

LES _____ DIFFERENTES APPROCHES D'ÉVALUATION DES PERFORMANCES DES METHODES DE DEPISTAGE

Exigences pour les méthodes de dépistage des résidus de médicaments vétérinaires

La méthode de dépistage est la première méthode qui est appliquée à l'échantillon à analyser, le but étant d'établir la présence ou l'absence de résidus de médicaments vétérinaires. Une méthode de dépistage doit être spécifique ou sélective en fonction de la cible, qualitative ou quantitative, simple, rapide et être la plus performante possible. En ce qui concerne les matrices complexes (i.e. le muscle, le rein), si nécessaire, l'extraction doit être très simple et rapide. Enfin, ces méthodes doivent avoir un faible coût, une capacité d'analyse élevée et sont utilisées pour passer au crible un grand nombre d'échantillons pour détecter des résultats non conformes potentiels. Ces méthodes sont spécialement conçues pour éviter les faux résultats conformes.

Les méthodes de dépistage pour les résidus de médicaments vétérinaires sont définies dans la décision 2002/657/CE de la Commission Européenne [13] comme "les méthodes utilisées pour détecter la présence d'un analyte ou une classe d'analytes au niveau d'intérêt ». Le "niveau d'intérêt" est généralement soit la limite maximale réglementaire (limite maximale de résidus (LMR), la limite minimale de performance requise (LMPR)) [13] ou un "niveau d'action. La décision européenne 2002/657/CE [13], en application de la directive 96/23/CE [2] définit la limite minimale de performance requise (LMPR) comme la teneur minimale en analyte présente dans un échantillon, qui doit au moins être détectée et confirmée. La LMPR est un paramètre de la performance technique auquel les laboratoires doivent se conformer lors de l'essai pour détecter la présence d'une substance interdite notamment, de son métabolite ou d'un marqueur. Cette limite est destinée à l'harmonisation des performances analytiques des méthodes applicables aux substances pour lesquelles aucune limite autorisée n'a été établie. Donc, la LMPR n'est pas une limite de sécurité, mais une limite analytique. Des LMPR ont été fixées pour le chloramphénicol, les métabolites de nitrofuranes et les colorants par exemple. Au contraire, la Limite Maximale de résidus (LMR) est une limite de sécurité. Elle est fixée en fonction de différents dangers (toxicologiques, microbiologiques, pharmacologiques, ...). La LMR est fixée pour une substance antibiotique et/ou ses métabolites et est spécifique de chaque matrice (lait, viande, etc). Lors de l'administration d'un médicament à un animal de rente, un temps d'attente est fixé sur la base des études de pharmacocinétique des résidus. Ce temps d'attente doit être respecté par l'éleveur avant l'abattage des animaux ou la commercialisation du lait par exemple. Quand le temps d'attente est bien respecté et à condition que les posologies aient été respectées aussi, la concentration en résidus de médicaments dans la matrice concernée sera inférieure à la LMR pour 95 % des animaux. Dans ce cas, il n'y aura pas de danger pour le consommateur (toxicité, allergies, etc) à ingérer ces aliments. Les 5 % restants expliquent que les produits d'origine animale doivent être contrôlés pour la présence de résidus d'antibiotiques (plans de surveillance).

Deux types de méthodes peuvent être développés en fonction des substances cibles. Dans le cas des substances interdites (eg. le chloramphénicol, les métabolites de nitrofuranes ou les colorants), la méthode doit être spécifique pour le dépistage d'un antibiotique seul et les concentrations

recherchées doivent être les plus faibles possibles. Dans le cas des substances à LMR, idéalement une méthode multi - résidus doit être utilisée, capable de détecter plusieurs familles d'antibiotiques en même temps et, si possible, au niveau des LMR respectives. De plus, les niveaux de résidus à rechercher pour les substances à LMR sont toujours plus élevés que les niveaux à détecter pour les substances interdites.

Même si, dans un cadre réglementaire, un résultat positif avec la méthode de dépistage est nécessairement confirmé par une méthode physico - chimique, une méthode de dépistage ne devrait pas produire de faux résultats positifs qui entraînent des analyses de confirmation coûteuses ; mais plus important encore, il ne devrait y avoir aucun résultat faux négatif puisqu'ils entraînent que des échantillons contaminés passent dans l'alimentation.

Avant la validation d'une méthode de dépistage et selon le principe de la méthode, il est nécessaire de définir la portée de la méthode :

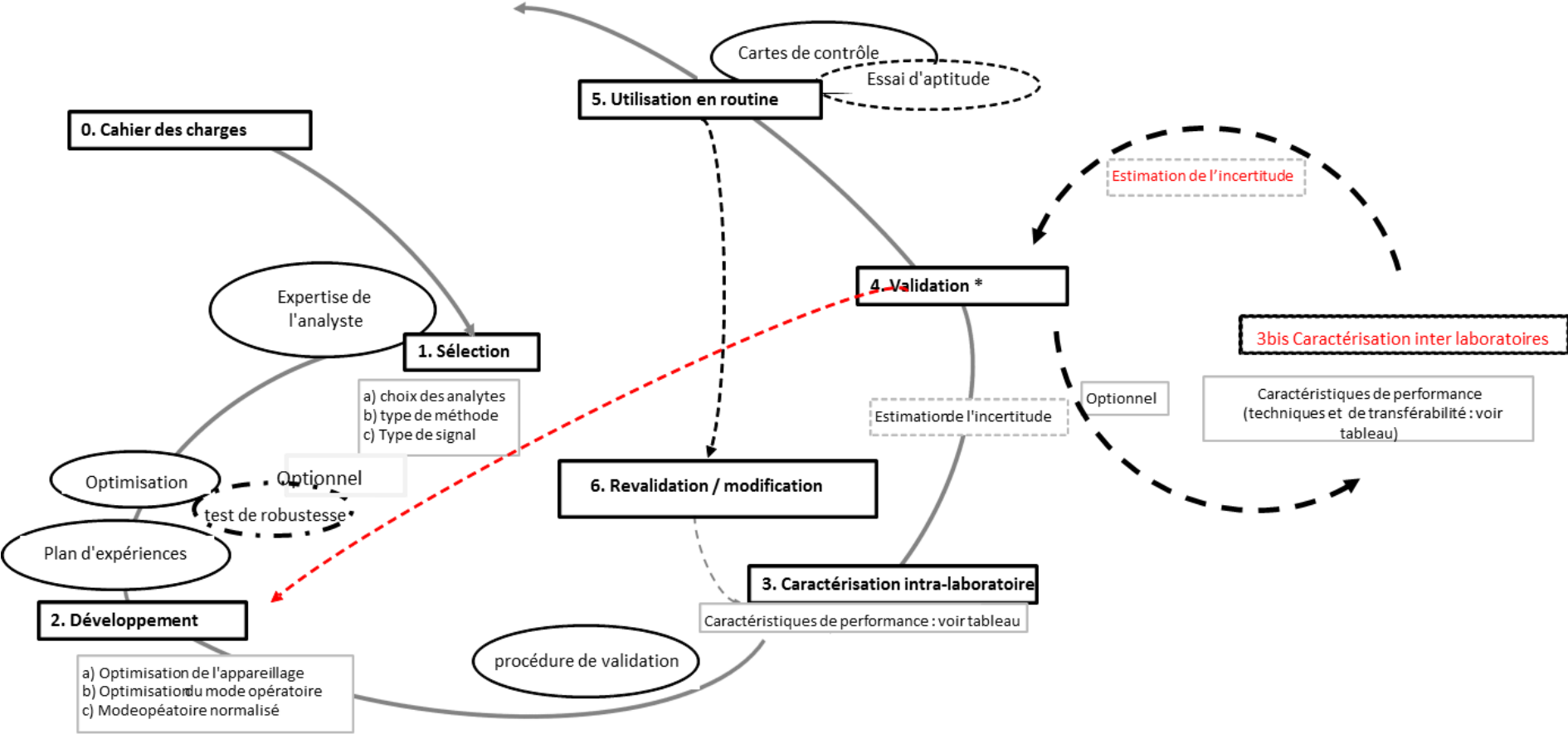
- l'analyte ou le groupe d'analytes à détecter,
- la gamme de concentration à détecter,
- la liste des matrices qui peuvent-être analysées.

Une méthode analytique doit pouvoir évoluer, suivant différentes étapes, qui sont représentées par le cycle de vie de la méthode (**Figure 27**) [243, 244]. En fonction de la demande d'un client (étape 0), une méthode va être développée et optimisée (étape 2), en commençant par sélectionner certains facteurs (étape 1), comme le type de méthode adapté pour la détection des analytes cibles. Ensuite, la méthode va être caractérisée en intra-laboratoire et, au besoin, en inter-laboratoires (étape 3). Enfin, la méthode va être validée, au regard de l'usage attendu (étape 4). À l'issue de ces étapes, l'utilisation en routine de la méthode peut être envisagée (étape 5). La revue périodique de la méthode peut donner lieu à un besoin de revalidation ou d'un nouveau développement (étape 6).

