiau-Edificio S, Lisboa 1699 cedex, Portugal; Prof. M.L. Stecchini, Dipartimento di Scienze Degli Alimenti, Via Marangoni 97, 33100 Udine, Italy; Dr. H. Asperger, Veterinarmedizinische Universitat Wien, Institute fur Milchhygiene und Milchtechnologie, Joseph Baumann-Gasse 1, A-1210 Wien, Austria; Mr. V. Young, EHB Public Analysts Laboratory, Sir Patrick Dun's, Lower Grant Canal Street, Dublin 2, Republic of Ireland; Prof. Debevere, Universiteit Gent, Faculteit Landbouwkundige En toegepaste Biologische Wetenschappen vakgroep Levensmiddele-technologie en Voeding, B-9000 Gent; Mrs. C. Allaert, Laboratory of Microbiology of the Food Technology Department, Avda. Alcalde Rovira Roure, 177, 25198 Lleida, Spain; Dr. T. Johansson, National Veterinary and Food Research Institute, Department of Food Microbiology, P.O. Box 368, Fin-00231 Helsinki, Finland; D.V.M. Marylene Bohnert, CNEVA Ploufragen B.P.53, Zoopole Beaucemarie, 22440 Ploufragan, France; Dr. D. Roberts, Central Public Health Laboratory, Food Hygiene Laboratory, 61 Colindale Avenue, London, NW9 5HT; J. Anderson, Unilever Research, Colworth Laboratory, Colworth House, Sharnbrook, Bedford, MK44 1LQ; Prof. P. Teufel/Dr. E. Bartelt, Federal Institute for Health Protection of Consumers, Unit Food Microbiology, Thielallee 88-92, 14195 Berlin, Germany; Dr. J.L. Cordier, Nestlé Research Center Lausanne, Quality and Safety Assurance Department, P.O. Box 44, 1000 Lausanne, Switzerland; Dr. K. Friedrich MUVA Kempten, Qualitats-Und Laborzentrum, Hirnbeinstrasse 10, D-87435 Kempten (Allgau), Postfach 2025, Allgau, Germany; Dr. A.

Steneryd/Dr. P. Norberg, National Food Administration, Biology Division, Box 622, 75126 Uppsala, Sweden; Mr. Vincent, Institut Pasteur de Lille, Service de Microbiologie et Hygiene des Aliments, 369 rue Jules Guesde, 59650 Villeneuve d'Ascq, France; Dr. Olli Pensala, Customs Laboratory, Tekniikantie 13, 02150 Espoo, Finland; Dr. S. Qvist, Danish Veterinary Service, Food Control Laboratory, Howitzvej 13, DK-2000 Frederiksberg, Denmark.

The authors also wish to acknowledge the expert statistical advice supplied by Dr. Remi Chevenement and also the technical assistance of all their laboratory staff in the preparation and analysis of the test materials used in this study.

References

- Anon., 1989a. EN 45001-General Criteria for the Operation of Testing Laboratories. Comité Européen de Normalization, Brussels.
- Anon., 1989b. ISO 4833 Microbiology of Food and Animal Feedingstuffs-General Guidance for the Enumeration of Microorganisms-Colony Count Technique at 30°C. International Organization for Standardization, Geneva.
- Anon., 1992. FIL 122b Milk and Milk Products-Preparation of Test Materials and Dilutions for Microbiological Examination. International Dairy Federation, Brussels.
- Anon., 1993. ISO 7402 Microbiology of Food and Animal Feedingstuffs-Enumeration of Enterobacteriaceae. International Organization for Standardization, Geneva.
- Anon., 1994. ISO 5725-2 Application of Statistics-Accuracy (Trueness and Precision) of Measurement Methods and Results: Part 2. Basic Method for the Determination of Repeatability of a Standard Measurement Method. International Organisation for Standardisation, Geneva.
- Anon., 1997. EN ISO 11290-1 Microbiology of Food and Animal Feedingstuffs-Horizontal Method for the Detection and Enumeration of *Listeria monocytogenes*: Part 1. Detection. International Organisation for Standardisation, Geneva.
- Anon., 1998. EN ISO 11290-2 Microbiology of Food and Animal Feedingstuffs-Horizontal Method for the Detection and Enumeration of *Listeria monocytogenes*: Part 2. Enumeration. International Organisation for Standardisation, Geneva.
- Anon., 1999. prEN ISO 16140 Microbiology of Food and Animal Feedingstuffs-Protocol for the Validation of Alternative Methods. International Organisation for Standardisation, Geneva.
- Rentenaar, I.M.F., 1996. MicroVal, a challenging Eureka project. Food Control 7 (1), 31–36.
- Rousseeuw, P.J., Bassett Jr., G.W., 1990. The mediant: a robust averaging method for large data sets. J. Am. Stat. Assoc. 85, 97–104.
- Rousseeuw, P.J., Croux, C., 1993. Alternatives to the median absolute deviation. J. Am. Stat. Assoc. 88, 1273–1283.
- Schulten, S.M., in't Veld, P.H., Nagelkerke, N.J.D., Scotter, S., de Buyser, M.L., Rollier, P., Lahellec, C., 2000. Evaluation of the ISO 7932 standard for the enumeration of *Bacillus cereus* in foods. Int. J. Food Microbiol. 57, 53–61.
- Scotter, S.L., Langton, S., Lombard, B., Lahellec, C., Schulten, S., Nagelkerke, N., in't Veld, P.H., Rollier, P., 2000. Validation of ISO method 11290 Part 1- detection of *Listeria monocytogenes* in foods. Int. J. Food Microbiol. 64, 294–306.

3.2 Annexe 2. Publication de l'essai de validation sur le dénombrement des staphylocoques

Validation of EN ISO standard methods 6888 Part 1 and Part 2 : 1999 – Enumeration of coagulase-positive staphylococci in foods.

De Buyser ML, et al. International Journal of Food Microbiology, 2003, No 83(2), p. 185-194.



Validation of EN ISO standard methods 6888 Part 1 and Part 2: 1999—Enumeration of coagulase-positive staphylococci in foods

M.L. De Buyser^{a,*}, B. Lombard^a, S.M. Schulten^{b,c}, P.H. In't Veld^{b,d},
S.L. Scotter^{e,1}, P. Rollier^f, C. Lahellec^g

^a Agence Française de Sécurité Sanitaire des Aliments, Laboratoire d'Etudes et de Recherches sur l'Hygiène et la Qualité des Aliments (AFSSA-LERHQA), 41 Rue du 11 Novembre 1918, 94700, Maisons-Alfort, France

^b Rijksinstituut voor Volksgezondheid en Milieu, Microbiologisch Laboratorium voor Gezondheidsbescherming (RIVM-MGB), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

^c Dutch Standardization Institute (NEN), P.O. Box 5059, 2600 GB Delft, The Netherlands

^d Inspectorate for Health Protection, Commodities and Veterinary Public Health, P.O. Box 2280, 5202 CG Den Bosch, The Netherlands

^e Ministry of Agriculture, Fisheries and Food, Central Science Laboratory (MAFF-CSL), Sand Hutton, York YO41 1LZ, UK
^f CECALAIT BP 129, 39802 Poligny cedex, France

^g Agence Française de Sécurité Sanitaire des Aliments, 23 Avenue du General de Gaulle, BP 19, 94701, Maisons-Alfort cedex, France

Accepted 12 August 2002

Abstract

The methods of European and International Organisations for Standardization for the enumeration of coagulase-positive staphylococci (CPS, *Staphylococcus aureus* and other species) described in EN ISO 6888 Part 1 and Part 2: 1999 were validated by order of the European Commission (Standards, Measurement and Testing Fourth Framework Programme Project SMT4-CT96-2098). EN ISO 6888-1 prescribes the use of Baird–Parker (BP) agar whereas EN ISO 6888-2 prescribes the use of Rabbit Plasma Fibrinogen Agar (RPFA). The objective was to determine the precision of each method in terms of repeatability (r) and reproducibility (R) using three different food types inoculated with various levels of *S. aureus* and a typical background flora. The results are intended for publication in the associated standards. Cheese, meat and dried egg powder were examined by 24 laboratories from 16 countries in Europe. Each participant received eight test materials per food type: blind duplicates at four inoculum levels (0 , 10^3 , 10^4 to 10^5 , 10^5 to 10^6 cfu/g). In addition, two reference materials (RM) (capsules containing milk powder inoculated with *S. aureus*) were included in the study. All test materials were subjected to stringent homogeneity and stability testing before being used in the collaborative trial. Two statistical methods were used to calculate the precision parameters. Draft EN ISO 16140: 2000 method appeared more appropriate to the case of microbiological data than ISO 5725-2: 1994 method and was retained to calculate the precision data.

Concerning EN ISO 6888-1, overall values for repeatability (r) when used with food test materials was $r = \log_{10} 0.28$ (expressed as an absolute difference between \log_{10} -transformed test results). For the reference materials, $r = \log_{10} 0.19$. Overall values for reproducibility (R) when used with food test materials were $R = \log_{10} 0.43$. For the reference materials, $R = \log_{10} 0.39$.

* Corresponding author. Tel.: +33-149-77-11-10; fax: +33-149-77-11-02.

E-mail address: m.debuyser@afssa.fr (M.L. De Buyser).

¹ Present address: Blackthorn House, Farlington, York YO61 1NW, UK.

Concerning EN ISO 6888-2, overall values for repeatability (r) when used with food test materials were $r = \log_{10} 0.22$. For the reference materials, $r = \log_{10} 0.17$. Overall values for reproducibility (R) when used with food test materials were $R = \log_{10} 0.33$. For the reference materials, $R = \log_{10} 0.31$.

These results were presented to the ISO technical committee and to the Comité Européen de Normalisation (CEN). Both committees agreed to incorporate the precision data obtained with food materials as two amendments to EN ISO 6888-1 and -2, and to give an equal status to each part of the standard.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: *Staphylococcus aureus*; EN ISO 6888-1; EN ISO 6888-2; Validation; Precision parameters; Food

1. Introduction

European Committee for Standardisation or Comité Européen de Normalisation (CEN) requires methods which are to become CEN standards to be validated by collaborative study. Project SMT4-CT96-2098, financed by the European Commission under the fourth Framework Standards, Measurements and Testing Program (SMT4)), was elaborated to determine the precision data in terms of repeatability (r) and reproducibility (R) or performance characteristics (qualitative methods) of six ISO (International Organisation for Standardisation) methods. These were prioritised for acceptance by CEN in the order of *Bacillus cereus* (enumeration), *Listeria monocytogenes* (detection and enumeration), *Staphylococcus aureus* (enumeration), *Clostridium perfringens* (enumeration) and *Salmonella* (detection). The determined performance characteristics are intended for publication by CEN in the corresponding CEN methods and subsequently by ISO via the Vienna Agreement.

The design of this EU-funded project and the results of the first validation carried out on *B. cereus* enumeration have been described by Schulten et al. (2000). The results of the second and third validations carried out respectively on detection and enumeration of *L. monocytogenes* have been described by Scotter et al. (2001a,b).

In this paper, the results of the fourth collaborative study in the project to validate EN ISO 6888 Part 1 (Anon., 1999a) and Part 2 (Anon., 1999b) for the enumeration of coagulase-positive staphylococci (*S. aureus* and other species) in foods are reported. The trial, led by AFSSA-LERHQA, was carried out in May–June 1999. Two statistical methods were also compared for the determination of precision data, a traditional approach for a general use, ISO 5725-2

(Anon., 1994), and a new approach developed especially for microbiological analysis, Draft EN ISO 16140 (Anon., 2000).

2. Materials and methods

2.1. Design of the trial

Twenty-four laboratories from 16 countries in Europe participated in the collaborative trial. All laboratories were accredited or in the process of being accredited according to EN 45001 standard (Anon., 1989a). The design of the trial was described to the participants in a detailed trial chronology, standard operating procedure (SOP) and a test report on which to record the results and return to the trial leader for analysis. Two SOPs were written, respectively, based on EN ISO 6888-1 and -2, because additional information needed to be provided to ensure that each participant in the trial treated all test materials the same way. In the test report, participants were required to detail all supplementary information that could have influenced their results. Prior to each collaborative trial, a ‘pre-trial’ was organised between the three contractors (AFSSA-LERHQA, MAFF-CSL, RIVM-MGB) to verify that the SOPs as written contained no potential ambiguities and to establish that the test materials were fit for purpose.

The method was challenged with three food types (fresh cheese, meat, egg powder) and each food type was inoculated at four different inoculum levels with *S. aureus* strains:

- negative sample (not containing *S. aureus*)
- low level (ca 10^3 cfu *S. aureus*/g)

- medium level (10^4 to 10^5 cfu *S. aureus*/g)
- high level (10^5 to 10^6 cfu *S. aureus*/g)

together with the same level of *S. epidermidis* or other coagulase-negative *Staphylococcus* spp. In addition, cheese and meat samples contained a simulated autochthonous flora.

Acceptance of all batches for use in the trials was made on the basis of achieving satisfactory homogeneity and stability. All test materials were tested extensively prior to the pre-trial and collaborative trial to ensure that they could be maintained in a stable and homogeneous state for the duration of the distribution and trial period.

Each participant received eight test materials for each food type comprising blind duplicates at each inoculum level. In addition, two reference materials (RM) were included in the trial to determine the highest achievable level of precision of the method under test. The test materials were shipped to the participants by courier in polystyrene insulated boxes containing temperature monitoring devices and several ice packs to prevent exposure of the test materials to high temperatures during transport. Participants provided all media and reagents needed for the collaborative trial. Test material codes were randomised for each participant to prevent collusion between laboratories. The examination of the test materials was carried out within a specified time period as prescribed in the trial chronology.

2.2. Methods under collaborative trial

The enumeration of *S. aureus* is performed with two methods, one using Baird–Parker (BP) medium (EN ISO 6888-1) and the other using Rabbit Plasma Fibrinogen Agar, RPFA (EN ISO 6888-2). They required the following successive stages:

- (i) Preparation of the initial suspension in sodium citrate solution for cheese samples, in buffered peptone water (BPW) for meat samples, and in peptone saline solution for egg powder samples and reference material.
- (ii) Surface plating (BP medium) or pour plating (RPFA medium) of the initial suspension and further decimal dilutions on or in the solid isolation medium in duplicate.

- (iii) Incubation of the plates at 37 °C and examination after 24 and 48 h.
- (iv) Confirmation of presumptive coagulase-positive staphylococci (CPS) typical and atypical colonies (BP medium only).
- (v) Calculation of the final number of CPS per gram.

2.3. Preparation of test materials

Cheese test materials were prepared by CECA-LAIT, Poligny, France. All batches were prepared according to the following procedure: liquid cheese was inoculated with the test strains comprising two *S. aureus* strains (T4, showing a typical appearance on BP medium, and B, showing an atypical appearance on BP medium) and two strains of coagulase-negative *Staphylococcus* spp. (D and E, both of atypical appearance on BP medium) all of which had been isolated from milk by CECALAIT. In addition, a background flora comprising *Lactococcus lactis* subsp. *lactis*, *Enterococcus faecalis*, *Lactobacillus paracasei* and *L. plantarum*, all of which have been isolated from cheese, was also added at a level of approximately 10^4 cfu/g. Staphylococcal strains were isolated on Baird–Parker agar (Biokar Diagnostics), and then on Plate-Count agar (PCA, Difco) for the strains T4, B and D, and on Trypticase–Soya–Yeast Extract agar (TSYEA, Merck) for strain E. *Lactococcus* and *Enterococcus* were isolated on M17 agar (M17 Biokar Diagnostics broth with 1% agar) and *Lactobacillus* on de Man, Rogosa and Sharp agar (MRS Biokar Diagnostics broth with 1% agar). Prior to inoculation, the staphylococcal strains were cultivated in Brain–Heart–Infusion broth (BHI, Oxoid) for 18 h at 37 °C, and lactic bacteria in MRS broth or M17 broth for 18 h at 30 °C. Then, the cultures were mixed: coagulase-negative staphylococci together, *Lactococcus* and *Enterococcus* together, and *Lactobacillus* together. The cultures were diluted in 1/4 Ringer’s solution (Merck), and inoculated in the liquid cheese. Subsequently, the inoculated cheese was clotted after dispensing aliquots of 100 g into individual vials by the addition of rennet. The microflora within the cheese was stabilised by the addition of an undisclosed bacteriostatic mixture. The bacteriostatic effect is negated when the sample is diluted during examination. The inoculum levels for the target organisms were verified using a spiral

plating system at a single dilution. Batches of cheese test materials were prepared for the collaborative trial 1 day prior to dispatch to participants due to their relatively short stability.

Meat test materials were prepared by MAFF-CSL, UK. The test strains used to inoculate the meat test materials were *S. aureus* NCDO 949 (showing a typical appearance on BP medium) and *S. epidermidis* NCTC 10518 (showing an atypical appearance on BP medium). All test materials were additionally inoculated with a simulated autochthonous meat flora comprising equal numbers of *L. plantarum* NCDO 1752 and *Micrococcus luteus* NCTC 8340. Batches of test materials were prepared using retail fresh minced beef which was freeze-dried in bulk, sub-sampled into vials (portions of 10 g equivalents) and irradiated at 25 kGy. The test strains and background flora were cultivated into nutrient broth (Oxoid), diluted to the appropriate level and added as a 'cocktail' to each vial. Finally, the vials were freeze-dried a second time to stabilise the microflora and stored at 4 °C until use.

Egg powder test materials were prepared by the RIVM-MGB, The Netherlands. Each sample as presented to participants consisted of a vial containing commercially available whole egg powder. Two series of 10 test materials of egg powder were tested for aerobic plate count according to the method ISO 4833 (Anon., 1989b) and for coagulase-positive staphylococci (Anon., 1999a). The test strains used in the egg test materials for the collaborative trial were *S. aureus* K 153 (showing a typical appearance on BP medium) and *S. epidermidis* ALM 109 (showing an atypical appearance on BP medium). Both strains were cultivated in BHI with 10% NaCl for 48 h at 42 °C. The cultures were centrifuged and each pellet was suspended into 100 ml of milk, supplemented with 2 M saccharose. Approximately 10 ml of this suspension was dried onto 100 g of sterile milk powder (Carnation, Nestlé) by means of a fluid bed spray granulation (STREA-1, Niro-Aeromatic). This resulted in two highly contaminated milk powders. The required contamination levels were prepared by mixing these powders with sterile milk powder in several steps (to optimise homogeneity of the test material) until the required contamination level was reached. It was not possible to obtain homogeneous mixtures of egg and milk powders, due to the stickiness of the egg powder. Therefore, the complete content of each vial had to be

analysed, so that homogeneity within a vial was not necessary.

For RM, RIVM-MGB used a highly contaminated milk powder prepared in 1993, HCMP 4-1, contaminated with *S. aureus* strain ALM 25 (showing a typical appearance on BP medium). To get the correct level (ca. 5000 cfu/capsule), this powder was mixed with sterile milk powder. After testing homogeneity, the powder was filled into gelatin capsules (ca. 0.3 g/capsule) in a laminar flow cabinet.

2.4. Homogeneity and stability of the test materials

The homogeneity and stability of the test materials were tested according to the usual procedures of the laboratory expert in their preparation. Acceptance of all production batches for use in the trials was made on the basis of achieving satisfactory homogeneity and stability according to the laboratories' own criteria.

For the cheese test materials, a batch that contained only the two *S. aureus* strains at a final level of 10^4 /g and no coagulase-negative staphylococci or lactic bacteria was prepared. For homogeneity, 10 test samples from this batch were examined in duplicate 3 days after preparation. For each analysis, 10 g of sample was diluted in 90 ml K_2HPO_4 solution (Anon., 1996) and blended for 3 min. This suspension was inoculated on two plates of BP agar using a spiral-plating system. Typical and atypical colonies were enumerated. The CECALAIT computer program was used for the statistical treatments. χ^2 (Chi-square) was calculated on the total count (\log_{10}) of colonies per sample to compare the dispersion of the contamination to the dispersion obtained in a Poisson series. The stability at 4 °C was measured by analysing, over several days, two test materials in duplicate on BP agar. Stability at 20 °C was also tested for the batch used in the pre-trial.

For the meat test materials, the homogeneity was determined by enumerating in duplicate on BP agar 10 vials from each batch representing the different inoculum levels and assessing the "within vial" and "between vial" heterogeneity using a standard *F*-test and also a *T*₁ test and a *T*₂ test as described for the egg powder samples. The enumerations were carried out both immediately after production (day 0) and at the same time as testing by participants (day 20). The stability of *S. aureus* in minced beef stored at 4 °C was verified over a 30-day period. Ten replicate vials

were enumerated on BP agar at each interval at days 0, 11, 17, 25 and 30 for the three contamination levels.

For the egg powder samples, the homogeneity was tested in each of the three milk powder batches, contaminated with *S. aureus*. Ten samples per level and per strain were examined in duplicate on BP agar. The samples were examined to determine the variation between duplicate counts, i.e. within sample variation (T_1 test) and also the variation between test materials (T_2 test), as described by Schulten et al. (2000). A value for $T_2/(I-1) < 2$ (where I is the number of test material tested), is regarded as good homogeneity. The stability of the egg samples at $-20\text{ }^\circ\text{C}$ was assessed at the medium level. To be able to use BP agar, the *S. aureus* and *S. epidermidis* levels were monitored separately by preparing special samples contaminated with either *S. aureus* or *S. epidermidis*. Five samples per week were examined by plating on BP agar and incubating for 48 h at $37\text{ }^\circ\text{C}$ over a period of 10 weeks. Additionally, stability tests were carried out at elevated temperatures to estimate the effect of transport. Two batches were tested at days 0, 2, 4 and 8 at -20 , 5 and $25\text{ }^\circ\text{C}$. At each interval, 10 samples were analysed as described above. The counts obtained were \log_{10} -transformed and subsequently analysed using a linear regression to determine the regression coefficient and corresponding t -value, which was used to determine whether or not the regression coefficients differed significantly from zero.

For RM, the homogeneity was determined by analysing three times 10 capsules as for egg samples. The stability at $-20\text{ }^\circ\text{C}$ was tested by examining 10 capsules each week over a period of 10 weeks, as for egg samples. The stability at elevated temperatures was assessed at days 0, 2, 5 and 7, as for egg samples.

2.5. Statistical analysis of the data

This trial was the opportunity to compare two ways of expressing precision data for quantitative microbiological methods:

- the “traditional” approach for determination of precision characteristics, repeatability and reproducibility, developed for chemical analysis according to ISO 5725-2 (Anon., 1994). AFSSA-LERHQA performed the statistical tests and calculated the

precision data using the software AIL 5725, developed in France according to this ISO Standard.

- the new approach of the draft standard EN ISO 16140 (Anon., 2000), specifically developed for microbiological analysis by CEN, based on the outcome of the European Eurêka “MicroVal” project (Rentenaar, 1996). This draft introduces robust estimators for calculating repeatability and reproducibility. Three main estimators are used:

- the median,
- the standard-deviation of repeatability derived from the median of the duplicate standard-deviations,
- the spread between the duplicate means, including the reproducibility, based on the recursive median (Rousseuw and Bassett, 1990; Rousseuw and Croux, 1993).

AFSSA-LERHQA performed the calculations using Excel spread-sheets developed according to the draft standard. The raw data were \log_{10} -transformed before performing any calculations.

3. Results and discussion

3.1. Stability and homogeneity of the test materials

The majority of the test materials were considered sufficiently stable and homogeneous for the trial. Cheese test materials stored at $4\text{ }^\circ\text{C}$ were stable and homogeneous for up to 7 days. Meat test materials were generally stable and homogeneous when stored at $4\text{ }^\circ\text{C}$ for up to 1 month. Egg samples and RM were generally stable and homogenous over a 10-week period when stored at $-20\text{ }^\circ\text{C}$, and over 8 days when stored at 5 and $25\text{ }^\circ\text{C}$. For some batches of egg samples, a significant value of $T_2/(I-1)$ was obtained, but these batches were considered as sufficiently homogeneous for use in the trial.

3.2. General results of the trial

A total of 24 laboratories from 16 countries in Europe participated in the collaborative trial. Due to the level of work required in this trial, all participants were given the option of whether they wished to

examine all test samples by both methods, or only by one. Most laboratories participated for the full study.

The cheese test materials were dispatched directly by CECALAIT separately from the other food types and RM, due to their shorter shelf life. All participants except two commenced examination of the cheese on the prescribed day. Two participants commenced the examinations 1 day later but the results of the stability tests enabled to take into account their results. No temperature abuse above 20 °C was recorded. For the meat, egg powder test materials and RM, all commenced their examinations within the prescribed period. Of these, three laboratories received packages which had been temperature abused above 20 °C but the storage temperature is not crucial for this type of dried materials

for a short period of time, as shown above for egg samples and RM.

3.3. Results excluded from further analysis

Some results were excluded from further statistical analysis for the following reasons (other than statistical):

- (i) Deviations from the protocol. The results from laboratory 21 for meat samples were excluded due to the extended time between preparation of initial suspension and end of inoculation of plates. All the results from laboratory 22 for BP medium were excluded due to the incomplete confirmation tests. Results from laboratory 11 for meat samples

Table 1

Precision data on the enumeration of Coagulase-Positive Staphylococci using EN ISO 6888-1 and -2 methods, calculated according to ISO 5725-2

Sample type	Method	Level	Number of participating labs	Number of labs for calculation	Mean (log cfu/g)	r (log) ^a	RSD _r ^b	R (log) ^c	RSD _R ^d
Cheese	BP ^e	low	22	19	3.33	0.40	4.25	0.55	5.88
		medium	22	19	5.12	0.24	1.63	0.54	3.70
		high	22	19	6.06	0.29	1.71	0.54	3.13
	RPFA ^f	low	19	18	3.25	0.21	2.24	0.25	2.73
		medium	19	18	5.03	0.19	1.36	0.38	2.70
		high	19	18	6.00	0.23	1.34	0.41	2.41
Meat	BP	low	23	18	3.27	0.36	3.86	0.78	8.40
		medium	23	18	4.20	0.31	2.58	0.50	4.18
		high	23	18	6.19	0.28	1.58	0.41	2.37
	RPFA	low	20	17	3.16	0.30	3.33	0.36	4.06
		medium	20	17	4.13	0.17	1.43	0.32	2.72
		high	20	17	6.08	0.24	1.41	0.35	2.05
Egg	BP	low	23	20	3.17	0.23	2.54	0.34	3.81
		medium	23	20	4.10	0.23	2.00	0.33	2.85
		high	23	20	5.23	0.23	1.57	0.36	2.46
	RPFA	low	20	19	3.25	0.34	3.72	0.40	4.29
		medium	20	19	4.20	0.19	1.57	0.30	2.49
		high	20	19	5.32	0.23	1.51	0.40	2.68
RM ^g	BP		23	20	3.79	0.18	1.72	0.58	5.49
	RPFA		20	19	3.85	0.18	1.62	0.61	5.61

^a r (log)=repeatability value in log₁₀ units, expressing the maximum occurring difference (alpha=0.05) between two counts, in repeatability conditions.

^b RSD_r=100 × S_r /global mean (for the given sample type/medium/level), expressing the value of r relative to the contamination level.

^c R (log)=reproducibility value in log₁₀ units, expressing the maximum occurring difference (alpha=0.05) between two counts, in reproducibility conditions.

^d RSD_R=100 × S_R /global mean (for the given sample type/medium/level), expressing the value of R relative to the contamination level.

^e Baird–Parker (EN ISO 6888-1).

^f Rabbit Plasma Fibrinogen Agar (EN ISO 6888-2).

^g Reference material.

were excluded since the standing period of 30 min during pre-treatment was omitted.

- (ii) Repeatability conditions. Two operators have partly analysed the same sample type, therefore, the results from laboratory 13 (meat and egg samples) and from laboratory 14 (cheese and RM) were excluded.
- (iii) Blank samples. High numbers of CPS were found by laboratory 21 when testing blank cheese samples by BP, so the results of this set of samples were excluded.
- (iv) Deviating counts. Only the results found to be much lower than expected by chance have been excluded: meat, BP, all levels for laboratories 9

and 12; egg, BP, all levels for laboratory 1 and RM, BP for laboratory 12.

3.4. Colony appearance on selective isolation media

Whereas no difficulty has been encountered for RPFA agar, some difficulties have arisen for BP medium. Mainly for meat samples, the distinction between typical and atypical colonies differed from one laboratory to the other one, maybe due to the difficulty to classify them according to the given definitions, or to the different appearance of colonies according to the brand of the media used. Among atypical

Table 2

Precision data on the enumeration of Coagulase-Positive Staphylococci using EN ISO 6888-1 and -2 methods, calculated according to the draft EN ISO 16140

Sample type	Method	Level	Number of participating labs	Number of labs for calculation	Median (log cfu/g)	r (log) ^a	RSD _r ^b	R (log) ^c	RSD _R ^d
Cheese	BP ^e	low	22	19	3.32	0.50	5.36	0.52	5.61
		medium	22	19	5.13	0.17	1.16	0.47	3.24
		high	22	19	6.04	0.33	1.96	0.66	3.91
	RPFA ^f	low	19	18	3.25	0.25	2.74	0.27	2.94
		medium	19	18	5.04	0.13	0.88	0.33	2.33
		high	19	18	6.04	0.17	0.98	0.32	1.91
Meat	BP	low	23	18	3.26	0.33	3.64	0.48	5.25
		medium	23	18	4.21	0.19	1.58	0.47	3.99
		high	23	18	6.20	0.29	1.67	0.39	2.26
	RPFA	low	20	17	3.12	0.25	2.85	0.32	3.66
		medium	20	17	4.15	0.21	1.79	0.28	2.38
		high	20	17	6.09	0.21	1.22	0.33	1.96
Egg	BP	low	23	20	3.20	0.25	2.78	0.32	3.57
		medium	23	20	4.10	0.25	2.17	0.29	2.55
		high	23	20	5.24	0.21	1.41	0.30	2.08
	RPFA	low	20	19	3.24	0.33	3.66	0.38	4.20
		medium	20	19	4.20	0.17	1.41	0.32	2.74
		high	20	19	5.30	0.25	1.68	0.39	2.64
RM ^g	BP		23	20	3.79	0.19	1.76	0.39	3.68
	RPFA		20	19	3.90	0.17	1.52	0.31	2.81

^a r (log)=repeatability value in log₁₀ units, expressing the maximum occurring difference (alpha=0.05) between two counts, in repeatability conditions.

^b RSD_r=100 × S_r /median of the mean of duplicates for the given sample type/medium/level, expressing the value of r relative to the contamination level.

^c R (log)=reproducibility value in log₁₀ units, expressing the maximum occurring difference (alpha=0.05) between two counts, in reproducibility conditions.

^d RSD_R=100 × S_R /median of the mean of duplicates for the given sample type/medium/level, expressing the value of R relative to the contamination level.

^e Baird–Parker (EN ISO 6888-1).

^f Rabbit Plasma Fibrinogen Agar (EN ISO 6888-2).

^g Reference material.

colonies, the proportion of coagulase-positive colonies differed also from one laboratory to the other one, due to the different selection of colonies for confirmation.

3.5. Comparison between BP and RPFA method

The significant differences between CPS counts using BP and RPFA methods were determined through an *F*-test, per sample type and level of contamination. A statistical difference, in terms of accuracy (comparison of the means between the two methods), was observed between the two methods for some combinations, but these differences appeared to be limited for such colony-count techniques. There was no difference in terms of precision (variability of results between duplicates and between laboratories), between the two methods, apart for two combinations where the precision was better with the RPFA method.

3.6. Determination of *r* and *R* values

The repeatability *r* of a method is defined as the value below which the absolute difference between two single test results obtained with the same method on identical test materials under the same conditions, may be expected to lie with a probability of 95% (Anon., 2000).

The reproducibility *R* of a method is defined as the value below which the absolute difference between two single test results obtained with the same method on identical test materials under different conditions, may be expected to lie with a probability of 95% (Anon., 2000).

Table 3
Average repeatability (*r*) and reproducibility (*R*) values^a obtained using EN ISO 6888-1 and -2 methods

Sample type	EN ISO 6888-1 method				EN ISO 6888-2 method			
	<i>r</i> (log)	<i>r</i>	<i>R</i> (log)	<i>R</i>	<i>r</i> (log)	<i>r</i>	<i>R</i> (log)	<i>R</i>
Cheese	0.33	2.1	0.55	3.6	0.18	1.5	0.31	2.0
Meat	0.27	1.9	0.45	2.8	0.22	1.7	0.31	2.0
Egg	0.24	1.7	0.30	2.0	0.25	1.8	0.36	2.3
All food samples	0.28	1.9	0.43	2.7	0.22	1.7	0.33	2.1
RM ^b	0.19	1.5	0.39	2.4	0.17	1.5	0.31	2.0

^a Calculated according to the draft EN ISO 16140.

^b Reference material.

Table 1 presents the mean and precision data obtained for each combination (sample type, method, level), and calculated according to the standard ISO 5725-2.

Table 2 gives the median and precision data obtained for each combination (sample type, method, level), and calculated according to the draft EN ISO 16140.

Since the two methods of calculation gave similar results, it was decided to express the precision data according to the draft EN ISO 16140, this approach being preferable for the domain of microbiology. Microbiological data do not always display a Gaussian distribution and the rejection of outlying results based only on statistical grounds is often controversial, e.g. are the values truly abnormal or do they reflect the reality of the method variability? The draft Standard EN ISO 16140, using robust statistics being less sensitive to extreme values, permits the statistical outliers to be retained. They also do not depend heavily on the distribution of the data.

Table 3 presents the average repeatability and reproducibility values per sample type, and per method, as well as overall values per method, calculated according to the draft EN ISO 16140. The values obtained with the cheese samples gave larger precision values than the other types of samples, but it was the most complex case, with a simulated flora of a raw milk cheese, including typical and atypical CPS colonies.

4. Conclusion

4.1. Repeatability

For the EN ISO methods 6888 Part 1 and Part 2, a general value of respectively $r = \log_{10} 0.28$ and 0.22 (expressed as an absolute difference between \log_{10} -transformed test results) can be used when examining food samples or $r = 1.9$ and 1.7 (expressed as an absolute ratio between test results on the normal scale).

4.2. Reproducibility

For the EN ISO methods 6888 Part 1 and Part 2, a general value of respectively $R = \log_{10} 0.43$ and 0.33 (expressed as an absolute difference between \log_{10} -

transformed test results) can be used when examining food samples, or $R=2.7$ and 2.1 (expressed as an absolute ratio between test results on the normal scale).

4.3. General conclusions from the trial

The precision data of this trial indicate a good precision of the two methods, RPFA being even better than BP. These values were similar to those obtained with the trial on enumeration of *B. cereus* (Schulden et al., 2000), and better than those obtained with the trial on enumeration of *L. monocytogenes* (Scotter et al., 2001b), both trials performed in the frame of the same project. This trial showed that it was possible to use the new standard EN ISO 16140 for the calculations of precision data obtained on an enumeration method in microbiology. It also enabled to give an equal status to both parts of the EN ISO standard 6888, whereas before the trial, part 1 was given priority. It finally highlighted the need to better describe the appearance of typical/atypical/other colonies (background flora) on Baird–Parker medium.

The outcome of this study was presented at the annual meeting of ISO/TC 34/SC 9 “Food products—Microbiology” (Vienna, June 2000). The committee has decided to include the precision data generated by this study and calculated according to draft EN ISO 16140 as amendments to the Standards EN ISO 6888-1 and -2, as well as to give an equal status to the BP and RPFA methods and to improve the description of typical/atypical/other colonies on Baird–Parker medium.

Acknowledgements

The validation of International Standards EN ISO 6888-1 and -2 has been carried out under the framework of Standards, Measurement and Testing project No. SMT4-CT96-2098. The authors wish to thank the following laboratories for their participation and cooperation in this collaborative trial: Mr. H. Stegeman, RIKILT-DLO, PO Box 230, 6700 AE Wageningen, The Netherlands; Mr. B. Wojton, National Veterinary Institute, Al Partyzantow 57, 24-100-Pulawy, Poland; Mrs. G. Vlaemynck, Dairy Research Station, Microbiology Laboratory, Brusselsesteenweg 370, 9090 Melle, Belgium; Mrs. IM. Vicente da Cruz, INETI-

IBQTA-DTIA, Estrada do Paco do Lumiau-Edificio S, Lisboa 1699 cedex, Portugal; Mrs. ML. Stecchini, Dipartimento di Scienze Degli Alimenti, Via Marangoni 97, 33100 Udine, Italy; Mr. H. Asperger, Veterinarmedizinische Universität Wien, Institute für Milchhygiene und Milchtechnologie, Joseph Baumann-Gasse 1, A-1210 Wien, Austria; Mr. V. Young, EHB Public Analysts Laboratory, Sir Patrick Dun’s, Lower Grant Canal Street, Dublin 2, Republic of Ireland; Mrs. C. Allaert, Lab of Microbiology of the Food Technology Department, Avda. Alcalde Rovira Roure, 177, 25198 Lleida, Spain; Mrs. T. Johansson, National Veterinary and Food Research Institute, Dept of Food Microbiology, PO Box 368, Fin-00231 Helsinki, Finland; Mrs. D. Roberts, Central Public Health Laboratory, Food Hygiene Laboratory, 61 Colindale Avenue, London, NW9 5HT, United Kingdom; Mrs. J. Anderson, Unilever Research, Colworth Laboratory, Colworth House, Sharnbrook, Bedford, MK44 1LQ, United Kingdom; Mrs. E. Bartelt, Federal Institute for Health Protection of Consumers, Unit Food Microbiology, Thielallee 88-92, 14195 Berlin, Germany; Mrs. E. Heck, Nestlé Research Center Lausanne, Quality and Safety Assurance Department, PO Box 44, 1000 Lausanne, Switzerland; Mr. K. Friedrich MUVA Kempten, Qualitäts-Und Laborzentrum, Hirnbeinstrasse 10, D-87435 Kempten (Allgäu), Postfach 2025, Allgäu, Germany; Mr. C. Wiberg, National Food Administration, Biology Division, Box 622, 75126 Uppsala, Sweden; Mr. J.P. Vincent, Institut Pasteur de Lille, Service de Microbiologie et Hygiène des Aliments, 369 rue Jules Guesde, 59650 Villeneuve d’Ascq, France; Mr. O. Pensala, Customs Laboratory, Tekniikantie 13, 02150 Espoo, Finland; Mr. S. Qvist, Danish Veterinary Service, Food Control Laboratory, Howitzvej 13, DK-2000 Frederiksberg, Denmark; Mrs. M. Lambiri, National School of Public Health, Bacteriology Department, 196 Alexandras avenue, 11521 Athens, Greece; Mr. C. Baylis, Campden and Chorleywood Food Research Association, Chipping Campden, Gloucestershire GL 556LD, United Kingdom; Mr. R. Havestadt, Sticking COKZ, PO Box 250, 3830 AG Leusden, The Netherlands, Mr. H. Becker, Lehrstuhl für Hygiene und Technologie der Milch, Tierärztliche Fakultät, Veterinar Strasse 13, 80539 München, Germany.

The authors also wish to acknowledge the technical assistance of all their laboratory staff in the

preparation and analysis of the test materials used in this study.

References

- Anon., 1989a. EN 45001 General Criteria for the Operation of Testing Laboratories. Comité Européen de Normalisation, Brussels.
- Anon., 1989b. ISO 4833 Microbiology of Food and Animal Feedingstuffs—General Guidance for the Enumeration of Micro-Organisms—Colony Count Technique at 30 degrees C International Organisation for Standardization, Geneva.
- Anon., 1994. ISO 5725-2 Application of Statistics—Accuracy (Trueness and Precision) of Measurement Methods and Results: Part 2. Basic Method for the Determination of Repeatability of a Standard Measurement Method. International Organisation for Standardisation, Geneva.
- Anon., 1996. FIL 122C Milk and Milk Products—Preparation of Test Materials and Dilutions for Microbiological Examination. International Dairy Federation, Brussels.
- Anon., 1999a. EN ISO 6888-1 Microbiology of Food and Animal Feeding Stuffs—Horizontal Method for the Enumeration of Coagulase-Positive Staphylococci (*Staphylococcus aureus* and other species): Part 1. Technique Using Baird–Parker Agar Medium. International Organisation for Standardisation, Geneva.
- Anon., 1999b. EN ISO 6888-2 Microbiology of Food and Animal Feeding Stuffs—Horizontal Method for the Enumeration of Coagulase-Positive Staphylococci (*Staphylococcus aureus* and other species): Part 2: Technique Using Rabbit-Plasma Fibrinogen Agar Medium. International Organisation for Standardisation, Geneva.
- Anon., 2000. Draft EN ISO 16140 Microbiology of Food and Animal Feeding Stuffs—Protocol for the Validation of Alternative Methods. International Organisation for Standardisation, Geneva.
- Rentenaar, I.M.F., 1996. MicroVal, a challenging eureka project. Food Control 7 (1), 31–36.
- Rousseeuw, P.J., Bassett Jr., G.W. 1990. The median: a robust averaging method for large data sets. J. Am. Stat. Assoc. 85, 97–104.
- Rousseeuw, P.J., Croux, C., 1993. Alternatives to the median absolute deviation. J. Am. Stat. Assoc. 88, 1273–1283.
- Schulten, S.M., In't Veld, P.H., Nagelkerke, N.J.D., Scotter, S., De Buyser, M.L., Rollier, P., Lahellec, C., 2000. Evaluation of the ISO 7932 standard for the enumeration of *Bacillus cereus* in foods. Int. J. Food Microbiol. 57, 53–61.
- Scotter, S.L., Langton, S., Lombard, B., Lahellec, C., Schulten, S., Nagelkerke, N., In't Veld, P.H., Rollier, P., 2001a. Validation of ISO method 11290: Part 1. Detection of *Listeria monocytogenes* in foods. Int. J. Food Microbiol. 64, 295–306.
- Scotter, S.L., Langton, S., Lombard, B., Lahellec, C., Schulten, S., Nagelkerke, N., In't Veld, P.H., Rollier, P., 2001b. Validation of ISO method 11290: Part 2. Enumeration of *Listeria monocytogenes* in foods. Int. J. Food Microbiol. 70, 121–129.

3.3 Annexe 3. Publication de l'essai de validation sur la détection de *L.monocytogenes*

Validation of ISO method 11290 Part 1 - Detection of *Listeria monocytogenes* in foods.

Scotter S. *et al.* *International Journal of Food Microbiology*, 2001, No 64, p. 295-306



ELSEVIER

International Journal of Food Microbiology 64 (2001) 295–306

INTERNATIONAL JOURNAL OF
Food Microbiology

www.elsevier.nl/locate/ijfoodmicro

Validation of ISO method 11290 Part 1 — Detection of *Listeria monocytogenes* in foods

S.L. Scotter^{a,*}, S. Langton^b, B. Lombard^c, S. Schulten^d, N. Nagelkerke^d, P.H. In't Veld^e,
P. Rollier^f, C. Lahellec^g

^aFood Microbiology Consultancy Service, Blackthorn House, Farlington, York YO61 1NW, UK

^bMinistry of Agriculture, Fisheries and Food, Central Science Laboratory, Sand Hutton, York YO41 1LZ, UK

^cAgence Française de Sécurité Sanitaire des Aliments, Lab. d'Etudes et de Recherches sur l'Hygiène et la Qualité des Aliments,
41 Rue du 11 Novembre 4918, 94700 Maison Alfort, France

^dMicrobiological Laboratory for Health Protection, National Institute of Public Health and the Environment, P.O. Box 1,
3720 BA Bilthoven, The Netherlands

^eInspectorate for Health Protection, Commodities and Veterinary Public Health, P.O. Box 2280, 5202 CG Den Bosch, The Netherlands

^fCECALAIT INRA-SRTAL, BP 89, 39801 Poligny cedex, France

^gAgence Française de Sécurité Sanitaire des Aliments, Direction Hygiène des Aliments, 23 Avenue du General de Gaulle, BP 19,
94701 Maison Alfort cedex, France

Received 5 July 2000; received in revised form 19 September 2000; accepted 26 October 2000

Abstract

The European and International Standard method for the detection of *Listeria monocytogenes*, described in EN ISO 11290 Part 1: 1997 (International Organisation for Standardisation, Geneva) was validated by order of the European Commission (Standards, Measurement and Testing Fourth Framework Programme Project SMT4-CT96-2098). Nineteen laboratories in 14 countries in Europe participated in a collaborative trial to determine the performance characteristics of the method, which are intended for publication in the corresponding standard. An additional objective of this project was to devise a new series of parameters to indicate the 'precision' of microbiological qualitative methods. The method was challenged with three food types, namely fresh cheese, minced beef and dried egg powder and a reference material. Inoculation levels ranged from 5 to 100 cfu/25 g. Each participant examined five replicates of each food type at three inoculum levels and five reference materials. Both PALCAM and Oxford media were assessed. All test materials were subjected to stringent homogeneity and stability testing before being used in the collaborative trial. The results demonstrated that the method prescribed in EN ISO 11290-1 had an overall sensitivity of 85.6% and a specificity of 97.4%. *L. monocytogenes* was detected in most cases after primary enrichment, although secondary enrichment often yielded further positives. However, a significant number of false-negative results were obtained with all food types when large numbers of *L. innocua* were present in the test materials. *L. innocua* tended to dominate *L. monocytogenes* during the selective enrichment stages and thus masked small numbers of colonies of *L. monocytogenes* on the isolation media. There was no evidence from this collaborative study to demonstrate a significant difference in performance between Oxford and PALCAM media. Due to the problem of false-negative results with this method as highlighted in this trial, recommendations have been made to ISO to launch a revision of the standard to improve the detection of low numbers of *L. monocytogenes* in foods. New statistical methods devised to advance the

*Corresponding author. Fax: +44-1347-811-636.

E-mail address: sue.scotter@newscientist.net (S.L. Scotter).

measurement of the performance of qualitative microbiological methods are also described. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: *Listeria monocytogenes*; EN ISO 11290-1; Validation; Performance characteristics; Food

1. Introduction

It is widely accepted that the use of validated methods is an essential part of any sound laboratory quality assurance programme. It is particularly important to have documented performance characteristics for methods that are utilised for statutory purposes or in cases of trade disputes. It is also the policy of the European Committee for Standardisation (CEN) to require methods that are to become European standards to be validated by collaborative study. Until recently, many internationally recognised reference methods were only ‘historically proven’ and had no defined precision parameters. In order to rectify this deficiency, project SMT4-CT96-2098, financed by the European Commission under the Fourth Framework Standards, Measurements and Testing Program (SM&T), was elaborated to determine the precision data in terms of repeatability (r) and reproducibility (R) or performance characteristics (qualitative methods) of six ISO (International Organisation for Standardisation) methods. These were prioritised for acceptance by CEN as follows: *Bacillus cereus* (enumeration), *Listeria monocytogenes* (detection and enumeration), *Staphylococcus aureus* (enumeration), *Clostridium perfringens* (enumeration) and *Salmonella* (detection). The determined performance characteristics are intended for publication by CEN in the corresponding CEN methods and subsequently by ISO via the Vienna Agreement.

The objectives of this EU-funded project and the results of the first validation carried out on ISO 7932 1993 ‘Enumeration of *Bacillus cereus*’ have been described previously by Schulten et al. (2000)

In this paper, we report the results of the second collaborative study in the project to validate EN ISO 11290 Part 1 (Anon., 1997) for the detection of *L. monocytogenes* in foods. The trial was carried out in March/April 1998. We also describe new statistical methods to enable the precision of qualitative meth-

ods to be assessed in the same way that repeatability and reproducibility are used with quantitative methods.

2. Materials and methods

2.1. Design of the trial

Nineteen laboratories from 14 countries in Europe participated in the collaborative trial. All laboratories were accredited or in the process of being accredited according to EN 45001 standard (Anon., 1989a). The design of the trial was described to participants in a detailed trial chronology, standard operating procedure (SOP) and a test report on which to record the results and return to the trial leader for analysis. A SOP was written, based on EN ISO 11290-1, because additional information was required to ensure that all test materials were treated the same way by each participant in the trial. In the test report, participants were required to detail all supplementary information that could have influenced their results. Prior to each collaborative trial, a ‘pre-trial’ was organised among the three contractors to verify that the SOP, as written, contained no potential ambiguities and to establish that the test materials were fit for purpose.

The method was challenged with three types of food, i.e. a food product from the dairy group (fresh cheese), meat group (minced beef) and dried food group (egg powder). Test materials used to validate the methods were artificially inoculated to achieve the desired inoculum levels and homogeneity. For each food type, three inoculum levels were prepared with target values of 0 colony forming units (cfu) of *L. monocytogenes* per 25 g, 5–10 cfu/25 g and 50–100 cfu/25 g. Additionally, each test material contained *L. innocua* at a level of 50–100 cfu/25 g plus a typical background flora at various levels depending upon the test material type. Each partici-

ant received 15 test materials for each food type, five replicates at each inoculum level. In addition, five reference materials (RM) were included in the trial to identify any serious errors in a participant's performance. All test materials were tested extensively prior to the pre-trial and collaborative trial to ensure that they could be maintained in a stable and homogeneous state for the duration of the distribution and trial period. The test materials were shipped to the participants by courier in polystyrene insulated boxes containing several ice packs to prevent exposure of the test materials to high temperatures during transport. Participants provided all media and reagents needed for the collaborative trial. Test material codes were randomised for each participant to prevent collusion among laboratories. The examination of the test materials was carried out within a specified time period as prescribed in the trial chronology.

2.2. Method under collaborative trial

The detection of *L. monocytogenes* according to EN ISO 11290-1 required the following successive stages:

- (a) Pre-treatment of each food type and RM according to a standardised procedure to obtain an initial sample for primary enrichment.
- (b) Primary enrichment in a selective liquid enrichment medium with reduced concentration of selective agents (half-Fraser broth) with incubation at 30°C for 24 h.
- (c) Secondary enrichment of a culture obtained from (b) in a selective liquid enrichment medium with full concentration of selective agents (Fraser broth) with incubation at 37°C for 48 h.
- (d) Plating out of cultures obtained in (b) and (c) and identification on either Oxford or PALCAM media with incubation at 30°C (Oxford) (Curtis, 1989) or 37°C (PALCAM) and examination after 24 h and, if necessary, after 48 h to check for the presence of characteristic colonies which are presumed to be *Listeria* spp.
- (e) Confirmation of the presence of *L. monocytogenes* by means of appropriate morphological, physiological and biochemical tests carried out on five presumptive colonies.

2.3. Preparation of test materials

Cheese test materials were prepared by CECALAIT (Poligny, France). Liquid cheese was inoculated with the test strains comprising *L. monocytogenes* (serovar 1/2a) and *L. innocua*, both of which had been isolated from milk by CECALAIT. In addition, a background flora isolated from cheese comprising *Lactococcus lactis* subsp. *lactis*, *Enterococcus faecalis*, *Lactobacillus paracasei* and *Lactobacillus plantarum* was also added at a level of approximately 10^4 cfu/g. *Listeria* strains were isolated on Oxford (Oxoid) and PALCAM agars (Merck), and then on Trypticase-Soya-Yeast Extract agar (Merck), *Lactococcus* and *Enterococcus* on M17 (Biokar Diagnostics) and *Lactobacillus* on de Man, Rogosa and Sharpe (MRS) (Biokar Diagnostics). Before inoculation, the *Listeria* strains were cultivated in Brain-Heart-Infusion broth (BHI) (Oxoid) for 18 h at 37°C and lactic bacteria in MRS or M17 broth for 18 h at 30°C. After mixing, cultures were diluted in 1/4 Ringer's solution (Merck) and inoculated into the liquid cheese. The inoculated cheese was clotted by the addition of rennet after dispensing aliquots of 100 g into individual vials. The cheese microflora was stabilised by the addition of an undisclosed bacteriostatic mixture. The bacteriostatic effect diminished when the test material was diluted during examination. Batches of cheese test materials were prepared for the collaborative trial immediately prior to dispatch to participants due to their relatively short stability.

Meat test materials were prepared by MAFF-CSL, UK. The test strains used to inoculate the test materials were *L. monocytogenes* NCTC 5214 and *L. innocua* NCTC 11288. All test materials were additionally inoculated with a simulated autochthonous meat flora comprising approximately 1×10^4 cfu/g *B. subtilis* and *Pseudomonas* sp. and also 1×10^5 cfu/g *Lactobacillus* sp. and 1×10^6 cfu/g *Micrococcus* sp. and 1×10^4 cfu/g *Pseudomonas* sp. Batches were prepared using retail fresh minced beef, which was freeze-dried in bulk, subsampled into vials (portions of 25 g equivalents) and irradiated at 25 kGy (Isotron, UK). The test strains and background flora were cultivated into Nutrient broth (Oxoid), diluted to the appropriate level and added as a 'cocktail' to each vial. Finally, the vials were freeze-dried a second time to stabilise the microflora and

stored at 4°C until use. Different batches were prepared for the pre-trial and collaborative trial.

Egg powder test materials were prepared by the RIVM-MGB, The Netherlands. Each test material was prepared by the participants immediately prior to examination by compositing a vial containing whole egg powder with an autochthonous flora at approximately 10^2 cfu/g and a vial with two capsules containing the *Listeria* species suspended in milk powder. For the 'negative' test materials, capsules contained only milk powder sterilised after filling by gamma irradiation (10 kGy). A batch of 25 kg of egg powder was obtained from an egg powder factory (Enthoven bv, Bennebroek, NL). Two series of 10 test materials of egg powder were tested for aerobic plate count (Anon., 1989b), *Enterobacteriaceae* count (Anon., 1993), *Listeria* enumeration (Anon., 1998) and *Listeria* detection (Anon., 1997). The test strains used in the egg test materials for the collaborative trial were *L. monocytogenes* Scott A, ALM 92 and *L. innocua* ALM 105. Both strains were cultivated in BHI for 24 h (shaken at 100 rpm) at 37°C. Each culture was centrifuged and the pellet suspended in 100 ml of milk, supplemented with 68.42 g saccharose (2 M). Approximately 10 ml of this suspension was dried onto sterile milk powder (Melkproduct 17, Nestlé) by means of a fluid bed spray granulation (STREA-1, Niro-Aeromatic). This resulted in two highly contaminated milk powders. The required contamination levels were prepared by mixing these powders with sterile milk powder in several steps (to optimise homogeneity of the test material) until the required contamination level was reached. For each prepared batch, the mixed powder was filled into gelatin capsules (ca. 0.3 g/capsule) in a laminar flow cabinet using an aluminium filling apparatus.

For the RM, capsules containing ca. 23 cfu *L. monocytogenes* Scott A, ALM 92/capsule were prepared by RIVM.

2.4. Homogeneity and stability of the test materials

The homogeneity and stability of the test materials were tested according to the procedures of the laboratory expert in their preparation. Acceptance of all production batches for use in the trials was made on the basis of achieving satisfactory homogeneity

and stability according to the laboratories' own criteria.

For the cheese test materials, stability at 4°C was measured by analysing, over several days, three test materials in duplicate. For each analysis, 10 g of cheese was diluted in 90 ml K_2HPO_4 solution (Anon., 1992) and blended for 3 min. This suspension was inoculated onto two plates of Oxford agar using a spiral plating system. Only a batch containing 10^4 cfu *L. monocytogenes* was tested. For homogeneity, 10 test materials of the same batch were examined in duplicate. The CECALAIT computer program was used for the statistical treatments. χ^2 (Chi square) was calculated on the total count (\log_{10}) of colonies per sample to compare the dispersion of the contamination to the dispersion obtained in a Poisson series.

For the meat test materials, stability of *L. monocytogenes* in minced beef stored at 4°C was verified over a 30 day period. Ten replicate vials were enumerated on Oxford agar at each interval at days 0, 11, 17, 25 and 30. The homogeneity of the batch was verified by enumerating 10 replicate vials immediately after production and again at the same time that participants examined the test materials. An *F*-test ($\alpha = 0.05$) was applied using \log_{10} transformed counts and the stability and homogeneity considered satisfactory if the *F*-test was not significant.

The homogeneity of both the egg powder test materials and the capsules was determined. For the capsules containing the target organisms at the low and high contamination levels, 25 capsules were examined to determine the variation between duplicate counts, i.e. within test material (T_1 test) and between test materials (T_2 test) (Schulten et al., 2000). The stability of the capsules at -20°C was assessed for two batches (*L. monocytogenes* 50–100 cfu/25 g and *L. innocua* 50–100 cfu/25 g). Ten test materials per week were examined by plating on PALCAM agar and incubating for 48 h over a period of 10 weeks.

2.5. Statistical analysis of the data

As there is no internationally agreed protocol for the statistical treatment of qualitative microbiological data, both traditional and new applications of routine statistical methods were applied to the data generated

from this trial to try to obtain useful measures of performance such as those used for quantitative methods. The first parameter to be determined was accuracy; for positive test materials this is termed sensitivity, defined as the percentage of known positive test materials that are correctly identified as positives. For negative test materials it is termed specificity, defined as the percentage of known negative test materials that are correctly identified as negatives. In addition, two new parameters were derived termed 'accordance' and 'concordance' which were designed to be equivalent to the more familiar parameters of repeatability and reproducibility, respectively, as determined for quantitative methods. Accordance is defined as the percentage chance of finding the same result (i.e. both positive or both negative) from two identical test materials analysed in the same laboratory, under standard repeatability conditions. To calculate it from the results of an inter-laboratory test, the probability that two samples give the same result is calculated for each participating laboratory in turn, and this probability is then averaged over all laboratories. Similarly, the concordance is defined as the percentage chance of finding the same result from two identical test materials analysed in different laboratories. To calculate it from the results of an inter-laboratory test, each observation in each participating laboratory is taken in turn, pairing it with all results from similar test materials from all the other laboratories. The concordance is the percentage of all pairings giving the same results over all possible pairings of data.

3. Results and discussion

3.1. Stability and homogeneity of the test materials

All test materials were considered sufficiently stable and homogeneous for the trial. Cheese test materials stored at 4°C were stable and homogeneous for up to 7 days. Meat test materials were stable and homogeneous when stored at 4°C for up to 1 month. For the egg test materials, separate examinations were carried out on the egg powder and capsules containing *L. innocua* and *L. monocytogenes*. Both the egg powder and *L. innocua* capsules were stable and homogeneous over a 10 week period when stored at -20°C. However, for *L. monocytogenes* 50–100

cfu/capsule, the regression coefficient was significantly decreasing (\log_{10} 0.0065 per week), but as the test materials were analysed within a 3 week period this decrease is unlikely to have had a significant effect on the results.

3.2. General results of the trial

Cheese test materials were received by most participants within 48 h of dispatch. One laboratory could not commence their examinations until 4 days after dispatch because their test materials did not arrive in time to start before a weekend. However, the test materials were stored at 4°C and as they were known to be stable for up to 7 days, the laboratory was instructed to continue with the trial. Two laboratories elected not to examine cheese test materials. No problems were encountered with distribution of the meat, egg powder test materials and RMs. Temperature control during shipment was satisfactory as no test materials had been subjected to any excessive temperature abuse according to data obtained from the temperature monitoring devices included in the packages.

Due to the level of work required in this trial, participants were instructed to detect *L. monocytogenes* using only one of the two media prescribed in EN ISO 11290-1, i.e. Oxford or PALCAM (whereas EN ISO 11290-1 requires both media to be used). Nine laboratories used Oxford agar and 11 used PALCAM agar.

3.3. Results excluded from further analysis

For this qualitative trial it was agreed that results from laboratories would not be excluded unless they fell into three specified categories: (a) the test materials had received a significant temperature abuse during shipment, (b) the laboratory had clearly deviated from the specified standard operating procedure or (c) the performance of the laboratory was questionable as indicated by large numbers of false-positive or false-negative results more than would be expected by chance.

Two laboratories fell into the above categories. One returned a significantly high proportion of false-positive results and thus their competence in confirmation of *L. monocytogenes*, relative to data returned by other participants, was questionable. A

second laboratory failed to rehydrate the meat test materials correctly prior to commencing examination and, as their results for the meat test materials were generally poor, their data (for the meat test materials only) was excluded from further analysis.

3.4. Statistical analysis of data

3.4.1. Summary of participants' results

A summary of participants' results depending upon the isolation medium used is given in Tables 1 and 2.

3.4.2. Specificity

A total of 15 negative test materials were examined by each laboratory (five per food type). After the exclusion of data from two laboratories, a total of eight false-positive results were reported. The percentage of false-positive results is shown in relation to food and isolation medium in Table 3. Although there was a higher percentage of false positives with PALCAM, analysis by logistic regression revealed no significant differences between media, food types or laboratories in the proportion of false positives observed. The overall specificity of the method was 97.4%.

Table 1
Summary of participants' results on PALCAM agar: No. test materials correctly identified from a possible maximum of five^a

Test material	Inoculum level	Laboratory code										
		1	2	3	5	8	10	14	18	19	20	22
Cheese	B	5	4	NT	5	0	5	5	NT	4	5	5
	L	5	4	NT	4	5	5	5	NT	5	5	5
	H	5	5	NT	5	5	5	5	NT	5	5	5
Meat	B	5	5	5	3	0	5	5	5	5	5	5
	L	4	1	5	4	5	5	4	5	3	3	5
	H	5	3	5	5	5	5	5	5	5	5	5
Egg	B	5	5	5	5	4	5	5	5	5	4	5
	L	4	5	1	2	4	2	2	3	3	4	3
	H	5	5	5	5	5	3	5	3	5	5	5
RM		5	5	5	5	4	5	5	5	5	5	3

^a B, negative for *L. monocytogenes*; L, 5–10 cfu/25 g *L. monocytogenes* plus 50–100 cfu/25 g *L. innocua*; H, 50–100 cfu/25 g *L. monocytogenes* plus 50–100 cfu/25 g *L. innocua*. NT, not tested.

Table 2
Summary of participants' results on Oxford agar: No. test materials correctly identified from a possible maximum of five^a

Test material	Inoculum level	Laboratory code									
		7	9	11	13	14	17	21	22	23	
Cheese	B	5	4	5	5	5	5	5	5	5	5
	L	4	1	3	4	5	4	4	4	5	5
	H	5	5	5	5	5	5	5	5	5	5
Meat	B	5	5	5	5	5	5	5	5	5	5
	L	5	5	5	4	4	5	5	1	3	
	H	5	5	5	5	5	5	5	5	5	5
Egg	B	5	5	5	5	5	4	5	5	5	5
	L	2	2	2	1	3	3	4	3	2	
	H	4	5	5	3	5	2	5	5	4	
RM		3	4	2	5	5	5	5	3	5	

^a B, negative for *L. monocytogenes*; L, 5–10 cfu/25 g *L. monocytogenes*; H, 50–100 cfu/25 g *L. monocytogenes* plus 50–100 cfu/25 g *L. innocua*.

Table 3
Percentage of false-positive results in relation to food and medium^a

Food type	Medium					
	Oxford		PALCAM		All	
	<i>n</i>	Mean %	<i>n</i>	Mean %	<i>n</i>	Mean %
Cheese	9	2.2	8	5.0	17	3.5
Meat	9	0.0	9	4.4	18	2.2
Egg	9	2.2	10	2.0	19	2.1
All	27	1.5	27	3.7	54	2.6

^a 'n' refers to the number of sets each of five replicates, e.g. *n* = 9 implies 45 test materials in total.

3.4.3. Sensitivity

The main objective of this collaborative trial was to establish the overall performance characteristics of the method and to determine the influence of specific factors, such as the effect of the isolation medium or concentration of target organism, on method performance. As for the negative test materials, the effects of the various factors on the probability of correctly identifying a positive sample were investigated using chi-squared statistics from a logistic regression analysis. Tables 4 and 5 show the sensitivity of the method in relation to food type/medium and food type/inoculum level, respectively.

Large differences in performance of the method in relation to inoculum levels and food types were observed with a significantly lower sensitivity for egg powder test materials. It is not clear why the method showed a poor sensitivity with egg powder test materials as the significant level of homogeneity and stability testing undertaken prior to the trial verified that the materials could be produced to the

Table 4
Percentage of correct results (sensitivity) in relation to medium used^a

Food type	Medium					
	Oxford		PALCAM		All	
	<i>n</i>	Mean %	<i>n</i>	Mean %	<i>n</i>	Mean %
Cheese	18	88.9	16	97.5	34	92.9
Meat	18	91.1	18	92.2	36	91.7
Egg	18	66.7	20	75.0	38	71.1
RM	9	82.2	10	96.0	19	89.5
All	63	82.2	64	88.7	127	85.6

^a 'n' refers to the number of sets each of five replicates, e.g. *n* = 9 implies 45 test materials in total.

Table 5
Percentage of correct results (sensitivity) in relation to inoculum level^a

Food type	Inoculum level					
	Low		High		All	
	<i>n</i>	Mean %	<i>n</i>	Mean %	<i>n</i>	Mean %
Cheese	17	85.9	17	100.0	34	92.9
Meat	18	83.3	18	100.0	36	91.7
Egg	19	53.7	19	88.4	38	71.1
RM					19	89.5
All	54	73.7	54	95.9	127	85.6

^a 'n' refers to the number of sets each of five replicates, e.g. *n* = 9 implies 45 test materials in total.

required quality standards. However, it was noted that, at the higher level, *L. monocytogenes* showed a slight decline in viability with time and it is possible that this may have been more manifest during the trial. There was also evidence that laboratories differed in their relative performance with the different foods. Data indicated that the method was marginally more sensitive when PALCAM agar was used for isolation, but this result was not highly significant ($0.05 < P < 0.1$). The overall sensitivity of the method was 85.6%.

3.4.4. Effect of primary and secondary enrichment on method sensitivity

One of the other factors which was of interest in relation to the method under test was whether the sensitivity was influenced by the primary and/or secondary enrichment stages. Table 6 shows the numbers of test materials found positive after primary and secondary enrichment. This variable was examined excluding results from five laboratories as they did not always carry out the secondary enrichment if the presence of *L. monocytogenes* had already been confirmed after the primary enrichment.

Most of the food test materials inoculated at the higher level which were found positive were detected after the primary enrichment (95.5%), whereas 30.0% of those inoculated at the low level required the secondary enrichment. This difference was highly significant ($P < 0.001$). There was also a marked difference in the sensitivity for the different foods ($P < 0.001$), with more cheese test materials being detected after the primary enrichment (97.5% of those found positive) compared to 73.8% for meat and 82.2% for egg. However, this observation should

Table 6
Number of positive samples after primary and/or secondary enrichment

Food	Inoculum level	No. test materials	Total No. positives	1° Positive 2° Positive	1° Positive 2° Negative	1° Negative 2° Negative	1° Negative 2° Positive
Cheese	Low	65	56	16	37	9	3
	High	65	65	36	29	0	0
	Both	130	121	52	66	9	3
Meat	Low	70	56	27	2	14	27
	High	70	70	59	5	0	6
	Both	140	126	86	7	14	33
Egg	Low	70	38	12	11	32	15
	High	70	63	39	21	7	3
	Both	140	101	51	32	39	18
RM	Rm	70	60	53	0	10	7
All	Low	205	150	55	50	55	45
	High	205	198	134	55	7	9
	All	480	408	242	105	72	61

be treated with some caution because of the significant interaction between laboratories and foods ($P < 0.001$). There was no significant difference in the performance of the isolation media from most of the data.

Where test materials were positive after primary enrichment, there were large differences between foods in the proportion that remained positive after secondary enrichment (Table 6). Most meat test materials (92.5%) remained positive, whereas only 44.1% of cheese test materials and 61.4% of egg test materials were also positive after secondary enrichment.

3.4.5. Performance assessment

Tables 7–10 show the overall performance characteristics of ISO 11290-1 in relation to food type and inoculum level. The data obtained from Oxford and PALCAM agars are combined. The specificity of the method did not vary greatly between food types, ranging from 96.5 to 97.9%. With respect to sensitivity, for minced meat and fresh cheese, similar results (83.3 and 85.9%, respectively) were obtained at the lower inoculum level (5–10 cfu *L. monocytogenes* per 25 g). At the higher inoculum level (50–100 cfu *L. monocytogenes* per 25 g) the sensitivity increased in both cases to 100%. For the reference

Table 7
Method performance characteristics with dried egg powder

	Sample/level of contamination		
	Dried egg powder (negative)	Dried egg powder (low level)	Dried egg powder (high level)
Year of inter-laboratory test	1998	1998	1998
No. of laboratories with valid results	19	19	19
No. of samples per laboratory	5	5	5
No. of accepted samples	90	90	90
Accuracy (specificity) (%)	97.9	–	–
Accuracy (sensitivity) (%)	–	53.7	88.4
Accordance (%)	95.8	48.4	83.2
Concordance (%)	95.8	49.8	79.1

Table 8
Method performance characteristics with minced meat

	Sample/level of contamination		
	Minced meat (negative)	Minced meat (low level)	Minced meat (high level)
Year of inter-laboratory test	1998	1998	1998
No. of laboratories with valid results	18	18	18
No. of samples per laboratory	5	5	5
No. of accepted samples	85	85	85
Accuracy (specificity) (%)	97.8	–	–
Accuracy (sensitivity) (%)	–	83.3	100.0
Accordance (%)	96.7	76.7	100.0
Concordance (%)	95.6	71.7	100.0

Table 9
Method performance characteristics with fresh cheese

	Sample/level of contamination		
	Fresh cheese (negative)	Fresh cheese (low level)	Fresh cheese (high level)
Year of inter-laboratory test	1998	1998	1998
No. of laboratories with valid results	17	17	17
No. of samples per laboratory	5	5	5
No. of accepted samples	90	90	90
Accuracy (specificity) (%)	96.5	–	–
Accuracy (sensitivity) (%)	–	85.9	100.0
Accordance (%)	92.9	80.0	100.0
Concordance (%)	93.1	75.2	100.0

Table 10
Method performance characteristics with reference materials

	Sample/level of contamination Reference material (capsules containing 23 cfu <i>L. monocytogenes</i>)
Year of inter-laboratory test	1998
No. of laboratories with valid results	19
No. of samples per laboratory	5
No. of accepted samples	90
Accuracy (specificity) (%)	–
Accuracy (sensitivity) (%)	89.5
Accordance (%)	85.3
Concordance (%)	80.8

materials containing ca. 23 cfu pure culture of *L. monocytogenes*, a slightly higher sensitivity of 89.5% was obtained. However, the sensitivity of the method, when challenged with egg powder, was poor, with only 53.7% of samples correctly identified as positive at the lower inoculum level and 88.4% at the higher level. The reasons for this disparity in this study are uncertain.

With respect to the accordance value, i.e. the percentage chance of obtaining the same result on two identical test samples under repeatability conditions, broadly similar results were obtained for each of the negative food types, ranging from 92.9 to 96.7%. For samples containing only a low level of *L. monocytogenes*, results in excess of 76% were obtained for meat, cheese and reference materials. However, an accordance of only 48.4% was achieved for the egg powder. This value improved to 83.2% at the higher inoculum, but was still poor compared to meat and cheese at 100%. It is observed from this data that there is a direct relationship between sensitivity and accordance/concordance values in that, when the sensitivity is high, there is also a high value for the accordance/concordance.

With respect to concordance, i.e. the percentage chance of obtaining the same result on two identical test samples under reproducibility conditions, again broadly similar results were obtained for each of the negative food types, ranging from 93.1 to 95.8%. For samples containing only a low level of *L. monocytogenes*, results in excess of 71% were obtained for meat, cheese and reference materials. However, a concordance of only 49.8% was achieved for the egg powder. This value improved to 79.1% at the higher inoculum, but was still poor compared to meat and cheese which again had excellent concordance values of 100% at the higher inoculum level. It is unclear why the results obtained with dried egg powder were poor compared to the other food types as the test materials were known to be stable and homogeneous for the duration of the trial period. Their method of preparation was different from the meat and cheese test materials as participants were provided with capsules containing the required levels of *L. monocytogenes* and *L. innocua* in skimmed milk powder and used these to inoculate the base product immediately before testing, whereas the meat and cheese test materials were directly inoculated, using broth cultures, by the responsible contractor. It is possible that

the target organisms in the egg powder were exposed to a higher degree of stress and therefore the method did not perform as well with these test materials.

3.5. Confirmation test results

The method for the detection of *L. monocytogenes* requires a number of procedures to be carried out to confirm presumptive colonies isolated on PALCAM or Oxford media. These tests are carried out in a series of stages. Three or four preliminary tests (catalase, Gram staining, motility and Henry illumination) are carried out in order to select *Listeria* spp. colonies. If these tests are indicative of *Listeria* spp., haemolysis, carbohydrate utilisation and the CAMP tests must be performed to confirm the presence of *L. monocytogenes* and also differentiate between some other *Listeria* spp.

As large numbers of presumptive colonies were expected in this trial, participants were allowed some flexibility in the confirmatory tests used and thus the Gram stain, motility and CAMP tests were considered optional. Participants were also permitted to use one of two types of haemolysis tests; a microwell test using a 2% suspension of sheep erythrocytes suspended in PBS or the more traditional sheep blood agar plates. However, participants were instructed that where haemolysis testing was carried out using only blood agar plates, the performance of the CAMP test became obligatory. The haemolysis test is the critical test for the differentiation of *L. monocytogenes* and *L. innocua* and, as both strains were present in the trial test materials, it was this test that caused most problems for some laboratories. Some laboratories obtained poor results with sheep blood agar plates and did not always supplement this test with the CAMP test as instructed. Most laboratories used the microwell method (despite the current discrepancy in the correct concentration of sheep erythrocytes to be used in the test). It is probable that many of the false-negative results were due to difficulties in interpretation of the haemolysis reactions. However, of greater concern is the performance of the method with respect to its selectivity for *L. monocytogenes*. It has become apparent both during the preparation of test materials for use in the trial and also from the trial results that when *L. monocytogenes* is in association with *L. innocua* in a food, the dual enrichment procedure allows *L. in-*

nocua to become dominant. Consequently, the picking of only five colonies for confirmation may result in the reporting of false-negative results, as *L. monocytogenes* colonies will always be fewer in number. Several participants reported that the picking of more colonies resulted in a higher detection rate.

4. Conclusion

The method prescribed in EN ISO 11290-1 has an overall sensitivity of 85.6% and a specificity of 97.4%. As a general indication of accordance and concordance, respectively, values of 95.1% (negatives), 81.4% (positives), 85.3% (RM) and 94.8% (negatives), 79.3% (positives) and 80.8% (RM) can be used.

In most cases, proceeding with a secondary enrichment will yield a higher sensitivity. However, when *L. monocytogenes* is known to occur in a foodstuff in close association with *L. innocua*, it is possible that *L. monocytogenes* will not be detected using this method due to the overgrowth of *L. innocua* during selective enrichment thus masking the presence of *L. monocytogenes* on both Oxford and PALCAM agars. The importance of picking off sufficient colonies for confirmation and carrying out the haemolysis tests correctly should not be underestimated. The value of the motility test and the tumbling action being highly characteristic for *L. monocytogenes* should also not be forgotten. Additionally, isolation from both the primary and secondary enrichment should be carried out as it is possible, with some food types, that false-negative results may be obtained if isolation is effected only after secondary enrichment.

From the overall results of this collaborative study it has been recommended to CEN and ISO to:

- Incorporate the performance criteria derived from this trial, together with appropriate explanations of how to interpret the various parameters, as an amendment to the Standard EN ISO 11290-1.
- Include in this amendment the correct value (2%) of the percentage of the sheep erythrocytes for the microwell haemolysis test in line with the enumeration method EN ISO 11290-2.
- Launch a revision of EN ISO 11290-1 in order to

address the problem of potential false-negative results. For example, by improving the isolation stage by replacing one of the current media with a medium which can differentiate *L. monocytogenes* from other species, especially *L. innocua* (e.g., a medium which reveals haemolytic colonies at the isolation stage). Moreover, the possibility of using molecular typing should be introduced as an alternative to the classical confirmation tests, since several commercial sources are now available.

Acknowledgements

The validation of International Standard EN ISO 11290-1 has been carried out under the framework of a Standards, Measurement and Testing project No. SMT4-CT96-2098. The authors wish to thank the following laboratories for their participation and cooperation in this collaborative trial: Mr ir H. Stegeman, RIKILT-DLO, PO Box 230, 6700 AE Wageningen, The Netherlands; Mr P.H. in't Veld, National Institute of Public Health and Environment, Microbiological Laboratory for Health Protection, Antonie Van Leeuwenhoeklaan, 93720 BA Bilt-hoven, The Netherlands; Mrs M. Lambiri, National School of Public Health, Bacteriology Department, 196 Alexandras Avenue, 115 21 Athens, Greece; Dr Boleslaw Wojton, National Veterinary Institute, Al Partyzantow 57, 24-100-Pulawy, Poland; G. Vlaemyneck, Dairy Research Station, Microbiology Laboratory, Brusselsesteenweg 370, 9090 Melle, Belgium; I.M. Vicente da Cruz, INETI-IBQTA-DTIA, Estrada do Paco do Lumiau-Edificio S, Lisboa 1699 cedex, Portugal; Prof. M.L. Stecchini, Dipartimento di Scienze Degli Alimenti, Via Marangoni 97, 33100 Udine, Italy; Dr H. Asperger, Veterinarmedizinische Universitat Wien, Institute fur Milchhygiene und Milchtechnologie, Joseph Baumann-Gasse 1, A-1210 Wien, Austria; Mr V. Young, EHB Public Analysts Laboratory, Sir Patrick Dun's, Lower Grant Canal Street, Dublin 2, Republic of Ireland; Prof. Debevere, Universiteit Gent, Faculteit Landbouwkundige En toegepaste Biologische Wetenschappen wak groep Levensmittelentechnologie En Voeding, B-9000 Gent; Mrs C. Allaert, Lab of Microbiology of the Food Technology Department, Avda. Alcalde Rovira Roure, 177, 25198 Lleida, Spain; Dr T.

Johansson, National Veterinary and Food Research Institute, Dept of Food Microbiology, PO Box 368, Fin-00231 Helsinki, Finland; DVM Marylene Bonhert, CNEVA Ploufragen, B.P. 53, Zoopole Beaucemarie, 22440 Ploufragan, France; Dr D. Roberts, Central Public Health Laboratory, Food Hygiene Laboratory, 61 Colindale Avenue, London NW9 5HT; J. Anderson, Unilever Research, Colworth Laboratory, Colworth House, Sharnbrook, Bedford MK44 1LQ; Prof. P. Teufel/Dr E. Bartelt, Federal Institute for Health Protection of Consumers, Unit Food Microbiology, Thielallee 88–92, 14195 Berlin, Germany; Dr J.L. Cordier, Nestlé Research Center Lausanne, Quality and Safety Assurance Department, PO Box 44, 1000 Lausanne, Switzerland; Dr K. Friedrich, MUVA Kempten, Qualitäts-Und Laborzentrum, Hirnbeinstrasse 10, D-87435 Kempten (Allgau), Postfach 2025, Allgau, Germany; Mr J.C. van Dijk, Stichting COKZ, PO Box 250, 3830 AG Leusden, The Netherlands. The authors also wish to acknowledge the expert statistical advice supplied by Dr Remi Chevennement and also the technical assistance of all their laboratory staff in the preparation and analysis of the test materials used in this study.

References

- Anon., 1989a. EN 45001 — General criteria for the operation of testing laboratories. Comité Européen de Normalization, Brussels.
- Anon., 1989b. ISO 4833 Microbiology of food and animal feedingstuffs — General guidance for the enumeration of micro-organisms — Colony count technique at 30 degrees C. International Organization for Standardization, Geneva.
- Anon., 1992. FIL 122b Milk and milk products — Preparation of samples and dilutions for microbiological examination. International Dairy Federation, Brussels.
- Anon., 1993. ISO 7402 Microbiology of food and animal feedingstuffs — Enumeration of Enterobacteriaceae. International Organization for Standardization, Geneva.
- Anon., 1997. EN ISO 11290-1 Microbiology of food and animal feedingstuffs — Horizontal method for the detection and enumeration of *Listeria monocytogenes* — Part 1: Detection. International Organisation for Standardisation, Geneva.
- Anon., 1998. EN ISO 11290-2 Microbiology of food and animal feedingstuffs — Horizontal method for the detection and enumeration of *Listeria monocytogenes* — Part 2: Enumeration. International Organisation for Standardisation, Geneva.
- Curtis, G.D., 1989. Selective agents for *Listeria* can inhibit their growth. Lett. Appl. Microbiol. 8, 169–172.
- Schulzen, S., in't Veld, P.H., Nagelkerke, N.J.D., Scotter, S., de Buyser, M.L., Rollier, P., Lahellec, C., 2000. Evaluation of the ISO 7932 standard for the enumeration of *Bacillus cereus* in foods. Int. J. Food Microbiol. (in press).

3.4 Annexe 4. Rapport de l'essai de validation sur la détection de *Salmonella*

(sans l'annexe B : modes opératoires et l'annexe C : rapport d'essai)

Validation of draft Standard EN ISO/DIS 6579 : 2000 – Detection of *Salmonella* in foods - Final Report

Bohnert ML. *et al.* Ploufragan : AFSSA, 2001.