3.2. Caractérisation des performances et validation des méthodes de dépistage

La validation des méthodes de dépistage est d'une grande importance puisque ces méthodes constituent la première étape des plans de surveillance et plans de contrôle dans tous les pays. Seuls les résultats positifs à l'étape de dépistage doivent être confirmés, identifiés et quantifiés [13]. La validation des méthodes analytiques d'essais est une exigence normative pour l'accréditation. En effet, la norme NF EN ISO/CEI 17025 [245] appliquée dans le cadre de l'accréditation des laboratoires d'essais exige que les méthodes non normalisées ou les méthodes développées par le laboratoire soient validées avant emploi. La définition de la validation de cette norme est la suivante : « La validation est la confirmation par examen et l'apport de preuves objectives du fait que les exigences particulières en vue d'une utilisation prévue déterminée sont remplies ». De plus, la validation des méthodes de dépistage pour les résidus de médicaments vétérinaires est une exigence réglementaire européenne [13].

Figure 27. Cycle de vie d'une méthode analytique.



Il existe deux types de validation :

- La validation inter-laboratoires (EILV) ou étude collaborative,
- La validation intra-laboratoire.

L'EILV permet de vérifier les paramètres de performance obtenus lors de la validation intra-laboratoire et de déterminer la reproductibilité, et si besoin la robustesse de la méthode. Contrairement au domaine de la microbiologie alimentaire, la validation inter-laboratoires est très peu utilisée dans le domaine du dépistage (en dehors des études de validation pour la certification AFNOR) et de la confirmation des antibiotiques dans les aliments, pour deux raisons principales.

Tout d'abord, ce type de validation est coûteux et demande de la part du laboratoire organisateur certaines compétences, différentes de celles nécessaires à une validation intra-laboratoire. En effet, la validation intra-laboratoire demande des compétences techniques et analytiques pour la mise en place de la méthode et des expériences nécessaires à la validation. La validation inter-laboratoires exige des compétences organisationnelles et de coordination.

De plus, quand la méthode est destinée à être utilisée dans un seul laboratoire, la validation inter-laboratoires est inutile. Un domaine où la validation inter-laboratoires est plus pertinente est l'industrie agro-alimentaire. En effet, plusieurs sites d'une même entreprise peuvent utiliser la même méthode. Dans le domaine du contrôle officiel et dans le cadre d'un réseau de laboratoires, comme celui des LNR, les méthodes peuvent parfois être transférées du laboratoire développeur vers les autres laboratoires. Dans ce cadre, une validation inter-laboratoires peut présenter un intérêt. Toutefois, en pratique, les laboratoires de référence (LNR ou LRUE) dans le domaine des résidus de médicaments vétérinaires, organisent très peu de validations inter-laboratoires. Après un transfert de méthodes au réseau, tout d'abord, le laboratoire qui reçoit la méthode doit réaliser une validation réduite, aussi appelée de transfert. Ce type de validation est décrit dans le guide de validation des méthodes de dépistage (2010). De plus, le choix est fait de vérifier la compétence de ces laboratoires grâce à l'organisation d'essais d'aptitude, et non l'organisation d'essais inter-laboratoires de validation (EILV). De même qu'un EILV, ces deux étapes permettent d'assurer la comparabilité des résultats entre plusieurs laboratoires utilisant la même méthode.

Nous allons nous focaliser sur la validation intra-laboratoire, qui est la pratique de loin la plus courante dans le domaine du dépistage des résidus de médicaments vétérinaires, et ses différentes approches. Etant donné que la norme NF EN ISO/CEI 17025 ne fait aucune recommandation concernant le ou les protocoles de validation à utiliser, le choix du protocole est lié à d'autres documents réglementaires ou normatifs. Les exigences de validation peuvent être différentes selon les pays, selon les domaines d'application et selon le type de méthodes (*eg.* qualitatives ou quantitatives). En outre, les protocoles de validation peuvent présenter des statuts différents : norme, réglementation nationale, européenne ou internationale, ou encore lignes directrices. Quelques exemples de protocoles de validation sont présentés dans le **Tableau 13**.

Tableau 13. Protocoles de validation au niveau national, européen et international.

Niveau	R/G/N ¹	Document	Année	QL et/ou QT ²	CP ou MA/MR ³	Analytes	Référence
National	N	AFNOR NF V 03-110	2010	QT	CP	Produits agricoles et alimentaires	[246]
	N	AFNOR XP V 03-111	1995	QL	MA/MR		[247]
	G	Rapport d'une commission SFSTP ⁴	1992	QT	CP	Produits pharmaceutiques	[248, 249]
	G	Rapports d'une commission SFSTP	2003 à 2008	QT	CP		[250-253]
Europe	G	EURACHEM	1998	QL/QT	CP	Tous domaines d'analyse	[254]
	G	Procédure NMKL ⁵	2009	QL/QT	MA/MR	Pathogènes microbiens	[255]
	R	Décision 2002/657/CE	2002	QL/QT	CP	Annexe I de la directive 96/23/CE	[13]
	G	Guide de validation	2010	QL/QT	CP		[256]
USA	R	FDA ⁶	2013	QT	MA/MR et CP	Principes actifs et/ou métabolites	[257]
International	G	IUPAC ⁷ /AOAC ⁸	2002	QT	CP	Tous domaines d'analyse	[258]
	G	ICH ⁹	1996	QT	CP	Médicaments humains	[259]
	G	VICH ¹⁰	1999	QT	CP	Médicaments vétérinaires	[260]
	N	ISO ¹¹ 13969 (IDF ¹² 183)	2003	QL	CP	Résidus d'antimicrobiens	[261]
	N	ISO 18330 (IDF 188)	2003	QL/QT	CP		[262]
	N	(N) ISO 16140	2003	QL/QT	MA/MR	Pathogènes microbiens	[263, 264]
	N	NF EN FDIS 16140-2	2015	QL/QT	MA/MR et CP		[265, 266]

¹ Réglementation (R), guide (G) ou norme (N) ; ² Méthodes qualitatives (QL) ou quantitatives (QT) ; ³ Caractéristiques de performance (CP) ou comparaison d'une méthode alternative (MA) avec une méthode de référence (MR) ; ⁴ Société Française des Sciences et Techniques Pharmaceutiques (SFSTP) ; ⁵ Nordic Committee on Food Analysis (NMKL) ; ⁶ Food and Drug Administration (FDA) ; ⁷ International Union of Pure and Applied Chemistry (IUPAC) ; ⁸ Association of Official Analytical Chemists (AOAC) ; ⁹ International Conference on Harmonisation (ICH) ; ¹⁰ Veterinary International Conference on Harmonization (VICH) ; ¹¹ International Standard Organization (ISO) ; ¹² International Dairy Federation (IDF).

En fonction du statut du protocole, l'application des règles établies est plus ou moins contraignante. Par exemple, une décision européenne, comme la décision 2002/657/CE, doit être appliquée de manière complète. Beaucoup d'autres documents sont des guides ou des lignes directrices. Ces documents proposent des recommandations, mais ils peuvent être appliqués de manière partielle. Même lorsqu'il s'agit de normes, elles ne sont contraignantes que si une législation ou un référentiel (eg. AFNOR et NF EN FDIS 16140-2 [266]) oblige dans un domaine spécifique à utiliser cette norme. Aux États-Unis, la réglementation de la FDA, telles que les bonnes pratiques de laboratoire (BPL) exige que les méthodes d'analyse soient validées avant et pendant l'utilisation en routine. Il n'y a pas de réglementation spécifique sur les validations de méthodes, mais la FDA, ainsi que d'autres organismes et groupes de travail de l'industrie ont élaboré des lignes directrices pour la validation des méthodes.

En Europe, il existe plusieurs approches pour la validation des méthodes de dépistage des résidus de médicaments vétérinaires. Les approches peuvent être différentes en fonction du type de méthodes (qualitatives ou quantitatives), mais aussi en fonction du cadre dans lequel la validation doit être réalisée (eg. cadre réglementaire, certification). Trois approches utilisées pour la validation de méthodes pour le dépistage des résidus d'antibiotiques dans les aliments vont être présentées. Nous allons voir que deux grands types d'approche peuvent être utilisés : une approche relative (par comparaison de deux méthodes) et une approche absolue (une seule méthode).

3.2.1.1. La comparaison de méthodes :

Pourquoi valide-t-on les méthodes alternatives ? Le plus souvent, en microbiologie alimentaire, mais aussi en résidus, les méthodes dites « de référence » sont des méthodes anciennes, longues et parfois avec des limites de détection élevées. Les méthodes alternatives sont des méthodes plus innovantes, le plus souvent commerciales, plus rapides et moins lourdes à mettre en place, parfois automatisées. Le choix a été fait en microbiologie alimentaire de ne pas normaliser les méthodes alternatives, ce qui représente un processus long (3-4 ans), mais de les valider en comparaison avec la méthode de référence, validation réalisée par une tierce partie, en un à deux ans. En France, l'approche par comparaison d'une méthode alternative à une méthode de référence est préconisée par l'Association Française de Normalisation (AFNOR) pour la certification des kits commerciaux en microbiologie alimentaire et par extension pour les kits commerciaux pour le dépistage des résidus d'antibiotiques et autres substances à effet antibactérien dans les denrées alimentaires issues des animaux. Les études de validation pour obtenir la certification sont donc basées sur la norme EN NF EN FDIS 16140-2 [266]. A ce jour, les systèmes de certification AFNOR (France), MicroVal (NEN, Pays-Bas) et NordVal (NMKL, pays nordiques) appliquent la norme NF EN FDIS 16140-2 [266].