Validation of ISO Microbiological Methods

Detection of *Salmonella* according to draft Standard pr EN ISO/DIS 6579: 2000

Final report, February 2001

Marylène BOHNERT, Florence HUMBERT

Bertrand LOMBARD

Agence Française de Sécurité Sanitaire des Aliments

AFSSA-Site de Ploufragan

AFSSA-Site de Maisons-Alfort

This investigation has been performed by order and for the account of the European Commission, DG XII, Brussels (EU contract number: SMT4-CT96-2098).

Table of contents

Abstract

Summary

1. Introduction
2. Materials and Methods
 - 2.1. Preparation of trial test materials
 - 2.2. Homogeneity and stability of the test materials
 - 2.3. Design of the collaborative trial
3. Results and discussion
 - 3.1. Stability and homogeneity of the test materials
 - 3.2. General results of the trial
 - 3.3. Statistical analysis of data
4. Conclusions and Recommendations to CEN and ISO
5. References
6. Participants in the project EU Project SMT4-CT96-2098
7. European participants in the collaborative trial

ANNEX A TIME SCHEDULE FOR THE COLLABORATIVE TRIALS:

Detection of Salmonella by use the ISO 6579 Standard and Detection of Salmonella by use the AOAC Standard

ANNEX B STANDARD OPERATING PROCEDURES, SOP-1 & SOP-2:

ANNEX C TEST REPORT FOR COLLABORATIVE TRIAL

ANNEX C CONTINUED: Example of TEST REPORT of sample examined by SOP-1 (like ISO)

Example of TEST REPORT of sample examined by SOP-2 (like AOAC)

ANNEX D : Summary of collaborative trial results,

Reference materials and cheese samples + capsule

ANNEX D CONTINUED: diced poultry samples

ANNEX D CONTINUED: egg powder samples

ANNEX D CONTINUED: cheese samples

Abstract

The method for the detection of *Salmonella*, described in draft Standard prEN ISO/DIS 6579: 2000 was validated by order of the European Commission, DG XII (Standards, Measurement and Testing Program). Cheese, meat, dried egg powder and a reference material were examined by 17 laboratories in 12 countries in Europe. In addition 10 US laboratories participated in the trial

Statistical analyses were carried out to determine the following:

overall performance of the test method for detecting *Salmonellas* in foods;

significant differences in performance of the 2 standards;

whether differences in performance can be related to food type and/or inoculum level;

overall performance of laboratories using the test method.

The results of this collaborative study are reported herein.

M.Bohnert, B. Lombard, F. Humbert, P. Rollier, K. Mooijman, R. Leuschner, C. Lahellec

February 2001

SUMMARY

The method for the detection of *Salmonella* described in draft Standard prEN ISO/DIS 6579: 2000 has been validated under the framework of a Standards, Measurement and Testing project (EU S, M & T project no. SMT4-CT96-2098). The results are intended for publication by the European Committee for Standardization (CEN) and International Organisation for Standardization (ISO) in the corresponding CEN/ISO Standard.

The first objective of the collaborative study was to determine the performance characteristics of the method for the detection of *Salmonella* as described in the draft Standard prEN ISO 6579: 2000. A second objective, specific to this trial, was to seek recognition of the Standard by AOAC, thus the need to compare the draft Standard with AOACI Official Method 967.25, 967.26, Detection of *Salmonella*. In this frame, 10 US laboratories took part in the trial.

Egg powder, meat and fresh cheese samples were used as food matrices respectively contaminated with *S. Panama*, *S. Typhimurium* and *S. Montevideo*. For each food type, 15 samples were examined: 5 replicates at 3 inoculum levels. In addition, two negative samples of cheese were to be contaminated by the laboratories with a capsule containing a typical *Salmonella* Typhimurium. Also 5 reference materials (RM's, capsules containing milk powder, artificially contaminated with *Salmonella*) were included in the trial. Samples were examined by both ISO and AOAC methods or only by ISO, depending on the choice of participants. All test materials were verified to ensure fitness for purpose.

The results demonstrated the following method as in draft Standard prEN ISO 6579 is generally satisfactory (overall sensitivity *ca* 92%).

Thus, the SMT project recommends to CEN & ISO:

- To include performance characteristics (possibly completed with the data generated from the additional trials) into the new CEN/ISO Standard 6579 with including explanatory notes on the origin of the data (data derived only from a few types of strains and from 3 food matrices, lactose positive strain used for the cheese samples).
- To give explanations in the Standard on how to interpret these performance characteristics;
- Enrichment stage : to reconsider the choice of the 2nd enrichment broth, since several participants had encountered difficulties when preparing and using MKTTn;
- Isolation stage : To introduce a note stating that a 2nd 90 mm plate is particularly necessary for the analysis of heavy contaminated samples;
- Confirmation stage :
 - to precise the confirmation procedure and interpretation (refer to Table 2 of the ISO Standard);
 - to restrict the serological confirmation, in the precise frame of the detection of *Salmonella* in foods, to the use of O polyvalent sera (A to O:67)⁵
- To aim at defining a common AOAC & ISO method for detection of *Salmonella* in foods.

⁵ The exact wording of this recommendation (in particular the use of only one or several polyvalent sera) would be investigated by K. Moojman within RIVM as CRL for *Salmonella*.

1. Introduction

For acceptance of standards by the European Committee for Standardization (CEN) and the International Organisation for Standardization (ISO), methods need to be validated. Project no. SMT4-CT96-2098, financed by the European Commission, DG XII (4th Framework Standards, Measurements and Testing Program (SM&T)), was elaborated to determine the precision data in terms of repeatability ® and reproducibility ® or performance characteristics (for qualitative methods) of six ISO methods. These were prioritised for acceptance by CEN and ISO as follows: *Bacillus cereus* (enumeration), *Listeria monocytogenes* (detection and enumeration), *Staphylococcus aureus* (enumeration), *Clostridium perfringens* (enumeration) and *Salmonella* (detection). The results are intended for publication by ISO in the corresponding ISO methods and subsequently by CEN via the Vienna Agreement.

Three contractors are involved in this project: Agence Française de Sécurité Sanitaire des Aliments (AFSSA, France), also co-ordinator of the project; the Ministry of Agriculture, Fisheries and Food, Central Science Laboratory (MAFF-CSL, York, UK) and National Institute of Public Health and the Environment, Microbiological Laboratory for Health Protection (RIVM-MGB, Bilthoven, The Netherlands). Each contractor is responsible as ‘trial leader’ for the validation of 2 of the 6 ISO methods.

All methods are validated by collaborative trial according to a similar design which conforms to the rules of the International Harmonised Protocol for Collaborative Studies [1]. Approximately 20 laboratories from various countries are invited to participate in each trial. Participants receive a detailed trial chronology, standard operating procedure (SOP) and a test report on which to record their results and return to the trial leader for analysis. Prior to each collaborative trial a “pre-trial” is organised between the three contractors to verify the SOP as written, identify any potential ambiguities and to establish that the test materials are fit for purpose. For quantitative studies a preliminary estimation of the precision of the test method is also possible from the pre-trial. For each validation, the method is challenged with three types of food from the meat, dairy and dried food groups. In addition, a reference material is included to identify any serious errors in a participant’s performance and to determine the maximum precision possible in the case of quantitative methods. Test materials used to validate the methods have to be artificially inoculated to achieve the desired inoculum levels and homogeneity. These materials are tested extensively prior to the pre-trial and collaborative trial to ensure that they can be maintained in a stable and homogeneous state for the duration of the distribution and trial period. Data from the collaborative trials is used to calculate 1) the repeatability ® and reproducibility ® of the quantitative methods in relation to food type and inoculum level and 2) the performance characteristics relating to an ‘equivalent value’ of r and R appropriate to qualitative data using model statistical systems developed as one of the objectives of this project.

This is the last report from this EU project describing the validation of the method for the detection of *Salmonella* as described in draft Standard prEN ISO/DIS 6579: 2000. The study was organised by AFSSA-site de Ploufragan.

Simultaneously to this validation, a comparison was made with the AOAC Official Method 967.25, 967.26, with the aim to get a full recognition from USA of the EN ISO Standard (= to seek validation of the Standard by AOAC for the foods submitted to the collaborative study). For this purpose, laboratories were asked to use both ISO and AOAC methods, and 10

US laboratories also participated in the collaborative study. The US commitment was co-ordinated by Mr Phil FELDSINE, BioControl, Bellevue.

This trial and the calculations were conducted according to the same design than the trial on the detection of *Listeria* [2].

For both ISO and AOAC methods, the test sample is subjected to a pre-enrichment followed by an enrichment in two broths. Isolation after enrichment is conducted for ISO method on two selective media, XLD and a second left to choice, and for AOAC method on 3 media, XLD, Hektoen and Bismuth sulfite.

Due to the extensive nature of this trial, laboratories were left the choice to examine samples of 1, 2 or 3 food matrices with ISO or both ISO and AOAC methods. Fresh cheese curd, diced poultry and whole egg powder were selected as the samples for the trial. As for the previous trials, cheese samples were prepared by CECALAIT (Poligny, France), meat samples by MAFF-CSL and egg powder by RIVM-MGB. The reference materials were also prepared by RIVM.

The collaborative trial was carried out during March/May 2000.

2. Materials and Methods

2.1. Preparation of trial test materials

2.1.1. Introduction

The three different food types were each inoculated with a different test strain of *Salmonella*. Three serovars of *Salmonella* were chosen:

- an atypical lactose positive *Salmonella enterica* subsp *enterica* serovar Montevideo for cheese curd;
- *Salmonella enterica* subsp *enterica* serovar Typhimurium for poultry;
- *Salmonella enterica* subsp *enterica* serovar Panama for egg powder (“unusual” strain to highlight possible cross-contamination in the participating laboratories)

These 3 strains are referred as *S. Montevideo*, *S. Typhimurium* and *S. Panama* in this report.

In addition, the samples contained either a natural or ‘simulated’ autochthonous flora.

For each food type, 3 inoculum levels were prepared:

- negative sample (containing only autochthonous flora)
- low level (5-10 cfu *Salmonella* /25g plus autochthonous flora)
- high level (50-100 cfu *Salmonella* /25g plus autochthonous flora).

Acceptance of all batches for use in the trials was made on the basis of achieving satisfactory homogeneity and stability.

2.1.2. Cheese samples

Cheese samples were prepared by CECALAIT, Poligny, France.

All batches were prepared according to the following procedure. Liquid cheese was inoculated with the test strains *S. Montevideo* which had been isolated from dehydrated soup mix in USA. This strain was provided by BioControl. A background lactic acid flora comprising *Lactococcus lactis* subsp. *lactis*, *Enterococcus faecalis*, *Lactobacillus paracasei* and *Lactobacillus plantarum* all isolated from cheeses, were added. In addition, a Gram negative background flora comprising *Escherichia coli*, *Klebsiella oxytoca*, *Citrobacter freundii* and *Pseudomonas aeruginosa* all isolated from milk by CECALAIT, were also added.

Salmonella strains were isolated on Rambach agar and then on Plat-Count-Agar (PCA); *Lactococcus* and *Enterococcus* on M17; *Lactobacillus* on MRS; *E. coli* on PTX and then on PCA; other coliforms on Violet-Red-Bile-Glucose agar (VRBG) and then on PCA; and *Pseudomonas* on PCA. Before inoculation, the *Salmonella* and other Gram negative strains were cultivated in Brain-Heart-Infusion broth (BHI) 18 hours at 37°C for *Salmonella* and at 30°C for the other strains, and lactic bacteria in MRS or M17 broth 18 hours at 30°C. Then, the cultured were mixed: background Gram negative strains together, *Lactococcus* and *Enterococcus* together, and *Lactobacillus* together. The cultures were diluted in ¼ Ringer's solution, and inoculated in the liquid cheese.

Table 1 Composition of the final cheese samples

Batch :	A (blank)	B (low level)	C (high level)
<i>Salmonella</i> :	0	5 to 10 /25 g	50 to 100 /25 g
Lactic acid bacteria:	1000 / g	1000 / g	1000 / g
Gram negative flora:	1000 / g	1000 / g	1000 / g

Subsequently, the inoculated cheese was clotted after dispensing aliquots of 80 g into individual vials by the addition of rennet (Figure 1). The microflora within the cheese was stabilised by the addition of an undisclosed bacteriostatic mixture. The bacteriostatic effect is negated when the sample is diluted during examination. The inoculum levels for the target organisms were verified using a spiral plating system at a single dilution.

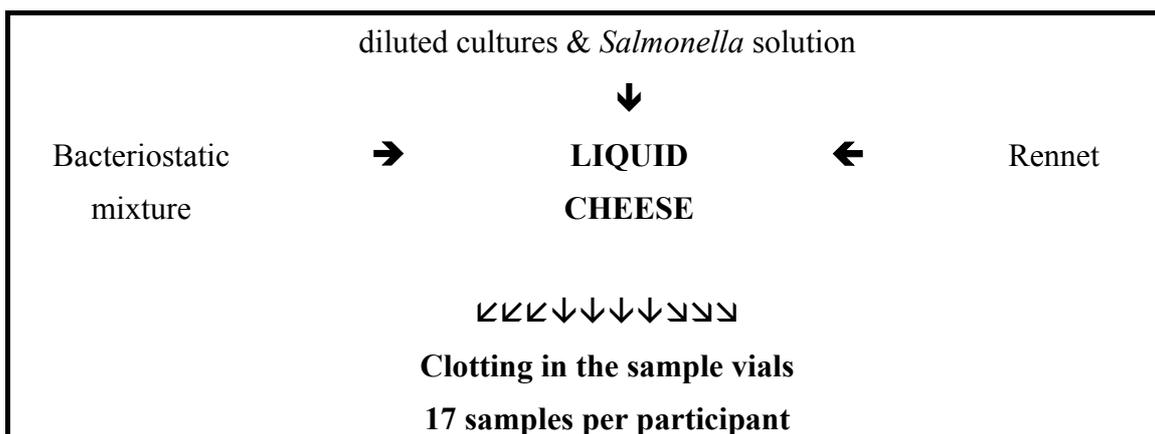


Figure 1 Preparation of cheese samples

17 samples are prepared for each laboratory (each level in 5 replicates plus 2 negative samples to be contaminated by the laboratories with a typical strain of *Salmonella* Typhimurium contained in capsules prepared by RIVM).

At the same time a batch with 10^4 *Salmonella* /g is prepared, for testing the homogeneity (cf. 2.2.1)

Batches of cheese samples were prepared for the collaborative trial one day prior to dispatch to participants.

Samples were stored at 1°C - 5°C until use.

2.1.3. Meat samples

Meat samples were prepared by CSL

Batches of test materials were prepared using retail diced poultry which was freeze-dried in bulk, sub-sampled into vials (portions of 25 g equivalents) and irradiated at 25 kGy.

The test strains and background flora were cultivated into broth, diluted to the appropriate level and added as a ‘cocktail’ to each vial. Finally, the vials were freeze-dried a second time to stabilise the microflora and stored at 4°C until use.

It was necessary to prepare a new batch for the collaborative trial to achieve the desired stability and homogeneity.

The test strains used to inoculate the meat samples were *S. Typhimurium* NCTC 0074. A high and a low level of *S. Typhimurium* was anticipated of 1-5 cfu/g and 10-50 cfu/g, respectively. All samples were additionally inoculated with a simulated autochthonous meat flora. The background flora was composed of a cocktail of strains with a concentration of 10^3 cfu/g. The cocktail contained *L. plantarum* NCDO 1369, *M. luteus* NCTC 8340, *Ps. Aeruginosa* NCDO 1369, *Citrobacter freundii* NCDO 1516.

2.1.4 Dried egg samples

Dried egg samples were prepared by RIVM.

Egg powder

A batch of 25 kg of egg powder was obtained from an egg powder factory (Enthoven BV, Bennebroek, NL): batch 29130 x 25 gram of egg powder have been tested for the presence of *Salmonella*. For this purpose the egg powder was cultured according to the ISO method, with pre-enrichment in BPW, enrichment in RV and plating on Brilliant Green Agar (BGA). All 10 samples were *Salmonella* negative.

The total number of aerobic colony forming units (cfu) in the egg powder was checked by dissolving 10 x 1 gram of egg powder into 10 ml peptone salt solution. 0.1 ml of this (-1) dilution was plated on Plate Count Agar (PCA) and incubated aerobically at (30 ± 1) °C for (68 ± 4) hours. The mean contamination level was $ca 4 \times 10^3$ cfu/g.

Preparation of *Salmonella* capsules

Salmonella Panama (ALM41) 5-10 and *Salmonella* Panama (ALM41) 50-100 were prepared from a highly contaminated milk powder (HCMP) with batch number 1-4, which was spray-dried in Wageningen, The Netherlands in 1978.

The background flora consisted of *Enterococcus faecium* (WR63), which was prepared from HCMP batch number LWL 34, spray dried at the RIVM in 1991.

To achieve the appropriate contamination level a small amount of these batches of HCMP were mixed with sterile milk powder in several steps (to optimise homogeneity of the test material) until the required contamination level was reached. Next, the mixed powder was filled into gelatine capsules ($ca 0.3$ g/capsule) in a laminar flow cabinet using an aluminium filling apparatus. The *Salmonella* powders were filled into blue/white capsules. The powder containing *E.faecium* was filled into green/white capsules.

For the pre-trial a low-level material containing $ca 14$ cfu/capsule of *S.Panama* was prepared. For the trial a new low-level material containing $ca 9$ cfu/capsule of *S.Panama* was prepared.

The final samples consisted of a vial with egg powder, 1 blue/white and 1 green/white capsule. For the blank sample, the blue/white capsule was filled with sterile milk powder (gamma irradiated at 10 kGy). An overview is given in Table 2.

Table 2 Composition of the final egg powder samples

Level	egg powder	<i>Salmonella Panama</i>	<i>Enterococcus faecium</i>
Blank	25 g	-	1 x 10 ³ cfu/g
Low	25 g	5-10 cfu/25 g	1 x 10 ³ cfu/g
High	25 g	50-100 cfu/25 g	1 x 10 ³ cfu/g

2.1.5 Reference materials (RMs)

Pre-trial: Capsules obtained from SVM: *S. Panama* (ALM 41) 5 (order number 701.10)

Trial: Capsules obtained from RIVM (CRM507r): *S. Typhimurium* (ALM40) 5.

Both batches have been tested and proven homogeneous and stable.

2.1.6 RMs for cheese

S. Typhimurium (ALM 40) at a level of ca 100 cfu/capsule was prepared by mixing HCMP batch 1-5 with sterile milk powder in several steps until the desired level of ca 100 cfu/capsule.

2.2. Homogeneity and stability of the test materials

2.2.1. Cheese samples

Homogeneity

The samples were examined in duplicate (method described below in the stability study), for the pre-trial 2 days after preparation, and for the trial 7 days after preparation (test day). 10 samples of the batch with 10⁴ *Salmonella*/g were examined.

The CECALAIT computer program was used for statistics. χ^2 (Chi square) was calculated on the total count of colonies per sample, to compare the dispersion of the contamination to the dispersion obtained in a Poisson series. From tables of χ^2 lower and upper limits were obtained: for example, with 10 samples (9 degrees of freedom) χ^2 value must be between 2.7 and 19.0.

Stability

The stability was tested by analysing 3 samples in duplicate. For each analyse, 10 g of cheese were diluted in 90 ml K₂HPO₄ solution (IDF 122B:1992), and blended during 3 minutes. This solution was inoculated on 2 plates of VRBG agar by the Spiral system.

2.2.2. Meat samples

Homogeneity

Homogeneity testing for the high and low level of *S. Typhimurium* was carried out during the initial 4 weeks work-up period before pre-trial (shown in (3.1.2) the tables 4 for the high and low level) and for the test material for the main trial at day 0 and day 17. All enumerations were carried out using the most probable number (MPN) technique. The MPN represents an estimate of levels and require therefore a more complex analysis of the data than the assumptions of the F-test.

A more complex analysis was therefore performed on the numbers of positives with each sample and each volume, using logistic regression.

Stability

The stability of *S. Typhimurium* in diced poultry stored at 4°C has been verified over a 30 day period during the four-week initial work-up period. All vials were enumerated using the most probable number (MPN) technique. Ten replicate vials were enumerated at each interval at days 0, 12, 19 and 28 for the initial work-up and on day 0 and 17 of the test material for the main collaborative study.

1.2.3. Egg powder samples

All materials were stored at -20 °C and regularly checked for contamination level and homogeneity. The measuring period varied from 3 months (*S. Panama* 5-10) till 26 months (*S. Typhimurium* 5). For complete information see (shown in (3.1.3)) the tables 5 for the high and low level.

During this measuring period the batches of capsules were analysed *ca* 8 times (*S. Typhimurium* 5, 16 times) in the following way: ten capsules containing *S. Panama* 5-10 or 50 capsules containing *S. Typhimurium* 5 were each dissolved in a Petri dish with (5 ± 0.5) ml peptone salt solution (pre-warmed to 38.5 °C) in a shaking incubator set at (38.5 ± 0.5) °C for (45 ± 5) min. To each dish with dissolved capsule (5 ± 0.5) ml double strength PCA was added and incubated at (37 ± 1) °C for (4 ± ½) h. An overlay of (10 ± 1) ml double strength VRBG was added and incubation was continued until a total incubation time of (24 ± 2) h.

Five to ten capsules containing *S. Panama* 50-100 or containing *S. Typhimurium* 100 were each dissolved in a tube with (10 ± 0.2) ml peptone salt solution at (38.5 ± 0.5) °C for 30 min. Two times (4 ± 0.08) ml of this capsule solution was pipetted into separate Petri dishes and each solution was mixed with (6 ± 0.5) ml double strength PCA. After (4 ± ½) h incubation at (37 ± 1) °C, an overlay of (10 ± 1) ml double strength VRBG was added to each dish and incubation was continued until a total incubation time of (24 ± 2) h.

The milk powder containing *E. faecium* was checked only twice in a period of 4.5 months. For this purpose 10 x 1 gram mixed milk powder was dissolved in (10 ± 0.2) ml peptone salt solution (this is the -1 dilution). After mixing on a whirlmixer for *ca* 10 seconds, 2 x 0.1 ml of the -2 dilution was spread over 2 plates of Kenner-Fecal agar with TTC [3](Anonymous, 1984) and incubated for (44 ± 4) h at (37 ± 1) °C.

2.3. Design of the collaborative trial

2.3.1 Procedure

The procedure for the evaluation and comparison of the ISO and AOAC methods for the detection of *Salmonella* was described in a trial chronology in combination with Standard Operating Procedures (SOP-1 and SOP-2) and test report. These documents are included in this report in Annexes A, B and C respectively.

For the detection of *Salmonella* several steps are involved for the two methods:

- Pre-treatment of the test material and preparation of pre-enrichment:
 - Cheese samples: twice 25 g of cheese from the same vial were weighed into each 225 ml pre-enrichment broth, to be examined by ISO and AOAC methods;
 - Meat samples (one vial per method): 20 ml buffered peptone water or lactose broth was added to each vial containing freeze-dried diced poultry. After rehydration for 30 min at room temperature the complete rehydrated sample was added to 205 ml pre-enrichment broth and the vial rinsed with 20 ml and added to the bulk;
 - Egg powder samples (one vial per method): 25g egg powder were transferred to a stomacher bag (or flask), 25 ml of pre-enrichment broth was added to the empty vial, then 2 capsules containing *Salmonella* and background flora were added. The preparation procedure was different for ISO and for AOAC (See annex B) ;
 - Reference materials: each capsule was added to 225 ml of pre-enrichment broth.
- Pre-enrichment

The pre-enrichments as prepared above were incubated at 37°C for 16h to 20 h (ISO) and at 35°C for 24h (AOAC)

- Enrichment

After pre-enrichment, 0.1 ml of culture was inoculated into 10 ml of Rappaport Vassiliadis broths (RVS for ISO and RV for AOAC) and 1 ml of culture into 10 ml of Tetrathionate broths (MKTTn. broth for ISO and TT for AOAC).

- Isolation

After both enrichments, loopfuls of broths were streaked onto agar plates (2 for ISO ; 3 for AOAC).

- Confirmation

Due to the extensive nature of this collaborative study it was not practicable to ask all participants to use all confirmed tests on the all number of suspect colonies to confirm. Thus, in order to make a comparison only of whole two standards, laboratories are allowed to stop confirmation step of suspect colonies when one of them are confirmed as *Salmonella*. So the comparison of enrichment media are not possible.

On the test report the participants were requested to report the following information:

- shipment conditions
- media formulations (manufacturer, code numbers, batch numbers, expiry dates)
- incubation conditions
- number of typical colonies per plate
- confirmation results per colony (until one is confirmed as *Salmonella*)
- all deviations from SOPs.

This detailed information enabled deviant or aberrant results to be correlated with technical errors i.e. incubator failure, incorrect use of media.

2.3.2 Test materials

In the collaborative trial, 15 samples per food type were examined comprising 5 replicates of negative, low and high contamination levels.

As positive controls, two additional cheese samples to be contaminated with *Salmonella* capsules (high level) by the participants, called RC1-RC2 were examined (RC1.1, RC1.2 for ISO ; RC2.1, RC2.2 for AOAC)

In addition, 5 reference capsules were examined per each method.

The sample codes assigned were C1-C15 for cheese samples (C1.1- C1.15 for ISO and C2.1-C2.15 for AOAC), M1-M15 for meat samples (M1.1- M1.15 for ISO and M2.1-M2.15 for AOAC), D1-D15 for egg powder samples (D1.1- D1.15 for ISO and D2.1-D2.15 for AOAC) and R1-R5 for the reference materials (R1.1- R1.5 for ISO and R2.1-R2.5 for AOAC). All samples were randomly coded for each participant so as to prevent falsification of data and collusion between laboratories.

Cheese samples were dispatched by CECALAIT on 21 March 2000 directly to European participants and to BioControl which then dispatched the samples to the 10 US participants. Meat, egg and reference samples were dispatched to the participants (directly for European participants and via BioControl for US participants) by the RIVM on 8 May 2000.

All samples were shipped by DHL Courier in insulated polystyrene boxes. Several ice packs, frozen at -18°C, were included in the parcel to prevent the samples from being subjected to high temperatures. To detect possible temperature abuse during transport, each parcel contained a time-temperature device from WarmMark (Blanken Controls, Loenen, The Netherlands) with 2 response temperatures of 10°C and 20°C.

Media needed for the analyses were not supplied by the organising laboratory

Cheese samples had to be examined on 27 March 2000. The examination of the meat, egg and reference samples had to commence respectively on 15 May 2000 and 22 May 2000. All the reference materials had to be examined together and concurrently with one of the food types at the choice of the participant.

2.4. Statistical analysis

The theoretical model which had been designed by statisticians from the three partners in advance of the *Listeria* detection trial and applied to the *Listeria* data, was also used for the analysis of qualitative data such as that produced in this *Salmonella* trial.

This model aims at providing a measure of performance of the test method which relates to those parameters which are familiar to most analysts *i.e.* accuracy, repeatability and reproducibility. The two latter being applicable to only quantitative methods, new parameters have been defined : accordance and concordance, as well as the odds ratio, which enables to characterise, together with the exact test, the significance of concordance and accordance (see appendix E to the final report on detection of *L.monocytogenes* [2] and [4].

3. Results and discussion

3.1. Stability and homogeneity of the test materials

3.1.1. Cheese samples

Homogeneity of cheese samples

Ten samples of the batch 10^4 cfu/g *Salmonella* were examined in duplicate, 2 days after preparation for the pre-trial and 7 days after for the trial (test day). By calculation of χ^2 on the total count of colonies per sample, agreement with a Poisson series was rejected at the 95% probability level, the homogeneity between samples was not therefore respected ($\chi^2 > 19$).

Table 3 Number of *Salmonella* CFU/ 25 g (confidence limits at 95%)

	Batch B	Batch C	χ^2	
pre-trial 2 days after preparation	7 (6 to 8)	68 (62 to 74)	45	Homogeneity criterion not respected
trial 7 days after preparation	11 (9 to 13)	110 (99 to 121)	78	homogeneity criterion not respected

But the variability is taken account for the calculation of the confidence limits of the *Salmonella* contamination level. Moreover, the homogeneity is not so much critical for a trial on a detection method.

Stability of cheese Samples

The results are given in Figure 2, 3 and 4. Only the batch 10^4 was tested . The results are given in log (cfu/ml).

The *Salmonella* contamination is stable until about 3 days at 4°C, and until 1 day at 12°C and

15°C.

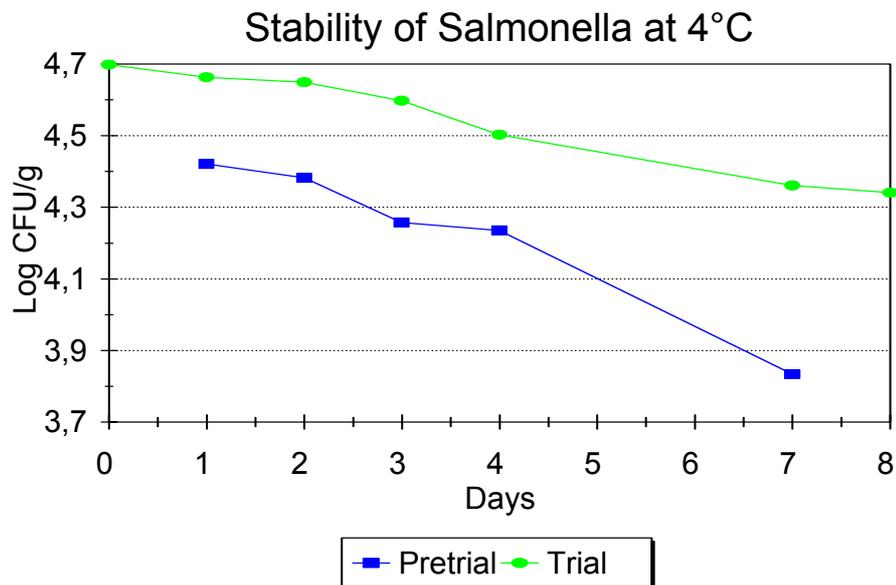


Figure 4 : Stability of *Salmonella* at 4°C during the pre-trial and the trial

3.1.2. Meat samples

The stability and homogeneity of the high and low levels of *S. Typhimurium* in diced poultry was tested.

The test material for the main collaborative trial was enumerated by using MPN on day 0 and day 17 (test day). All enumerations were carried out on 10 samples in duplicate and were tested for homogeneity. The results are displayed in the table 4.

Table 4 Numbers of *S. Typhimurium* present in meat samples immediately after preparation (day 0) and on test day (day 17) (enumerated by MPN)

Level	Day 0 cfu/25g		Day 17 cfu/25g	
	A	B	A	B
Low	4.3	4.3	0.92	0.92
Low	9.3	4.3	0.74	0.92
Low	4.3	4.3	0.92	0.92
Low	9.3	9.3	0.36	0.92
Low	2.3	4.3	0.92	0.36
Low	0.92	2.3	0.36	0.36
Low	4.3	4.3	0.92	0.74
Low	2.3	4.3	0.36	0.36
Low	4.3	4.3	0.92	0.92
Low	1.5	1.5	0.92	0.36
High	24	24	4.3	46
High	46	46	9.3	2.3
High	24	46	4.3	4.3
High	24	46	4.3	9.3
High	24	24	4.3	4.3
High	9.3	46	9.3	4.3
High	24	46	4.3	24
High	24	24	4.3	4.3
High	46	46	9.3	9.3
High	24	24	9.3	4.3

Low: Target level 5 - 10 cfu/g

High: Target level of 10 – 50 cfu/g

MPN: Most probable number technique

From the figures given above it is evident that the low level inoculum was slightly below the target level, whereby the high level was at the intermediate range of the target level. The differences between replicates at both the low and medium levels was more variable than desired. However, this variability is attributed to the need to use an MPN technique to enumerate the samples due to the low numbers of cells present which could not be enumerated using a plate count technique; it is recognised that the MPN method is imprecise.

A more complex analysis was therefore performed on the numbers of positives with each sample and each volume, using logistic regression. There are obviously massive differences in the number of positives between the different volumes. However one these differences have been allowed for there is no evidence for the consistent differences between samples that would be expected if they were not homogeneous ($\chi^2= 12.73$ with 9 d.f. for low, $\chi^2= 8.02$ with 9 d.f. for high, both not significant).

Hence these samples can safely be regarded as homogeneous, especially for the purpose of a trial on a detection method.

3.1.3. Egg powder samples

In Table 5 the average results calculated from the total measuring period are given. In this Table also the regression coefficients of the stability results are given.

The figures of the stability test results are given in Figure 5-8

Table 5 Homogeneity and stability results of tested capsules and powder

Batch	Measuring period (months)	Mean cfu/capsule	$T_2 / (I-1)$ ¹	Regression coefficient ²	Pr > χ^2
S.Panama 5-10 (a)	3	9,45	1,30	-0,0010	0,9227
S.Panama 50-100 (b)	7,5	93,3	1,80	-0,0007	0,7602
S.Typhimurium 100 ©	3,5	110,7	1,62	-0,0157	0,0002 ³
S.Typhimurium 5 (RM)	26	5,50	1,44	-0,0068	0,0015 ³
E.faecium (d)	4,5	9,0x10 ⁴ cfu/g	3,16	-0,0039	0,5078

(a): used for egg powder low level; (b): used for egg powder high level; (c): used as positive control in parallel with the cheese samples; (d): background flora ; (RM) : reference material.

¹: Criterion: $T_2 / (I-1) \leq 2$

²: rate of change in contamination level per week or per month (in case of *S.Typhimurium* 5) based on log₁₀ transformed counts

³ significant decrease ($\alpha = 0,05$).

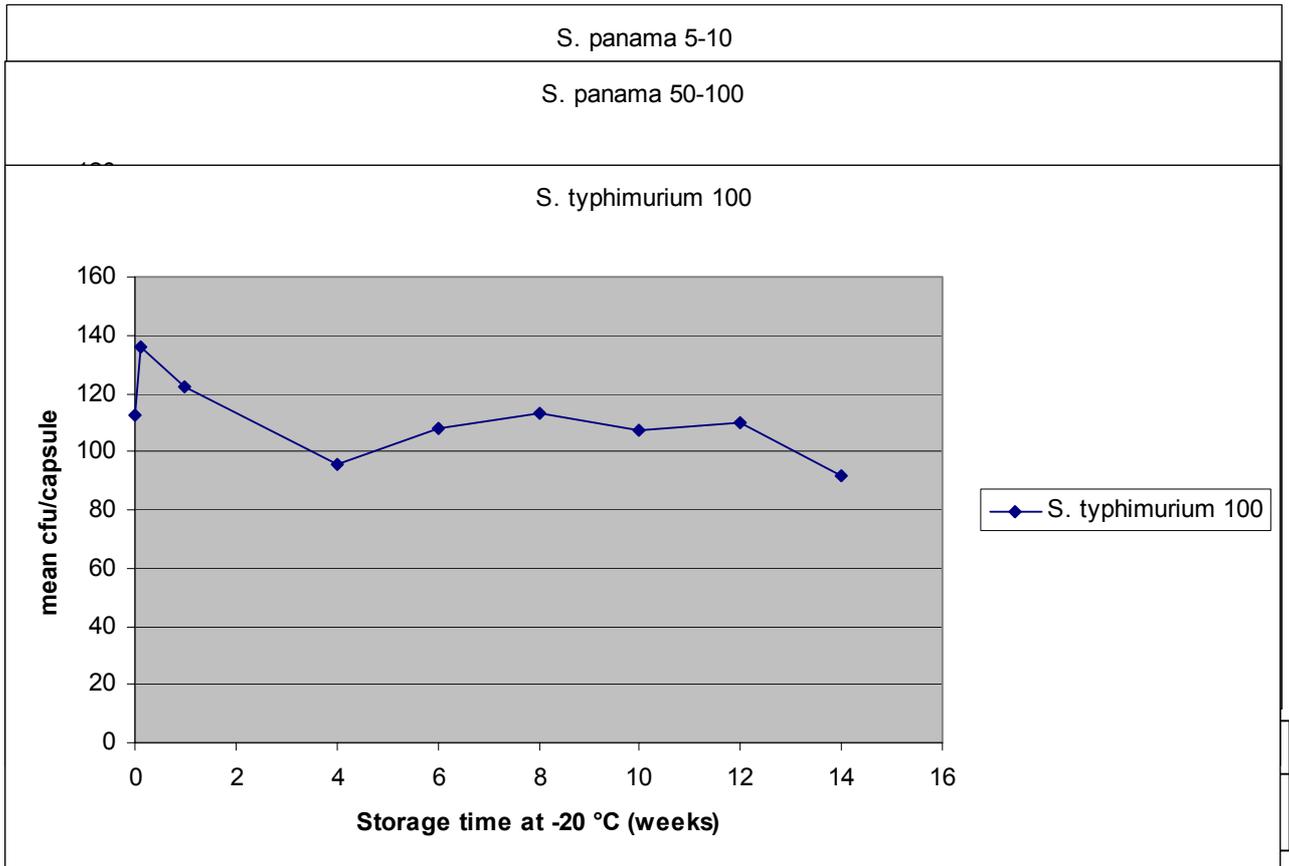


Figure 7 Stability of *S. Typhimurium* 100

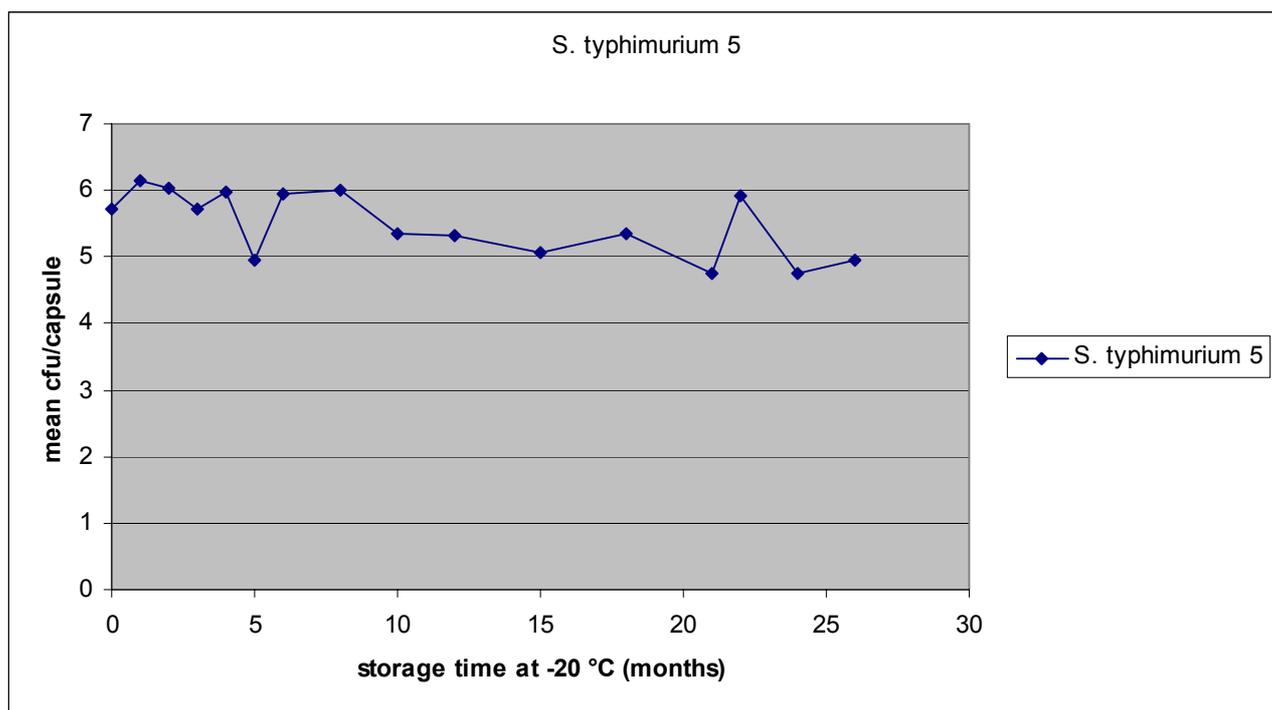


Figure 8 Stability of *S. Typhimurium 5*

The results of the homogeneity controls showed good homogeneity for all batches of capsules. The powder containing *E.faecium* showed somewhat high variation in results ($T_2/(I-1) = 3,16$). However this material was checked in a different way than the other materials. In case of *E.faecium* not a capsule but the powder was analysed directly. Furthermore dilutions were made from this powder solution which can cause extra variation in the final result. The powder containing *E.faecium* was used as (disturbing) background flora in the egg powder. For this purpose a somewhat higher variation in number of cfu is not a big problem as it has little influence on the final count of the target organism (*Salmonella*).