En 2005, un référentiel spécifique aux antibiotiques a été rédigé au sein de l'AFNOR par un groupe de travail, auquel l'Anses Fougères a participé [267]. Sur la base de l'ancienne norme ISO 16140 de 2003 [263], la validation d'une méthode alternative a été définie comme la « démonstration selon laquelle ladite méthode apporte un niveau de confiance approprié, sur le fait que les résultats obtenus

avec la MA sont comparables aux résultats obtenue avec la MR ». A cette date, le choix a été fait de conserver le principe de base de la comparaison de méthodes et les caractéristiques de performance à déterminer étaient similaires à celles de la norme. L'approche caractéristiques et critères de performance n'a pas été retenue. De plus, le référentiel pour les antibiotiques a aussi conservé les deux grandes étapes d'une certification d'un kit en microbiologie alimentaire: une étude préliminaire (intra-laboratoire) et une étude collaborative (inter-laboratoires). Les essais mis en œuvre dans le cadre d'une certification AFNOR sont réalisés par un laboratoire expert, indépendant, nommé par l'AFNOR. Cette approche est applicable aux méthodes qualitatives et quantitatives.

3.2.1.1.1. Etude préliminaire

L'objectif de l'étude préliminaire est de caractériser la méthode alternative et de comparer les performances de la méthode alternative avec celles d'une méthode prise en référence (**Figure 28**).

Praticabilité du kit

La praticabilité du kit qui permet de juger de la possibilité et de la facilité d'utilisation du kit en routine. Elle est déterminée suite à l'examen et au contrôle de 13 critères. Pour chacun de ces critères le mode de communication de ce critère auprès de l'utilisateur et le mode de contrôle de ce critère ont été définis. L'étude de praticabilité va consister à vérifier par exemple si le volume des réactifs fournis est suffisant et s'ils sont prêts à l'emploi ou non. La praticabilité est aussi liée à la nécessité ou non d'avoir des instruments ou des locaux spécifiques pour mettre en œuvre la méthode alternative.

Etude comparative de la méthode alternative et de la méthode de référence

Différentes caractéristiques vont être déterminées à partir des essais suivants pour chacune des méthodes, puis ces mêmes essais vont permettre de comparer les deux méthodes.

Caractéristiques absolues pour chacune des méthodes

Limite de détection / taux de faux négatifs

Pour chacune des méthodes, la (ou les) limite(s) de détection est (ou sont) déterminée(s) après l'analyse de matériaux contaminés (artificiellement ou non) avec une (ou des) molécule(s) à activité antibiotique à plusieurs niveaux de concentration. Le choix des analytes à valider ainsi que des concentrations doit se faire avant de démarrer la validation, lors de la présentation du projet d'étude préliminaire (**Figure 29**). Le choix est lié au type de méthode alternative, mais doit aussi tenir compte de leur fréquence d'utilisation en médecine vétérinaire ou du risque de survenue sous la forme de résidus dans la matrice concernée, si ces données sont connues. De plus, il faut veiller à respecter la proportionnalité entre la taille de chaque famille et le nombre d'antibiotiques testés dans cette famille. Le choix des concentrations doit se baser sur les limites réglementaires (eg. LMR) des différents antibiotiques dans la matrice d'intérêt, sur la (ou les) limite(s) de détection annoncée(s) par le fournisseur de kits et/ou sur des études précédentes réalisées dans d'autres laboratoires si elles existent.

Figure 28. Caractéristiques de performance à déterminer lors de l'étude préliminaire.

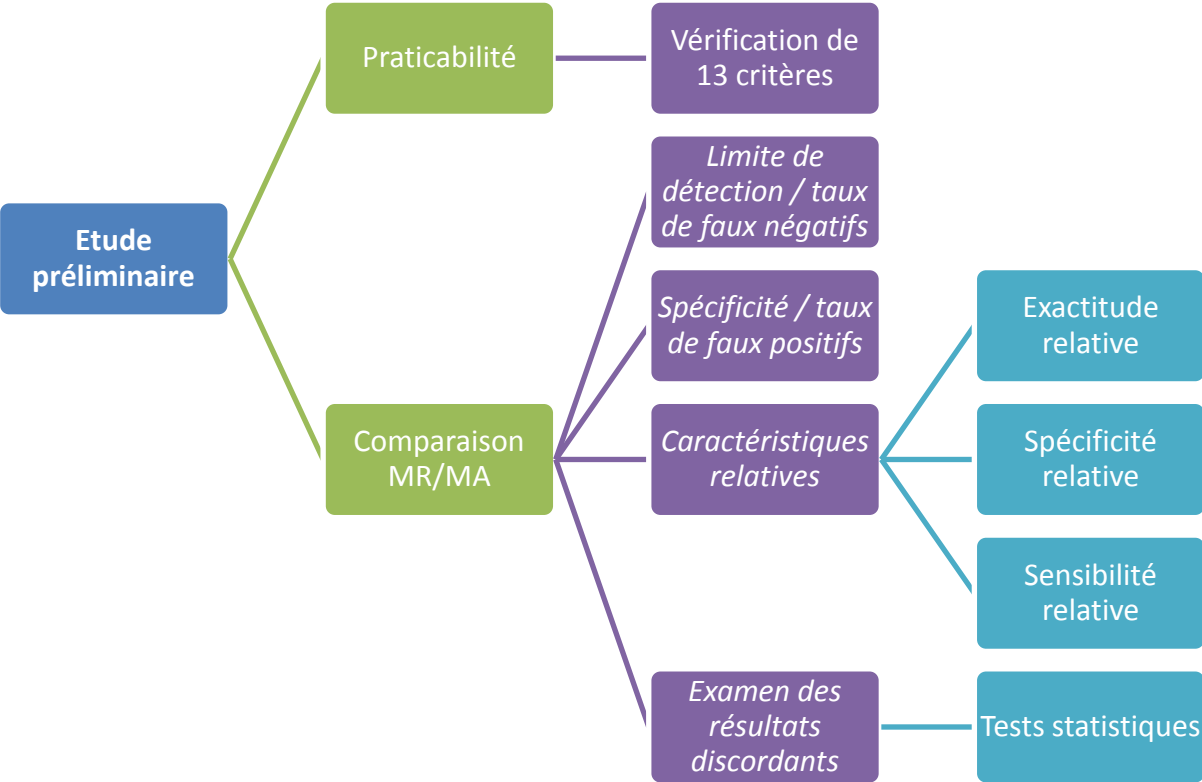
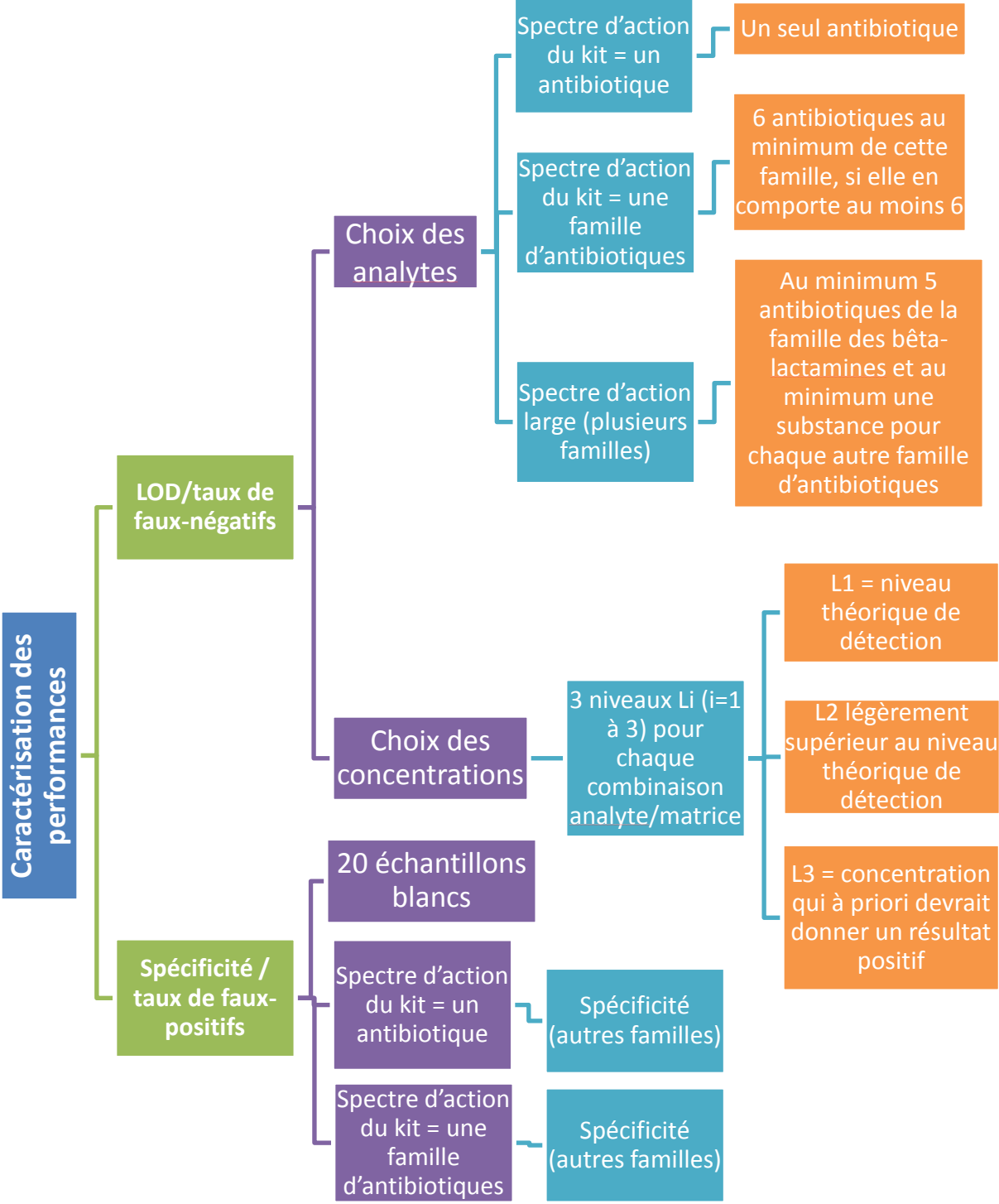