The results of the stability tests showed no significant decrease for the capsules containing *S.Panama 5-10*, *S.Panama 50-100* and *E.faecium*. The capsules containing *S.Typhimurium 5* showed a significant decrease, which was relatively small: only 0,7 % decrease in cfu per month. The capsules containing *S.Typhimurium 100* showed a significant decrease of ca 1,6 % cfu per week. This material was used as positive control in parallel with the cheese samples. Even with a decrease of 1,6 % per week sufficient cfu would be remained in the capsules to be used as positive control.

In the past, the mentioned HCMP-batches have been used for the preparation of several batches of ©RMs. Stability studies of these batches have been reported in [5] Manavakis *et al.*, 1995 (*S.Panama*), [6] Schulten *et al.*, 2000 (*S.Typhimurium*) and [7] Mooijman *et al.*, 1999 (*E.faecium*).

Information from Manavakis *et al* (1995) showed for RMs containing *S.Panama 5* a decrease in the mean contamination level of 1,3 % and 4,0 % per day when the material was stored at

22 °C and 30 °C respectively. Schulten *et al.*, 2000 described for RMs containing *S.Typhimurium* 5 a decrease in the mean contamination level of 5,3 % and 8,2 % per day when stored at 22 °C and 30 °C respectively. RMs containing *E.faecium* 500 showed better stability at elevated temperatures. According to Mooijman *et al* (1999) a decrease of the mean contamination level of only 0,2 % per day was visible when the material was stored at 22 °C and 1 % per day when stored at 30 °C.

In conclusion, the stability and homogeneity of these samples were satisfactory for the purpose of this trial.

3.2. General results of the trial

3.2.1 A summary of participants' data is given in Annex D.

3.2.2. Participants and number of results

17 laboratories from 12 countries in Europe and 10 US laboratories participated in the collaborative trial. Due to the level of work required for this trial, laboratories were left the choice to examine samples of 1, 2 or 3 food matrices with ISO or both ISO and AOAC methods.

After exclusion of data as rules reported in (3.2.7), laboratories have generated results per food type and per method, as described in tables [6] & [7].

Table 6 Results generated per food type and per method

Cheese		Egg Powder		Poultry		Number of Lab.
ISO	AOAC	ISO	AOAC	ISO	AOAC	
16 European Laboratories ¹						
+	+	+	+	+	+	5
+	+	+	+	-	-	2
-	-	+	+	+	+	1
+	+	-	-	-	-	1
+	-	+	-	+	-	3
+	-	+	-	-	-	1
+	-	-	-	+	-	1
-	-	+	-	+	-	2
10 US Laboratories						
+	+	+	+	+	+	5
-	-	+	+	+	+	1
+	+	+	+	-	-	1
+	+	-	-	+	+	1
+	+	-	-	-	-	1
-	-	-	-	+	+	1

¹ 17 participant laboratories but diced poultry's results of one laboratory were excluded, due to false positive results obtained for negative controls

Table 7 Number of European and US laboratories results for the combination
(method, food type) :

	Cheese	Egg Powder	poultry
ISO and AOAC	16	15	14
ISO	21	21	20
AOAC	16	15	14

3.2.3. Shipment of the test materials

Cheeses

Cheese samples were shipped on 21 March 2000 and all but one laboratory commenced examinations on the prescribed date of 27 March April. One laboratory commenced the examinations 2 days later.

For cheese samples, the dispatching and delivery to participants was performed by national and international express carriers. Samples were packaged in insulated boxes with ice. Shipment times were about 15 h for France, 24-48 h for Europe, and about 3 days for the coordinating laboratory in USA (BioControl), and then 2 days for dispatching from this laboratory to the other laboratories in USA. The temperature was controlled at reception with 2 temperature devices included in the package.

Other samples

Meat, egg powder samples and RM's were shipped on 8 May 2000 and all but 1 laboratory commenced their examinations on the 15 May and 22 May as instructed. One laboratory commenced its examinations on 9 May for poultry.

US laboratories received their samples 10 days after shipment from RIVM.

3.2.4. Temperature control during shipment

Temperature control during shipment into Europe was satisfactory. All temperature devices accompanying egg and poultry samples were red. According to the shipment delay, this was considered insufficient to have affected the results.

3.2.5. Media

Each participant was able to choose whether to use commercially available dehydrated media or use media prepared from the individual ingredients as specified in the SOPs. If the latter was chosen all ingredients and concentrations used had to be recorded on the test report.

Table 8 summarises media elected for the second isolation agar for SOP-1(European laboratories).

Table 8 ISO standard : second agar used by European labs

BGA	Brillant green agar manufactured by OXOID CM 329	3 labs
BGA	Brillant green agar ready poured plate manufactured by bioMerieux	1 lab
Rambach	Rambach agar manufactured by MERCK Art 7500	2 labs
MLCB	agar manufactured by OXOID CM 7839	1 lab
HE	Hektoen agar manufactured by Biorad-Sanofi-Diagnostics Pasteur Art 63894	1 lab
BS	Bismuth Sulfite agar manufactured by OXOID CM 201	2 labs
BS	Bismuth Sulfite agar manufactured by DIFCO 0073	4 labs
BS	Bismuth Sulfite agar manufactured by MERCK art 1. 05418	3 labs

Note: one lab used 3 agars = Rambach, BPLS, and XLT4

3.2.7. Results excluded from further analysis

Rules

For this qualitative trial it was agreed that results from laboratories would not be excluded unless they fell into these specified categories:

- a) the samples had received a sample with significant temperature abuse during shipment,
- b) the laboratory had clearly deviated from the specified standard operating procedures or
- c) the results of the laboratory were questionable as indicated by
 - a negative control found positive.

All results obtained with the matrix would be excluded.

Except : egg samples, for whose non sterile egg powder had been used, and which may have contained *Salmonella* ; the results would be kept if the *Salmonella* strain found would be different from the inoculated *Salmonella* ; all the results obtained with the matrix would be excluded if the *Salmonella* strain found would be the same than the inoculated *Salmonella* (cross-contamination).

- an identified *Salmonella* strain differing from the strain used.

All results obtained with the matrix would be excluded.

- 4 out of 5 RMs found negative.

All results obtained with the RMs and with the matrix analysed at the same time than the RMs would be excluded, these false negative results would be more than expected by chance.

- only 1 of 5 RMs found negative.

The results would be kept (may be due to absence of inoculation of the RM).

Results excluded

The results with cheese, poultry and egg were excluded respectively for 2, 5 and 5 laboratories. 1 laboratory did not finish poultry sample analysis.

3.3 Statistical analysis of data Data obtained with the ISO method Data

Table 9 provides a variety of statistical parameters for the ISO method, for the different sample types and levels.

- a. The column “Nb of labs” gives the number of laboratories whose results (5 replicates per laboratory/food type/level) have been taken into account for the data analysis (see 3.2.7).
- b. The following columns provides *statistical parameters* which can be defined as follows (for more details, see Appendix E to the final report on detection of *L.monocytogenes* [2] and [4] (S. Langton, *et al.*, submitted to Int. J of Food Micro.)

Accuracy :

- for positive samples (low & high levels) : **sensitivity**
The percentage of samples that are correctly found to be positive.
- for negative samples (blank) : **specificity**
The percentage of samples that are correctly found to be negative.

Accordance

The percentage chance of finding the same result (i.e. both positive or both negative) for two identical samples analysed in the same laboratory, under standard repeatability conditions (same conditions : apparatus, operator and the shortest feasible intervals of time).

Concordance

The percentage chance of finding the same results (i.e. both positive or both negative) for two identical samples analysed by two different laboratories, thus under standard reproducibility conditions.

Concordance odds ratio (COR)

$$COR = \frac{\text{accordance} * (100 - \text{concordance})}{\text{concordance} * (100 - \text{accordance})}$$

This parameter is introduced in order to assess the degree of between-laboratory variation in the results : it reflects the relative magnitude of the accordance and concordance figures.

A value of COR of 1,00 or less indicates that the accordance is equal to or less than the concordance, so two samples sent to different laboratories will probably produce the same result as the two samples analysed by the same laboratory.

A value of COR of more than 1.00 indicates that the accordance is greater than the concordance, suggesting that the variability between laboratories is larger than the intra-laboratory variation, and the larger the COR is, the more inter-laboratory variation is present.

Table 9 Statistics (%) for the different sample types and levels relating to the ISO method

FOOD	LEVEL	Nb labs	Sensitivity	Specificity	Accordance	Concordance	Odds ratio
CHEESE	Blank	21	–	100 %	100 %	100 %	–
	Low	21	74.3 %	–	83.8 %	60.5 %	3.38 (*)
	High	21	83.8 %	–	95.2 %	71.7 %	7.83 (*)
	RC (**)	23	100 %	-	100 %	100 %	-
POULTRY	Blank	20	–	100 %	100 %	100 %	–
	Low	20	98 %	–	96.9 %	96 %	1.3
	High	20	100 %	–	100 %	100 %	–
EGG POWDER	Blank	21	–	100 %	100 %	100 %	–
	Low	21	98.1 %	–	96.2 %	96.2 %	1.32
	High	21	99 %	–	98.1 %	98.1 %	–
REF. MAT.		25	94.4 %	-	88.8 %	89.1 %	1

(*) : significant value
(**) : Reference capsules of *Salmonella* added by each participant to uninoculated samples prior to analysis (2 per laboratory), standing for positive controls.

Interpretation

From these figures, it can be deduced that the ISO method displayed good performances for meat and egg powder samples, as well as for reference materials, with comparable intra- and inter-laboratory variabilities. Logically, all the parameters (in particular for meat samples) have better values for the high level than for the low level : this can be easily explained by the fact that the homogeneity of the contamination is better at higher levels. Surprisingly, the performances of the method for the reference materials revealed to be somewhat lower than for the two food matrices.

For cheese samples, the values of all parameters are not so good, especially at the low level, and the significant values of the odds ratio mean that the inter-laboratory variability is relatively larger than the intra-laboratory one. It may be explained by the fact that the ISO method does not target the lactose + *Salmonella*, whereas the cheese samples had been contaminated with *S. Montevideo*, but the detailed results of the participants do not clearly confirm such hypothesis. More probably the cheese samples are, compared to the other samples, the more “difficult” ones from a microbiological point of view, since they are in a fresh state (thus less stable) and they contain a lot of competitive flora. They are also the ones nearer to the “real life”, especially the difficult food samples.

Data obtained with the AOAC method

Data

Table 10 provides a variety of statistical parameters for the AOAC method, for the different sample types and levels.

The definition of the different parameters are given in 3.3.1.

Table 10 : Statistics (%) for the different sample types and levels relating to the AOAC method

FOOD	LEVEL	Nb labs	Sensitivity	Specificity	Accordance	Concordance	Odds ratio
CHEESE	Blank	16	-	100 %	100 %	100 %	-
	Low	16	83.8 %	-	88.8 %	71.6 %	3.14 (*)
	High	16	91.3 %	-	88.8 %	83.6 %	1.56
	RC (**)	17	97 %	-	94 %	94 %	1
POULTRY	Blank	14	-	100 %	100 %	100 %	-
	Low	14	55.1 %	-	56.6 %	49.5 %	1.33
	High	14	94.3 %	-	92.9 %	88.8 %	1.65
EGG POWDER	Blank	15	-	100 %	100 %	100 %	-
	Low	15	100 %	-	100 %	100 %	-
	high	15	100 %	-	100 %	100 %	-
REF.MAT.		18	96.7 %	-	93.3 %	93.4 %	1

(*) : significant value
(**) : Reference capsules of *Salmonella* added by each participant to uninoculated samples prior to analysis (2 per laboratory), standing for positive controls.

Interpretation

From these figures, it can be said that the AOAC method perfectly performed with egg powder samples, very well with reference materials, with comparable intra-laboratory and inter-laboratory variability.

For cheese and meat samples, the method did not performed so well, especially for the low level of meat samples, for whose sensitivity, accordance and concordance rank at about 50 % (about one chance out of two to miss a contaminated sample). It was not possible to find a convincing explanation of the latter results, except a lack of performance of the AOAC method for such meat samples.

3.3.3 Comparison of the ISO and AOAC methods

Data

The following table provides a comparison of the sensitivity for each method, as well as the χ^2 test for assessing the degree of significance of the difference between the two methods.

Table 11 : Percentage of correct results (sensitivity) in relation to the method used

METHOD FOOD	ISO METHOD		AOAC METHOD		ALL		χ^2	PROBA
	n	Mean	n	Mean	n	Mean		
CHEESE	42	79 %	32	87.5 %	74	82.7 %	4.54	3.13 (*)
POULTRY	40	99 %	28	74.8 %	68	89 %	49.1	0.00 (*)
EGG	42	98.6 %	30	100 %	72	99.2 %		13.8
ALL	124	92.1 %	90	87.7 %	214	90.2 %	5.58	1.73 (*)

(*) : significant value

Interpretation

From this table, it can be seen that from an overall point of view (all food types considered), there is a significant difference between the two methods, as well as for the cheese samples and the meat samples considered separately.

The difference in sensitivity is rather limited (8 %) for the cheese samples, the AOAC method performing better than the ISO method. This can be easily understood, since the AOAC method is in principle better suited to find out lactose positive *Salmonella*, with the use of BS agar for isolation. It is meanwhile interesting to note that the difference in performances is limited.

On the opposite, this difference is much larger (24 %) for the meat samples, the ISO method performing better than the AOAC one. A die-off factor of 4 has been recorded between the CSL counts and the BioControl counts on the samples when they reached the US, thus suspicion that samples have not been stable on the way to the US, but this could not explain a difference of performances between methods. Moreover, this difference has been found by all laboratories, either European or US ones. As already mentioned above, it was not possible to find a convincing explanation of these results, except a lack of performance of the AOAC method for such meat samples.

Finally, it is important to note that US and European laboratories have found similar results for each combination matrix/SOP : there is no noticeable difference in proficiency between US and European laboratories when using both methods, whereas it was clear that on one side, European participants had not a large experience of the use of the AOAC method and on the other side, US laboratories had not a large experience of the use of the ISO method.

Additional trials, egg and cheese samples

Purely for the purpose of gaining recognition by AOAC of the ISO method, it was decided to launch supplementary trials.

For cheese samples, it was assumed that the equivalence of the 2 methods could be accepted, due to the limited differences.

For the egg samples, it was necessary to launch a limited trial, involving 5/6 laboratories, since the low level archived for the full trial (5-10 cfu/25 g) was too high for the AOAC requirements, which requests the testing of a partial recovery. The level further tested was therefore lower than 5 to 10 cfu/25 g. Samples were again prepared by RIVM and dispatched in November 2000 to both European laboratories and US ones through BioControl.

For the poultry samples, the 2 methods revealed to have a clearly significant difference in performances, thus the need to conduct a further trial in order to try to have no more a significant difference between the two methods. Naturally contaminated samples were prepared by BioControl and dispatched in November to US laboratories, as well as to European ones through AFSSA-Ploufragan.

It was agreed that data generated by these supplementary trials may be incorporated into tables displaying the precision characteristics of the ISO Standard, but clearly stating that they had been derived from a supplementary trial.

The results are still being exploited, but it can be already announced that they would enable to prove the equivalence of the two methods for both egg and poultry samples.

3.3.5 Additional trials, inclusivity and exclusivity

Still with the aim of getting recognition of the ISO method by AOAC, AFSSA-Ploufragan has performed inclusivity and exclusivity studies with pure cultures on the ISO method (about 120 *Salmonella* strains and 30 non-*Salmonella* strains).

The results are still being collected, but looks promising.

4. CONCLUSIONS AND RECOMMENDATIONS TO CEN and ISO

- ✓ The method as in draft Standard EN ISO/DIS 6579 : 2000 is generally satisfactory (overall sensitivity *ca* 92%).
- ✓ Thus, the SMT Project recommends to CEN & ISO:
- ✓ To include performance characteristics (possibly completed with the data generated from the additional trials) into the new CEN/ISO Standard 6579 with including explanatory notes on the origin of the data (data derived only from a few types of strains and from 3 food matrices, lactose positive strain used for the cheese samples).
- ✓ To give explanations in the Standard on how to interpret these performance characteristics;
- ✓ **Enrichment stage** : to reconsider the choice of the 2nd enrichment broth, since several participants had encountered difficulties when preparing and using MKTTn;
- ✓ **Isolation stage** : To introduce a note stating that a 2nd 90 mm plate is particularly necessary for the analysis of heavy contaminated samples;
- ✓ Confirmation stage :
 - to precise the confirmation procedure and interpretation (refer to Table 2 of the ISO Standard);
 - to restrict the serological confirmation, in the precise frame of the detection of *Salmonella* in foods, to the use of O polyvalent sera (a to 67)⁶
 - ✓ To aim at defining a common AOAC & ISO method for detection of *Salmonella* in foods.

⁶ The exact wording of this recommendation (in particular the use of only one or several polyvalent sera) would be investigated by K. Moojman within RIVM as CRL for *Salmonella*.

5. REFERENCES

1. Horwitz W. (1995), International Harmonised Protocol for the Design, Conduct and Interpretation of method Performance Studies (Technical Report). IUPAC/AOAC/ISO. Pure and Applied Chemistry, **67**, 2, 331-343
2. Scotter S. and Langton S.(1999), EU Projet SMT4-CT96-2098. Validation of ISO Microbiological Method, Detection of *Listeria monocytogenes* by EN ISO 11290 part 1: 1997, Final report. Ministry of Agriculture, Fisheries and Food, Central Science Laboratory, York, England. Report no FD 98/15
3. Anonymous. ISO 7899/2 Water quality – Detection and enumeration of faecal streptococci – part 2: Method by membrane filtration. 1984. International Organization for Standardization, Switzerland.
4. S. Langton, R. Chevennement, N. Nagelkerke, B. Lombard : Analysing collaborative trials on microbiological qualitative methods: Accordance and concordance, submitted to Int. J of Food Micro.)
5. Manavakis M, Dommelen JA van, Mooijman KA and Havelaar AH. Collaborative study with reference materials containing *Salmonella Panama* for water microbiology (trial 7). National Institute of Public Health and the Environment, Bilthoven, The Netherlands. Report no.: 281008010, July 1995.
6. Schulten SM, in 't Veld PH, Ghameshlou Z, Schimmel H. The certification of the number fraction of negative capsules and number of colony forming particles of *Salmonella typhimurium* from artificially contaminated milk powder. CRM507r. EUR Report 196222 EN Commission of European Communities, Community Bureau of reference, Brussels, Luxembourg, in press
7. Mooijman KA, Havelaar AH, van Strijp-Lockfeer NGWM, Branger D and Schimmel H. Recertification of number of colony forming particles of *Enterococcus faecium* in 1 ml suspension of reconstituted artificially contaminated milk powder. CRM 506. EUR 19281 EN. Commission of European Communities, Community Bureau of reference, Brussels, Luxembourg, 1999.

6. PARTICIPANTS IN THE PROJECT EU Project SMT4-CT96-2098

- | | | |
|---|-----------|----|
| - I de Froidment-Görtz, European Commission,
Directorate-Générale XII / C-5
(responsible commission scientific officer) | Brussels | B |
| - C. Lahellec, AFSSA - Direction Générale
(project coordinator) & B. Lombard, AFSSA - LERHQA | Paris | F |
| - L. Cox / R. Leuschner, MAFF-CSL Food Science Laboratory | Norwich | UK |
| - K. Mooijman / S..Schulten, RIVM-MGB | Bilthoven | NL |

Organisation *Salmonella* detection trial

- | | | |
|---|------------|-----|
| - M. Bohnert/ F. Humbert , AFSSA - Site de Ploufragan | Ploufragan | F |
| - Ph. Feldsine/ R. Forgey / L. Mui, BioControl | Bellevue | USA |

Preparation of test material

- | | | |
|------------------------------------|-----------|----|
| - CECALAIT INRA SRTAL | Poligny | F |
| - MAFF-CSL Food Science Laboratory | Norwich | UK |
| - RIVM-MGB | Bilthoven | NL |

Participants pre-trial

- | | | |
|---|------------|----|
| - AFSSA - Ploufragan, Laboratoire d'Etudes et de Recherches
Avicoles et Porcines | Ploufragan | F |
| - PHLS - Central Public Health Laboratory | London | UK |
| - RIVM-MGB | Bilthoven | NL |

7. EUROPEAN PARTICIPANTS IN THE COLLABORATIVE TRIAL

Dr H Asperger
Veterinärmedizinische Universität Wien
Institute für Milchhygiene und Milchtechnologie
Joseph Baumann-Gasse 1
A-1210 Wien
Austria

Mrs M Lambiri
National School of Public Health
Bacteriology Department
196 Alexandras Avenue
115 21 Athens
Greece

Mr V Young
East Coast Area Health Board
Public Analysts Laboratory
Sir Patrick Dun's, Lower Grant Canal Street
Dublin 2
Republic of Ireland

Prof M L Stecchini
Dipartimento di Scienze Degli Alimenti
Via Marangoni 97
33100 Udine
Italy

Mrs C Allaert
Lab of Microbiology of the Food Technology
Department
Avda. Alcalde Rovira Roure, 177
25198 Lleida
Spain
Mr Vincent
Institut Pasteur de Lille
Service de Microbiologie et Hygiene des Aliments
369 rue Jules Guesde
59650 Villeneuve d'Ascq
France

Dr E Bartelt
Federal Institute for Health Protection of
Consumers and Vet.
Unit Food Microbiology
Thielallee 88-92
14195 Berlin
Germany

Mrs S. Schulten and Mrs K. Mooijman
National Institute of Public Health and
Environment
Microbiological Laboratory for Health
Protection
Antonie Van Leeuwenhoeklaan 9
3720 BA Bilthoven
The Netherlands

Dr A Steneryd/Dr C. Wiberg
National Food Administration
Biology Division
Box 622
75126 Uppsala
Sweden

Mr ir H Stegeman
RIKILT-DLO
PO Box 230
6700 AE Wageningen
The Netherlands

Dr S Qvist / Mrs M. Espersen
Institute of Food Safety and Toxicology
Mørkhøj Bygade 19
DK-2860 Søborg
Denmark

Mrs E. Heck
Nestlé Research Center Lausanne
Quality and Safety Assurance Department
PO Box 44
1000 Lausanne
Switzerland

Dr D Roberts
Central Public Health Laboratory
Food Hygiene Laboratory
61 Colindale Avenue
London
NW9 5HT
England

J Anderson
Unilever Research
Colworth Laboratory
Colworth House
Sharnbrook
Bedford MK44 1LQ
England

Dr. H. Becker
 Lehrstuhl für Hygiene und Technologie der Milch
 Tierärztliche Fakultät
 Veterinärstrasse 13
 D 80539 München
 Germany

Dr K Friedrich/ Mrs M. Knödlseeder
 Milchwirtschaftliche Untersuchungs und
 Versuchsanst
 P.O. Box 2025
 87410 Kempten (Allgau)
 Germany
 Mrs Marylene Bohnert/ F. Humbert
 Afssa Ploufragen B.P.53
 Zoopole Beaucemarie
 22440 Ploufragan
 France

US PARTICIPANTS IN THE COLLABORATIVE TRIAL

Contact name	Laboratory Name
Jim Agin	Ohio Departement of Health
Thomas Hammack	FDA – Washington, D.C.
Marilyn Smith	FDA – Jamaica, N.Y.
Lurlene Dixon	FDA – Atlanta, GA
Lorraine Humes	FDA – Alameda, CA
Kathy jost-Keating	Silliker Labs – Garwood, N.J.
Linda Mui	BioControl Systems – Bellevue, WA
Richard Ruby	FDA – Los Angeles, CA
Richard Rude	FDA – Bothell, WA
Vidhya Gangar	Silliker Labs – Carson, CA

ANNEX A TIME SCHEDULE FOR THE COLLABORATIVE TRIALS:

Detection of Salmonella by use the ISO 6579 Standard and

Detection of Salmonella by use the AOAC Standard

Introduction

This time schedule in combination with the standard operating procedures (SOPs) and test reports describes the procedure for:

- The evaluation of method for the detection of Salmonella as described in ISO 6579
- The comparison with method for the detection of Salmonella as described in AOAC

The aim is to determine the performance characteristics of the ISO method and to compare the 2 methods. The trials will be carried out at the same time. For the purpose of these trials 3 food types artificially inoculated with Salmonella will be examined for the 2 methods or ISO method only as you have chosen. Enclosed you will find a reply form, on which you can confirm your participation in the trial, if you agree to the time schedule, and whether you will perform both methods and three matrix, or only one method and which matrix.

Please return this reply form as soon as possible to the AFSSA-Ploufragan.

Outline of the trial

Cheese samples will be prepared by CECALAIT (France), diced poultry samples by CSL (UK), and egg powder samples by RIVM (Netherlands). For each method 15 samples of each food will be examined as well as 5 reference capsules + 2 cheese samples to be contaminated by yourself with capsules (52 samples in total per method). The 3 sample types will be dispatched separately by CECALAIT, CSL and RIVM.

The 3 sample types will be dispatched separately from CECALAIT, CSL and RIVM.

For the cheese sample, one sample will be sent for analysis by the 2 methods; 25 g to be taken for the ISO method and another 25 g for the AOAC method. Two flasks of the same sample (same level, same matrix) will be sent for egg powder and diced poultry, one flask will be examined by the ISO method and the other one by the AOAC method. The samples must be examined on the date indicated hereafter in the chronological description of the trials.

Each participating laboratory will receive by mail the documents listed below

- the *time schedule*, (see enclosed)
 - ✓ the complete *Standard Operating Procedure SOP-1* and *SOP-2* according to your degree of participation.

In order to order the necessary media, their composition is enclosed in this mail for the laboratories who have not received the draft SOPs.

The number of colonies picked for confirmation are described in SOP-1 & 2 are reduced compared to the ISO & AOAC standards. In SOP-1 one typical or suspect colony per plate (*ie* 4 per sample) is purified then confirmed. If none of them is confirmed as Salmonella, picked

one another typical or suspect colony per plate until if necessary 5 suspect colonies per plate. In SOP-2, 2 typical or suspect colonies are picked for confirmation from the 3 selective isolation plates. If none of them are confirmed as Salmonella, repeat this process and stop after 2 suspect colonies per plate (*ie* 12 colonies per samples)

➤ and test reports for both methods ; Shipments and Media used, Examination of samples, SOP-1 results and SOP-2 results

The methods prescribed in the SOPs must be strictly adhered to and any deviations must be reported. The results pro-forma must be returned to the organising laboratory not later than the specified date for statistical analyses and discussion.

Chronological description of the collaborative trials on Salmonella

Date 2000	Action
first week of January	Mailing of draft SOP-1 and -2 for training exercise by AFSSA-Ploufragan
January 17 th	Reference capsules dispatched by RIVM for training exercise Store reference capsules at $-20^{\circ}\text{C} \pm 5^{\circ}\text{C}$ until examination
February or March	Training exercise
February	Mailing of confirmation of participation, time schedules, SOPs and test report to participants
March 21 st	Cheese samples dispatched by CECALAIT. Upon arrival note date, temperature, condition and record on test report. Store cheese samples at $1^{\circ}\text{C} - 5^{\circ}\text{C}$ until examination
May 8 th	Egg samples and reference capsules dispatched by RIVM. Upon arrival note date, temperature, condition and record on test report. Store egg samples and reference capsules at $-20^{\circ}\text{C} \pm 5^{\circ}\text{C}$ until examination
May 8 th	Poultry samples dispatched by CSL. Upon arrival note date, temperature, condition and record on test report. Store poultry samples at $1^{\circ}\text{C} - 5^{\circ}\text{C}$ until examination
May 22 th	Commence examination of egg samples *
March 27 th	Commence examination of cheese samples and the 2 cheese samples contaminated with capsules *
May 15 th	Commence examination of poultry samples *
June 30 th	Closing date for receipt results by AFSSA-Ploufragan for analysis
* Note :in addition the 5 reference capsules must be examined together concurrently with any of the food types and is the discretion of the participant.	

Trial Organiser :

Florence Humbert and Marylène Bohnert

AFSSA-Ploufragan

Zoopôle - Beaucemaine

B.P. 53

Fax : 33 (0)2 96 01 62 23

22440 PLOUFRAGAN

E.Mail

f.humbert@ploufragan.afssa.fr

France

m.bohnert@plougragan.afssa.fr

Amendment to sample shipment: each lab will receive one parcel. RIVM will send meat samples + egg samples + the reference capsules to the 17 labs in Europe. RIVM will receive the meat samples in the week of 2TH May from CSL, and RIVM will ship the samples to the European participants on the 8th May.

ANNEX D : Summary of collaborative trial results, Reference materials and cheese samples + capsule

Number of Salmonella positive sample among the 5 RMs and among the 2 RC samples
 italic & underlined results = excluded

R capsules examined with diced poultry				R capsules examined with egg powder	
N° Lab	R code of sample			RC code of sample	
	ISO	AOAC		ISO	AOAC
<i>1</i>	<u>1</u>	<u>5</u>	<i>RM's results and diced poultry's results excluded (4 out of 5 RMs found negative)</i>	2	2
2	5	5		2	1
3	4	//		2	//
4	5	5		2	2
5	5	5		2	2
6	5	//		2	//
7	5	//		//	//
8	5	//		2	//
9	5	//		2	//
10	4	//		2	//
11	5	5		2	2
12	5	5		//	//
13	5	5		2	2
14	5	4		2	2
15	4	//		2	//
16	5	5		//	//
17	5	5		2	2
18	4	5		//	//
19	5	5		2	2
20	5	5		2	2
21	5	4		2	2
22	4	5		2	2
23	4	5		2	2
24	5	5		2	2
25	5	5		2	2
26	4	4		2	2
27	did not report capsule data			2	2

R capsules contained about 5 cfu / capsule of *S. Typhimurium*

The capsules used to contaminate cheese sample contained about 100 *S. Typhimurium* per capsule

The lab N°15 did not received capsules to contaminate the 2 cheese samples called RC1 and RC2.

Capsules contained 3-5 *Salmonella* Panama made by RIVM were used ; 2 capsules per sample

Cheese samples

N° Lab	ISO (C1. x)		AOAC (C2. x)	
	Low level	High level	Low level	High level
1	4	5	5	5
2	1	1	0	3
3	1	2	//	//
4	5	5	5	5
5	3	5	5	5
6	5	5	//	//
8	5	5	//	//
9	5	5	//	//
10	5	5	//	//
11	5	5	5	5
13	5	5	5	5
14	5	5	5	5
15	<u>4</u>	<u>5</u>	//	//
17	4	5	5	5
19	5	5	5	5
20	2	5	5	5
21	3	5	5	5
22	<u>2</u>	<u>5</u>	<u>5</u>	<u>5</u>
23	0	0	3	2
24	5	5	2	5
25	5	5	5	5
26	5	5	5	5
27	0	0	2	3

Results excluded = a negative control found positive (all results obtained with the matrix)

Results excluded = a negative control found positive (all results obtained with the matrix)

Diced poultry samples

N° Lab	ISO (M1. ...)		AOAC (M2. ...)		
	Low level	High level	Low level	High level	
1	<u>5</u>	<u>5</u>	<u>1</u>	<u>4</u>	<i>Results excluded (all results obtained with the SOP , 4 out of 5 RMs found negative)</i>
2	<u>5</u>	<u>5</u>	<u>3</u>	<u>5</u>	<i>Results excluded (a negative control found positive)</i>
4	<u>4</u>	<u>5</u>	<u>2</u>	<u>4</u>	<i>Results excluded (a negative control found positive)</i>
12	<u>3</u>	<u>5</u>	<u>3</u>	<u>4</u>	<i>Results excluded (an identified Salmonella strain differing from the strain used)</i>
20	<u>5</u>	<u>5</u>	<u>2</u>	<u>5</u>	<i>Results excluded (a negative control found positive)</i>
27	<i>did not finish sample analysis</i>				

N° Lab	ISO (M1. ...)		AOAC (M2. ...)	
	Low level	High level	Low level	High level
3	5	5	//	//
5	3	5	4	5
6	5	5	//	//
7	5	5	//	//
9	5	5	//	//
10	5	5	//	//
11	5	5	1	5
13	5	5	5	5
14	5	5	2	5
15	5	5	//	//
16	5	5	1	4
17	5	5	3	5
18	4/4 ¹	5	3/4 ¹	5
19	5	5	4	5
21	5	5	1	2
22	5	5	1	5
23	5	5	4	5
24	5	5	4	5
25	5	5	2	5
26	5	5	3	5

¹one sample not tested due to sample bag leakage

Egg powder samples

N° Lab	ISO (D1. ...)		AOAC (D2. ...)		
	Low level	High level	Low level	High level	
3	<u>5</u>	<u>5</u>	//	//	<i>Results excluded (a negative control found positive)</i>
4	<u>5</u>	<u>5</u>	<u>2</u>	<u>4</u>	<i>Results excluded (a negative control found positive)</i>
22	<u>5</u>	<u>5</u>	<u>4</u>	<u>5</u>	<i>Results excluded (have uninoculated positives -same serotype as contaminating Salmonella)</i>
25	<u>5</u>	<u>5</u>	<u>5</u>	<u>5</u>	
27	<u>5</u>	<u>5</u>	<u>5</u>	<u>5</u>	

N° Lab	ISO (D1. ...)		AOAC (D2.)	
	Low level	High level	Low level	High level
1	3	5	5	5
2	5	5	5	5
5	5	4	5	5
6	5	5	//	//
7	5	5	//	//
8	5	5	//	//
9	5	5	//	//
10	5	5	//	//
11	5	5	5	5
13	5	5	5	5
14	5	5	5	5
15	5	5	//	//
16	5	5	5	5
17	5	5	5	5
18	5	4/4 ¹	5	4/4 ¹
19	5	5	5	5
20	5	5	5	5
21	5	5	5	5
23	5	5	5	5
24	5	5	5	5
26	5	5	5	5

¹ one sample not tested due to sample bag leakage

3.5 Annexe 5. Etude des estimateurs de fidélité

**Estimation de la répétabilité et de la reproductibilité des méthodes
de dénombrement.**

Approches statistiques non robustes et robustes

3.5.1 Approche non robuste selon l'ISO 5725-2

3.5.1.1 Tests d'identification des valeurs aberrantes

Test de Cochran

Pour un ensemble de p écarts-types s_i , tous calculés avec le même nombre de répétitions (n), la statistique C du test de Cochran est la suivante :

$$C = \frac{s_{\max}^2}{\sum_{i=1}^p s_i^2} \quad \text{Eq. 27}$$

où s_{\max} est l'écart-type maximal de l'ensemble.

Si C est inférieur ou égal à la valeur critique de la statistique à 5 % (voir les tables de Cochran fournies en annexe de l'ISO 5725-2), s_{\max} est considéré comme acceptable. Si C est supérieur à la valeur critique à 5 % mais inférieur à la valeur critique à 1 %, s_{\max} est considéré comme douteux. Si C est supérieur à la valeur critique à 1 %, s_{\max} est considéré comme aberrant.

Ce test peut être répété sur les valeurs restantes après avoir éliminé un écart-type, mais une utilisation trop répétée peut conduire à des rejets excessifs (c'est ce qu'ont voulu éviter l'AOAC ou la FIL en fixant un ratio maximal de rejet).

Test de Grubbs

a) *Test simple*

Pour un ensemble de moyennes de répétitions \bar{x}_i , $i = 1$ à p , ordonnées dans un ordre croissant, la statistique G_p du test de Grubbs est calculée afin de tester si la valeur la plus grande \bar{x}_p est aberrante :

$$G_p = (\bar{x}_p - \bar{\bar{x}}) / s \quad \text{Eq. 28}$$

où :

$$\bar{\bar{x}} = \frac{1}{p} \sum_{i=1}^p \bar{x}_i \quad \text{et} \quad s = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (\bar{x}_i - \bar{\bar{x}})^2} \quad \text{Eq. 29}$$

Pour tester la signification de la plus petite valeur \bar{x}_1 , la statistique G_1 du test de Grubbs est calculée :

$$G_1 = (\bar{\bar{x}} - \bar{x}_1) / s \quad \text{Eq. 30}$$

Si la statistique G_p ou G_1 est inférieure ou égale à sa valeur critique à 5 % (voir les tables de Grubbs, fournies en annexe de ISO 5725-2), \bar{x}_p ou \bar{x}_1 est considérée comme satisfaisante. Si la statistique G_p ou G_1 est entre sa valeur critique à 5 % et sa valeur critique à 1 %, \bar{x}_p ou \bar{x}_1 est considérée comme valeur douteuse. Si la statistique G_p ou G_1 est supérieure à sa valeur critique à 1%, \bar{x}_p ou \bar{x}_1 est considérée comme valeur aberrante.

b) Test double

Pour tester si les deux plus grandes moyennes de répétitions \bar{x}_p et \bar{x}_{p-1} sont aberrantes, la statistique G du test de Grubbs est calculée ainsi :

$$G = s_{p-1,p}^2 / s_0^2 \tag{Eq. 31}$$

avec:

$$s_0^2 = \sum_{i=1}^p (\bar{x}_i - \bar{\bar{x}})^2, \quad s_{p-1,p}^2 = \sum_{i=1}^{p-2} (\bar{x}_i - \bar{\bar{x}}_{p-1,p})^2 \quad \text{et} \quad \bar{\bar{x}}_{p-1,p} = \frac{1}{p-2} \sum_{i=1}^{p-2} \bar{x}_i \tag{Eq. 32}$$

Pour tester si les deux plus petites moyennes de répétitions \bar{x}_1 , \bar{x}_2 sont aberrantes, la statistique G du test de Grubbs est calculée ainsi:

$$G = s_{1,2}^2 / s_0^2 \tag{Eq. 33}$$

où :

$$s_{1,2}^2 = \sum_{i=3}^p (\bar{x}_i - \bar{\bar{x}}_{1,2})^2 \quad \text{et} \quad \bar{\bar{x}}_{1,2} = \frac{1}{p-2} \sum_{i=3}^p \bar{x}_i \tag{Eq. 34}$$

L'interprétation est équivalente à celle du test simple.

c) Utilisation des tests de Grubbs

Le test simple est appliqué aux moyennes de répétitions. Si une moyenne est considérée comme aberrante, après l'avoir exclue, le test simple est répété sur l'autre moyenne extrême. Si le test simple n'indique pas de moyenne aberrante, le test double est utilisé.

3.5.1.2 Estimation de écarts-types de répétabilité et de reproductibilité

Les calculs sont menés selon les étapes suivantes :

Moyenne générale $\bar{\bar{x}}$ pour une combinaison (matrice x niveau de contamination) :

$$\bar{x} = \frac{\sum_{i=1}^p n_i \bar{x}_i}{\sum_{i=1}^p n_i} \quad \text{Eq. 35}$$

où : n_i est le nombre de répétitions pour le laboratoire i , \bar{x}_i est la moyenne des répétitions pour le laboratoire i et p est le nombre de laboratoires participants.

- Variance de répétabilité :

$$s_r^2 = \frac{\sum_{i=1}^p (n_i - 1) s_i^2}{\sum_{i=1}^p (n_i - 1)} \quad \text{Eq. 36}$$

Variance intra-laboratoire du laboratoire i , s_i^2 ,

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (x_j - \bar{x}_i)^2}{n_i - 1} \quad \text{Eq. 37}$$

Variance inter-laboratoires s_L^2 :

$$s_L^2 = \frac{s_d^2 - s_r^2}{n} \quad \text{Eq. 38}$$

Variance des moyennes de répétitions

$$s_d^2 = \frac{1}{p-1} \sum_{i=1}^p n_i (\bar{x}_i - \bar{x})^2 \quad \text{Eq. 39}$$

$$\bar{n} = \frac{1}{p-1} \left[\sum_{i=1}^p n_i - \frac{\sum_{i=1}^p n_i^2}{\sum_{i=1}^p n_i} \right] \quad \text{Eq. 40}$$

Variance de reproductibilité s_R^2 :

$$s_R^2 = s_r^2 + s_L^2 \quad \text{Eq. 41}$$

3.5.2 Approche robuste selon l'EN ISO 16140

3.5.2.1 Écart-type de répétabilité

Un estimateur robuste de l'écart-type de répétabilité s_r est déduit de la médiane des écarts-types intra-laboratoires s_i (notée $MED\{s_i\}$), de la façon suivante :

$$s_r = 1,483 \text{ MED}\{s_i\} \quad \text{Eq. 42}$$

où 1,483 est un facteur d'échelle utilisé pour obtenir un estimateur non biaisé de l'écart-type. Ce coefficient provient du fait que $1,483 \approx \sqrt{\nu/\chi^2}$ avec $\nu = n - 1$ le nombre de d.d.l. des répétitions et le χ^2 pour $p = 0,5$ qui vaut $\chi^2 = 0,455$ (Analytical Methods Committee, 1989a)

Ecart-type de reproductibilité

Un estimateur robuste de l'écart-type des moyennes de répétitions s_d est déduit de l'estimateur d'échelle de Rousseeuw S_n (Rousseeuw & Croux, 1993). S_n consiste en deux médianes successives :

- à partir de chacune des p moyennes de répétitions (\bar{x}_i), la médiane de ses $(p-1)$ différences absolues aux autres moyennes \bar{x}_j est calculée ;
- la médiane de ces p médianes de différences est ensuite calculée.