Figure 29. Construction du plan d'expérience pour la détermination de la spécificité et de la limite de détection.



Les échantillons sont codifiés aléatoirement pour réaliser toutes les analyses en aveugle. Chaque concentration d'antibiotique (pour une matrice) sera préparée en 5 exemplaires et chacun sera analysé une fois par la méthode alternative et une fois par la méthode de référence.

La limite de détection correspond à la concentration la plus basse qui donne un résultat positif ou douteux pour chacun des 5 réplicats des essais.

Le taux de faux négatifs est égal au nombre de résultats négatifs obtenus pour des échantillons contaminés (faux-négatifs FN) divisé par le nombre total d'échantillons négatifs (ce même nombre plus le nombre d'échantillons blancs donnant un résultat négatif (vrais négatifs VN)) et multiplié par 100.

$$\text{Taux de faux négatifs} = \frac{FN}{FN+VN} * 100 \quad \text{Equation n°1}$$

Spécificité / taux de faux positifs

L'objectif de ces essais est de démontrer, pour chacune des méthodes, la capacité à garantir que des échantillons donnant des résultats négatifs sont de vrais négatifs (exempt de résidus). Pour cela, 20 échantillons de matrice blanche (dépourvue de résidus d'antibiotiques ou autres substances à activité antibactérienne) seront analysés en double par chacune des méthodes (**Figure 29**). Dans le cas où le spectre d'action du kit est limité à une famille d'antibiotiques, il faudra en plus préparer et analyser des matrices blanches contenant des substances à activité antibiotique issue d'autres familles.

La spécificité correspond à la probabilité que la méthode donnera un résultat négatif pour un échantillon blanc (reconnu exempt de substances à activité antibiotique ou contaminé avec des substances qui ne sont pas détectées par cette méthode, si elle est spécifique d'un antibiotique ou d'une famille).

Le taux de faux positifs correspond au nombre de résultats positifs (faux-positifs FP) obtenus pour des échantillons blancs divisé par le nombre total d'échantillons positifs (ce même nombre plus le nombre d'échantillons contaminés donnant un résultat positif (vrais positifs VP)) et multiplié par 100.

$$\text{Taux de faux positifs} = \frac{FP}{FP+VP} * 100 \quad \text{Equation n°2}$$

Caractéristiques relatives

A partir de l'ensemble des résultats des essais présentés ci-dessus (matériaux contaminés et matériaux blancs), la comparaison des performances des 2 méthodes va être réalisée au moyen de caractéristiques relatives. Le **Tableau 14** présente la synthèse des résultats des deux méthodes, qui va permettre la comparaison.

Tableau 14. Couples de résultats des méthodes de référence et alternative.

Réponse	Méthode de référence positive (R+)	Méthode de référence négative (R-)
Méthode alternative positive (A+)	Accord positif +/+ (PA)	Déviaton positive -/+ (PD) (R-/A+)
Méthode alternative négative (A-)	Déviaton négative +/- (ND) (A-/R+)	Accord négatif -/- (NA)

L'accord positif (PA) et l'accord négatif (NA) correspondent à un résultat identique avec la MR et avec la MA (positif et négatif respectivement). Une déviation positive (PD) est observée quand la méthode alternative donne un résultat positif tandis que la méthode de référence donne un résultat négatif. Une déviation positive devient un résultat **faux positif** quand on prouve que le résultat vrai est négatif. Une déviation positive devient un vrai résultat positif quand on prouve que le résultat vrai est positif. Une déviation négative (ND) est observée quand la méthode alternative donne un résultat négatif tandis que la méthode de référence donne un résultat positif. Une déviation négative devient un résultat **faux négatif** quand on prouve que le vrai résultat est positif. Une déviation négative devient un vrai résultat négatif quand on prouve que le vrai résultat est négatif.

Trois caractéristiques relatives sont alors déterminées, à partir de ce tableau de contingence.

L'exactitude relative (AC) est définie comme le niveau de correspondance entre la réponse obtenue avec la méthode de référence et la réponse obtenue avec la méthode alternative sur les mêmes échantillons.

$$AC = \frac{PA+NA}{N^1} * 100 \quad \text{Equation n°3}$$

La sensibilité relative (SE) est la capacité de la méthode alternative à détecter l'analyte quand la méthode de référence le détecte.

$$SE = \frac{PA}{N+2} * 100 \quad \text{Equation n°4}$$

La spécificité relative (SP) est la capacité de la méthode alternative à ne pas détecter l'analyte quand la méthode de référence ne le détecte pas.

$$SP = \frac{NA}{N-3} * 100 \quad \text{Equation n°5}$$

Où : ¹N = NA + PA + PD + ND : nombre total d'échantillons

²N- est le nombre total d'échantillons négatifs obtenus avec la méthode de référence (NA + PD)

³N+ est le nombre total d'échantillons positifs échantillons obtenus avec la méthode de référence (PA + ND).

Ensuite, les intervalles de confiance (IC) sont calculés pour chaque pourcentage p de AC, de SE et de SP et associés au nombre d'échantillons soumis à l'essai. Deux calculs différents sont proposés, en fonction de la valeur de la caractéristique.

- Si 10% < p < 90%, l'intervalle de confiance **bilatéral** à 95% approximatif est calculé :

$$IC\ 95\% \approx p \mp 2 * \sqrt{p(1-p)/n} \quad \text{Equation n°6}$$

avec n=N, N+, N- respectivement pour p (%) = AC, SE, SP.

- Si p ≥ 90%, la Limite de confiance inférieure (Low Confidence Limit LCL) à 95% (limite **unilatérale**) est déterminée à partir de la table de la loi binomiale pour un n et un p donné.

Examen des résultats discordants

Le test de Mc Nemar est utilisé afin de déterminer si les méthodes sont différentes ou non, en termes de sensibilité et spécificité. Ce test permet de comparer deux échantillons appariés. Ici, les deux échantillons représentent les résultats de chacune des deux méthodes.

Le nombre de résultats discordants Y entre les deux méthodes correspond à la somme des déviations positive et négative ($Y = PD + ND$). Quand Y est inférieur à 6, aucun test statistique n'est disponible. A l'inverse, quand les valeurs de PD et ND sont élevées et quasiment équivalentes, le test de Mc Nemar ne peut montrer aucune différence statistique entre les 2 méthodes. Toutefois, le laboratoire expert doit tenter d'expliquer les causes de ces valeurs de PD et ND élevées.

Ensuite, deux cas se présentent :

- Si $6 \leq Y \leq 22$, m est déterminée comme la plus petite des 2 valeurs de PD et ND. Pour une valeur de Y donnée, un $M = \text{Max}(m)$ avec $\alpha < 0.05$ est déterminé à partir de la table de la loi binomiale. Si $m \leq M$, les 2 méthodes sont différentes à $\alpha < 0.05$ (bilatéral).
- Si $Y > 22$, le test de McNemar avec la distribution de Chi deux pour 1 degré de liberté (ddl) est appliqué.

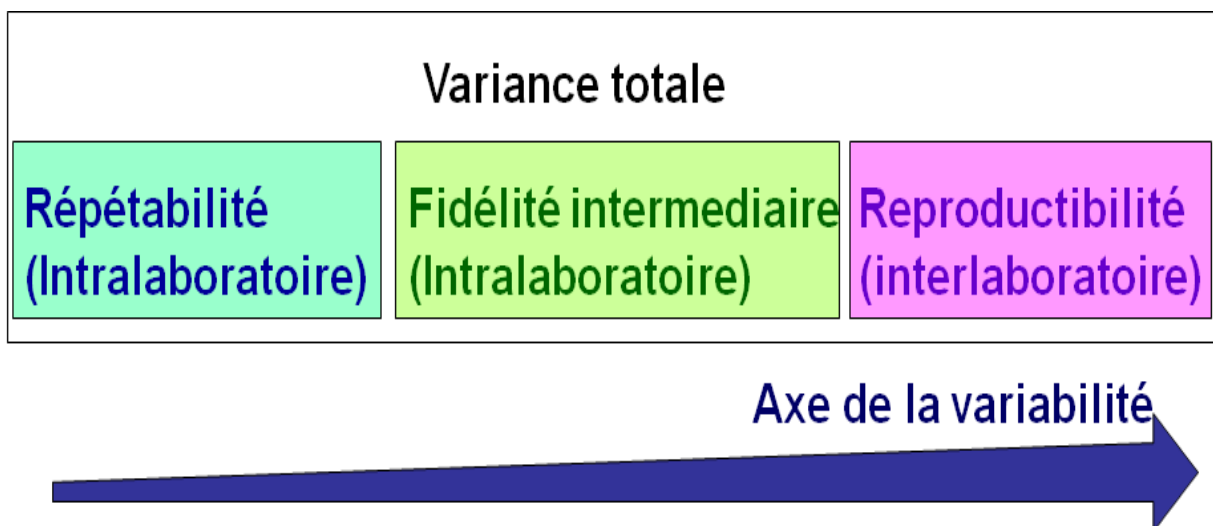
$$\chi^2 = \frac{d^2}{Y}, \text{ avec } d = |PD - ND| \quad \text{Equation n°7}$$

Pour un Y donné, d doit être supérieur ou égal à la valeur donnée dans une table pour conclure que les 2 méthodes sont différentes pour $\alpha < 0.05$ (bilatéral).

3.2.1.1.2. Etude collaborative

Dans cette deuxième phase, seule la méthode alternative (le kit) est mise en place par les laboratoires participants. L'objectif de l'étude collaborative (ou EILV) est de déterminer la fidélité de la méthode, dans des conditions de répétabilité et de reproductibilité inter-laboratoires (**Figure 30**). Comme nous avons choisi dans cette thèse de nous focaliser sur la validation intra-laboratoire, cette deuxième phase de l'étude de validation d'une méthode alternative ne sera pas développée ici.

Figure 30. Participation de la répétabilité, de la fidélité intermédiaire et de la reproductibilité dans la détermination de la fidélité.



3.2.1.1.3. Conclusions sur l'approche comparaison de méthodes

Nous n'avons trouvé aucun avantage à utiliser l'approche de validation par comparaison de méthodes pour les méthodes de dépistage des résidus d'antibiotiques.

Quel que soit le domaine d'application des méthodes, l'inconvénient majeur de l'approche par comparaison de méthodes est le suivant : la méthode alternative est caractérisée par comparaison avec une autre méthode. Elle n'est donc pas caractérisée individuellement, et les probabilités d'erreur de cette méthode (*eg.* risque β de produire des résultats faux-négatifs) ne sont pas déterminées.

Les inconvénients spécifiques au domaine des résidus de cette approche sont multiples :

- il n'existe pas de méthodes normalisées, dites de référence, par opposition à la microbiologie alimentaire. Le choix de la méthode de référence est donc fait par le laboratoire expert. En général, la méthode de référence utilisée est la méthode officielle française, utilisée dans les plans de contrôle et plans de surveillance (PSPC) pour la matrice concernée. Le principe de la méthode de référence est souvent différent de celui du kit, la préparation de l'échantillon est parfois aussi différente, ce qui complique la mise en place de l'étude comparative.
- le problème se pose de l'obsolescence des méthodes de référence. L'évolution des méthodes analytiques, y compris en dépistage, est continue ; nous l'avons vu avec les méthodes de type biocapteurs. Ces méthodes dites alternatives sont de fait plus performantes que les MR. Le plus souvent, la MA est plus performante que la MR. La détermination de caractéristiques de performance relatives et donc le désaccord obtenu entre les deux méthodes peut pénaliser une méthode alternative plus performante que la MR.
- Aucun critère n'est fixé pour statuer sur la validation de la méthode en ce qui concerne la détermination de la spécificité, de la sensibilité, du taux de faux-négatifs et du taux de faux-positifs de la méthode alternative. La seule possibilité est de comparer ces paramètres déterminés pour la MR avec ceux obtenus avec la MA. En outre, dans la pratique, pour le dépistage des résidus d'antibiotiques, le laboratoire expert dans son rapport compare les paramètres obtenus, soit avec les critères de performance réglementaires (*eg.* Taux de faux-négatifs inférieur ou égal à 5 % [13]), soit avec les limites réglementaires européennes (*eg.* LMR).
- Dans le domaine de la microbiologie alimentaire, si un kit est certifié AFNOR, le laboratoire utilisateur n'a pas besoin de le valider. Ce n'est pas le cas dans le domaine des résidus puisque le protocole de validation de l'AFNOR est complètement différent des exigences de la décision 2002/657/CE [13] et du guide de validation Européen (2010) [256].

Après l'approche relative, basée sur la comparaison de méthodes, nous allons nous intéresser à l'approche absolue. Dans cette approche, la validation comporte deux étapes :

- Tout d'abord la caractérisation des performances de la méthode,
- Ensuite l'étape de validation proprement-dite, qui consiste à comparer les valeurs des paramètres de performances obtenus pour les caractéristiques étudiées, aux critères de validation prédéterminés, soit de manière globale (3.2.1.2.), soit critère par critère (3.2.1.3.).

3.2.1.2. L'approche globale : le profil d'exactitude :

L'approche globale consiste à déterminer une caractéristique de performance. Ensuite, la validité de la méthode est statuée par rapport à un intervalle d'acceptabilité (eg. réglementaire). Une commission de la Société Française des Sciences Techniques et Pharmaceutiques (SFSTP) a développé un concept de validation basé sur l'erreur totale et le profil d'exactitude pour les méthodes quantitatives [250-253, 268].

Le profil d'exactitude est un modèle statistique utilisé pour valider des **méthodes quantitatives**.

En outre, une approche globale a été développée pour les méthodes qualitatives binaires. Elle est fondée sur la probabilité de détection (POD). Cette approche a été présentée pour la première fois par Wilrich en 2009 [269], puis Wehling en 2011 [270].

3.2.1.2.1. Le profil d'exactitude :

La caractéristique de performance pour cette approche est l'exactitude, c'est-à-dire la combinaison de la justesse (erreur systématique) et de la fidélité (erreurs aléatoires). En utilisant les mesures de répétabilité et de fidélité intermédiaire, le profil d'exactitude permet de calculer un intervalle où sera situé une proportion connue de mesures. Lorsque cet intervalle est comparé à un intervalle d'acceptabilité défini par l'utilisateur ou réglementairement, il est possible de décider simplement si une méthode est valide ou non [271].

a. Construction des plans d'expérience

Deux plans d'expérience sont construits (**Tableau 15**) :

- Un plan d'étalonnage (standards en solution et/ou en matrice), qui sert à estimer la fonction de réponse de la méthode.

Le plan d'étalonnage n'est nécessaire que pour les méthodes indirectes, c'est-à-dire les méthodes pour lesquelles un étalonnage préalable est nécessaire pour calculer les concentrations des échantillons inconnus.

Pour réaliser le plan d'étalonnage, il faut définir les niveaux de concentration k' , le nombre de séries I et le nombre de répétition par niveau J' . Le choix de ces différents paramètres doit s'approcher au maximum des conditions des analyses de routine ultérieures. Le nombre k' peut être différent du nombre k du plan de validation et égal à 1 si on est sûr que la droite passe par 0. Le nombre k' dépend du type de fonction de réponse de la courbe d'étalonnage. Le modèle d'étalonnage le plus souvent utilisé pour les méthodes de dépistage immunologiques (e.g. ELISA) est la fonction logistique à 4 paramètres. Dans ce cas, le nombre de niveaux de concentration k' pour le plan d'étalonnage est de 5 niveaux minimum. Il est recommandé de prendre au moins un niveau de plus que le nombre de paramètres de la fonction d'étalonnage, afin d'avoir la meilleure estimation statistique des paramètres de la fonction de réponse. Ensuite, il faut faire I séries de mesures. Enfin, pour chaque série, il faut effectuer J' répétitions. Le nombre d'essais à réaliser pour le plan d'étalonnage est donc égal à $I \cdot J' \cdot k'$ essais.

- Un plan de validation (standards supplémentés dans la matrice), qui sert à déterminer les caractéristiques de validation.

Tableau 15. Construction des plans d'expérience (étalonnage et validation).