Soit :

$$S_n = MED_i \left\{ MED_j \left| \bar{x}_i - \bar{x}_j \right| \right\} \quad \text{Eq. 43}$$

On en déduit s_d de la façon suivante :

$$s_d = 1,1926 S_n \quad \text{Eq. 44}$$

où : 1,1926 est un facteur déduit d'une distribution de χ^2 centrée sur 0,75, afin d'obtenir un estimateur non biaisé de l'écart-type, sous l'hypothèse de normalité.

L'écart-type inter-laboratoires s_L et l'écart-type de reproductibilité s_R sont ensuite déduits comme dans l'approche classique de l'ISO 5725-2 (voir ci-dessus, A.1).

3.5.3 Approche robuste selon l'ISO 5725-5

3.5.3.1 Algorithme S

Cet algorithme combine de façon robuste les écarts-types intra-laboratoires s_i de chaque laboratoire pour estimer l'écart-type de répétabilité s_r .

Soit un ensemble de p écarts-types intra-laboratoires s_i , classés dans un ordre croissant, s^* désigne leur valeur combinée robuste et ν le nombre de d.d.l. associé à chaque s_i (pour n répétitions, $\nu = n-1$).

Une valeur initiale de s^* est calculée ainsi:

$$s^* = \text{MED} \{s_i\} \text{ (médiane des } \{s_i\}\text{), pour } i = 1 \text{ à } p \quad \text{Eq. 45}$$

La valeur de s^* est mise à jour de la façon suivante :

- soit $\psi = \eta s^*$, où η est un facteur limite dépendant de ν (décrit en annexe B de l'ISO 5725-5) ;
- pour chaque s_i ($i=1$ à p), $s_i^* = \psi$ si $s_i > \psi$, ou s_i sinon ;
- la valeur mise à jour de s^* est calculée ainsi :

$$s^* = \xi \sqrt{\sum_{i=1}^p (s_i^*)^2 / p}, \quad \text{Eq. 46}$$

où ξ est un facteur d'ajustement (décrit en Annexe B de l'ISO 5725-5)

L'estimateur robuste s^* peut être déduit d'un calcul par itérations, jusqu'à ce que le changement de valeur de s^* soit petit. Il peut être également déduit, comme nous l'avons choisi, en exprimant s^* ainsi :

$$(s^*)^2 = (\xi^2 / p) \times \left[\sum' (s_i^*)^2 + u_U \times (\eta s^*)^2 \right] \quad \text{Eq. 47}$$

où : \sum' est la sommation des $s_i \leq \psi$ et u_U est le nombre des $s_i > \psi$.

Nous avons d'abord suivi l'approche itérative pour obtenir une valeur approximative de s^* , puis utilisé l'équation ci-dessus pour obtenir la valeur exacte.

L'écart-type de répétabilité s_r est égal à la valeur finale de s^* .

3.5.3.2 Algorithme A

Cet algorithme estime l'écart-type inter-laboratoires s_L à partir des moyennes de double de chaque laboratoire et utilise un estimateur initial d'échelle s^* qui est la Médiane des Déviations Absolues (MAD).

Soit un ensemble de p moyennes de doubles \bar{x}_i , classées en ordre croissant, x^* et s^* désignent la moyenne et l'écart-type robustes de ces données. Les valeurs initiales de x^* et s^* sont calculées ainsi:

$$x^* = \text{MED} \{ \bar{x}_i \} \quad \text{Eq. 48}$$

$$s^* = 1,483 \text{ MED} \{ |\bar{x}_i - x^*| \} \quad \text{Eq. 49}$$

Le coefficient 1,483 est le même facteur d'échelle que celui de EN ISO 16140 (voir §0)

$\text{MED} \{ |\bar{x}_i - x^*| \}$ est appelée la Médiane des Déviations Absolues (MAD).

Les valeurs de x^* et s^* sont mises à jour ainsi :

- $\varphi = 1,5 s^*$ où 1,5 est une valeur seuil [Analytical Method Committee, 1989 a] ;
- pour chaque \bar{x}_i ($i = 1$ à p), $x_i^* = x^* - \varphi$ si $\bar{x}_i < x^* - \varphi$; $x_i^* = x^* + \varphi$ si $\bar{x}_i > x^* + \varphi$; \bar{x}_i sinon ;
- les nouvelles valeurs de x^* et s^* sont calculées ainsi:

$$x^* = \sum_{i=1}^p x_i^* / p \quad \text{Eq. 50}$$

$$s^* = 1,134 \sqrt{\sum_{i=1}^p (x_i^* - x^*)^2 / (p-1)} \quad \text{Eq. 51}$$

Le coefficient $1,134 \approx 1/\sqrt{0,778}$, 0,778 est un facteur choisi pour obtenir une réponse correcte pour des données distribuées selon la loi Normale [Analytical Method Committee, 1989a].

Les estimateurs robustes x^* et s^* peuvent être déduits par itérations, jusqu'à ce les changements de valeurs soient faibles. Il est également possible, comme nous l'avons fait, d'exprimer x^* et s^* ainsi :

$$x^* = x' + 1,5 \times (u_U - u_L) s^* / (p - u_L - u_U) \quad \text{Eq. 52}$$

$$(s^*)^2 = (p - u_L - u_U - 1) \times (s')^2 / [(p-1)/1,134^2 - 1,5^2 (pu_L + pu_U - 4u_L u_U) / (p - u_L - u_U)] \quad \text{Eq. 53}$$

où:

- u_L est le nombre de $x_i < x^* - \varphi$,
- u_U est le nombre de $x_i > x^* + \varphi$,
- x' et s' sont les moyennes et écarts-types des $(p - u_L - u_U)$ \bar{x}_i tels que $|\bar{x}_i - x^*| \leq \varphi$.

Comme pour l'algorithme S, nous avons d'abord suivi la méthode itérative pour obtenir des valeurs approximatives de x^* et s^* et nous avons ensuite utilisé les équations ci-dessus pour obtenir les valeurs exactes.

L'écart-type des moyennes de doubles s_d est égal à la valeur finale de s^* . L'écart-type inter-laboratoires s_L et l'écart-type de reproductibilité s_R sont ensuite déduits comme dans l'approche classique de l'ISO 5725-2 (voir ci-dessus, A.1).

3.6 Annexe 6. Publication sur la fidélité des méthodes qualitatives

Analysing collaborative trials for qualitative microbiological methods : accordance and concordance.

Langton S.D., et al. *International Journal of Food Microbiology*,
2002, 79, p 175-181.



ELSEVIER

International Journal of Food Microbiology 79 (2002) 175–181

INTERNATIONAL JOURNAL OF
Food Microbiology

www.elsevier.com/locate/ijfoodmicro

Analysing collaborative trials for qualitative microbiological methods: accordancy and concordance

S.D. Langton^{a,*}, R. Chevenement^b, N. Nagelkerke^c, B. Lombard^d

^aDEFRA Central Science Laboratory, Sand Hutton, York, YO41 1LZ, UK

^bEcole Nationale d'Industrie Laitière et des Biotechnologies (ENILBIO), Poligny, BP: 49, 39800, Poligny, France

^cNational Institute of Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

^dAgence Française de Sécurité Sanitaire des Aliments, Laboratoire d'Etudes et de Recherches en Hygiène et Qualité des Aliments, 41 Rue du 11 Novembre 1918, F-94 700 Maisons-Alfort, France

Received 5 March 2001; received in revised form 25 January 2002; accepted 26 February 2002

Abstract

In qualitative (detection) food microbiology, the usual measures of repeatability and reproducibility are inapplicable. For such studies, we introduce two new measures: accordancy for within laboratory agreement and concordance for between laboratory agreement, and discuss their properties. These measures are based on the probability of finding the same test results for identical test materials within and between laboratories, respectively. The concordance odds ratio is introduced to present their relationship. A method to test whether accordancy differs from concordance is discussed. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Collaborative trials; Qualitative microbiological methods; Accordancy; Concordance

1. Introduction

The concepts of repeatability (r) and reproducibility (R) (Anonymous, 1994; IDF, 1991) are widely used in the analysis of data from collaborative trials in quantitative microbiology (see for example, Szita et al., 1998; Schulten et al., 2000). Repeatability provides a measure of the variability between analyses conducted on identical test materials by the same technician in the same laboratory, under conditions as similar as possible (e.g. by using the same apparatus and reagents within

the shortest possible interval of time), whilst the reproducibility measures the variability when the analyses are conducted by different technicians at different laboratories. The concepts were originally borrowed from the literature relating to chemical analysis, but they are equally applicable to microbiological data, provided the counts are transformed to logarithms before analysis. Formally, there is a 95% probability that the absolute difference between two test results should be less than r and R under repeatability and reproducibility conditions, respectively.

These measures of repeatability and reproducibility do not measure departure from the “true value”, i.e. accuracy. They only indicate to what extent results can be repeated or reproduced—either correctly or not. In microbiology, where the objective is to count the

* Corresponding author. Tel.: +44-1904-455100; fax: +44-1904-455254.

E-mail address: s.langton@csl.gov.uk (S.D. Langton).

number of colony forming units (cfu), accuracy may depend on many factors, such as the (food) matrix, the strain of organism or the presence and type of competing flora. A particular quantification method may perform differently under different circumstances. However, it should not perform randomly. Random behaviour indicates a lack of “control”, e.g. due to ambiguities in the definition of the method. It is this random behaviour that r and R are designed to measure. In addition, (substantial) differences between r and R may give clues about the origin of randomness. From the wide use that is being made of these concepts, one may infer that they must be useful and aid microbiologists in decision making and in improving and standardizing their procedures. The increasing need to standardize microbiological methods (Lahellec, 1998; Leclercq et al., 2000) provides an additional argument for the development and use of (statistical) methods and the organization of collaborative trials (e.g. Scotter et al., 2001; Schulten et al., 1998; De Buyser and Lombard, 2000) that promote further standardization.

A major limitation of repeatability and reproducibility is that they are only applicable to quantitative data and cannot be applied to qualitative analyses, i.e. detection methods. This is because both parameters are expressed in terms of the likely difference in log units between two test results; clearly this can have no meaning for qualitative data where two results are either the same or are different. In this paper, we describe two new measures, accordance and concordance, which can be regarded as analogous quantities to repeatability and reproducibility for qualitative data.

These new parameters have been developed as part of European project SMT 4-CT 96-2098 funded by European Commission, which aimed to validate the six main reference methods (ISO/CEN Standards) in food microbiology by means of interlaboratory trials. They have been in particular applied to the validation of the reference method EN ISO 11290-1 for the detection of *Listeria monocytogenes* in foods (Scotter et al., 2001).

2. Example data

The methods will be illustrated by means of the data in Table 1, which is taken from a collaborative study on *L. monocytogenes*. These data are just for

Table 1

Example data taken from a collaborative study on *L. monocytogenes*

Laboratory	Replicate number					Number positive (out of 5)
	1	2	3	4	5	
1	+	+	+	+	+	5
2	+	+	+	+	+	5
3	+	+	+	+	+	5
4	+	+	+	+	+	5
5	–	–	+	+	+	3
6	+	+	+	+	+	5
7	–	–	+	+	+	3
8	+	+	+	+	+	5
9	+	+	+	+	+	5
10	+	+	+	+	+	5

one level of one food type using one particular medium. In practice, a collaborative trial would usually aim to involve at least 20 laboratories (in this case, there was a similar number using a different medium), but a smaller data set makes it easier to explain the calculations.

3. Statistics for qualitative data

3.1. Accuracy

In a collaborative trial with quantitative data, whenever possible, results will be compared with the “true” contents of the sample(s) in order to demonstrate the accuracy of the method. The equivalent statistics for qualitative data are straightforward. For positive samples this is known as *sensitivity* and is the percentage of samples correctly identified as positives; in the example this is 46 (eight laboratories with five correct plus two laboratories with three correct) out of 50 or 92.0%. For the purposes of this calculation it must be assumed that all supposedly positive samples do in fact contain the organism; in some cases with very low levels, this may not be the case in reality. As the sensitivity can depend on circumstances such as the food matrix, a reported sensitivity only applies to the set of circumstances under which it was measured.

For blank samples, the percentage of samples correctly identified as being negative is recorded; this is known as the *specificity*.

Approximate standard errors and confidence limits for these figures can be based on standard methods for

binomial (positive/negative) data (Armitage and Berry, 1987). For example, using their methods to calculate an exact 95% confidence interval for the sensitivity, we obtain 80.8–97.8%. However, such figures may be misleading if there is any heterogeneity among laboratories and should not be used where this is clearly the case. Also, it is assumed that all samples are in fact positive and that negative results are therefore incorrect.

3.2. Accordance

We will define the qualitative equivalent of repeatability as *accordance*; this is the (percentage) chance that two identical test materials analysed by the same laboratory under standard repeatability conditions will both be given the same result (i.e. both found positive or both found negative). To calculate the accordance, we take each laboratory in turn and calculate the probability that two samples will give the same result and then average this probability over all laboratories. For those laboratories (such as laboratory 1) where all samples were found positive, the best estimate of the probability of getting the same result is clearly 1.00 or 100%; all 10 out of 10 possible pairs are both positive. For the others (in the example, laboratories 5 and 7) we must count the number of pairs that are either both negative (1 pair) or both positive (3 pairs). Thus a total of 4 out of 10 pairs (40%) give the same result. In general, when a laboratory has n results and k of these are positive, then the accordance for that laboratory is estimated as

accordance for laboratory

$$= \{k(k-1) + (n-k)(n-k-1)\} / n(n-1)$$

The accordance for the collaborative trial as a whole is the average (mean) of these probabilities for each laboratory, which is 88% in this case, as can be seen from Table 2. The calculation of the standard errors and confidence intervals is discussed in Appendix A.

3.3. Concordance

The equivalent of reproducibility is *concordance*, which is the percentage chance that two identical test materials sent to different laboratories will both be

Table 2

Calculating the accordance for the example data; the accordance is the average of the probabilities in the final column

Laboratory	Number positive	Positive pairs	Negative pairs	Total pairs the same	Proportion the same
1	5	10	0	10	100%
2	5	10	0	10	100%
3	5	10	0	10	100%
4	5	10	0	10	100%
5	3	3	1	4	40%
6	5	10	0	10	100%
7	3	3	1	4	40%
8	5	10	0	10	100%
9	5	10	0	10	100%
10	5	10	0	10	100%
				Total = 88	Mean = 88%

given the same result (i.e. both found positive or both found negative result). The most intuitive way to calculate concordance is simply to enumerate all possible between laboratory pairings in the data.

To enumerate all pairings, each observation in each laboratory is considered in turn, starting with the first observation of laboratory 1, which is positive. This can be paired with any of the 45 observations from other laboratories, and all but four of these pairings (those with samples 1 and 2 of laboratories 5 and 7) match (i.e. give a pair with both positive); there are thus 41 pairs giving the same result. Similarly, the other four samples from laboratory 1 also each have 41 out of 45 pairings giving the same result, and so there are a total of 225 (5×45) between laboratory pairings involving laboratory 1, of which 205 (5×41) give the same result. The same applies to all other laboratories with all samples positive. For laboratory 5, with three out of five positives, the two negative samples each match with just two other negative samples at other laboratories, whilst the three positive samples each match with 43 positive samples. Thus the total number of pairs with the same result for laboratory 5 (and also laboratory 7) is 133 ($2 \times 2 + 3 \times 43$). These calculations are summarised in Table 3.

The concordance is the percentage of all pairings giving the same result; in this example this is 84.7% ($1906/2250 \times 100$). (The alert reader will have noticed that there are actually only 1125 pairings as we have double counted them all by this procedure; this does not affect the results as it applies to all pairs

Table 3
Calculating the concordance for the example data

Laboratory	Number positive	Between-laboratory pairings with same results	Total between-laboratory pairings
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

and is easier than allowing for it during the calculations.)

An alternative way to calculate concordance is to first calculate all pairs with the same results from all laboratories irrespective whether the result is from the same laboratory or not. As there are four negative results and 46 positive results, we have 1041 (1035 + 6) pairs with the same result (note that the number of pairings possible between n items is given by $n(n-1)/2$, for example for the 46 positives $46 \times 45/2 = 1035$). From the calculation of accordance, we know that 88 are from the same laboratory (Table 2), and so 953 (1041 – 88) pairs with the same result are from different laboratories (counting each pair only once). Similarly, it is easily found that there are a total of 1125 (1225 – 100) pairs from different laboratories. Again, we find 84.7% as the value of concordance.¹ The calculation of the standard errors and confidence intervals is discussed in Appendix A, and an Excel spreadsheet to perform these calculations is available by email from the authors.

¹ Concordance can also be calculated from accordance using the formula

$$(\text{estimated}) \text{ concordance} = \frac{\{2r(r - nL) + nL(nL - 1) - AnL(n - 1)\}}{\{(n^2)L(L - 1)\}}$$

where r = total number of positives, L = number of laboratories, n = replications per laboratory, and A = accordance, expressed as a proportion.

3.4. Measuring and testing between laboratory variation

With quantitative data, variation between laboratories can be tested for statistical significance and its magnitude can be estimated by examining the laboratory component of variance or, alternatively, simply by comparing the repeatability with the reproducibility. A similar comparison between accordance and concordance can be used to assess the magnitude of between laboratory variation; if the concordance is smaller than the accordance (i.e. samples analysed at the same laboratory are more likely to give the same result than ones sent to different laboratories) between laboratory variation is present. This may be due to, for example, variations in media quality across laboratories, differences in the likelihood of cross-contamination or different interpretations of guidelines by different laboratory technicians.

Unfortunately, the magnitude of the concordance and accordance is strongly dependent on the sensitivity, making it difficult to assess easily the degree of between laboratory variation. It is therefore helpful to calculate the concordance odds ratio (COR)

$$\text{COR} = \frac{\text{accordance}^*(100 - \text{concordance})}{\text{concordance}^*(100 - \text{accordance})}$$

where concordance and accordance are expressed as percentages. This ratio is less sensitive to the level of sensitivity than either accordance, concordance, or a simple ratio between the two, and therefore provides a convenient measure of the degree to which results vary between laboratories. The magnitude of the ratio can be interpreted as the relative chance ('odds' in betting terms) of getting the same result when two samples are analysed in the same laboratory compared to if they are sent to the different laboratories. Thus an odds ratio of 1.3 indicates that the samples are 1.3 times more likely to produce the same result (both positive or both negative) if they are analysed in the same laboratory than if they are analysed in different laboratories. Ideally an odds ratio should be close to 1.0, indicating that results are just as likely to be the same irrespective of whether the two samples are analysed at the same or different laboratories. However, in practice, the variation between laboratories will generally be larger than within laboratories. The

larger the odds ratio, the greater is the level of inter-laboratory variation in the data.

The COR equals infinity in the case when accordance is 100% (i.e. when all laboratories report either all positives or all negatives), indicating serious differences between laboratories. When both accordance and concordance are 100%, there is no variability whatsoever in the data and the COR can be considered to equal the ideal value of 1.0.

Values of the concordance odds ratio above 1.00 may occur by chance variation, and so a statistical significance test should be used to confirm whether the evidence for extra variation between laboratories is convincing. Such tests are based on the idea that if accordance equals concordance, then each laboratory has the same probability of finding a positive test. A cross tabulation of laboratory by result (number positive, number negative) should then be homogeneous. The best test to use is an ‘exact test’, which can be obtained using statistical packages such as SAS or StatXact. The philosophy behind such tests is that the probabilities of occurrence are calculated for all sets of results that could have produced the overall numbers of positives and negatives. In the example, with 46 positives and four negatives the arrangements shown in Table 4 are possible.

The laboratories are not labelled in this table, as any permutation would give the same overall result (e.g. for the right hand column, the 1 could be for any of the laboratories). The test adds up the probabilities

of all those possible arrangements that show at least as much evidence for between laboratory variation as the real result—here that means all permutations of the three columns on the right. If this probability is less than the conventional value of 0.05 or 5%, it is unlikely that this degree of between laboratory variation could have occurred by chance and hence we conclude that there is significant variation in performance between laboratories. In the example above $P=0.039$ indicating that variation between laboratories is significant at the 5% level. Except in very simple examples like this one, specialist software is needed to calculate exact tests.

Where software for exact tests is not available, an ordinary chi-squared analysis for contingency tables (see e.g. Armitage and Berry, 1987) can be used as an alternative. The results of this test will be less reliable than the exact test with the number of replicates usually used in collaborative trials, but simulations suggest that the results provide a reasonable guide to the significance of between laboratory differences.

With either test, it must be remembered that the ability to detect between laboratory differences is dependent on the number of laboratories and the number of replicate samples analysed at each laboratory. A non-significant test result should not be taken to mean that performance does not vary between laboratories, but rather that such differences have not been proved; this is particularly true where the P -value is only just above 0.05. The ideal solution is to quote the odds ratio with a standard error or confidence limits, but the distribution of the odds ratio is highly skewed making it very difficult to produce reliable limits.

Table 4
Possible arrangements of the example data with 46 positives and four negatives

	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
Accordance (%)	84.0	86.0	88.0	90.0	96.0
Concordance (%)	85.1	84.9	84.7	84.5	84.0
COR	0.92	1.09	1.32	1.65	4.57

The observed arrangement is shown in bold. For convenience, the laboratories are shown in ascending order of positives.

4. Discussion

We have presented two new statistics, accordance and concordance, that are aimed to assume the role of repeatability and reproducibility for qualitative (detection) studies. Like the latter two measures, accordance and concordance do not measure departure from the “true” value, i.e. presence or absence of an organism. They only indicate whether the particular procedure used is sufficiently “standardised”. Departure from the true value can be measured using concepts such as sensitivity (probability of giving a positive result

when the organism is present) and specificity (probability of giving a negative result when the organism is absent). These concepts are different: for example, specificity may be low (e.g. due to the presence of a cross-reacting competitive organism, or to cross-contamination) while both accordance and concordance are high.

Although this is strictly speaking not part of the description of accordance and concordance, we feel it is appropriate to say a few words on which data should be included in the calculation of these measures. As with the validation of alternative methods (Anonymous, 2000), data should not be excluded solely due to a statistical test for outliers (e.g. Dixon, 1951; Grubbs, 1969). This is because routine removal of 'outliers' may give an unrealistic picture of the method's performance by removing the extremes of variation in the data. Data should only be rejected where there is a clear protocol violation or where a laboratory is clearly not producing acceptable results. The latter criterion might include instances where a laboratory reports a very high proportion of positives amongst blank samples. However, if such data would have been deemed "acceptable" outside the context of the trial (i.e. under routine circumstances), then they should be retained in the analysis. In any case, exclusion of data, and on what grounds, should always be clearly reported.

One option where there are results which are considered unusual (perhaps identified using Fisher's Exact Test), but cannot be ascribed to a protocol violation, is to analyse the data with and without the data from that laboratory. The value including all laboratories should still be regarded as the definitive figure, but comparison of the two figures will indicate whether the one doubtful observation is having a major influence on the overall statistics and what scope there is for improvement in accordance and concordance.

Acknowledgements

This study was supported by the European Union's Standards, Measurements and Testing Fourth Framework Programme Project SMT4 CT 96 2098 (Coordinator C. Lahellec, Agence Française de Sécurité Sanitaire des Aliments).

Appendix A. Standard errors and confidence intervals

The standard error of both the accordance and the concordance depend on the way the data has been collected and the interpretation we want to give to the data.

The most common situation is that laboratories should be considered a representative sample of a larger "population" of laboratories, e.g. all laboratories in Europe recognised under a certain programme. Similarly, they may represent all laboratories in a certain region or represent, for example, high quality laboratories. For example, in order to assess the variability in performance of a test and to certify its performances, a certification body might randomly select a small sample of the laboratories it has approved and ask them to take part in a collaborative trial.

When laboratories are considered representative, results from a trial have implications for all laboratories in the "population" of laboratories, not just the participating ones. In the above example of a certification body examining the performance of a test, the results of the trial have implications for all laboratories which it has approved and therefore *could* have been included in the trial. Standard errors then provide an indication of how different the accordance and concordance values might have been if, instead of the actual sample, another sample of laboratories had been taken.

Less frequently, the participating laboratories may be "fixed" in the sense that they are not considered a random sample of a population of laboratories. They only represent themselves. For example, a group of laboratories may have joined forces to examine (and improve) the performance of certain tests in their laboratories. The results of the trial have no implications for non-participating laboratories. Standard errors then refer to how variable the data would have been under replication of the trial in the *same* laboratories.

The easiest way to calculate standard errors is by using a statistical device called the "bootstrap" (Davison and Hinkley, 1997). From the actual observations a number M (usually a number of between 20 and 100 suffices) of bootstrap samples is created. The standard deviation of accordance, concordance or COR (or any

other measure) among bootstrap samples then estimates the standard error of these measures. A bootstrap sample is constructed by sampling with replacement from the actual observations. The sample size should be exactly equal to the original data set. Where the L laboratories can be considered as a random sample of a larger population, we proceed as follows to obtain a single bootstrap sample. We give laboratories consecutive numbers 1, 2, 3, ..., L . We write these numbers on pieces of paper and put them in a box. We then draw L times a number, write this number down, and after each draw we put the piece of paper back in the box. Some laboratories are thus “sampled” more than once, while others are not included at all. Of course, randomly drawing from a box can be done efficiently by simulating it on a computer using a random number generator in, for example, Excel. Standard errors then have the interpretation of showing how the value of accordance and concordance might fluctuate under replication of the same collaborative trial in different samples of laboratories from the same “population”.

For *fixed* laboratories, it makes little sense to resample laboratories, as these are considered fixed. Instead, each bootstrap sample is obtained by bootstrapping *within* each laboratory. The idea behind this is that while laboratories are fixed, results obtained within each laboratory are not. All laboratories are present (once) in each bootstrap sample, but observations from within a laboratory are bootstrapped and thus may be either absent, occur once, or more than once in each bootstrap sample.

Once bootstrap standard errors have been estimated, approximate 95% confidence intervals can be obtained by taking the actual number plus or minus two standard deviations.

An Excel application to calculate all the statistics described in this paper, including the bootstrapped standard errors, can be obtained by sending an email to s.langton@csl.gov.uk.

References

- Anonymous, 1994. ISO 5725 Accuracy (Trueness and Precision) of Measurement Methods and Results International Organization for Standardization, Geneva, Switzerland.
- Anonymous, 2000. prEN ISO 16 140 Microbiology of Food and Animal Feeding Stuffs - Protocol for the Validation of Alternative Methods International Organization for Standardization, Geneva, Switzerland.
- Armitage, P., Berry, G., 1987. Statistical Methods in Medical Research, 2nd edn. Blackwell, Oxford.
- Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application Cambridge Univ. Press, Cambridge.
- De Buyser, M.L., Lombard, B., 2000. Validation of ISO Microbiological Methods: Enumeration of Coagulase-Positive Staphylococci According to EN ISO 6888 Part 1 and Part 2: 1999. Agence Française de Sécurité Sanitaire des Aliments, Maisons-Alfort.
- Dixon, W.J., 1951. Ratios involving extreme values. *Ann. Math. Stat.* 22, 68–78.
- Grubbs, F.E., 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11, 1–21.
- IDF, 1991. Precision characteristics of analytical methods—outline of collaborative study procedure. International IDF Standard 135B.
- Lahellec, C., 1998. Development of standard methods with special reference to Europe. *Int. J. Food Microbiol.* 45, 13–16.
- Leclercq, A., Lombard, B., Mossel, D.A., 2000. Standardisation of food microbiological analysis methods: an asset or a constraint. *Sci. Aliments* 20, 179–202.
- Schulten, S.M., vd Lustgraaf, B.E.B., Nagelkerke, N.J.D., In't Veld, P.H., 1998. Validation of Microbiological Methods: Enumeration of *Bacillus cereus* According to ISO 7932 (2nd ed., 1993). RIVM, Bilthoven.
- Schulten, S.M., in't Veld, P.H., Nagelkerke, N.J.D., Scotter, S., de Buyser, M.L., Rollier, P., Lahellec, C., 2000. Evaluation of the ISO 7932 standard for the enumeration of *Bacillus cereus* in foods. *Int. J. Food Microbiol.* 57, 53–61.
- Scotter, S.L., Langton, S., Lombard, B., Lahellec, C., Schulten, S., Nagelkerke, N., In't Veld, P.H., Rollier, P., de Buyser, M.L., 2001. Validation of ISO method 11290: Part 1. Detection of *Listeria monocytogenes* in foods. *Int. J. Food Microbiol.*, 64, 295–306.
- Szita, G., Tabajdi, V., Fabian, A., Biro, G., Reichart, O., Kormoczy, P.S., 1998. A novel, selective synthetic acetamide containing culture medium for isolating *Pseudomonas aeruginosa* from milk. *Int. J. Food Microbiol.* 43, 123–127.

3.7 Annexe 7. Publication sur la fidélité des méthodes de comptage

Experimental evaluation of different precision criteria applicable to microbiological counting methods

Lombard B., Cornu M., Lahellec C. & Feinberg M.
Version révisée, acceptée pour publication dans *Journal of AOAC INTERNATIONAL*, 2004.

Experimental evaluation of different precision criteria applicable to microbiological counting methods

Bertrand Lombard¹, Marie Cornu¹, Cécile Lahellec² & Max H. Feinberg³

1) AFSSA-LERQAP (Agence Française de Sécurité Sanitaire des Aliments – Laboratoire d’Etudes et de Recherches sur la Qualité des Aliments et sur les Procédés agro-alimentaires), 23 avenue du Général De Gaulle, 94 706 Maisons-Alfort, France.

2) AFSSA Direction Générale, 17 avenue du Général Leclerc, 94 700 Maisons-Alfort.

3) Institut National de la Recherche Agronomique (INRA) 16 rue Claude Bernard, 75005 Paris, France.

Abstract

The dispersion of microbiological counting measurements, when repeating the analysis on the same material both within a laboratory (repeatability) and between laboratories (reproducibility) can be characterized by the organization of inter-laboratory studies, where several sets of identical test materials are sent to several laboratories. Using the example of data generated by an inter-laboratory study on enumeration of *Listeria monocytogenes* in foods by the standardized reference method (colony-count technique), two types of robust estimators of reproducibility standard deviations, based on the median, are examined, in comparison with the classical estimators, based on the mean. Experimental evaluation indicates that the three approaches give consistent results for most of the combinations.

Before conducting these calculations, the usual \log_{10} transformation of the enumeration results is also questioned.

Introduction

In order to estimate within-laboratory and between-laboratory dispersion characteristics applicable to measurements obtained by microbiological counting methods for food analysis, inter-laboratory studies are routinely organized: they enable the calculation of the precision criteria of the method. The precision criteria usually chosen (repeatability and reproducibility) correspond to two extreme experimental conditions: (i) repeatability is defined as the precision of a given method in a set of conditions of minimal variability (independent test results obtained with the same method on identical test materials, within the same laboratory, by the same operator, using the same equipment/reagents and within the shortest feasible interval of time); (ii) reproducibility is defined as the precision of a given method in another set of conditions, this time of maximal variability (test results obtained with the same method on identical test materials in different laboratories, with different operators using different equipment/reagents).

Inter-laboratory studies and the related precision criteria have been described and thoroughly defined for any type of analysis in the different parts of the ISO 5725 standard series, especially parts 1 and 2 (i, ⁱⁱ), as well as in the AOAC International guidelines for food analysis (iii).

To calculate the repeatability and the reproducibility of a method which enumerates a target micro-organism by a colony-count technique (results expressed as colony forming units or CFU), the usual procedure, accepted in particular by AOAC International (iv), is (i) to convert raw results into \log_{10} , in order to supposedly approximate the Normal distribution law, a prerequisite to calculate the repeatability and reproducibility in their standard format; (ii) to identify outlying results by using statistical tests against a Normal distribution; and (iii) to calculate standard deviations of repeatability and reproducibility, based on the mean of data.

Alternative robust procedures have also been introduced in the part 5 of the ISO 5725 Standard (^v) for any quantitative methods. In the specific case of microbiological enumeration methods, other robust techniques have been developed in ISO 16140 Standard (^{vi}). Both these standards introduce robust estimators of the repeatability and reproducibility standard deviations, based on the median of data. The most interesting property of robust estimators is to be less sensitive to extreme values than a classical estimator based on the arithmetic mean. A robust estimator can then be directly calculated without excluding so-called outliers identified by statistical rejection tests.

These different robust and non-robust estimators of reproducibility have been calculated for the set of data obtained from an inter-laboratory study which enabled to generate precision data of the standardized reference method ISO 11 290-2 (^{viii}) on enumeration of *Listeria monocytogenes* in foods.

This particular set of data has been chosen since the outcome of the study had been particularly poor, in terms of precision: the robust and non robust calculations could be compared in such an ill-conditioned case.

Materials and Methods

Experimental data

The data used were obtained by an inter-laboratory study organized in the frame of a European project (Project SMT4/CT96-2098, funded by the European Commission, which aimed at validating six main ISO/CEN standardized reference methods in food microbiology). Its purpose was to determine the precision of the standardized method ISO 11 290-2 for the enumeration of *L. monocytogenes* ^(ix). Out of the 24 laboratories who initially registered for the study, 21 sent back results, and the study was performed on three different food types (fresh cheese curd, minced meat, dried egg powder) artificially inoculated at a low, a medium and a high level (10^2 , 10^3 and 10^4 CFU/g), plus negative control, and a certified reference material (capsules prepared by RIVM, Bilthoven, The Netherlands). Each participant received the samples in blind duplicates. The upper part of Table 1 gives the results (expressed in CFU/g) of the participating laboratories per food type. The table maintains the original numbering of the 24 registered participants. It only includes results used for calculating the precision data, i.e. after rejecting data for technical reasons, such as samples having reached the laboratory in bad shape or preserved under too high temperature, reported large deviations to the operating protocol more details on data rejection are given in (ix). The results reported by the participants as “0 CFU/g” or “below 10 CFU/g”, or “below 100 CFU/g”, are noted “<LoE”, i.e. below the limit of enumeration. Before the calculation step these data were conventionally replaced by 0. Moreover Table 1 neither includes the results for negative control, not used for the calculation of precision data, nor the results for the reference material.

Normality tests

The normality of the data sets, each consisting of the non \log_{10} transformed enumeration results of a combination <food types x contamination level>, was assessed through the Shapiro-Wilks test, using the software Statgraphics Plus 5.1™ (Manugistics). This test was repeated after exclusion of duplicates identified as outliers according to the Grubbs and Cochran tests. Detailed procedure is given in Method A section.

The same procedure was applied after \log_{10} transformation.

For this test, the confidence level was chosen at $\alpha = 5\%$. However, as $2 \times 2 \times 3 \times 3 = 36$ tests were performed, the question of multiple comparisons had to be taken into account. We therefore used in addition the very conservative Bonferroni procedure, and the less conservative false discovery rate procedure, according to Curran-Everett ^(x).

Estimates of precision data

Three standardized protocols have been applied for the calculation of precision criteria: Method A, classical calculations based on the exclusion of statistical outliers and on the means, as in Standard ISO 5725-2; Method B, a robust approach described ISO 5725-5; and Method C, another robust technique proposed in ISO 16140.

The algorithms are detailed in Appendix.

As mentioned before, the common starting point of the 3 procedures (Method A to C) is to transform raw data into \log_{10} (hereafter called also “raw” data, for the calculations of precision characteristics) from the laboratories having participated to the inter-laboratory study after exclusion of results for technical reasons. For comparison purposes and given the

shape of the data distribution, the calculations have been also performed without transformation into \log_{10} .

All procedures and calculations have been performed for each combination of the types of test material and the inoculation levels of *Listeria monocytogenes*.

Method A

It is the basic approach defined in ISO 5725-2, and consists in a preliminary detection of outliers before performing a one-way random effects Analysis of Variance. The identification of outlying results is based on two different statistical tests: (i) the Cochran test, which enables to test the hypothesis that only small differences exist between the intra-laboratory variances of the participating laboratories, and (ii) the Grubbs test, which enables to identify outlying laboratory means.

We have performed the calculations according to this procedure, with the software EasyVal™ (add-in to Excel).

Method B

This is the first robust approach, detailed in ISO 5725-5. Two algorithms, called A and S, are used in this iterative procedure, which is based on works of the UK Royal Society of Chemistry (xi,xii). Algorithm A estimates the inter-laboratory standard deviation from the duplicate means of each laboratory, and uses as initial estimator of scale (s^*) consisting in the Median of Absolute Deviates (MAD). Algorithm S combines the duplicate standard deviations of each laboratory to estimate the repeatability standard deviation.

We have developed Excel spreadsheets to perform the calculations according to this protocol.

Method C

This is the second robust approach, described in ISO 16140. A robust estimate of the repeatability standard deviation is derived from the median of the duplicate standard deviations of each laboratory. A robust estimate of the inter-laboratory standard deviation of the duplicate means of each laboratory is based on S_n , an estimator of scale which was introduced by Rousseeuw (^{xiii}) as an alternative to the MAD with better properties. S_n consists in two phases of mediating: (i) from each of the duplicate means, the median of its absolute differences with the other means is calculated; (ii) the median of these deviate medians is then taken.

We have also developed Excel spreadsheets and an Excel add-in to conduct the calculations according to this protocol.

Comparison of Methods A, B, and C

The main objective of this work was to compare the three previously described standardized procedures for the estimation of reproducibility standard deviation. The usual statistical tests could not be used, since they cannot apply to the comparison of variances estimated on the same sample from the same population. A graphical comparison has been performed, and the consistency assessed according to a pragmatic criterion: the difference between two standard deviations has been considered as acceptable up to $0.1 \log_{10}$ (based on the usual expression of standard deviation to one significant figure), or up to 10 % for the relative standard deviations.

Results

Data distribution

In a preliminary step, the nine distributions of enumeration results, one per combination <food type×contamination level>, were investigated, to check their normality before \log_{10} transformation, or log-normality after \log_{10} transformation. The results are presented in Table 1. On the basis of the complete data sets, the log-normality was rejected ($P < 5\%$) for eight combinations out of nine. **Note that only the P -value obtained for <meat×low> (0.04 at $\alpha = 0.05$), was found not significant when the multiple comparisons procedures were applied.** The normality **of the complete data sets** was rejected for five combinations out of nine.

As recommended by method A, outlying duplicate results (before and after \log_{10} transformation) were then identified by the Cochran and Grubbs for their standard deviation and their mean, respectively (see Table 1). After exclusion of outliers on log-transformed data, the log-normality was still rejected for four combinations out of nine. On the contrary, after exclusion of outliers on non log-transformed data, the normality was rejected for only one combination <egg powder x medium level> out of nine.

In conclusion, the hypothesis of normality, without \log_{10} transformation, appears to be more satisfactory than the hypothesis of log-normality.

Estimates of precision data

The procedures of methods A, B and C were applied. Statistical outliers, identified according to method A, were excluded from the calculations only for this method.

Table 1 gives the reproducibility standard deviations (s_R), calculated for each combination <food type×contamination level> according to the 3 methods, with and without \log_{10} transformation of the data. The repeatability standard deviations were not included in Table 1, since similar conclusions could be drawn, in terms of comparison between the different precision estimators.

The mean s_R per level, calculated on log-transformed data, are plotted in Figure 1.a. The mean relative standard deviations of reproducibility (RSD_R) per level, calculated on non \log_{10} transformed data, are plotted in Figure 1.b. RSD_R are obtained by dividing s_R by the mean level for method A, and by the median level for methods B and C.

With calculations on log-transformed data, it can be seen in Figure 1.a that the three approaches globally give consistent results for all combinations <food type x contamination level> of this inter-laboratory trial, except for <cheese x high>. The pragmatic criterion chosen for consistency is respected by all combinations, except 2: (i) clearly the combination <cheese x high>, for which the difference between the non robust standard deviation and a robust one (method C) is 0.84, and (ii) the combination <egg x low>, for which the difference between the non robust standard deviation and a robust one (method C) is slightly over 0.1 (0.13).