	Plan d'étalonnage	Plan de validation
Niveaux de concentration	k' niveaux, nombre $k' \geq 1$	k niveaux, $k \geq 3$
Nombre de séries	l séries réparties sur l jours, nombre $l \geq 3$	l séries réparties sur l jours, nombre $l \geq 3$
Nombre de répétitions	J' répétitions, nombre recommandé $J' \geq 2$, au choix pour chaque étalon et pour chaque série	J répétitions, nombre $J \geq 2$, constant pour chaque série
Nombre total d'essais	$l * J' * k'$ essais	$l * J * k$ essais

Pour réaliser le plan de validation, il faut d'abord choisir les niveaux de concentration à valider et qui vont couvrir le domaine d'application de la méthode. Ensuite, il faut faire des séries de mesures et pour chaque série, il faut effectuer des répétitions. Ensuite, il faut faire I séries de mesures, autant que pour le plan d'étalonnage. Enfin, pour chaque série, il faut effectuer J répétitions. Ce nombre J peut être différent du nombre J' du plan d'étalonnage.

Le nombre d'essais à réaliser pour le plan de validation est donc égal à $I \cdot J \cdot k$ essais. Plus le nombre d'essais est élevé, plus la confiance statistique dans l'estimation des critères de validation est élevée. De même, plus le nombre d'essais est élevé, plus les intervalles de tolérance sont diminués et plus la confiance dans les résultats de la validation augmente.

Les essais de validation doivent être réalisés les mêmes jours que les essais d'étalonnage, afin d'utiliser les fonctions de réponse correspondantes. De plus, toutes les mesures doivent être réalisées dans des conditions de fidélité intermédiaire ou de répétabilité.

b. Modèle d'étalonnage et construction du profil d'exactitude

Les résultats bruts sont utilisés pour choisir le modèle d'étalonnage optimal et construire le profil d'exactitude.

- Choix du modèle d'étalonnage :

Les paramètres du modèle d'étalonnage sont estimés pour chaque série I, en représentant graphiquement la réponse instrumentale Y en fonction des concentrations des standards x. Le calcul des estimations des paramètres du modèle d'étalonnage peut être basé sur différentes techniques statistiques (eg. régression par la méthode des moindres carrés). Le modèle d'étalonnage doit être le même pour toutes les séries I, mais les paramètres estimés peuvent être différents (inter-séries = inter-jours).

A partir des résultats du plan d'étalonnage, il est possible de tester plusieurs modèles d'étalonnage et donc de construire plusieurs profils d'exactitude pour les mêmes données [272]. Le choix du modèle le plus adapté est réalisé à partir des indices suivants : indice d'exactitude (accuracy) I_A , indice de fidélité (precision) I_P , indice d'intervalle de dosage (dosing range) I_{DR} et indice de justesse (trueness) I_T . Le critère appliqué aux indices est le suivant : plus l'indice est proche de 1, plus le modèle est adapté pour cette caractéristique. Ce modèle optimal n'est pas forcément le modèle classiquement utilisé pour ce type de méthodes.

A partir des données du plan de validation dans la matrice, il faut calculer les concentrations prédites inverses z (ou concentrations retrouvées), pour chaque série I, à partir du modèle d'étalonnage correspondant. Pour cela, la fonction inverse du modèle d'étalonnage est utilisée.

- Construction du profil d'exactitude :

Pour construire le profil d'exactitude, il faut tout d'abord fixer les objectifs de la méthode :

- Les limites de tolérance (%) sont liées à la proportion β . Cette proportion est fixée arbitrairement, généralement entre 80 et 95 %, en fonction du champ d'application de la méthode. Elle correspond au pourcentage de résultats qui seront compris en moyenne dans l'intervalle de tolérance calculé.

- Les limites d'acceptabilité ($\pm \lambda$) correspondent aux performances exigées pour la méthode. Elles sont fixées en fonction de l'utilisation prévue en routine (eg. limites réglementaires (eg. critères de justesse dans la décision 2002/657/CE [13]), demande d'un client), parfois en fonction du niveau de concentration. Elles sont exprimées dans la même unité que le mesurande et encadrent la valeur de référence ; elles permettent de délimiter l'intervalle de dosage.

Ensuite, les calculs de justesse (biais ou recouvrement) et de fidélité sont réalisés, à partir des concentrations retrouvées z , par niveau de concentration k , pour chaque modèle d'étalonnage qui va être testé.

Justesse

La justesse peut être exprimée sous la forme d'un critère, au choix parmi les critères suivants :

$$\text{Biais absolu : } b_{ijk} = z_{ijk} - x_{ijk} \quad \text{Equation n°8}$$

$$\text{Biais relatif : } b_{ijk} \% = \frac{(z_{ijk} - x_{ijk})}{x_{ijk}} * 100 \quad \text{Equation n°9}$$

$$\text{Recouvrement : } \frac{z_{ijk}}{\bar{x}_k} * 100 \quad \text{Equation n°10}$$

Le biais permet d'estimer l'erreur systématique de mesure (ou la différence systématique entre la valeur assignée quantitative et la moyenne des résultats de mesure de réplicats) [265]. Le biais (absolu et relatif) doit être proche de zéro, alors que le recouvrement doit être proche de 100%.

Fidélité

Les écart-type de répétabilité (S_{kr}), inter-séries (S_{kB}) et de fidélité intermédiaire (S_{FI}) sont calculés pour chaque niveau de concentration k , de façon indépendante.

$$S_{kFI} = \sqrt{S_{kr}^2 + S_{kB}^2} \quad \text{Equation n° 11}$$

Intervalle de tolérance

β est la proportion du contenu de l'intervalle de tolérance. Si β est égal à 80%, un résultat sur cinq risque d'être en dehors de l'intervalle de tolérance. La valeur de β est généralement choisie entre 80 % et 95 %. Si β est égal à 95%, un résultat sur vingt risque d'être en dehors de l'intervalle de tolérance. Quand β diminue, il y aura plus de chances d'avoir des mesures à l'extérieur des intervalles de tolérance.

L'intervalle de tolérance est un intervalle symétrique autour de la concentration moyenne retrouvée \bar{z} , pour chaque niveau de concentration k . Pour chaque niveau, il faut calculer les limites basses et hautes de l'intervalle de tolérance.

$$\text{Limite de tolérance basse : } = \bar{z} - k_{tol} * S_{IT} \quad \text{Equation n°12}$$

$$\text{Limite de tolérance haute : } = \bar{z} + k_{tol} * S_{IT} \quad \text{Equation n°13}$$

k_{tol} est le facteur de couverture de l'intervalle de tolérance. Il est déterminé en fonction du seuil de tolérance β et du nombre de degrés de liberté (ddl).

$$k_{tol} = t_{v, \frac{1+\beta}{2}} \quad \text{Equation n°14}$$

$t_{v, \frac{1+\beta}{2}}$ est déterminé à partir de la table de Student, pour v ddl.

Le nombre de ddl est déterminé grâce à l'équation de Satterthwaite [273], qui prend en compte le nombre de répétitions J , le nombre de séries I et le nombre R , qui est le rapport entre la variance inter-séries (inter-jours S_B^2) et la variance de répétabilité (intra-jour S_r^2).

$$R = \frac{S_B^2}{S_r^2} \quad \text{Equation n°15}$$

R reflète l'effet de la série (jour). Si la série n'a pas ou peu d'influence sur le résultat, R sera proche de 1. Par contre, si la série a un effet important, R sera élevé et le nombre de ddl diminuera. Dans ce cas, l'intervalle de tolérance sera élargi. Donc, la méthode aura moins de chance d'être valide.

L'impact de l'augmentation du R sur l'augmentation de l'intervalle de tolérance va diminuer si on augmente le nombre de séries I . Le Tableau 9 dans la norme V03/110 montre l'influence du nombre d'essais et de R sur la largeur de l'intervalle de tolérance (pour un β fixe de 80%) [246].

L'écart-type de l'intervalle de tolérance est calculé avec les équations suivantes :

$$S_{IT} = S_{FI} * \sqrt{1 + \frac{1}{I * J * B^2}}, \quad \text{Equation n°16}$$

$$\text{avec } B = \sqrt{\frac{R+1}{J * R+1}} \quad \text{Equation n°17}$$

L'intervalle de tolérance est lié à l'écart-type de répétabilité et de fidélité intermédiaire. Moins la répétabilité et la fidélité intermédiaire sont bonnes (écarts-type faibles), plus l'intervalle de tolérance va augmenter et donc moins la méthode aura de chance d'être valide. De plus, l'intervalle de tolérance va augmenter si le facteur B diminue, c'est-à-dire plus on augmente le nombre de répétitions J .

Intervalle d'acceptabilité

L'intervalle d'acceptabilité est représenté sur le profil d'exactitude, avec une limite inférieure et une limite supérieure.