Discussion

Log-transformation of the data

As mentioned in the introduction, the usual approach developed by AOAC or ISO 16140 consists in previously transforming the enumeration data into \log_{10} . Reasons for this transformation are discussed in this section.

The logarithmic transformation is often justified by the hypothesis that it would enable to approximate a Gaussian distribution, i.e. that enumeration results would be log-normally distributed. However, this hypothesis is questionable both for theoretical considerations and experimental observation. From a theoretical point of view, errors associated with colony-count techniques have been studied for several decades^(xiv, xv). They mostly include errors of technical nature, such as pipetting errors, and distribution errors, due to sampling of the test portion in the food sample where the microorganisms have a given distribution, as well as to the sampling during the successive dilutions, and for the plating on the Petri dishes. Recently, Hedges^(xvi) has conducted a complete theoretical investigation, taking into account most of these errors, but neglecting others originating, for example, from the heterogeneity of the contamination between samples, or from the confirmation of 5 colonies randomly selected on the Petri dishes. One of the conclusions of this work was that enumeration results were expected to be normal, without any need for a logarithmic transformation. From a more practical point of view, in an experimental study of naturally contaminated products, Corradini *et al.*^(xvii) observed that industrial records of microbial counts in commercial products (e.g. anaerobes in dairy-based products, yeasts in frozen apple concentrate...) were not really right-skewed, and could often be described satisfactorily by a normal distribution, without any logarithmic transformation. Finally, in the present study, the normality of non-transformed enumeration results was also accepted in most cases (see before).

Another important justification of the \log_{10} transformation is to gain homoscedasticity, i.e. to obtain repeatability and reproducibility standard deviations globally stable over the different contamination levels. Enumeration results are indeed obtained by multiplying a total count on plates by a dilution factor. The enumeration results obtained in such a way are expected to follow a distribution, whose mean is the mean of the plate counts' distribution multiplied by the dilution factor, and whose variance is the variance of the plate counts' distribution multiplied by the squared dilution factor. Thus, the standard deviation of the enumeration results is the standard deviation of the plate counts multiplied by the dilution factor, and is then proportional to the contamination level.

That is why either the \log_{10} -transformation or the use of the relative standard deviation are necessary for comparison purposes between contamination levels.

In the example that we have chosen, with \log_{10} transformation of the data, it can be seen in Figure 1.a that the reproducibility standard deviation remains globally stable over the different contamination levels (except from the combination <cheese x high> that will be discussed later), with a slight tendency to decrease: from 0.40-0.45 (depending on the calculation's method) at a level of 1.95 \log_{10} CFU/g (cheese/low) to 0.32-0.35 at a level of 4.34 (meat/high), and even down to 0.17-0.18 at a level of 4.39 (egg/high), but the reproducibility standard deviation appears to be particularly low for that food type.

As anticipated, the same shape is exactly observed for the plot of the relative standard deviations of reproducibility, without logarithmic transformation of the data (Figure 1.b). It

reflects the similarity of the effect to calculate a standard deviation of \log_{10} transformed data, and to take the relative value of this standard deviation.

It can be finally said that the logarithmic transformation enables as expected to stabilize the variance over the contamination levels. In other words, the reproducibility, expressed either in log units or in relative terms, is constant over the range studied.

Robust versus non-robust precision estimates

Interest of robust estimates

From a theoretical point of view, the “classical” Method A, with estimators using the mean, is based on the assumption that errors are normally distributed. This assumption is often sound when referring to the central limit theorem, which well applies to measurement errors lying in an acceptable range, since they result from sums of many small independent errors of equivalent weight. On the contrary, this assumption cannot be applied to outliers, which usually result from single large errors (11).

As already noted in the introduction, robust statistics *accommodate* with outliers, instead of *rejecting* them. This represents an advantage from several points of view (v, xii):

- (i) There is no general agreement on the most appropriate procedure for rejecting outliers. For example, ISO 5725-2 states that the risk level for rejecting an outlier using either Cochran or Grubbs tests must be less or equal to 1%, whereas AOAC (iii) takes the 2.5% quantile values as critical values. Moreover, the ISO 5275-2 and AOAC sequences of application of both these tests are different, possibly leading to different conclusions.
- (ii) The decision to what extent test results shall be excluded is not always easy to make, and introduces a degree of subjectivity. For example, a laboratory may give one replicates' mean classified as outlier for a given material, but three other means classified as straggler: in this case, shall all the results of this laboratory be excluded, or only the outlying set of data? It is rather usual to have data being at the boundary between outlier and straggler, so that the organizer of the trial is to take a decision of exclusion/non exclusion that will affect the result of the precision characteristics' calculation.
- (iii) The classical approach assumes that it is reasonable to establish a common value for the repeatability standard deviation for all laboratories having taken part into the trial. This may not be verified in practice, and in such cases when some laboratories display poorer repeatability than the others (especially when these laboratories have a too limited experience of the method tested), the use of robust statistics would be preferable.
- (iv) Extensive exclusion of data as outliers may seriously underestimate the true variability of the method itself when used in practice.

ISO 16140 clearly states that data shall be excluded from calculation only for technical reported reasons (samples having reached the laboratory in a bad state, with a temperature abuse during transportation, deviations to the protocol reported by the participating laboratory...).

Consistency between robust and non-robust estimates

Using the example of the inter-laboratory trial that was investigated, with calculation on log-transformed data, **a generally good consistency can be observed in Figure 1.a between the three robust and non-robust estimates. Our pragmatic criterion for consistency** is respected in a clear majority of cases: seven out of nine combinations. This is a positive aspect for the use of robust statistics, since it has been objected that precision characteristics derived with robust statistics may hide the true variation of a method, by down weighting the extreme values.

It is reported (xii) that robust estimators can only be applied when the data of an inter-laboratory study display only a small proportion of outliers. As a matter of fact, the estimators x^* and s^* (see Appendix, Method B) measure the mean and the standard deviation of data **in the center of the dataset, and, at the condition that these data form the majority**, such estimators represent consensus values near to the true values. In our example, this condition seems to be verified, since no statistical outliers have been identified for 5 combinations out of 9, and only 1 set of duplicates (out of 16 or 17) has been identified as outlier in each of the 4 combinations: cheese/medium, meat/high, egg/low and egg/high. For the 3 last levels, the duplicates are “obvious” outliers, since one of the duplicate result is null.

Let us consider the particular case of the combination cheese/high, where the value of the non robust standard deviation is clearly larger (1.15) than the ones of the robust procedures (mainly with method C (0.31), and to a lesser extent with method B (0.81)). For this combination, no statistical outlier has been identified, but the data are highly scattered (large variations between duplicates, and even more between laboratories). Here the drawback reported on robust statistics appears to be verified: their use underestimates the true variation of the method. Since most of the laboratories have obtained highly variable results (compared to the low and medium levels of the same food type), the quality of the experiment itself may be questioned (such as the samples used for the trial). It may even be appropriate to decide to not derive precision criteria in such conditions; a value of more than 1 \log_{10} for the reproducibility standard deviation is not satisfactory for an enumeration method which is intended to check the conformity of samples to a low limit of 100 CFU/g.

Without transformation into \log_{10} of the data, a consistency of the three approaches can be also observed in Figure 1.b for the relative standard deviations. In particular, the inconsistency for the combination cheese/high discussed before is no more encountered. But the use of our pragmatic criterion for consistency is not respected for 4 combinations, including the 3 combinations at low level: for the combination cheese/low, a difference of 11 % between the 2 robust values, for the combinations meat/low, egg/low and egg/medium, a difference of respectively 14 %, 15 % and 23 % between the non robust value and a robust value (method C). In 2 of these cases, the non robust standard deviation is lower than the robust ones, and in the 2 other cases, this is the contrary: no conclusion can be drawn on the effect of using robust or non robust estimates. According to our practical criterion, the consistency between non robust and robust estimates is therefore better on \log_{10} transformed data. It seems that the log-transformation tends to reduce the difference between the different estimates of standard deviation: it may be due to the reduction, by this transformation, of the dispersion of data.

Recommendations on robust estimates

The interest of using robust statistics for calculating the precision criteria of a method can therefore be confirmed, given the illustration of the example, and the advantages recalled

before. An additional advantage is that the median is not influenced by the convention chosen to report the “null” results, which are actually below the limit of enumeration of the method (reporting conventions other than “0” are possible).

Meanwhile, it should be noted that, unlike some robust statistics, the robust dispersion estimators used in ISO 5725-5 and ISO 16140 (Median Absolute Deviation –MAD, and Rousseeuw’s S_n) are heavily dependent on the normality of the distribution: they are robust against outliers (i.e. a good resistance against “polluting” data) but they lack robustness (or efficiency) against non normal distributions, such as a χ^2 one (simulations of Wilrich P., non published).

Moreover, it can be discussed why two different robust approaches have been retained in ISO 5725-5 and in ISO 16140. Rousseeuw and Croux (xiii) discussed the drawbacks of the MAD. The MAD has a low efficiency at Gaussian distributions (asymptotic efficiency of 37 % against 64 % for the median), and requires the distribution of data to be symmetric (it estimates at first a central value, the median, then gives an equal weight to positive and negative deviations to this central value).

ISO 16140 has adopted the Rousseeuw’s first alternative to the MAD as a robust scale estimator, S_n . S_n is based on a typical distance between data, still valid for asymmetric distributions: this property is particularly important for microbiological counts, whose distribution is often asymmetric. S_n has the highest possible finite sample breakdown point (50 %) for a scale estimator and a better asymptotic efficiency (58 %) than the MAD. But the influence function of S_n has discontinuities.

Rousseeuw proposed a second alternative to the MAD, Q_n , better than S_n : Q_n is equally suitable for asymmetric distributions, has a 50 % breakdown point, but has a smooth influence function and a very high asymptotic efficiency (82 %). It can therefore be questioned why the authors of ISO 16140 have not chosen Q_n instead of S_n .

A possible use of the reproducibility standard deviations

The justification to calculate standard deviations on log-transformed data have been already discussed (normality of distribution of enumeration results, stabilization of variance between contamination levels), as well as the effect of log-transformation on the comparison between robust and non robust calculations of the reproducibility standard deviation.

The standard deviations expressed in log and in non log units have been further compared through a possible use of standard deviation, the estimation of **measurement uncertainty**. **According to the principles of GUM^(xxii), a laboratory can associate an interval of measurement uncertainty to an individual result z . According to ISO/TS 21748^(xxiii), when the result has been obtained using a given enumeration method characterized by a reproducibility standard deviation s_R , the uncertainty interval can be expressed, under certain conditions and with a 95 % confidence level, as : $z \pm 2 s_R$. It corresponds to the interval of expanded measurement uncertainty.**

If the reproducibility standard deviation s_R is calculated without logarithmic transformation, and if the individual result x is expressed in CFU/g or mL without logarithmic transformation, then the **uncertainty interval** for x is $[x - 2 s_R ; x + 2 s_R]$ with a 95 % confidence level.

If the reproducibility standard deviation s_R is calculated after logarithmic transformation, and denoting $y = \log_{10} x$, then the **uncertainty interval** for y is $[y - 2 s_R ; y + 2 s_R]$, and the **uncertainty interval** with a 95 % confidence level for x is:

$$\left[10^{y-2s_R}; 10^{y+2s_R}\right] = \left[\frac{x}{10^{2s_R}}; x \cdot 10^{2s_R}\right] \quad (1)$$

In Table 2, the **uncertainty intervals** for the individual results x of laboratory 1, duplicate 1 (as an example), have been calculated according to these two different ways. The comparison has been undertaken for only one robust calculation of s_R (method C).

The anti-log and non-log **uncertainty intervals** appear to be of the same order, which confirms that the two approaches are not contradictory. However, the anti-log **uncertainty intervals** are shifted towards higher values and almost all the theoretical lower limits of the non-log **uncertainty intervals** are negative (except two), at the difference of the anti-log **uncertainty intervals**. From a biological point of view, any positive enumeration result clearly proves that the 'true' contamination level of the product is neither negative nor null: null or negative lower confidence limits are thus unexpected.

Therefore, **when log-normality of the data's distribution can be accepted**, an additional advantage for the log-transformation is that it automatically excludes null and negative contamination levels and then leads to realistic lower limits of the **uncertainty intervals**. However, when calculations are performed without log-transformation, it is still possible to replace by convention negative lower limits by 0, as in Table 2.

Conclusion

From the example of the inter-laboratory study on enumeration of *Listeria monocytogenes*, and given the reported advantages of robust statistics, it can be concluded that robust calculation of precision characteristics for microbiological quantitative methods is preferred. The ISO 16140 procedure seems better suited to microbiological counts, but another robust scale estimator can be recommended (Q_n instead of S_n).

However, the use of robust statistics should not result in failure to carefully consider the data before calculation: a graphical representation of the data, as well as a statistical identification of outliers are still needed, in order to investigate with the participants of the inter-laboratory study the possible reasons for their results to deviate, and to correct aberrant data, if appropriate (if for example an inconsistent result can be traced back to a transcription mistake).

This work has enabled also to question the usual logarithmic transformation of data before conducting calculations. In this study, calculations based on non-transformed data appeared to be a satisfactory alternative, at the condition to express standard deviations in relative terms, and to replace by convention lower limits of **uncertainty intervals** by zero. However, this study cannot be generalized, given the homogeneity of the samples used in the example. The log-normality or normality of the microbial counts should not be taken for granted, and should ideally be investigated on a case-by-case basis.

Acknowledgements

We would like to thank Mr Philip Feldsine and Mrs Maritta Ko for their valuable improvements to the draft.

Appendix: Details of the calculations of repeatability/reproducibility standard deviations, according to Methods A, B and C

Method A

(a) Tests for outliers

(i) Cochran test

For a set of p standard deviations s_i , all calculated from the same number of replicates (n), the statistic C of the Cochran test is:

$$C = \frac{s_{\max}^2}{\sum_{i=1}^p s_i^2}$$

Where s_{\max} is the largest standard deviation of the set.

If C is less or equal to the critical value of the statistic at 5% (see Cochran tables, such as in (2)), s_{\max} is considered as correct. If C is larger than the critical value at 5% but less or equal to the critical value at 1%, s_{\max} is considered as isolated (or doubtful). If C is larger than the critical value at 1%, s_{\max} is considered as outlier.

This test can be repeated on the remaining values after having excluded the outlying standard deviation, but a too repeated use may conduct to excessive rejections.

(ii) Grubbs test

Single-value test

For a set of means of replicates \bar{x}_i , $i = 1$ to p , ordered in increasing rank, the statistic G_p of the Grubbs test is calculated in order to test whether the largest value \bar{x}_p is an outlier:

$$G_p = (\bar{x}_p - \bar{\bar{x}}) / s$$

Where: $\bar{\bar{x}} = \frac{1}{p} \sum_{i=1}^p \bar{x}_i$ and $s = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (\bar{x}_i - \bar{\bar{x}})^2}$

For testing the significance of the smallest value \bar{x}_1 , the statistic G_1 of the Grubbs test is calculated:

$$G_1 = (\bar{\bar{x}} - \bar{x}_1) / s$$

If the statistic G_p or G_1 is less or equal to its critical value at 5% (see Grubbs tables, such as in (2)), \bar{x}_p or \bar{x}_1 is considered as correct. If the statistic G_p or G_1 is between its critical value at 5% and its critical value at 1%, \bar{x}_p or \bar{x}_1 is considered as isolated (or doubtful). If the statistic G_p or G_1 is larger than its critical value at 1%, \bar{x}_p or \bar{x}_1 is considered as an outlier.

Pair-value test

For testing whether the two largest means of replicates \bar{x}_p , \bar{x}_{p-1} are outliers, the statistic G of the Grubbs test is calculated:

$$G = s_{p-1,p}^2 / s_0^2$$

Where: $s_0^2 = \sum_{i=1}^p (\bar{x}_i - \bar{\bar{x}})^2$, $s_{p-1,p}^2 = \sum_{i=1}^{p-2} (\bar{x}_i - \bar{\bar{x}}_{p-1,p})^2$, and $\bar{\bar{x}}_{p-1,p} = \frac{1}{p-2} \sum_{i=1}^{p-2} \bar{x}_i$

For testing whether the two smallest means of replicates \bar{x}_1, \bar{x}_2 are outliers, the statistic G of the Grubbs test is calculated:

$$G = s_{1,2}^2 / s_0^2$$

Where: $s_{1,2}^2 = \sum_{i=3}^p (\bar{x}_i - \bar{\bar{x}}_{1,2})^2$, and $\bar{\bar{x}}_{1,2} = \frac{1}{p-2} \sum_{i=3}^p \bar{x}_i$

The interpretation is equivalent to the single-value test.

Use of the Grubbs tests

The single-value test is applied to the means of replicates. If it shows that a mean is an outlier, after excluding it, the single-value test is repeated to the other extreme mean. If the single-value test doesn't indicate any outlying mean, the paired-value test is applied.

(b) Calculation of repeatability and reproducibility standard deviations

The general mean $\bar{\bar{x}}$ for a given level is calculated:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^p n_i \bar{x}_i}{\sum_{i=1}^p n_i}$$

Where: n_i is the number of replicates for the laboratory i , \bar{x}_i is the mean of replicates of the laboratory i , and p is the number of participating laboratories.

The repeatability variance is calculated as follows:

$$s_r^2 = \frac{\sum_{i=1}^p (n_i - 1) s_i^2}{\sum_{i=1}^p (n_i - 1)}$$

Where: s_i^2 is the intra-laboratory variance of laboratory i , $s_i^2 = \frac{\sum_{j=1}^{n_i} (x_j - \bar{x}_i)^2}{n_i - 1}$

The inter-laboratory variance s_L^2 is calculated as follows:

$$s_L^2 = \frac{s_d^2 - s_r^2}{n}$$

Where: $s_d^2 = \frac{1}{p-1} \sum_{i=1}^p n_i (\bar{x}_i - \bar{\bar{x}})^2$ (variance of the means of replicates),

$$\text{and } \bar{n} = \frac{1}{p-1} \left[\sum_{i=1}^p n_i - \frac{\sum_{i=1}^p n_i^2}{\sum_{i=1}^p n_i} \right]$$

The reproducibility variance is then deduced:

$$s_R^2 = s_r^2 + s_L^2$$

Method B

(a) Algorithm S

This algorithm is applied to the intra-laboratory variances and provides a robust combined value of standard deviation (the repeatability standard deviation).

For a set of p intra-laboratory standard deviations s_i , ordered in increasing rank, s^* designates their combined robust value, and ν the number of degrees of freedom associated to each s_i (for n replicates, $\nu = n-1$).

An initial value of s^* is calculated as follows:

$$s^* = \text{MED } \{s_i\} \text{ (median of the } \{s_i\}), \text{ for } i = 1 \text{ to } p$$

Then the value of s^* is updated as follows.

For each s_i ($i=1$ to p), $s_i^* = \psi$ if $s_i > \psi$, or s_i otherwise,

where: $\psi = \eta s^*$ (η is a limit factor depending on ν , see Annex B in (5))

The updated value of s^* is calculated as follows:

$$s^* = \xi \sqrt{\sum_{i=1}^p (s_i^*)^2 / p}, \text{ where } \xi \text{ is an adjustment factor (see Annex B in (5))}$$

The robust estimator s^* can be deduced from an iterative calculation, until the change of s^* is small from an iteration to the following one. It can also be done, as we have chosen, by expressing s^* as follows:

$$(s^*)^2 = (\xi^2 / p) \times \left[\sum' (s_i^*)^2 + u_U \times (\eta s^*)^2 \right]$$

Where \sum' is the summation of the $s_i \leq \psi$, and u_U is the number of the $s_i > \psi$.

We have at first followed an iterative method to get an approximate value for s^* , and then used the above equation to get the exact value.

The repeatability standard deviation s_r is equal to the final value of s^* .

(b) Algorithm A

The algorithm A is applied to the means of replicates, and provides a robust value of the standard deviation of the means of replicates (thus the reproducibility standard deviation).

For a set of p means of replicates \bar{x}_i , ordered in increasing rank, \bar{x}^* and s^* designate the robust mean and the robust standard deviation of these data.

The initial values of x^* and s^* are calculated as follows:

$$x^* = \text{MED} \{ \bar{x}_i \}$$

$$s^* = 1.483 \text{ MED} \{ |\bar{x}_i - x^*| \}$$

Where 1.483 is a scale factor used to obtain a correct answer for normally distributed data ($\approx \sqrt{v/\chi^2}$ with v the number of degrees of freedom of the replicates ($v = n-1$), and χ^2 the chi-square for $p = 0.5$; 0.455 here, see (11)).

$1.483 \text{ MED} \{ |\bar{x}_i - x^*| \}$ is called the median absolute deviation, or MAD.

The values of x^* and s^* are updated as follows.

For each \bar{x}_i ($i = 1$ to p), $x_i^* = x^* - \varphi$ if $\bar{x}_i < x^* - \varphi$; $x^* + \varphi$ if $\bar{x}_i > x^* + \varphi$; \bar{x}_i otherwise,

where: $\varphi = 1.5 s^*$ (1,5 is a largely used value of the cut-off factor c , see (11)).

The new values of x^* and s^* are calculated as follows:

$$x^* = \sum_{i=1}^p x_i^* / p$$

$$s^* = 1.134 \sqrt{\sum_{i=1}^p (x_i^* - x^*)^2 / (p-1)}$$

Where: $1.134 \approx 1/\sqrt{0.778}$; 0.778 is a factor (β) chosen to obtain a correct answer for normally distributed data (11).

The robust estimators x^* and s^* can be deduced from iterative calculations, until the change of x^* and s^* are small from an iteration to the following one. It can also be done, as we have chosen, by expressing x^* and s^* as follows:

$$x^* = x' + 1.5 \times (u_U - u_L) s^* / (p - u_L - u_U)$$

$$(s^*)^2 = (p - u_L - u_U - 1) \times (s')^2 / \left[(p-1) / 1.134^2 - 1.5^2 (pu_L + pu_U - 4u_L u_U) / (p - u_L - u_U) \right]$$

Where: u_L is the number of $x_i < x^* - \varphi$, u_U is the number of $x_i > x^* + \varphi$, x' and s' are the mean and standard deviation of the $(p - u_L - u_U)$ data \bar{x}_i such as $|\bar{x}_i - x^*| \leq \varphi$.

We have at first followed an iterative method to get approximate values for x^* and s^* , and then used the above equations to get the exact values.

The standard deviation of the means of replicates s_d is equal to the final value of s^* . The inter-laboratory standard deviation s_L , as well as the reproducibility standard deviation s_R are then calculated as in the non-robust approach according to ISO 5725-2.

Method C

(a) Repeatability standard deviation

A robust estimator of the repeatability standard deviation is directly derived from the median of the intra-laboratory standard deviations s_i ($\text{MED}\{s_i\}$) as follows:

$$s_r = 1.483 \text{ MED}\{s_i\}$$

Where 1.483 is the same factor than in algorithm A of ISO 5725-5 approach.

(b) Reproducibility standard deviation

A robust estimator of the standard deviation of the means of replicates s_d is derived from the Rousseeuw's scale estimator S_n (13).

The scale estimator S_n consists in two successive phases of mediating: (i) from each of the p values of the means of replicates \bar{x}_i , the median of its ($p-1$) absolute differences to the other values \bar{x}_j is calculated; (ii) the median of these p deviate medians is then calculated.

$$S_n = \text{MED}_i \left\{ \text{MED}_j \left| \bar{x}_i - \bar{x}_j \right| \right\}$$

The standard deviation of the means of replicates s_d is then deduced as follows:

$$s_d = 1.1926 S_n$$

Where: 1,1926 is derived from a non-central χ^2 distribution centered on 0.75.

The inter-laboratory standard deviation s_L , as well as the reproducibility standard deviation s_R are then calculated as in the non-robust approach according to ISO 5725-2.

References

- (i) International Organization for Standardization (1994) *Accuracy (trueness and precision) of measurement methods and results. Part 1. General principles and definitions*, Standard ISO 5725-1, Geneva, Switzerland
- (ii) International Organization for Standardization (1994) *Accuracy (trueness and precision) of measurement methods and results. Part 2. Basic method for the determination of repeatability and reproducibility of a standard measurement method*, Standard ISO 5725-2, Geneva, Switzerland

-
- (iii) AOAC INTERNATIONAL Guidelines for Collaborative Study Procedures To Validate Characteristics of a Method of Analysis (1995) *J. AOAC Int.* **78**(5), 143A-160A
- (iv) Feldsine, P.T., Abeyta C., Andrews W.H., AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Qualitative and Quantitative Food Microbiological Official Methods of Analysis (2002) *J. AOAC Int.* **85**(5), 1187-1200
- (v) International Organization for Standardization (1998) *Accuracy (trueness and precision) of measurement methods and results. Part 5. Alternative methods for the determination of the precision of a standard measurement method*, Standard ISO 5725-5, Geneva, Switzerland
- (vi) International Organization for Standardization (2003) *Microbiology of food and animal feeding stuffs. Protocol for the validation of alternative methods*, Standard ISO 16140, Geneva, Switzerland
- (viii) International Organization for Standardization (1998) *Microbiology of food and animal feeding stuffs. Horizontal method for the detection and enumeration of Listeria monocytogenes. Part 2: Enumeration method*, Standard ISO 11290-2, Geneva, Switzerland
- (ix) Scotter, S. L., Langton, S., Lombard, B., Lahellec, C., Schulten, S., Nagelkerke, N., in't Veld, P. H., Rollier, P. (2001) *Int. J. Food Microbiol.* **70**, 121-129
- (x) **Curran-Everett, D. (2000) *Am J Physiol Regul Integr Comp Physiol.* **279**, R1-R8.**

-
- (^{xi}) Analytical Methods Committee, Royal Society of Chemistry (1989) *Analyst* **114**, 1697-1698
- (^{xii}) Analytical Methods Committee, Royal Society of Chemistry (1989) *Analyst* **114**, 1699-1702
- (^{xiii}) Rousseeuw, P.J., Croux, C. (1993) *J. Am. Stat. Assoc.* **88**, 1273-1283
- (^{xiv}) Hedges, A. J. (1967). *Biometrics.* **23**, 158-159
- (^{xv}) Jarvis, B. (1989). *Statistical aspects of the microbiological analysis of the foods. In: Progress in industrial microbiology*, M.E. Bushell (Ed.), Elsevier, Netherlands.
- (^{xvi}) Hedges, A. J. (2002), *Int. J. Food Microbiol.* **76**, 207-214
- (^{xvii}) Corradini, M. G., Normand, M. D., Nussinovitch, A., Horowitz, J. Peleg, M. (2001) *J. Food Protec.* **64**, 674-681
- (^{xviii}) Hedges, A. J. (1967). *Biometrics.* **23**, 158-159
- (^{xix}) Jarvis, B. (1989). *Statistical aspects of the microbiological analysis of the foods. In: Progress in industrial microbiology*, M.E. Bushell (Ed.), Elsevier, Netherlands.
- (^{xx}) Hedges, A. J. (2002), *Int. J. Food Microbiol.* **76**, 207-214
- (^{xxi}) Corradini, M. G., Normand, M. D., Nussinovitch, A., Horowitz, J. Peleg, M. (2001) *J. Food Protec.* **64**, 674-681
- (^{xxii}) International Organization for Standardization (1995) *Guide to the expression of uncertainty in measurement*, Geneva, Switzerland.

^(xxiii) **International Organization for Standardization (2004) *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation*, Technical Specification ISO/TS 21748, Geneva, Switzerland**

List of tables

Table 1. Results of the inter-laboratory study for *L. monocytogenes* enumeration (in CFU/g) – Normality tests - Calculations of reproducibility standard deviations according to methods A (ISO 5725-2), B (ISO 5725-5) and C (ISO 16140), on \log_{10} /non \log_{10} transformed results

Table 2. Comparison of **uncertainty intervals around the results of laboratory 1, duplicate 1, obtained on \log_{10} / non \log_{10} transformed data**

List of figures

Figure 1 : Standard deviations of reproducibility s_R /Relative standard deviations of reproducibility (RSD_R) for the inter-laboratory study on enumeration of *Listeria monocytogenes*, calculated according to method A (\diamond , ISO 5725-2), method B (\square , ISO 5725-5) and method C(Δ , ISO 16 140)

1.a s_R on \log_{10} transformed data

1.b RSD_R on non \log_{10} transformed data (log scale for the x-axis)

Table 2. Results of the inter-laboratory study for *L. monocytogenes* enumeration (in CFU/g) – Normality tests - Calculations of reproducibility standard deviations according to methods A (ISO 5725-2), B (ISO 5725-5) and C (ISO 16140), on log₁₀/non log₁₀ transformed results

	Fresh Cheese Curd						Minced Meat						Dried Egg Powder						
	Low		Medium		High		Low		Medium		High		Low		Medium		High		
	# 1 ^a	# 2 ^a	# 1	# 2	# 1	# 2	# 1	# 2	# 1	# 2	# 1	# 2	# 1	# 2	# 1	# 2	# 1	# 2	
Lab 1	200	227	1909	1636	14545	30454	105	536	1182	1091	12273	15909	350	409	1864	818	12000	23182	
Lab 2	9	182	909	677	6364	23182	432	400	2864	4500	31364	15909	250	182	773	6818	7273	15455	
Lab 4	250	182	2250	1350	18182	21500	841	1095	4273	4591	54091	33182	25	310	818	909	22727	25909	
Lab 5	<LoE ^b	23	7000 ^c	1200 ^c	<LoE	5000	405	86	1600	2650	36000	25000	330	255	1500	1700	31500	21500	
Lab 9	32	45	600	1500	455	3636	223	364	4000	2300	10000	35909	427	282	3227	2000	33182	32273	
Lab 10	82	123	1364	1636	24091	20000	427	514	545	1364	19545	16364	136	277	1455	2136	20909	18636	
Lab 11	45	9	45 ^d	500 ^d	909	455	691	327	2636	3773	35909	32273	377	205	4136	3045	23636	20000	
Lab 12	65	75	2000	3500	17000	14000	735	430	1850	4350	4500	28000	265	90	1850	1800	26000	18000	
Lab 13	77	227	1150	1850	17273	30455	573	827	3182	5910	85000 ^f	85000 ^f	2200 ^h	40 ^h	955	2227	27727	33182	
Lab 14	77	109	1500	1545	33636	20000	323	295	4227	3773	6818	14545	218	400	1591	4409	32273 ^g	<LoE ^g	
Lab 15	32	168	1273	1227	21818	17727	305	441	3227	5364	21818	31818	250 ^g	<LoE ^d	9136	1500	36364	18182	
Lab 16	55	86	500	727	<LoE	2273	736	673	5136	3545	36818 ^g	<LoE ^g	268	582	4091	2727	45909	31364	
Lab 17	145	59	1727	500	11818	455	1818 ^h	2909 ^h	7909	773	90455 ^f	100000 ^f	-	-	-	-	-	-	
Lab 18	- ^e	-	-	-	-	-	609	473	4864	3636	19545	39545	105	136	2182	1500	27727	26818	
Lab 19	336	205	2409	1909	26363	34545	364	50	591	3545	16364	11818	114	105	2000	727	25455	22727	
Lab 20	200	255	2136	2182	21364	30909	682	1005	7045	4636	23636	15000	523	100	2682	1409	47727	41818	
Lab 22	177	205	1182	955	24545	45000	418	718	4045	6682	33182	20909	327	168	2545	2409	27273	35000	
Lab 24	-	-	-	-	-	-	105	536	1182	1091	12273	15909	350	409	1864	818	12000	23182	
<i>P</i>	Shapiro Wilks normality test on log ₁₀ transformed results: <i>P</i> for the complete data set / <i>P</i> for the data set, outliers excluded																		
	2.10 ⁻⁴	2.10 ⁻⁴ / 0.28				9.10 ⁻⁹	0.04	5.10 ⁻³				2.10 ⁻¹² / 0.45		2.10 ⁻⁶ / 0.11		0.23	2.10 ⁻¹⁴ / 0.15		
<i>P</i>	Shapiro Wilks normality test on non-log ₁₀ transformed results: <i>P</i> for the complete data set / <i>P</i> for the data set, outliers excluded																		
	0.06	4.10 ⁻⁷ / 0.60				0.08	2.10 ⁻⁸ / 0.56				0.25	3.10 ⁻⁶ / 0.44		2.10 ⁻¹¹ / 0.81		5.10 ⁻⁷	0.88		
	Calculations of <i>s_R</i> on log ₁₀ transformed results																		
Robust meanⁱ	1.95	3.13				3.90	2.66				3.50	4.34		2.32	3.31			4.39	
<i>s_R</i> (A) ^j	0.40	0.24				1.15	0.34				0.30	0.32		0.43	0.27			0.17	
<i>s_R</i> (B) ^k	0.44	0.27				0.81	0.30				0.26	0.32		0.34	0.25			0.18	
<i>s_R</i> (C) ^l	0.45	0.25				0.31	0.27				0.24	0.35		0.30	0.24			0.17	
	Calculations of <i>s_R</i> on non-log ₁₀ transformed results																		
Robust meanⁱ	123	1497				16811	524				3624	25709		261	2320			25646	
<i>s_R</i> (A) ^j	89	729				12299	252				1845	12185		141	1890			10200	
<i>s_R</i> (B) ^k	95	847				13844	302				1826	14640		169	1358			9552	
<i>s_R</i> (C) ^l	81	846				13618	334				1819	15210		187	1294			10274	

^a # 1 or 2 = duplicate 1 or 2.

^b <LoE = less than the limit of enumeration (replaced by 0 CFU/g or 0 log(CFU/g) in calculations).

^c Duplicates identified (without log₁₀ transformation) by the Cochran test as outliers for their standard-deviation.

^d Duplicates identified (after log₁₀ transformation) by the simple Grubbs test as outliers for their mean.

^e - = the laboratory has not participated to the study for this material, or its results have been not taken into account for calculations.

^f Two series of duplicates identified (without log₁₀ transformation) by the double Grubbs test as outliers for their mean.

^g Duplicates identified (after log₁₀ transformation) by the Cochran and the simple Grubbs tests as outliers for both their standard-deviation and their mean.

^h Duplicates identified (without log₁₀ transformation) by the Cochran and the simple Grubbs tests as outliers for both their standard-deviation and their mean.

ⁱ According to method B

^j Reproducibility standard-deviation, according to method A

^k Reproducibility standard-deviation, according to method B

^l Reproducibility standard-deviation, according to method C

Table 2. Comparison of **uncertainty intervals around the results of laboratory 1, duplicate 1, obtained on log₁₀ / non log₁₀ transformed data**

Food type	Level	Results x of laboratory 1, duplicate 1	Uncertainty intervals^a on log ₁₀ transformed data, and transformed to antilog values	Uncertainty intervals^a on non log ₁₀ transformed data
Fresh cheese curd	Low	200	[25 ; 1585]	[38 ; 362]
	Medium	1909	[603 ; 6026]	[217 ; 3601]
	high	14545	[3467 ; 60256]]0 ; 41781]
Minced meat	Low	105	[30 ; 363]]0 ; 773]
	Medium	1182	[389 ; 3548]]0 ; 4820]
	high	12273	[2455 ; 61659]]0 ; 42693]
Dried egg powder	Low	350	[87 ; 1380]]0 ; 724]
	Medium	1864	[617 ; 5623]]0 ; 4452]
	high	12000	[5495 ; 26303]]0 ; 35548]

^a **Uncertainty intervals** = $x \pm 2s_R$ with s_R calculated according to method C

Figure 1.a

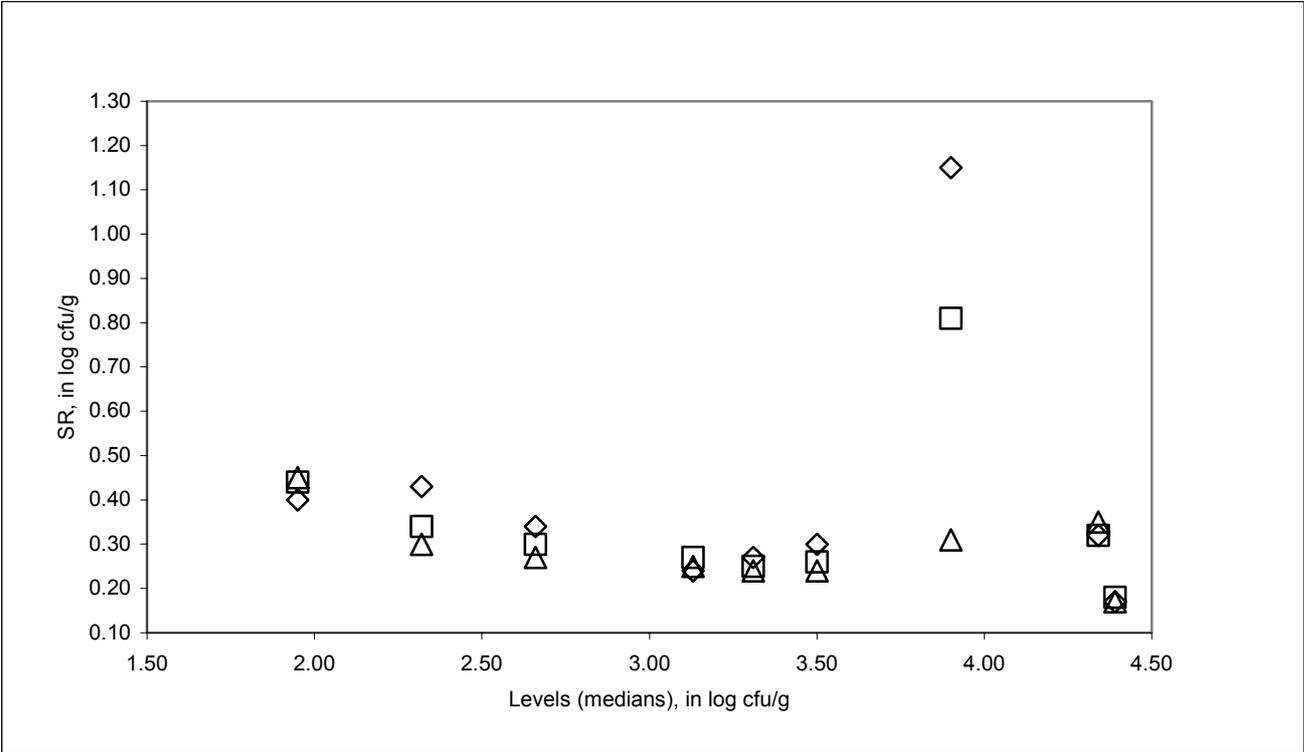
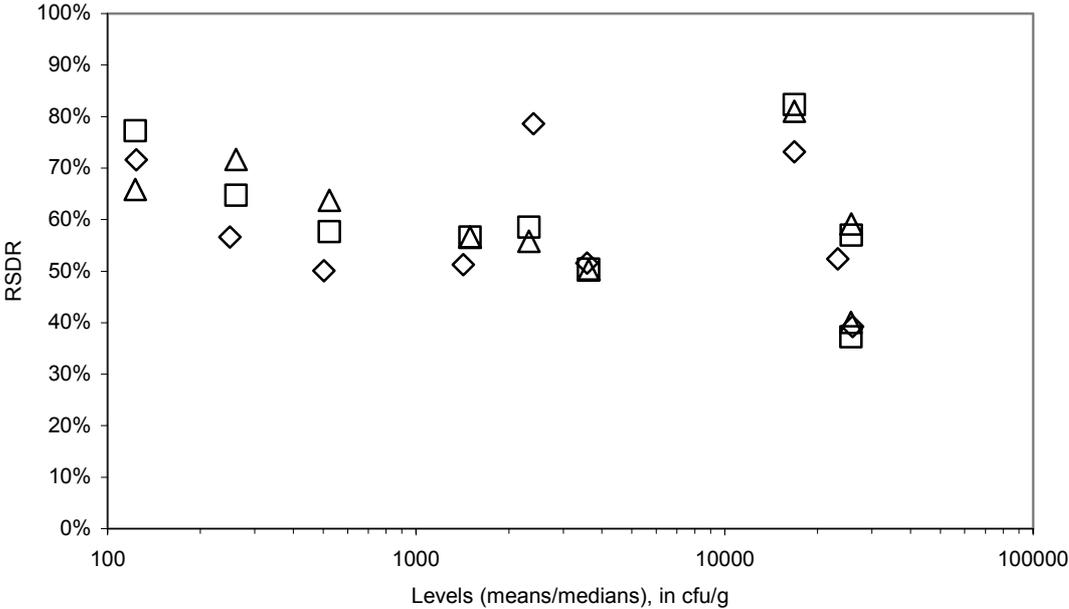


Figure 1.b



3.8 Annexe 8. Rapport de l'essai d'aptitude sur le dénombrement des coliformes

**Inter-laboratory proficiency trial on the enumeration of coliforms
in milk : final report of the CRL « Milk »**

Lafarge V.*et al.* Maisons-Alfort : AFSSA-LERQAP, 2000, 28 p.