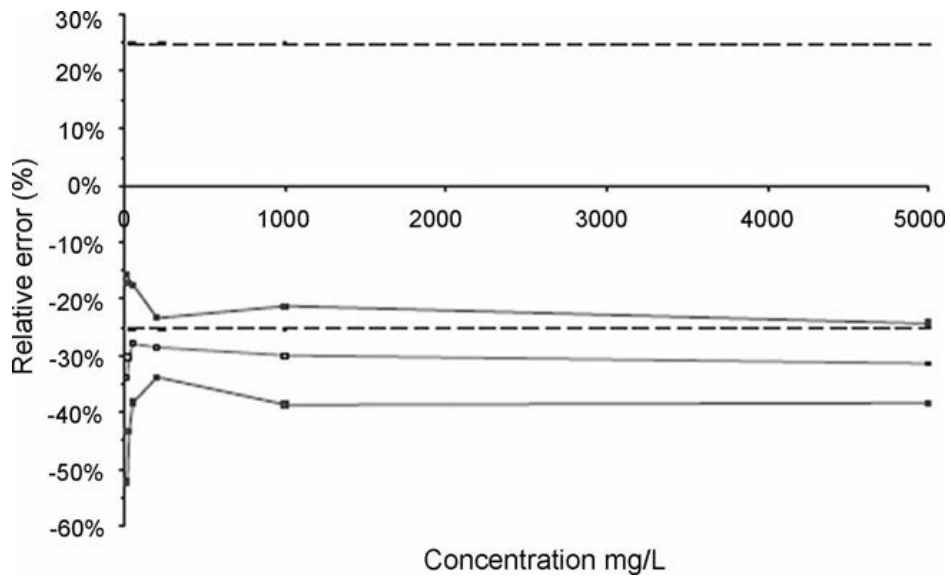
c. Interprétation du profil d'exactitude

Le profil d'exactitude est une représentation graphique des concentrations de référence moyennes (concentrations théoriques) en abscisse, des taux de recouvrement moyens (ou biais relatif), des limites de tolérance relatives (basses et hautes) et des limites d'acceptabilité relatives (basses et hautes) en ordonnées. La **Figure 31a** présente un profil d'exactitude.

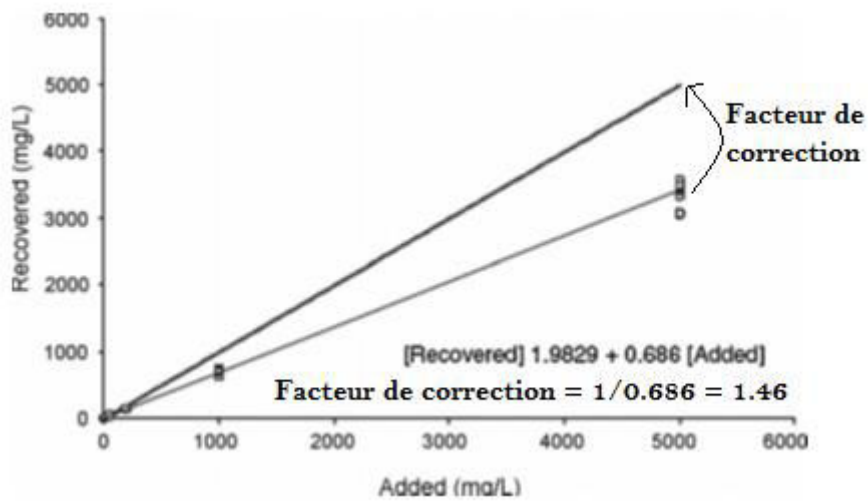
L'interprétation des résultats du profil d'exactitude se fait en fonction des objectifs fixés au départ, concernant la proportion β et les limites d'acceptabilité $\pm\lambda$. Dans notre domaine des résidus d'antibiotiques, les limites d'acceptabilité sont fixées par la décision européenne 2002/657/CE [13]. En effet, les critères de justesse minimale sont fixés dans un tableau de cette décision, en fonction de la concentration.

Figure 31. Profil d'exactitude extrait de la publication de Hubert *et al.* [253].

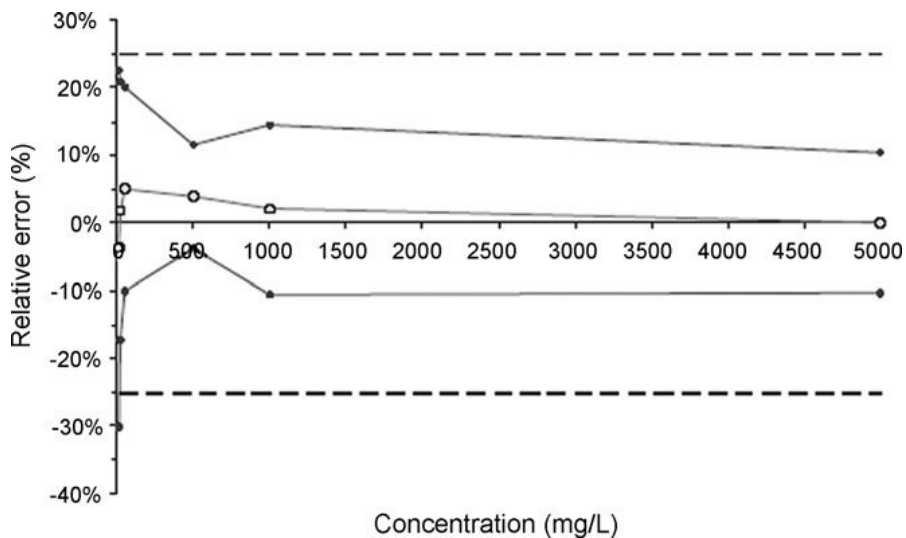
31a. Profil d'exactitude avec un biais.



31b. Représentation graphique de $z=f(x)$ pour le calcul du facteur de correction.



31c. Profil d'exactitude corrigé, à partir de celui du 30a.



La représentation graphique du profil permet une interprétation visuelle dans un premier temps.

La validité de la méthode va être déterminée en fonction des limites de tolérance et des limites d'acceptabilité.

- Pour conclure que la méthode est valide, la majeure partie des intervalles de tolérance doit être compris dans l'intervalle d'acceptabilité. Le domaine de validité de la méthode correspond au domaine dans lequel la proportion de résultats acceptables est au moins égale à β . La partie qui est comprise dans l'intervalle d'acceptabilité correspond à l'intervalle de dosage. De plus, une limite de quantification (LOQ) basse et une limite de quantification haute vont pouvoir être dérivées du profil d'exactitude. La LOQ basse correspond à la borne inférieure du domaine de validité et la LOQ haute à la borne supérieure du domaine de validité.

- Dans le cas contraire (la majeure partie des intervalles de tolérance n'est pas compris dans l'intervalle d'acceptabilité), la méthode n'est pas valide. Par exemple, la méthode dont le profil d'exactitude est représenté dans la **Figure 31a** n'est pas valide ; elle présente un biais systématique. Ce manque de justesse (biais), par exemple dû à un effet matrice, peut être corrigé, en appliquant un facteur de correction. En fait, l'effet matrice aura déjà été détecté pendant la phase de développement. Dans ce cas, il faudra réaliser deux plans de validation, à partir d'échantillons indépendants. Le premier plan servira à calculer le facteur de correction, facteur qui sera appliqué au deuxième plan. Pour calculer ce facteur de correction, une droite est tracée, en mettant la concentration ajoutée en abscisse et la concentration retrouvée (calculée) en ordonnée (**Figure 31b**). Le facteur de correction est égal à l'inverse de la pente de cette droite. Ce facteur de correction va être appliqué pour tracer un nouveau profil d'exactitude et sera aussi appliqué pour les analyses de routine (**Figure 31c**).

Le profil d'exactitude va permettre de choisir le modèle d'étalonnage optimal, qui pourra être choisi ensuite pour l'application de la méthode en routine. De plus, il peut permettre d'identifier des points aberrants. Toutefois, un grand nombre de valeurs aberrantes est le signe que la méthode n'est pas satisfaisante. Elle devrait être ré-optimisée et revalidée ou abandonnée.

Le profil d'exactitude permet de choisir des points de contrôle (contrôle qualité) pertinents pour suivre les performances de la méthode en routine, par exemple en utilisant une carte de contrôle.

En outre, l'incertitude de mesure peut être dérivée du profil d'exactitude, à partir de l'écart -type de l'intervalle de tolérance S_{IT} , [246].

Le profil d'exactitude peut s'appliquer à différents types de méthodes quantitatives : méthode CL-SM/SM [274, 275], PCR quantitative [276], comptage bactérien par une méthode microbiologique [277], biocapteur optique [278]. Dans ce dernier article, les conclusions de l'approche globale ont pu être comparées avec celles de l'approche conventionnelle critère par critère que nous avons aussi publiée [163]. En résumé, les paramètres de fidélité et de justesse étaient similaires pour le même modèle (logistique à 5 paramètres) et les conclusions identiques. Le meilleur modèle après analyse des différents critères I_A , I_{DR} , I_P et I_T était le modèle quadratique pondéré ($1/X^2$), alors que classiquement, pour les méthodes immunologiques, les modèles utilisés sont le modèle logistique à 4 ou 5 paramètres. Toutefois, la méthode était valide avec le modèle classiquement utilisé et intégré dans le logiciel du système biocapteur.

La norme NF EN ISO 16140 de 2003 [263] a été révisée et est remplacée par une nouvelle version NF EN FDIS ISO 16140-1 et NF EN FDIS ISO-16140-2 [265, 266]. Cette nouvelle version apporte une approche plus statistique de l'exploitation des résultats. A présent, l'approche relative par critère est combinée avec l'approche relative globale appelée probabilité de détection (POD).