***EU COMMUNITY REFERENCE
LABORATORY FOR MILK &
MILK PRODUCTS***

**INTERLABORATORY PROFICIENCY TRIAL
ON THE ENUMERATION OF COLIFORMS
IN MILK
April 2000**

Véronique LAFARGE, CRL « Milk », Unit HMPL
Véronique MALADEN, CRL « Milk », Unit HMPL
Bertrand LOMBARD, Co-ordinator of CRL « Milk »

INTRODUCTION

According to the EC Directive EEC/92/46, one of the tasks of the Community Reference Laboratory (CRL) for milk and milk products is to co-ordinate the application by National Reference Laboratories (NRLs) of the reference methods for the official control of milk & milk products by organising interlaboratory proficiency trials.

The interlaboratory trial on the enumeration of coliforms in milk was proposed and approved by NRLs at the NRLs Workshop held in Paris on 22 & 23 November 1999.

It was organised and conducted by the Unit HMPL of the CRL. The CRL had sent a circular letter on 17 February 2000 in order to launch this trial and to ask the NRLs to fill in a participation form.

The interlaboratory trial has been performed in April 2000.

LIST OF PARTICIPING LABORATORIES

Out of a total of 29 laboratories nominated as NRLs and contacted by circular letter of 17/02/00, 20 NRLs had accepted to take part in the trial and they received the contaminated samples (see annex n°1).

All these laboratories, except the one from Belfast, have analysed the samples. The NRL of Belfast has not received the samples.

The following NRLs have therefore not taken part in the trial: UK/ VLA-Surrey, IRL/ Limerick, IRL/ Cork, IRL/ Sligo, IRL/ Waterfork, S/ NFA-Uppsala, NL/ RIVM-Bilthoven, B/ Melle (DG 5 contrôle, Brusselsesteenweg 370A), B/ Gembloux (Station Laitière de l'Etat, chaussée de Namur 24).

SAMPLES

Eight samples of milk were sent to each NRL. Two samples of pasteurised milk were not contaminated by coliforms and six samples of raw milk were contaminated by coliforms at three levels of contamination ($5 \cdot 10^2 - 10^3$; $5 \cdot 10^3 - 10^4$ and $5 \cdot 10^4 - 10^5$ cfu/g) (see table n°1).

The samples were prepared by CECALAIT (Centre d'Etudes et de Contrôle des Analyses en industries LAITIÈRES, Poligny, France) and sent by express mail under refrigerated conditions on 11 April 2000. A bacteriostatic mixture was used to prevent bacteria growth and to stabilise inoculated flora (bacteriostatic effect disappears after dilution).

Stability and homogeneity testings of samples were performed by CECALAIT (see Annex n°2).

16 laboratories received the parcels within two days (see table n°2). Three laboratories (N°08, N°10 and N°11) did not indicate the date of arrival of the parcel.

All parcels arrived without damage.

Temperature of the samples at reception was below 10°C except for the laboratory N°03 (+13°C). The laboratory N°08 had not taken the temperature and the laboratory N°10 did not indicated the temperature of the samples at reception.

14 laboratories began the analyses on 13 April 2000. Three laboratories (N°03, N°04 and N°05) launched of the analyses on 14 April 2000. One laboratory (N°16) did not indicate the date of launching of the analyses.

METHOD

For this interlaboratory study, one method was prescribed the method ISO/ DIS 5541-1.

The aim of this study was to establish whether the NRLs apply the reference method and if they are able to apply it correctly.

15 laboratories used the ISO/ DIS 5541-1 method, 3 laboratories the IDF method (2 laboratories : IDF 73A and one laboratory : IDF 73B) and one laboratory the AFNOR routine method (AFNOR Standard V08-050). The IDF methods (IDF 73A or IDF 73B) and the ISO/DIS 5541-1 method are the same methods (see table n°3).

The laboratories which used the ISO method or the IDF method applied it correctly, except :

the laboratory N°15 which used only one Petri dish per dilution and confirmed colonies on BBL Enterotube,

the laboratory N°18 which incubated lactose bile brilliant green broth at 30°C for 48h.

Out of 18 laboratories which used the ISO method or the IDF method, 16 laboratories (N°01, N°02, N°03, N°04, N°05, N°06, N°07, N°08, N°10, N°11, N°12, N°13, N°14, N°17 , N°19, N°20) applied correctly the method.

RESULTS

The forms for the test results (see Annex 3) were sent to the NRLs by the circular letter of 17/03/00, which had asked to send back the forms completed by 15/05/00.

The laboratory N°09 (Belfast NRL) did not analyse the samples because it had not received them.

Analysis of results

a. Raw data produced by the laboratories

Enumeration of coliforms (in cfu/g and in log cfu/g) determined by the participants in blind duplicates for the four batches are reported in table n°4.

b. Choice of the assigned value

According to the ISO/CEI Guide 43-1, Annex A, one of the possibilities to determine the assigned value (or accepted reference value) is the consensus value derived from participating laboratories.

For each batch, this consensus value is taken as the median of the mean of duplicates of all participating laboratories (without excluding statistical outliers, since the median is a robust statistic, less sensitive to extreme values than the mean). See Table a.

Table a : Assigned values (X)

Batch	log X (cfu/g)	X (cfu/g)
A	3,10	$1,3 \cdot 10^3$
B	4,06	$1,1 \cdot 10^4$
C	4,95	$8,9 \cdot 10^4$

Distributions of laboratories data

A graphical representation of the data (duplicates and mean of duplicates) produced by the participating laboratories is given in Annex 4.

c. Statistical analysis

The statistical analysis was not performed according to the “traditional” procedure of the Standard ISO 5725, since a tool better adapted to microbiological quantitative determinations has been recently finalised : the draft Standard prEN ISO 16140 (final stage prior to publication). The precision data relating to quantitative methods (of enumeration of micro-organisms) are based on robust estimators (median instead of the mean), which display the advantages of being less sensitive to extreme values.

This was therefore not necessary to identify statistical outliers to calculate the global consensus value (see above), as well as the repeatability and reproducibility data (see under).

SUMMARY OF STATISTICS

Table b summarises the results of the overall statistical analysis which gives a good idea of the interlaboratory repeatability and reproducibility of the NRL network using the official reference method.

Table b : Statistical parameters of the proficiency testing “enumeration of coliforms in milk”, April 2000 (in log values)

Batch	n	X	SDr	RSDr %	r	SD_R	RSD_R %	R
B	19	3,10	0,04	1,18	0,1	0,06	1,82	0,16
C	19	4,06	0,04	0,98	0,11	0,06	1,60	0,18
D	19	4,95	0,04	0,72	0,1	0,09	1,74	0,24

Where :

- n is the number of laboratories having sent back results;
- X is the assigned value (here the median of the mean of duplicates);
- SDr & SD_R are the standard deviations for repeatability and reproducibility;
- RSDr & RSD_R are the relative standard deviations for repeatability and reproducibility;
- r & R are the repeatability and reproducibility values.

Globally, for a microbiological method enumerating micro-organisms, it can be said that these precision data are very satisfactory (r values of about 0,1 log and R values less than 0,3 in log). As a whole, the network of NRLs is therefore competent to enumerate coliforms in milk with using the official reference method.

INDIVIDUAL Z-SCORES

The individual z-score is one of the performance statistics recommended by the Guide ISO 43-1, Annex A.

For each laboratory i and each batch, an individual z-score was calculated as described in the Guide ISO 43-1 :

$$z = (m_i - X)/s$$

where :

m_i is the mean of duplicates of the laboratory i;

X is the assigned value (here the median of the mean of duplicates);

S is an appropriate measure of the variability (here taken as SD_R , the reproducibility standard deviation obtained by this interlaboratory study).

The values of individual z-scores are given in Table c.

Table c : Individual laboratory z-scores

Lab N°	Batch B	Batch C	Batch D
1	-1,00	-1,17	-1,11
2	-0,58	-1,42	-1,67
3	<u>7,50</u>	0,00	0,00
4	0,25	-0,33	-0,61
5	0,00	0,58	0,89
6	0,00	0,00	-0,11
7	0,25	0,25	0,11
8	-0,25	-1,25	-0,44
10	0,00	-0,33	-0,50
11	0,58	0,33	0,56
12	0,33	0,83	0,44
13	<u>-4,50</u>	<u>-2,33</u>	-1,67
14	0,58	-0,67	0,28
15	1,50	1,75	1,39
16	<u>-2,42</u>	0,00	<u>3,56</u>
17	1,17	0,58	1,00
18	0,00	-0,75	-0,67
19	-0,25	-0,67	0,00
20	<u>2,83</u>	<u>4,08</u>	-0,11

The performance of laboratories per batch can be evaluated by considering the absolute value of z (see Guide ISO 43-1, Annex A) :

$|z| \leq 2$: satisfactory

$2 < |z| \leq 3$: questionable

$|z| > 3$: unsatisfactory

In Table c, laboratories exhibiting questionable performances for a given batch have their z-score underlined in grey, or underlined in black for unsatisfactory performances.

CONCLUSION

Even if some laboratories showed z-score values unsatisfactory for one batch, it can be said that as a whole, given the very good precision data obtained, the network of NRLs is competent to enumerate coliforms in milk with using the official reference method.

Table n°1**Code and contamination of the samples**

Laboratory code	Sample code			
	Coliforms/ ml (target value)			
	Batch A 0 CFU / ml	Batch B $5 \cdot 10^2 - 10^3$ CFU/ ml	Batch C $5 \cdot 10^3 - 10^4$ CFU/ ml	Batch D $5 \cdot 10^4 - 10^5$ CFU/ ml
01	3 - 5	4 - 8	1 - 6	2 - 7
02	2 - 5	1 - 8	6 - 7	3 - 4
03	3 - 4	2 - 6	5 - 7	1 - 8
04	3 - 7	1 - 4	2 - 5	6 - 8
05	4 - 7	2 - 3	1 - 6	5 - 8
06	1 - 8	6 - 7	2 - 4	3 - 5
07	6 - 8	1 - 2	3 - 7	4 - 5
08	1 - 3	2 - 6	4 - 8	5 - 7
09	5 - 8	3 - 6	1 - 7	2 - 4
10	1 - 7	5 - 8	2 - 6	3 - 4
11	3 - 5	4 - 8	1 - 6	2 - 7
12	2 - 5	1 - 8	6 - 7	3 - 4
13	3 - 4	2 - 6	5 - 7	1 - 8
14	6 - 7	2 - 4	1 - 5	3 - 8
15	4 - 8	2 - 6	1 - 3	5 - 7
16	3 - 5	2 - 4	6 - 7	1 - 8
17	4 - 5	3 - 7	1 - 2	6 - 8
18	5 - 7	4 - 8	2 - 6	1 - 3
19	2 - 4	1 - 7	3 - 6	5 - 8
20	3 - 4	2 - 6	5 - 8	1 - 7

Table n°2
State of the samples at arrival

Laboratory code	Date of arrival of the parcel in the laboratory	State of the samples at arrival	Temperature of the samples at reception	Storage of the samples	Date of launching of the analyses
01	12 / 04 / 00	Good	+ 8°C	24 h / + 5°C	?
02	12 / 04 / 00	Good	+ 6°C	20 h / + 5°C	13 / 04/ 00
03	13 / 04 / 00	Good	+ 13°C	24 d / + 4°C	14 / 04/ 00
04	13 / 04 / 00	Good	+ 5°C	22 h / + 4°C	14 / 04/ 00
05	13 / 04 / 00	Good	+ 4°C	20 h / + 4°C	14 / 04/ 00
06	12 / 04 / 00	Good	+ 4°C	24 h / + 3°C	13 / 04/ 00
07	12 / 04 / 00	Good	+ 4°C	20 h / + 4°C	13 / 04/ 00
08	?	Good	not taken	? / + 4°C	13 / 04/ 00
09					
10	?	Good	?	24 h / + 5°C	13 / 04/ 00
11	?	Good	+ 1°C	24 h / + 4°C	13 / 04/ 00
12	12 / 04 / 00	Good	+ 8,3°C	20 h / + 4°C	13 / 04/ 00
13	12 / 04 / 00	Good	+ 5°C	18 h / + 4°C	13 / 04/ 00
14	12 / 04 / 00	Good	+ 5,5°C	21 h / + 3-5°C	13 / 04/ 00
15	12 / 04 / 00	Good	+ 7,9°C	28 h / + 6°C	13 / 04/ 00
16	11 / 04 / 00	Good	+ 7,4°C	48h / + 5°C	?
17	12 / 04 / 00	Good	+ 5,1°C	19 h / + 3,5°C	13 / 04/ 00
18	12 / 04 / 00	Good	+ 6°C	24 h / + 5°C	13 / 04/ 00
19	12 / 04 / 00	Good	+ 4,5°C	25 h / + 4°C	13 / 04/ 00
20	12 / 04 / 00	Good	+ 4°C	26 h / + 1,2°C	13 / 04/ 00

Table n°3

Details of the methods used

Laboratory code	Method used	INOCULATION – INCUBATION			CONFIRMATION	
		Nb of Petri dishes per dilution: two/ dilution	Inoculation agar VRBL agar	Incubation 30°C – 24h +/- 2h.	Confirmation medium Lactose bile brillant green broth	Incubation 30°C – 24h +/- 2h.
01	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
02	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
03	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
04	ISO/ DIS 5541 - 1	Y	Y	Y	N	N
05	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
06	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
07	ISO/ DIS 5541 - 1	Y	Y	Y	N	N
08	ISO/ DIS 5541 - 1	Y	Y	Y	N	N
10	ISO/ DIS 5541 - 1	Y	Y	Y	N	N
11	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
12	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
13	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
14	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
15	ISO/ DIS 5541 - 1	N*	Y	Y	N*	N*
16	AFNOR V08 – 050	Y	Y	Y	N	N
17	IDF 73A	Y	Y	Y	N	N
18	IDF 73B	Y	Y	Y	Y	N*
19	ISO/ DIS 5541 - 1	Y	Y	Y	Y	Y
20	IDF 73A	Y	Y	Y	N	N

*15 : - one Petri dish per dilution

*18 : - confirmation : incubation at 30°C – 48h

Table n°4

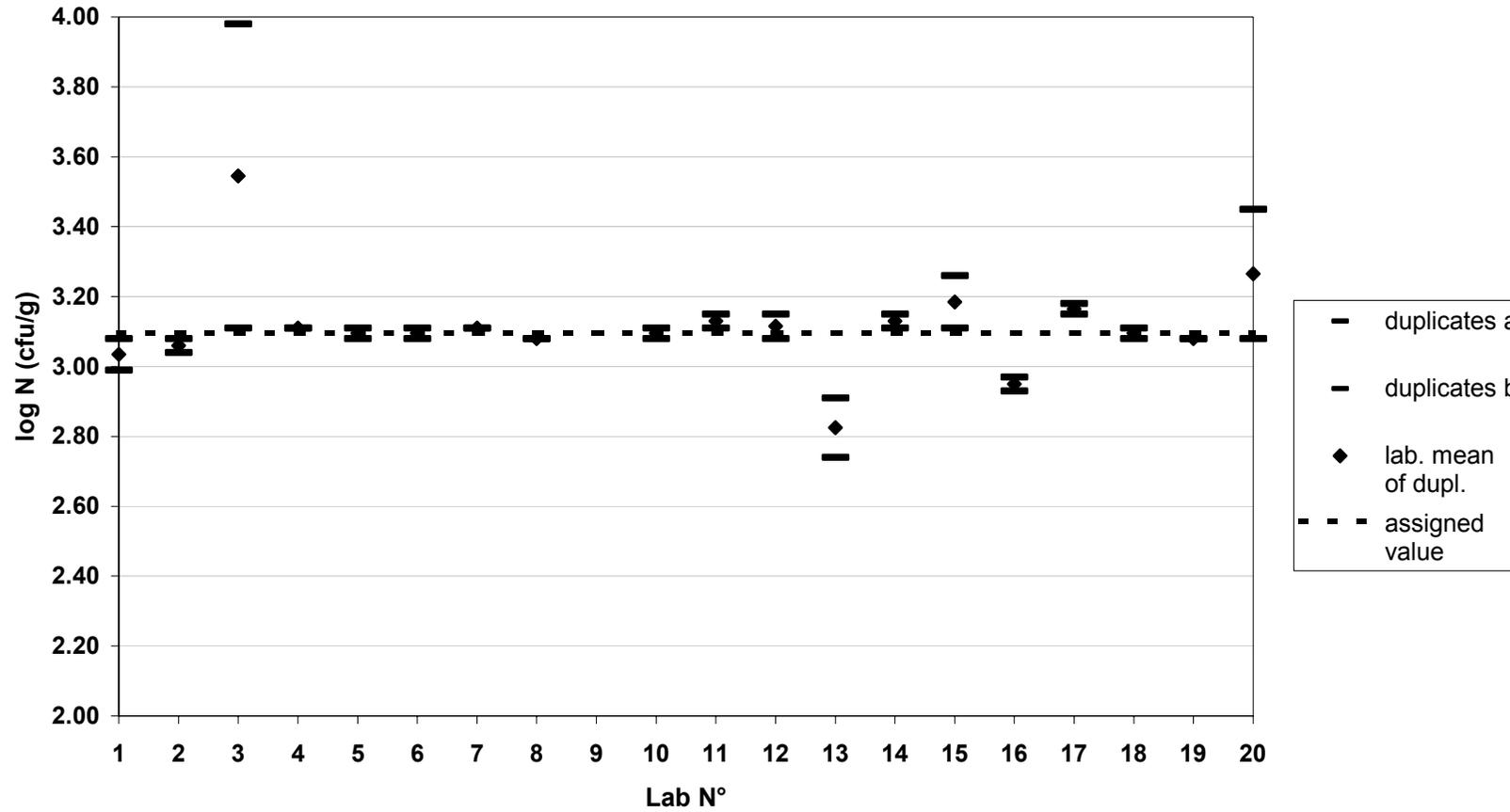
Raw data (in cfu/d & log cfu/g) produced by the participating laboratories

Lab.	Dupli.	A	B		C		D	
		N	N	log N	N	log N	N	log N
1	a	<1	970	2,99	9700	3,99	88000	4,94
	b	<1	1200	3,08	9800	3,99	57000	4,76
2	a	<1	1100	3,04	8900	3,95	65000	4,81
	b	<1	1200	3,08	10000	4,00	62000	4,79
3	a	<1	9600	3,98	12000	4,08	86000	4,93
	b	<1	1300	3,11	11000	4,04	93000	4,97
4	a	<1	1300	3,11	11000	4,04	82000	4,91
	b	<1	1300	3,11	11000	4,04	76000	4,88
5	a	<1	1200	3,08	13000	4,11	120000	5,08
	b	<1	1300	3,11	12000	4,08	96000	4,98
6	a	<1	1300	3,11	11000	4,04	88000	4,94
	b	<1	1200	3,08	12000	4,08	87000	4,94
7	a	<1	1300	3,11	11000	4,04	87000	4,94
	b	<1	1300	3,11	13000	4,11	95000	4,98
8	a	<1	1200	3,08	9600	3,98	83000	4,92
	b	<1	1200	3,08	9800	3,99	79000	4,90
10	a	<1	1300	3,11	12000	4,08	85000	4,93
	b	<1	1200	3,08	10000	4,00	76000	4,88
11	a	<1	1300	3,11	12000	4,08	100000	5,00
	b	<1	1400	3,15	12000	4,08	100000	5,00
12	a	<1	1200	3,08	13000	4,11	98000	4,99
	b	<1	1400	3,15	13000	4,11	97000	4,99
13	a	<1	820	2,91	7000	3,85	65000	4,81
	b	<1	550	2,74	9800	3,99	62000	4,79
14	a	<1	1400	3,15	10000	4,00	89000	4,95
	b	<1	1300	3,11	11000	4,04	100000	5,00
15	a	<1	1300	3,11	15000	4,18	110000	5,04
	b	<1	1800	3,26	14000	4,15	130000	5,11
16	a	<1	860	2,93	12000	4,08	160000	5,20
	b	<1	930	2,97	11000	4,04	220000	5,34
17	a	<1	1500	3,18	13000	4,11	120000	5,08
	b	<1	1400	3,15	12000	4,08	100000	5,00
18	a	<1	1200	3,08	9800	3,99	78000	4,89
	b	<1	1300	3,11	11000	4,04	78000	4,89
19	a	<1	1200	3,08	11000	4,04	90000	4,95
	b	<1	1200	3,08	10000	4,00	90000	4,95
20	a	<1	1200	3,08	26000	4,41	63000	4,80
	b	<1	2800	3,45	16000	4,20	120000	5,08

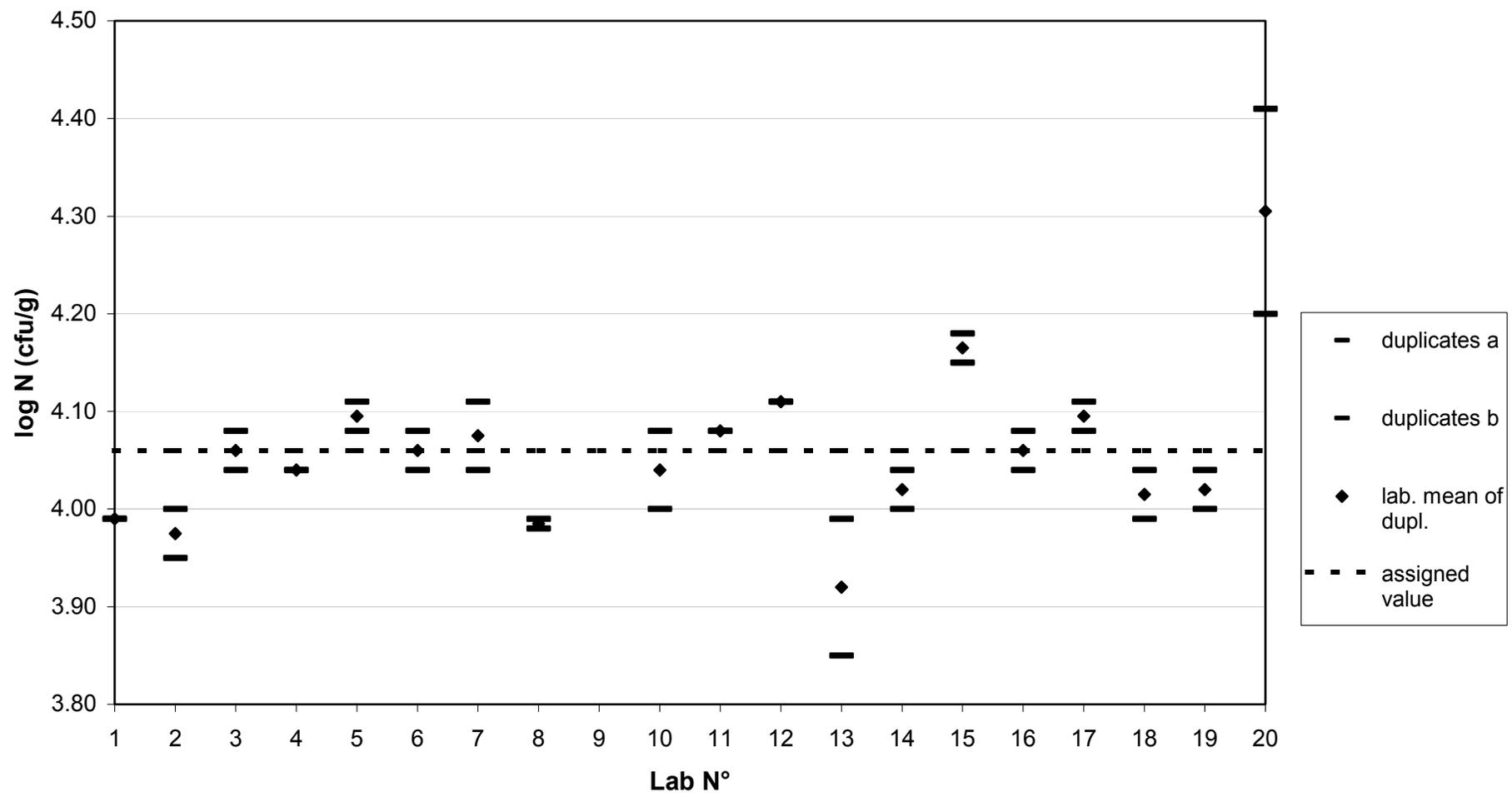
Annex 4

Graphical distribution of the data (duplicates and mean of duplicates) of the participating laboratories for batch A, B and C

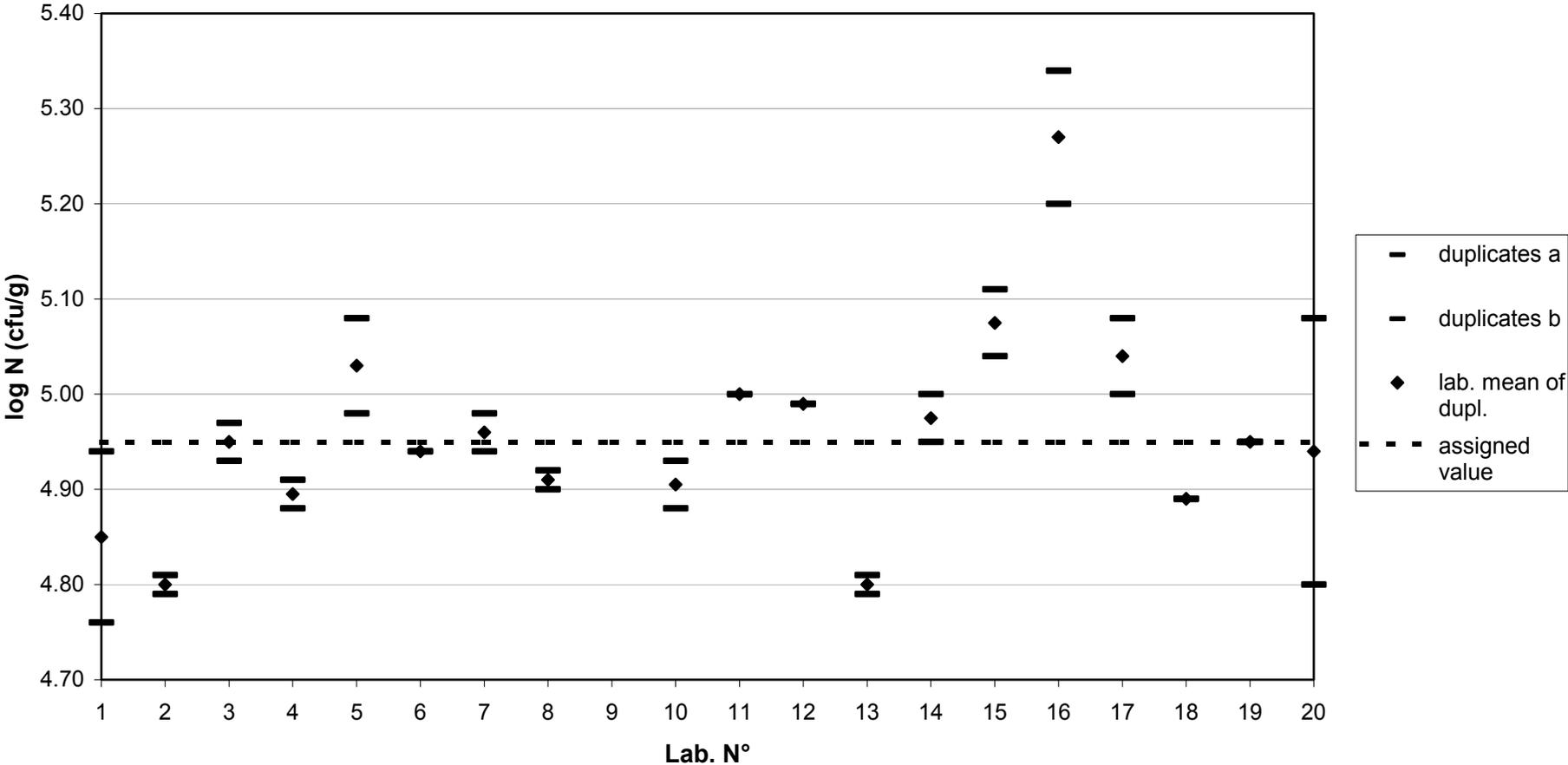
NRLs "Milk" Inter-laboratory trial on enumeration of coliforms in liquid milk, April 2000
 Data of laboratories for batch A



Data of laboratories for batch B



Data of laboratories for batch C



3.9 Annexe 9. Rapport de l'essai d'aptitude sur la détection de *Salmonella*

(sans les annexes 3 & 4 (modes opératoires), 5 & 6 (instructions et rapport d'essai))

Inter-laboratory trial on the detection of *Salmonella* in liquid milk : final report of the CRL « Milk »

Voitoux E., Lombard B. Maisons-Alfort : AFSSA-LERQAP, 2002,
33 p.



***EU COMMUNITY REFERENCE
LABORATORY FOR MILK &
MILK PRODUCTS***

INTER-LABORATORY TRIAL ON THE DETECTION OF *SALMONELLA* IN LIQUID MILK

OCTOBER 2002

Elodie VOITOUX, Unit HMPL

Bertrand LOMBARD, Co-ordinator CRL "Milk"

1.) INTRODUCTION

At the 4th Workshop of National Reference Laboratories for milk (NRL), 11 and 12 October 2001, it was decided that an inter-laboratory trial for the NRLs on the detection of *Salmonella* would be organized by the Community Reference Laboratory (CRL) for milk once the on-going revision of the horizontal Standard method EN ISO 6579 (detection of *Salmonella* in all foods and animal feeding stuffs) would be finalized.

Since the final Draft for this Standard was approved by formal vote of ISO and CEN in 2002, the CRL could organize the inter-laboratory trial in October 2002.

The participating laboratories were given the choice to use either this horizontal Standard, either the Sectorial Standard method, IDF 93B/ISO 6785 (only applicable to milk and milk products); both methods being currently selected for official controls of milk and milk products, within the frame of Directive 92/46.

2.) ORGANISATION

Out of the 24 NRLs contacted by the CRL in circular letter dated 23 May 2002 to invite them to take part to the trial, 21 NRLs decided to take part to the trial (see annex 1). A laboratory from Cyprus, to become the NRL for this country (candidate country for EU), also took part to the trial.

The participating laboratories received detailed instructions for the trial by circular letter of 12 september 2002 (see annex 5).

All these laboratories, except laboratory N°21, have performed the analysis.

14 laboratories (see table 3) have performed the trial using the EN ISO 6579 method.

11 laboratories (see table 3) have performed the trial using the Standard IDF 93B method.

3 laboratories used both methods.

3.) SAMPLES

Six samples of pasteurised milk (blind duplicates) were sent to each participating laboratory.

Three levels of contamination were used (blank, low and high) for performing the artificial contamination of the samples (see table 2).

Each lab received • 2 blank samples

2 low contaminated samples (5-15 cfu/25 ml)

2 high contaminated samples (50-70 cfu/25 ml)

The samples were prepared by Unit HMPL (see annex 2).

The samples were then sent by express carrier (DHL) under refrigerated conditions on Monday 14th October, using isolated boxes with refrigerating devices (ice packs).

Tests of stability and homogeneity of the samples were performed by the CRL (see annex 2).

The samples revealed to be sufficiently homogenous and stable over 4 days for the purpose of a trial on a detection test.

The labs were to confirm per fax to the CRL the good receipt of the samples, immediately after their reception.

16 laboratories have received the parcel the day after the dispatch, 5 within two days and one within 3 days (see table 1).

All parcels arrived without damage.

Temperature of the samples was checked by a water vial included in each parcel.

At reception, it was inferior to 8°C, except for laboratory N°21 (+16°C).

All the laboratories launched the analyses on the prescribed day, Wednesday 16 October, except laboratories N°13 & 22 (15/10/02) and N°21 (17/10/02).

4.) METHOD

The choice for the method to be used was left between the horizontal Standard EN ISO 6579 and the sectorial Standard IDF 93B.

Laboratories N°4, 5, 6, 7, 8, 9, 11, 13, 15, 16, 17, 19, 21 and 22, have performed the trial using the EN ISO 6579 method.

Laboratories N°1, 2, 3, 6, 9, 10, 12, 13, 14, 18 and 20, have performed the trial using the Standard IDF 93B method.

Three laboratories N°6, N°9 and N°13 used both methods.

The way each laboratory performed the analysis is detailed in tables 7, 8 and 9.

DETECTION PROCEDURES:

Detection Procedure of the EN ISO 6579 Method (See Annex 3)

Non selective pre-enrichment:

25 g or 25 ml of a test portion (pasteurised milk of the vial) was to be homogenized with 225 ml of Buffered Peptone Water using a stomacher and incubated at 37°C ± 1°C for 18 h ± 2 h.

Selective enrichment:

After incubation, the pre-enriched sample was to be homogenized and respectively 0.1ml or 1ml aliquot of the clear suspension was pipetted and transferred to a tube containing 10ml of RVS broth or MKTTn broth.

The RVS broth was to be incubated at 41.5°C ± 1°C for 24 h ± 3 h.

The MKTTn broth was to be incubated at 37°C ± 1°C for 24 h ± 3 h.

Isolation:

The culture was then to be plated by means of a loop onto XLD and a second isolation medium, left to the choice of the laboratory (both selective media).

The plates were to be incubated at 37°C ± 1°C for 24 h ± 3 h.

Confirmation of the presumptive *Salmonella* colonies:

The suspected colonies from each plate were to be isolated onto the nutrient agar at 37°C ± 1°C for 24h ± 3 h.

Isolated colonies on the agar were to be submitted to biochemical and serological confirmations.

Detection Procedure of the ISO 6785 / IDF 93B Method (See Annex 4)

Non selective pre-enrichment:

25 g or 25 ml of a test portion was to be homogenized with 225 ml of Buffered Peptone Water using a stomacher and to be incubated at $37^{\circ}\text{C} \pm 1^{\circ}\text{C}$ for $18 \text{ h} \pm 2 \text{ h}$.

Selective enrichment:

After incubation, the pre-enriched sample was to be homogenized and respectively 0.1ml or 10ml aliquot of the clear suspension was to be pipetted and transferred to a tube containing 10ml of RV broth or 90ml of SC broth.

The RV broth was to be incubated at $42^{\circ}\text{C} \pm 1^{\circ}\text{C}$ for two successive periods of 24 h.

The SC broth was to be incubated at $37^{\circ}\text{C} \pm 1^{\circ}\text{C}$ for two successive periods of 24 h.

Isolation:

The culture was then to be plated by means of a loop onto GVB and a second medium left to the choice of the laboratory (both selective media). The plates were to be incubated at $37^{\circ}\text{C} \pm 1^{\circ}\text{C}$ for $22 \text{ h} \pm 2 \text{ h}$.

Confirmation of the presumptive *Salmonella* colonies:

The suspected colonies from each plates were to be isolated onto the nutrient agar at $37^{\circ}\text{C} \pm 1^{\circ}\text{C}$ for $24 \text{ h} \pm 3 \text{ h}$.

Isolates on the agar were to be submitted to biochemical and serological confirmations.

5.) RESULTS

The template of test report (see annex 6) was sent to the NRLs by circular letter on 12/09/02 and was to be filled in and returned to Mrs E.Voitoux before 29/11/02 per fax and mail.

All the participating laboratories performed the analysis, filled in and sent back the test report, some of them with long delay.

ANALYSIS OF THE RESULTS (see table n°3, 5 and 6)

Blank samples:

Out of the 22 labs, 20 have found the blank samples negative as expected; whereas lab n°03 has found both blank samples positive and lab n°17 has found one blank sample positive.

Out of the 44 blank samples, 41 have been found negative: That is to say 93.2% of correct results (specificity).

Low contaminated samples:

Out of the 22 labs, 21 have found the expected result: Presence of *Salmonella*.

One lab (n°17) has found one of the samples negative.

Out of 44 low contaminated samples, 43 revealed to be positive: That is to say 97.7% of correct results (Sensitivity at low level).

High contaminated samples:

Out of the 22 labs, 22 have found the expected result: Presence of *Salmonella*.

Out of 44 low contaminated samples, 44 revealed to be positive: That is to say 100% of correct results (Sensitivity at high level).

It can be noted that the 2 labs (N°13 and 22) having received lately the samples; (especially lab 22, whom samples had moreover experienced an abusive T°), found however the right results.

6.) CONCLUSION

Out of 132 analysed samples, the overall percentage of correct results is 97% (accuracy).

Out of 22 labs, 20 labs have found the expected results.

Lab n°3 has found positive the two blank samples after 48h incubation of the plates.

Lab 3 had not strictly completely applied the method; but, the deviation of the incubation time at the enrichment step can not explain the results.

For lab n°17, there is a problem since one blank sample has been found positive, and it has not found positive one low contaminated sample.

This lab has not reported any deviation to the prescribed protocol.

The problem may be due to wrong identification of the samples by the laboratories when carrying out the analysis.

Moreover, the performance of each of the two methods revealed to be good (global accuracy of 97% for each of the two methods).

Finally, it can be derived from this inter-laboratory trial that the network of NRLs is globally competent to detect *Salmonella* in liquid milk, using one of the two reference methods selected for the official control of *Salmonella* in milk and milk products.

7.) ACKNOWLEDGMENTS

We wish to thank the scientists and technicians from each laboratory participating in this trial

Table 1: STATE OF THE SAMPLES AT ARRIVAL

Laboratory code	Date of arrival of the parcel in the laboratory	State of the samples at arrival	Temperature of the samples at reception	Storage of the samples
01	15/10/02	good	+4.6°C	22 h at +4°C
02	15/10/02	good	+5°C	? h at +5°C
03	15/10/02	good	+5°C	6 h at +1.7°C
04	15/10/02	good	+0°C	29 h at +1°C
05	15/10/02	good	+2.9°C	24 h at +4°C
06	16/10/02	good	+4°C	4 h at +0-5°C
07	15/10/02	good	< 0°C	1 d at +6°C
08	16/10/02	good	+6.2°C	No
09	15/10/02	good	+5°C	24 h at +4°C
10	15/10/02	good	+0°C	14 h at +5°C
11	15/10/02	good	+1°C	24 h at +1°C
12	16/10/02	good	+3.9°C	No
13	15/10/02	good	+2°C	1 h at +4°C
14	15/10/02	good	+0°C	28 h at +4°C
15	15/10/02	good	+6.5°C	25 h at +4°C
16	15/10/02	good	+0.8°C	24 h at +4°C
17	15/10/02	good	+2°C	24 h at +4°C
18	15/10/02	good	-3°C	24h at +6°C
19	16/10/02	good	+4°C	No
20	16/10/02	good	+8°C	4h at 5°C
21	17/10/02	good	+16°C	No
22	15/10/02	good	+0.4°C	3 h at 4°C

Table 2: IDENTIFICATION OF THE BLANK AND CONTAMINATED SAMPLES PER LAB.

Laboratory code	Blank samples 0 CFU/25 ml		Low level of contamination 10-20 CFU/25ml		High level of contamination 70-80 CFU/25ml	
01	1B	1C	1D	1E	1F	1A
02	2C	2D	2E	2F	2A	2B
03	3D	3E	3F	3A	3B	3C
04	4E	4F	4A	4B	4C	4D
05	5F	5A	5B	5C	5D	5E
06	6F	6E	6D	6C	6B	6A
07	7E	7D	7C	7B	7A	7F
08	8D	8C	8B	8A	8F	8E
09	9C	9B	9A	9F	9E	9D
10	10B	10A	10F	10E	10D	10C
11	11A	11F	11E	11D	11C	11B
12	12A	12D	12B	12E	12C	12F
13	13D	13B	13E	13C	13F	13A
14	14B	14E	14C	14F	14A	14D
15	15E	15C	15F	15A	15D	15B
16	16C	16F	16A	16D	16B	16E
17	17E	17B	17D	17A	17F	17C
18	18B	18D	18A	18F	18C	18E
19	19D	19A	19F	19C	19B	19E
20	20A	20F	20C	20E	20B	20D
21	21F	21C	21E	21B	21D	21A
22	22B	22C	22D	22E	22F	22A

Table 3: GLOBAL RESULTS PER LABORATORY

Laboratory code	Method used	Blank samples 0 CFU/25 ml		Low contaminated samples 10-20 CFU/25ml		High contaminated samples 70-80 CFU/25ml	
01	IDF 93B	-	-	+	+	+	+
02	IDF 93B	-	-	+	+	+	+
03	IDF 93B 24h	-	-	+	+	+	+
	IDF 93B 48h	+	+	+	+	+	+
04	ISO 6579	-	-	+	+	+	+
05	ISO 6579	-	-	+	+	+	+
06	ISO 6579	-	-	+	+	+	+
	IDF 93B	-	-	+	+	+	+
07	ISO 6579	-	-	+	+	+	+
08	IDF 93B	-	-	+	+	+	+
09	ISO 6579	-	-	+	+	+	+
	IDF 93B	-	-	+	+	+	+
10	IDF 93B	-	-	+	+	+	+
11	ISO 6579	-	-	+	+	+	+
12	IDF 93B	-	-	+	+	+	+
13	ISO 6579	-	-	+	+	+	+
	IDF 93B	-	-	+	+	+	+
14	IDF 93B	-	-	+	+	+	+
15	ISO 6579	-	-	+	+	+	+
16	ISO 6579	-	-	+	+	+	+
17	ISO 6579	-	+	+	-	+	+
18	IDF 93B	-	-	+	+	+	+
19	ISO 6579	-	-	+	+	+	+
20	IDF 93B	-	-	+	+	+	+
21	ISO 6579	-	-	+	+	+	+
22	ISO 6579	-	-	+	+	+	+

Table 4: Global results for each laboratory

	A	B	C	D	E	F
01	+	-	-	+	+	+
02	+	+	-	-	+	+
03	+	+	+	+	+	+
04	+	+	+	+	-	-
05	-	+	+	+	+	-
06	+	+	+	+	-	-
08	+	+	-	-	+	+
09	+	-	-	+	+	+
10	-	-	+	+	+	+
11	-	+	+	+	+	-
12	-	+	+	-	+	+
13	+	-	+	-	+	+
14	+	-	+	+	-	+
15	+	+	-	+	-	+
16	+	+	-	+	+	-
17	+	-	+	-	+	+
18	+	-	+	-	+	+
19	-	+	+	-	+	+
20	-	+	+	+	+	-
21	+	+	-	+	+	-
22	+	-	-	+	+	+

+ Presence of *Salmonella*

Absence of *Salmonella*

 wrong results

Table 5: SUMMARY OF RESULTS WITH BOTH METHODS: PER SAMPLE

	Blank samples	Low contaminated samples	High contaminated samples
Nb of correct results	41	43	44
Total Nb of results	44	44	44
%age of correct results	93.2	97.7	100

Table 6: SUMMARY OF RESULTS WITH THE BOTH METHODS: PER LABORATORY

Laboratory code	Nb of correct results	Total Nb of results	%age of correct results
01	6	6	100
02	6	6	100
03	4	6	66.7
04	6	6	100
05	6	6	100
07	6	6	100
06	6	6	100
08	6	6	100
09	6	6	100
10	6	6	100
11	6	6	100
12	6	6	100
13	6	6	100
14	6	6	100
15	6	6	100
17	4	6	66.7
18	6	6	100
19	6	6	100
20	6	6	100
21	6	6	100
22	6	6	100

Table 7: DETAILS OF THE WAY THE EN ISO 6579 METHOD WAS PERFORMED**PRE-ENRICHMENT & ENRICHMENT STEP**

Lab. code	PRE-ENRICHMENT STEP			ENRICHMENT STEP					
	Pre-enri. broth BPW	T° of BPW	Incubation	RVS	Manufac	Incubation	MKTTn	Manufac	Incubation
04	Y	20°C	19h 37°C	Y	OXOID	25h 42°C	Y	DIFCO	25h 37°C
05	Y	37°C	24h 37°C	Y	Lab M	* 41.5°C	Y	OXOID	* 37°C
06	Y	37°C	16-20h 37°C	Y	OXOID	24h 42°C	Y	BIOKAR	24h 37°C
07	Y	37°C	20h 37°C	Y	OXOID	22h 42°C	Y	OXOID	22h 37°C
09	Y	*	20h 37°C	Y	Mihajlovic	20h 42°C	Y	Mihajlovic	20h 37°C
11	Y	37°C	18h 37°C	Y	BioMérieux	24h 41.5°C	Y	BIOKAR	24h 37°C
13	Y	20°C	18h 37°C	Y	OXOID	24h 41.5°C	Y	Lab M	24h 37°C
15	Y	23°C	19h 37°C	Y	OXOID	24h 41.5°C	Y	home made	24h 37°C
16	Y	30°C	20h 35°C	Y	MERCK	24h 42°C	Y	MERCK	24h 37°C
17	Y	24°C	19h 37°C	Y	MERCK	24h 41.5°C	Y	BIOKAR	24h 37°C
19	Y	4°C	20h 37°C	Y	OXOID	24h 41.5°C	Y	OXOID	24h 37°C
21	Y	8°C	20h 37°C	Y	OXOID	22h 41.5°C	Y	MERCK	22h 37°C
22	Y	20°C	18h 37°C	Y	Lab M	18h 41.5°C	Y	OXOID	18h 37°C

Table 7 bis: DETAILS OF THE WAY THE IDF 93B METHOD WAS PERFORMED**PRE-ENRICHMENT & ENRICHMENT STEP**

Lab. code	PRE-ENRICHMENT			ENRICHMENT STEP							
	Pre-enri. broth BPW	T° of BPW	Incubat°	RV	Manufactu	Incub 1	Incub 2	SC	Manufactu	Incub 1	Incub 2
01	Y	37°C	18h 37°C	Y	OXOID	18-24h 42°	18-24h 42°	Y	OXOID	18-24h 37°	18-24h 37°
02	Y	25°C	16h 37°C	Y	BioTrading	24h 42°C		Y	BioTrading	24h 37°C	
03	Y	20°C	18h 37°	Y	BioMérieux	6h 42°C		Y	BioMérieux	6h 37°C	
06	Y	37°C	16-20h 37°	RVS	OXOID	24h 42°C	24h 42°C	Y	MERCK	24h 37°C	24h 37°C
08	Y	23°C	20h 37°C	Y	OXOID	20h 42°C	20h 42°C	Y	*	20h 37°C	
09	Y	*	20h 37°C	RVS	Mihaljlovic	24h 42°C	24h 42°C	TBG	*	24h 37°C	24h 37°C
10	*	*	*	Y	BioMérieux	24h 42°C		Y	*	24h 37°C	
12	Y	22°C	24h 37°C	Y	OXOID	24h 42°C	24h 42°C	Y	OXOID	24h 37°C	24h 37°C
13	Y	20°C	18h 37°C	Y	OXOID	24h 41.5°C	21h 41.5°C	Y	OXOID	24h 37°C	21h 37°C
14	Y	30°C	19h 37°C	Y	OXOID	23h 42°C	22h 42°C	Y	OXOID	23h 37°C	22h 37°C
18	Y	12°C	20h 37°C	Y	BioMérieux	24h 42°C	24h 42°C	Y	DIFCO	24h 37°C	24h 37°C
20	Y	21°C	17h 37°C	Y	MERCK	24h 41.5°C	23h 41.5°C	Y	MERCK	24h 37°C	23h 37°C

- **information not provided**

Table 8: DETAILS OF THE WAY THE EN ISO 6579 METHOD WAS PERFORMED**PLATING**

Lab. code	XLD manufacturer	Incubation of the first agar	2 nd agar name & manufacturer	Incubation of the 2 nd agar
04	OXOID	24h 37°C	MLCB OXOID	24h 37°C
05	Lab M	24h 37°C	ONoz MERCK	48h 37°C
06	MERCK	24h 37°C	BGA OXOID	24h 37°C
07	OXOID	22h 37°C	RAMBACH MERCK	22h 37°C
09	Mihajlovic	24h 37°C	BPLS Mihajlovic	24h 37°C
11	AES	24h 37°C	GVB AES	24h 37°C
13	OXOID	23h 37°C	Salmonella Chromogenic OXOID	23h 37°C
15	OXOID	24h 37°C	MLCB OXOID	24h 37°C
16	BioMérieux	24h 37°C	Hektden BioMérieux	24h 37°C
17	BioMérieux	24h 37°C	RAMBACH MERCK	24h 37°C
19	OXOID	24h 37°C	SS Agar OXOID	24h 37°C
21	OXOID	22h 37°C	BPLS Agar MERCK	22h 37°C
22	OXOID	*	BGA OXOID	*

Table 8 bis: DETAILS OF THE WAY THE IDF 93B METHOD WAS PERFORMED**PLATING**

Lab. code	E&K manufacturer	Incubation of the first agar	2 nd agar name & manufacturer	Incubation of the 2 nd agar
01	XLD OXOID	18-24h 37°C	BGA OXOID	18_ 24h 37°C
02	BGAM OXOID	24h 37°C	OSCM OXOID	24h 37°C
03	XLD OXOID	48h 37°C	Compass SLM BLOKAR	48h 37°C
06	BGA OXOID	24h 37°C	XLD MERCK	24h 37°C
08	E&K OXOID	24h 37°C	XLD *	24h 37°C
09	E&K Mihajlovic	24h 37°C	RAMBACH Mihajlovic	24h 37°C
10	Helctoen BioMérieux	48h 37°C	BGPRA BioMérieux	48h 37°C
12	E&K OXOID	24h 37°C	XLD OXOID	24h 37°C
13	E&K OXOID	23h 37°C	Salmo Chromogenic OXOID	23h 37°C
14	E&K OXOID	22h 37°C	Des OXOID	22h 37°C
18	E&K DIFCO	24h 37°C	XLD BioMérieux	24h 37°C
20	E&K BLOKAR	24h 37°C	RAMBACH MERCK	24h 37°C

Table 9: DETAILS OF THE WAY THE METHOD WAS PERFORMED**IDENTIFICATION**

Lab. code	Biochemical test	Serological test	Serotyping
01	API 20E		
02		Tests TECRA Salmonella	
03	Mini-API Rapid E20		
04	API 20E	Polyvalent O-antiserum Polyvalent H-antiserum	Salmonella Typhimurium (1,4,5,12: i: 1,2)
05	TSI & UREA	Polyvalent antisera I&II	Preliminary results 5 B,C,D&E antisera 0 4,5 positive
06	BBL E/NF Oxydase reaction	Agglutination reaction	Biochemical Salmonella Serological rough
07	BBL Enterotube™	Agglutination groups A-67	Salmonella typhimurium 7A, 7B, 7C, 7F
08	API 20E	OV, PV I, II, III	A,B,E,F Salmonella typhimurium
09	API 20E	Not possible: growth in rough forms	Salmonella
10	API 20E, Microbact 24E	Slide agglutination OMA/OMB	
11	API 20E, TSI		
12	API 20E		Antigenic structure O rough: i: 2 Antibiotic resistance ACSSUTSP Phage type DT104b
13	API 20E	Polyvalent O-antiserum Polyvalent H-antiserum	Salmonella unnamed O rough i: 1, 2
14	Kit API 20E	Pol O (A-I+Vi) & Pol O (pol A-S) Biolife	Salmonella typhimurium 14A, 14C, 14D, 14F
15	TSI, Urease, Lys/Dec, β-Gal, VP, Indole	Polyvalent O-antiserum (A-I+Vi) Difco	Salmonella typhimurium (4,5,12: i: 1,2) lisotype 104b
16	API 20E	Microscreen Salmonella MICROKIT	
17	API 20E	Salmonella O-antiserum polyvalent A-I +Vi	Salmonella enterica Serovar DiouBel 21: i: 12
18	API 32E & Oxydase	Polyvalent O-antiserum Polyvalent H-antiserum	Salmonella enterica Serovar DiouBel 21: i: 1, 2
19	TSI, UREA, ONPG, β-Gal, VP, Indole		Results are waiting
20	API 20E	Agglu O, H & Vi	S.typhimurium (4, 5, 12: i: 1, 2)
21	API 20E & TSI	Agglu O, H & Vi	Results are waiting
22	TSI, SLUMS, Mac Conkey broth		

Annex 1

LIST OF PARTICIPATING LABORATORIES

LNIV

Mrs Celcidina PIRES GOMES

Estrada de Benfica, 701

PT-1549 - 011 LISBOA PORTUGAL

BgVV

Mrs Juliana BRAEUNIG

Diedersdorfer Weg 1

DE - 12277 BERLIN GERMANY

Veterinary Laboratory of Larisa

M. Theophanis SAGRIS

Ad National Road of Larisa-Trikala 6km

GR - 41110 – LARISA GREECE

Veterinary Laboratory of Patras

Mrs Klytemnestra VELETA

15, Notara street

GR - 264 42 – PATRAS GREECE

Laboratorio Agroalimentario de Santander

Mrs Marian SANCHEZ MERINO

Ministerio de Agricultura, Pesca y Alimentacion

Prolongacion Marqués de la Hermida s/n

SP - 39011 - SANTANDER - CANTABRIA

Instituto de Salud Carlos III

Centro Nacional de Alimentacion

Dra Rosana REYMUNDO CUESTA

Carretera de Majadahonda-Pozuelo Km 2

SP - 28220- MAJADAHONDA - SPAIN

Istituto Superiore di Sanita

Mrs Alfonsina FIORE

Viala Regina Elena, 299

IT - 00161 – ROMA ITALY

RIKILT – DLO

Mr Henk STEGEMAN

Building no. 123

Bornesesteeeg 45

NL - 6708 PD - WAGENINGEN

CSL Food Science Laboratory

Mrs R.LEUSCHNER & Mrs J. BEW

Sand Hutton

UK – YO41 1LZ – YORK UNITED-KINGDOM

Dept of Agriculture for Northern Ireland

Food Science Centre

Mr John EARLY

Newforge Lane

BELFAST BT9 5PX UNITED-KINGDOM

Mrs Tuula JOHANSSON
National Veterinary and Food Research Institute
PO Box 45
FIN - 00581 – HELSINKI

Mr Gilbert MORIS
Laboratoire National De Santé
1A, rue Auguste Lumière
L – 1950 LUXEMBOURG

Austrian Agency for Health and Food Safety
Mr Horst FALZBERG
Kinderspitalgasse, 15
A - 1090 – WIEN AUSTRIA

National Reference Laboratory for Milk
Mr Christer WIBERG
Livsmedelsverket
PO Box 622
S - 751 26 – UPPSALA - SWEDEN

DGFCQA
Laboratorio Central de Qualidade Alimentar
Mrs Margarida ARAUSO
Av Conde de Valbom, 98
PT - 1050 - 070 LISBOA PORTUGAL

Bundesanstalt für Milchforschung
Institut für Hygiene
Dr Philipp HAMMER
Hermann - Weigmann - Straße 1
D - 24103 – KIEL - GERMANY

Dairy Science Laboratory
Mr Peter COLLINS
Harcourt Terrace Lane
IE - DUBLIN 2 IRELAND

FAM
Mr Urs SPAHR
Liebefeld
Schwartzenburgstrasse 161
CH - 3003 BERNE SWITZERLAND

Agence Fédérale pour la sécurité de la chaîne Ali.
DG5
Mrs Marie-Christine WILEN
Chaussée de Namur ,22
B - 5030 – GEMBLOUX BELGIUM

DVK – CLO
Mr Koen DE REU
Brusselsesteenweg 370
B- 9090 – MELLE BELGIUM

Laboratory of control (LCFAO)
Department Veterinary Services
Mr Charalambos KAKOYANNIS
1417 ATHALASSA – NICOSIA - CYPRUS

AFSSA-LERHQA
Mrs Elodie Voitoux
39, 41 rue du 11 Novembre 1918
FR - 94700 MAISONS-ALFORT FRANCE

Annex 2

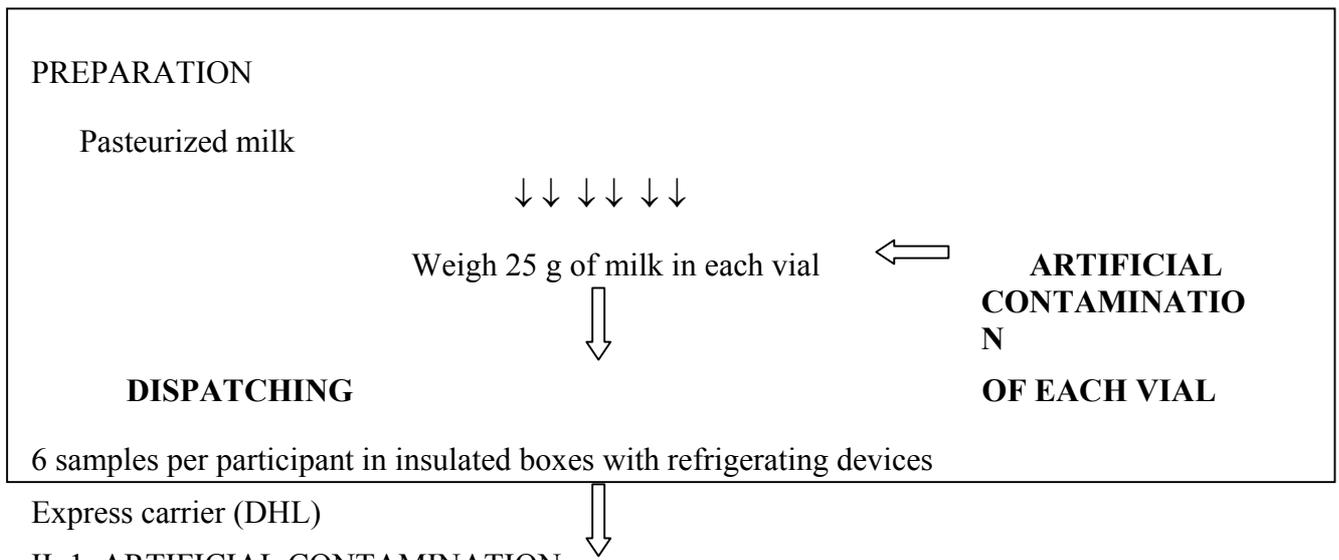
PREPARATION AND CHECK OF THE SAMPLES FOR THE INTER-LABORATORY TRIAL ON THE DETECTION OF *SALMONELLA* in LIQUID MILK

I. THE STRAIN

The bacterial strain used for this trial, *Salmonella* Typhimurium, was isolated from milk product.

II. PREPARATION AND DISPATCHING OF THE SAMPLES (at Jo:14/10/02)

Figure 1: Summary



II. 1. ARTIFICIAL CONTAMINATION

The *Salmonella* strain was cultured on TSAYE at $37^{\circ}\text{C}\pm 1^{\circ}\text{C}$ for 24 h.

Then, the strain was suspended in 10 ml of EPT and incubated at 37°C for 24h.

II. 1.1. Preparation of the inoculum

The optical density (Do) of the suspension was measured by using a densitometer.

The cell count (N) was determined by the formula $N=K.Do$; K being the slope of the calibration curve for the strain.

Serial dilutions of the inoculating suspension were prepared in TS to obtain the desired cell density.

So, 30 ml of two inoculating suspension targeted at 25 cfu/ml and 150 cfu/ml were prepared.

II. 1.2. Sample inoculation

Two test portions (25 g of milk) per lab were inoculated with 0.4 ml of the first diluted suspension (25 cfu/ml) to reach the low level of contamination 5-15 cfu per 25 g.

Two test portions (25 g of milk) per lab were inoculated with 0.6 ml of the second diluted suspension (150 cfu/ml) to reach the high level of contamination 50-70 cfu per 25 g.

Two non-inoculated samples per lab were prepared as negative control.

III. HOMOGENEITY BETWEEN THE SAMPLES

Since the contamination was performed separately for each sample, it was not necessary to test for the homogeneity between the contaminated samples.

It was only necessary to establish the precise level of contamination and the homogeneity of the inoculating suspensions. For that purpose, 0.4 ml of each of the two inoculating suspensions were pour plated in TSAYE ten times (10 petri dishes per inoculum) and incubated at $37^{\circ}\text{C}\pm 1^{\circ}\text{C}$ for 24 h.

It is important to highlight that this study was performed on the inoculating suspensions used to contaminate the samples.

	Numeration of the colonies from the 10 Petri dishes of TSAYE									
Low level (inoculum A)	9	9	14	15	8	8	5	12	14	clump
High level (inoculum B)	65	65	56	70	60	63	54	58	58	54

The mean for inoculum A was: 10 cfu for 0.4 ml, standard-deviation 3,4.

The mean for inoculum B was: 60 cfu for 0.4 ml, standard-deviation 5,3.

Therefore, the samples (of 25 ml each) were contaminated at a level of 10 cfu/25 ml for batch A, and at a level of 60 cfu/25 ml for batch B.

IV. STABILITY OF THE SAMPLES

Two flasks (A&B) containing 24 ml of milk were contaminated at J0 (with 1 ml of an inoculum to reach a contamination level of 500 cfu/ml).

The milk was the same than the one used for the dispatched samples. But the contaminating suspension was not the same, in order to reach a high contamination level and to be able to perform an enumeration of *Salmonella*.

The two contaminated samples were kept for 4 days at 4°C .

The stability was measured by surface plating 0.1 ml of each sample of contaminated milk on XLD (0.1ml per petri dish; 2 petri dishes per sample) at J0; J+1; J+2; and J+3.

	J0	J+1	J+2	J+3
A cfu/ml	410	150	110	70
	510	240	130	90
Mean for A	460	195	120	80
B cfu/ml	470	260	110	40
	500	350	300	90
Mean for B	485	305	205	65

An univariate ANOVA was performed using the software StatView (SAS) on the factors: day of analysis and sample A/B.

It showed that there was no significant difference between the sample A/B ($p=0.15$), thus the values of the two samples A and B have been considered together.

A significant difference was found between the days of the analysis ($p<0.001$, power:100%).

A clear decrease in the number of micro-organisms can be found between J0 and J+1: it may be due to the stress of the micro-organisms by the storage of the contaminated milk sample at refrigerated temperature. The contamination table, then goes on decreasing, but much more slightly.

Anyway, this stability was considered as enough for a detection method.

4. Bibliographie

4.1 Textes normatifs

(les normes AFNOR et ISO sont classées par ordre croissant du numéro de référence)

- AFNOR FD V01-000, 1999. *Fascicule de Documentation: Analyse des produits agricoles et alimentaires – Terminologie*. Paris : AFNOR, 1999, 26 p.
- AFNOR FD V03-115, 1996. *Fascicule de Documentation : Analyse des produits agricoles et alimentaires – Guide pour l'utilisation des matériaux de référence*. Paris : AFNOR, 1996, 33p.
- AFNOR Certification, 2002. *Validation AFNOR des Méthodes Alternatives – Application à l'agro-alimentaire – Règles de fonctionnement. 7^{ème} révision*. Paris : AFNOR, 2002, 38 p.
- AOAC, 1995. AOAC INTERNATIONAL Guidelines for Collaborative Study Procedures To Validate Characteristics of a Method of Analysis. *Journal of AOAC International*, 1995, Vol. 78, No 5, p. 143A-160A
- AOAC INTERNATIONAL, 2000. *AOAC Performance Tested Methods Program Manual*. Gaithersburg, MD : AOAC INTERNATIONAL, 2000, www.aoac.org
- AOAC INTERNATIONAL, 2003. *AOAC Official Methods Program Manual*. Gaithersburg, MD : AOAC INTERNATIONAL, 2003, www.aoac.org
- Codex Alimentarius, 2004. *Analytical Terminology for Codex Use, Circular Letter CL 2004/37-MAS*. Rome : Codex Alimentarius Office, 2004, 19 p., www.codexalimentarius.net
- FIL 135B, 1991. *Lait et produits laitiers – Caractéristiques de fidélité des méthodes analytiques – Schéma de conduite d'une étude collaborative*. Bruxelles : FIL, 1991, 11 p.
- FIL-IDF/ISO 2003. *Draft Standard: Microbiology of food and animal feeding stuffs – Protocol for the establishment of precision characteristics of quantitative methods by inter-laboratory studies*. Bruxelles, FIL, 2003, 17 p.
- ISO, 1993. *Guide BIPM/CEI/FICC/ISO/UICPA/UPPA/OIML pour l'expression de l'incertitude de mesure (GUM)*. Genève : ISO, 1993, 105 p.
- ISO Guide 30, 1992. *Termes et définitions utilisés en rapport avec les matériaux de référence*. Genève : ISO, 1992, 8p.
- ISO Guide 34, 2000. *Exigences générales pour la compétence des producteurs de matériaux de référence*. Genève : ISO, 2000, 27 p.
- ISO Guide 35, 1989. *Certification des matériaux de référence - Principes généraux et statistiques*. Genève : ISO, 1989, 32 p.
- ISO Guide 43-1, 1997. *Essais d'aptitude des laboratoires par intercomparaison – Partiel : Développement et mise en œuvre de systèmes d'essais d'aptitude*. Genève : ISO, 1997, 17 p.

- ISO 5725-1, 1994a. *Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 1 : Principes généraux et définitions*. Genève : ISO, 1994, 18 p.
- ISO 5725-2, 1994b. *Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 2 : Méthode de base pour la détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée*. Genève : ISO, 1994, 44 p.
- ISO 5725-3, 1994c. *Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 3 : Mesures intermédiaires de la fidélité d'une méthode de mesure normalisée*. Genève : ISO, 1994, 25 p.
- ISO 5725-4, 1994d. *Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 4 : Méthodes de base pour la détermination de la justesse d'une méthode de mesure normalisée*. Genève : ISO, 1994, 23 p.
- ISO 5725-5, 1998. *Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 5 : Méthodes alternatives pour la détermination de la fidélité d'une méthode de mesure normalisée*. Genève : ISO, 1998, 57 p.
- ISO 5725-6, 1994e. *Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 6 : Utilisation dans la pratique des valeurs d'exactitude*. Genève : ISO, 1994, 41 p.
- ISO 8259, 1991. *Cartes de contrôle de Shewhart*. Genève : ISO, 1991, 29p.
- ISO/DIS 13528, 2002. *Projet de norme internationale : Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires*. Genève : ISO, 2002, 70p.
- ISO 16140, 2003a. EN ISO. *Microbiologie des aliments – Protocole pour la validation des méthodes alternatives*. Genève : ISO, 2003, 70p.
- ISO/DTS 19036, 2003b. *Projet de Spécification Technique : Microbiologie des aliments – Guide pour l'expression de l'incertitude de mesure des déterminations quantitatives ; Document de travail ISO/TC 34/SC 9 N 605*. ISO, Genève, 2003, 11 p.
- ISO/TS 21748, 2004. *Lignes directrices relatives à l'utilisation d'estimations de la répétabilité, de la reproductibilité et de la justesse dans l'évaluation de l'incertitude de mesure*. Genève : ISO, 2004, 32 p.
- IUPAC, 1995. Protocol for the design, conduct and interpretation of method-performance studies. *Pure and Applied Chemistry*, 1995, vol. 67, No 2, p. 331-343
- NMKL, 2003. *Report 20. Guide for referees within microbiology – Elaboration of analytical methods within NMKL*. Oslo: NMKL, 2003, 25 p.

4.2 Publications et rapports

- Abdulmawjood A. *et al.*, 2004. Toward an international standard for PCR-based detection of foodborne *Escherichia coli* O157: validation of the PCR-based method in a multicenter interlaboratory trial. *Journal of AOAC International*, 2004, Vol. 87, No 4, p. 856-860.
- Analytical Methods Committee, Royal Society of Chemistry, 1989a. Robust statistics – How not to reject outliers. Part 1: Basic Concepts. *The Analyst*, 1989, 114, p. 1697-1698.
- Analytical Methods Committee, Royal Society of Chemistry, 1989b. Robust statistics – How not to reject outliers. Part 2: Inter-laboratory trials. *The Analyst*, 1989, 114, p. 1699-1702.
- Andrews W.H., 1996. Validation of modern methods in food microbiology by AOAC INTERNATIONAL collaborative study. *Food Control*, 1996, Vol. 7, No. 1, p. 19-29.
- Archer D.L., 1996. Editorial, Special Issue on Validation of Rapid Methods in Food Microbiology. *Food Control*, 1996, Vol. 7, No. 1, p. 3-4.
- Augustin J.C., Carlier V., 2002. French Laboratory Proficiency Testing Program for Food Microbiology. *Journal of AOAC International*, 2002, Vol. 85, No 4, p. 952-959.
- Blodgett R.J., Hitchins A.D., 2000. Evaluating Presence/Absence of Target Microbes in Microbiological Tests. *Journal of AOAC International*, 2000, Vol. 83, No 6, p. 1429-1433.
- Bugner E., 1992. *Chimométrie des Analyses Interlaboratoires – Application à la validation d'une méthode d'analyse des sucres simples dans les aliments*. Thèse chimie analytique, INA-PG, 19 mars 1992, 220 p.
- Bohnert M.L., Humbert F., **B. Lombard**, 2001. *Validation of draft Standard EN ISO/DIS 6579 : 2000 – Detection of Salmonella in foods - Final Report*, Ploufragan : AFSSA, 2001, 78 p.
- Corradini M.G., Normand M.D., Nussinovitch J., Horowitz J., Peleg M., 2001. Estimating the frequency of high microbial counts in commercial food products using various distribution functions. *Journal of Food Protection*, 2001, 64, p. 674-681.
- D'Agostino M. *et al.* A multicenter-validated non-commercial PCR-based method for the detection of *Listeria monocytogenes*, submitted.
- De Buyser M.L., **Lombard B.**, Schulten S.M., In't Veld P.H., Scotter S.L., Rollier P., Lahellec C., 2003. Validation of EN ISO standard methods 6888 Part 1 and Part 2 : 1999 – Enumeration of coagulase-positive staphylococci in foods. *International Journal of Food Microbiology*, 2003, 83(2), p. 185-194.
- Dixon W.J., Massey F.J., 1969. *Introduction to Statistical Analysis*, 3rd Ed. New York : McGraw-Hill Book Co, 1969.
- Feldsine Ph., Abeyta C., Andrews W., 2002. AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Qualitative and Quantitative Food Microbiological Official Methods of Analysis. *Journal of AOAC International*, 2002, Vol. 85, No 5, p. 1187-1200.

- Feldsine Ph. *et al.*, 2003. Detection of *Salmonella* in Fresh Cheese, Poultry Products, and Dried Egg Products by the ISO 6579 *Salmonella* Culture Procedure and the AOAC Official Method: Collaborative Study. *Journal of AOAC International*, 2003, Vol. 86, No 2, p. 275-295.
- Fleet G.H., 1996. Microbiological methods – the Australian perspective. *Food Control*, 1996, Vol. 7, No. 1, p. 41-46.
- Gnanou Besse N., Audinet N., Beaufort A., Colin P., Cornu M., **Lombard B.**, 2004. A contribution to the improvement of *Listeria monocytogenes* enumeration in cold-smoked salmon. *International Journal of Food Microbiology*, 2004, 91, p. 119-127.
- Gomy C., **Lombard B.** Normalisation et validation AFNOR des méthodes d'analyse. *Analisis Magazine*, 1992, Vol 20, n°9, p. 17-20.
- Hedges A.J., 1967. On the dilution errors involved in estimating bacterial numbers by the plating method. *Biometrics*, 1967, 23, p. 158-159.
- Hedges A.J., 2002. Estimating the precision of serial dilutions and viable bacterial counts. *International Journal of Food Microbiology*, 2002, 76, p. 207-214.
- Hitchins A.D., 1996. The International Dairy Federation's procedure for the validation of microbiological analytical methods for dairy foods. *Food Control*, 1996, Vol. 7, No. 1, p. 13-18.
- In't Veld P., 1998. *The development and evaluation of reference materials for food microbiology*. Thèse : Microb., Wageningen, Pays-Bas, 1998, 184 p.
- Jackson G.J., Wachsmuth I.K., 1996. The US Food and Drug Administration's selection and validation of tests for foodborne microbes and microbial toxins. *Food Control*, 1996, Vol. 7, No. 1, p. 37-39.
- Jarvis B., 1989. *Statistical aspects of the microbiological analysis of the foods* (coll. Progress in industrial microbiology, Vol. 21). Amsterdam : Elsevier, 1989, 179 p.
- Josefsen M.H., Lambertz S.T., Jensen S., Hoofar J., 2003. FOOD-PCR, Validation and Standardization of Diagnostic PCR for Detection of *Yersinia enterocolitica* and other foodborne Pathogens. In : *The Genus Yersinia*. New York : Kluwer Academic/Plenum Publishers, 2003, p. 443-449.
- JRC/IRMM, 2004. *Certified Reference Materials 2004*. Geel : JRC/IRMM, 2004, 78 p.
- Lafarge V., **Lombard B.** 1999. *Inter-laboratory study on the detection of L.monocytogenes in liquid milk : final report of the CRL « Milk »*. Maisons-Alfort : AFSSA-LERQAP, 1999, 43 p.
- Lafarge V., Maladen V., **Lombard B.**, 2000. *Inter-laboratory proficiency trial on the enumeration of coliforms in milk : final report of the CRL « Milk »*. Maisons-Alfort : AFSSA-LERQAP, 2000, 28 p.
- Lahellec C., 1998. Development of standard methods with special reference to Europe. *International Journal of Food Microbiology*, 1998, 45, p. 13-16.
- Langton S.D., Chevennement R., Nagelkerke N., **Lombard B.**, 2002. Analysing collaborative trials for qualitative microbiological methods : accordance and concordance. *International Journal of Food Microbiology*, 2002, 79, p. 175-181.

- Le Barillec K., Tyburski C., **Lombard B.**, 2003. *Inter-laboratory trial on the detection of L.monocytogenes : final report of the CRL "Milk"*. Maisons-Alfort : AFSSA-LERQAP, 2003, 21p.
- Leclercq A., **Lombard B.**, Mossel D.A., 2000. Normaliser les méthodes d'analyse dans le cadre de la maîtrise de la sécurité microbiologique française des aliments : atout ou contrainte. *Science des Aliments*, 2000, 20(2), p. 179-202.
- Lombard B.**, Gomy C., Catteau M., 1996. Microbiological analysis of foods in France : standardized methods and validated methods. *Food Control*, 1996, Vol. 7, No 1, p. 5-11.
- Lombard B.**, 2001 Aspects normatifs de la validation des méthodes. In : *L'assurance qualité dans les laboratoires agroalimentaires et pharmaceutiques*, Feinberg M., coordonnateur. Paris : Editions TEC & DOC, 2001, p. 81-94.
- Lombard B.**, 2004. Microbiological analysis of foods and animal feeding stuffs. *ISO Focus*, Vol. 1, No. 8, September 2004, p. 16-17.
- Lombard B.**, Cornu M., Lahellec C., Feinberg M. Experimental evaluation of different precision criteria applicable to microbiological counting methods. *Journal of AOAC International*, accepté.
- Lubeck P.S. *et al.*, 2003. Toward an International Standard for PCR-Based Detection of Food-Borne Thermotolerant *Campylobacter*: Validation in a Multicenter Collaborative Trial. *Applied and Environmental Microbiology*, 2003, 69, 9, p. 5670-5672.
- Malorny B. *et al.*, 2004. Multicenter Validation of PCR-Based Method for Detection of *Salmonella* in Chicken and Pig Samples. *Journal of AOAC International*, 2004, Vol. 87, No. 4, p. 861-866.
- McClure FD., 1990. Design and Analysis of Qualitative Collaborative Studies: Minimum Collaborative Program. *Journal of AOAC International*, 1990, Vol. 73, No 6, p. 953-960.
- NordVal, 2004. *Protocol for the validation of alternative microbiological methods*. Söborg : NordVal c/o Danish Institute for Food and Veterinary Research, 2004, 15 p.
- Piton C. et Grappin R., 1991. A Model for Statistical Evaluation of Precision Parameters of Microbiological Methods: Application to Dry Rehydratable Film Methods and IDF Reference Methods for Enumeration of Total Aerobic Mesophilic Flora and Coliforms in Raw Milk. *Journal of AOAC International*, 1991, Vol. 74, No. 1, p. 92-103.
- RAEMA, 2004. *RAEMA : état des lieux en 2004*. Maisons-Alfort : Animal Société Aliment, 2004, 83 p.
- Rentenaar I.M.F., 1996. MicroVal, a challenging Eureka project. *Food Control*, 1996, Vol. 7, No 1, p. 31-36.
- Robouch P., Younes N., Vermaercke P. The "Naji Plot", a simple graphical tool for the evaluation of inter-laboratory comparisons. *International Workshop Data Analysis of ILC*, Berlin, December 2002, non publié.
- Rollier P., Trossat Ph., Leray O., 1996. *Interlaboratory proficiency studies for enumeration and detection of pathogens in cheese*. In : Second International Sonthofen Symposium, Analytical Quality and Economic Efficiency in Dairy and Food Laboratories, Sonthofen (Allemagne), 20-22 mai 1996. 4 p.

- Rousseeuw P.J., Croux C., 1993. Alternatives to the median absolute deviation. *Journal of American Statistical Association*, 1996, 88, p. 1273-1283.
- Schulten S.M. *et al*, 2000. Evaluation of the ISO 7932 standard for the enumeration of *Bacillus cereus* in foods. *International Journal of Food Microbiology*, 2000, 57, p. 53-61.
- Schulten S.M., Benschop E., Nagelkerke N.J.D., Mooijman K.A., 2001. *Validation of Microbiological Methods: Enumeration of C.perfringens according to ISO 7937 (1997) – Final report*. Bilthoven : RIVM, 2001, 83 p.
- Scotter S., Aldridge M., Back J., Wood R., 1993. Validation of European Methods for Microbiological and Chemical Analysis of Raw and Heat-treated Milk. *Journal of the Association of Public Analysts*, 1993, 29, p. 1-32.
- Scotter S., Wood R., 1996. Validation and acceptance of modern methods for the microbiological analysis of foods in the UK. *Food Control*, 1996, Vol. 7, No 1, p.47-51.
- Scotter S., Langton S., **Lombard B.**, Lahellec C., Schulten S., Nagelkerke N., in't Veld P., Rollier P., Bohnert M., 2001. Validation of ISO method 11290 Part 1 - Detection of *Listeria monocytogenes* in foods. *International Journal of Food Microbiology*, 2001, 64, p. 295-306.
- Scotter S., Langton S., **Lombard B.**, Lahellec C., Schulten S., Nagelkerke N., in't Veld P., Rollier P., Bohnert M., 2001. Validation of ISO method 11290 Part 2 - Enumeration of *Listeria monocytogenes* in foods. *International Journal of Food Microbiology*, 2001, 70, p. 121-129.
- Struijk C.B., 1996. Guidelines for method validation techniques used in the microbiological examination of food samples. *Food Control*, 1996, Vol. 7, No 1, p. 53-58.
- Thompson M. et Wood R., 1993. International Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories. *Journal of AOAC International*, 1993, Vol. 76, No. 4, p. 926-940.
- Van der Voet H., 2004. *Mean squared error of random and fixed estimators for accordance and concordance*. Document de travail pour l'ISO/TC 34/SC 9/WG 2, non publié.
- Van der Voet H., van Raamsdonk L.W.D., 2004. Estimation of accordance and concordance in inter-laboratory trials of analytical methods with qualitative results. *International Journal of Food Microbiology*, 2004, 95, p. 231-234.
- Voitoux E., Lafarge V., **Lombard B.**, 2001. *Inter-laboratory trial on the detection of E.coli O157 in liquid milk : final report of the CRL "Milk"*. Maisons-Alfort : AFSSA-LERQAP, 2001, 28 p.
- Voitoux E., **Lombard B.**, 2002a. *Inter-laboratory trial on the detection of Salmonella in liquid milk : final report of the CRL « Milk »*. Maisons-Alfort : AFSSA-LERQAP, 2002, 33 p.
- Voitoux E., **Lombard B.**, 2002b. *Inter-laboratory trial on the counting of somatic cells in liquid raw milk : final report of the CRL « Milk »*. Maisons-Alfort : AFSSA-LERQAP, 2002, 35 p.
- Voitoux E., Lafarge V., Collette C., **Lombard B.**, 2002c. Applicability of the draft standard method for the detection of *Escherichia coli* O157 in dairy products. *International Journal of Food Microbiology*, 2002, 77, p. 213-221.

Wilrich P. 2004a. *Robust estimates of the theoretical standard deviation to be used in interlaboratory precision experiments*. Document de travail pour l'ISO/TC 34/SC 9/WG 2, non publié.

Wilrich P. 2004b. *The determination of precision of measurement methods with qualitative results by interlaboratory experiments*. Document de travail pour l'ISO/TC 34/SC 9/WG 2, non publié.

Youden W.J., Steiner E.H., 1987. *Statistical Manual of the Association of Official Analytical Chemists – Qualitative and Quantitative Checklist*. Gaithersburg, MD : AOAC, 1987, 88 p.