3.2.1.2.2. La POD :

La probabilité de détection POD (Probability of detection) est un modèle statistique utilisé pour valider des **méthodes qualitatives**. Cette approche POD a été présentée pour la première fois par Wilrich en 2009 [269], puis par Wehling [270]. Cette approche est à présent intégrée dans la dernière version de la norme NF EN FDIS 16140-2 [266].

L'approche globale permet de combiner 4 caractéristiques (sensibilité, spécificité, faux-positifs et faux-négatifs) en une seule, la POD.

Le premier objectif de la POD était d'harmoniser les approches de validation entre les méthodes qualitatives et quantitatives. Comme le profil d'exactitude, ce modèle permet :

- de réaliser une représentation graphique de la réponse de la méthode,
- de comparer les résultats d'une méthode alternative avec ceux d'une méthode de référence,
- de s'appliquer aussi bien à une validation intra-laboratoire qu'à une validation inter-laboratoires (étude collaborative),
- et de calculer la fidélité d'une méthode lors d'une étude collaborative.

Le deuxième objectif était de pouvoir valider une méthode qualitative, sans obligatoirement se comparer à une méthode de référence. En effet, l'approche statistique préconisée dans l'ISO 16140 (2003) [263] exigeait la mise en place d'une MR pour comparer la MA. Cette exigence posait problème dans le domaine de la microbiologie alimentaire (*eg.* quand aucune MR n'existe), mais aussi dans le domaine des résidus d'antibiotiques, comme nous l'avons observé plus haut (chapitre 3.2.1.1.). Au niveau des lignes directrices de l'AOAC, ce même problème se posait puisque le cas des méthodes pour lesquelles aucune méthode de référence n'existait n'était pas traité non plus [270]. En outre, les tests statistiques classiques n'étaient pas adaptés dans le cas de méthodes non appariées, c'est-à-dire que le protocole d'enrichissement de l'échantillon est différent entre la MA et la MR. Ce modèle a été développé pour être applicable aussi bien aux méthodes (ou analytes) chimiques que biologiques.

a. Le principe de la POD

La POD est une probabilité conditionnelle, avec la concentration en analyte ou en bactéries comme variable conditionnelle. La POD est définie comme « la proportion de résultats analytiques positifs pour une méthode qualitative sur une matrice donnée à une concentration en analyte ou à un niveau d'analyte donné » (NF EN FDIS 16140-2) [266]. Cette probabilité est assimilable à la concentration moyenne calculée pour une méthode quantitative. La POD est concentration dépendante. Plus les concentrations sont faibles, plus la probabilité est proche de zéro. A l'inverse, plus les concentrations sont élevées, plus la probabilité est proche de 1.

Figure 32. Les deux modèles théoriques de courbes de POD (Wehling *et al.* [270]).

Figure 31a. Courbe POD pour une méthode microbiologique (loi de Poisson).

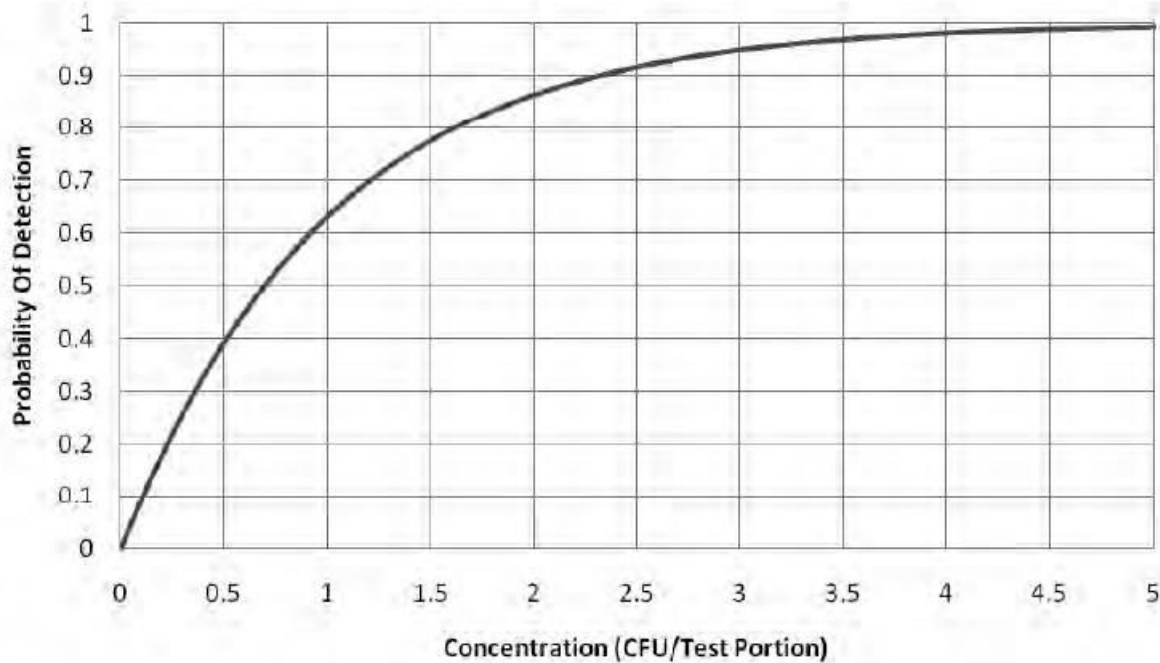
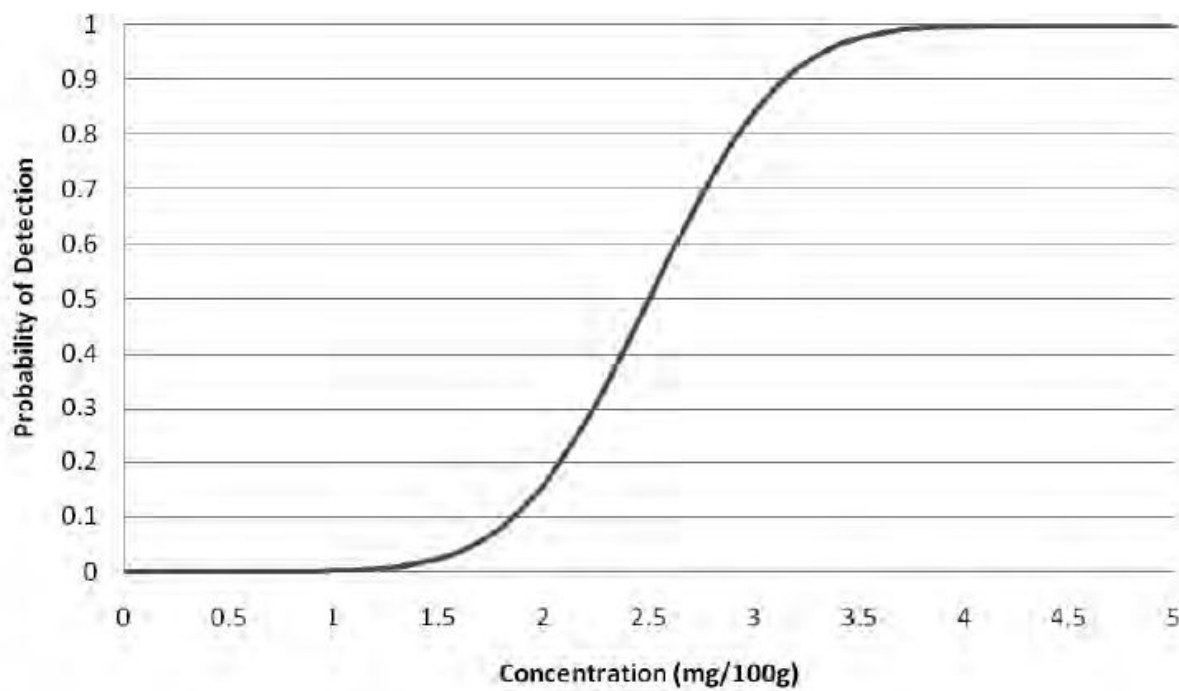


Figure 31b. Courbe POD pour une méthode qui utilise une valeur seuil (eg. ELISA).



Deux modèles théoriques de courbes de POD existent [270]:

- un pour les méthodes microbiologiques (avec une distribution selon la loi de Poisson) (**Figure 32a**),
- un pour les méthodes de type ELISA, pour lesquels une valeur seuil est fixée pour transformer les résultats quantitatifs (densités optiques) en variables qualitatives (+/-) (avec une distribution normale) (**Figure 32b**).

La variance de la POD va être liée aux concentrations en analyte, comme c'est le cas pour les méthodes quantitatives. En effet, quand la POD est proche de 0 ou de 1, la variabilité va diminuer. Par contre, dans l'intervalle de POD compris entre 0.15 et 0.85, la variabilité sera plus forte. Cette variabilité est corrélée à la variabilité entre les portions tests, mais aussi à la variabilité de la méthode dans certains intervalles de concentration. Le maximum de la variabilité sera observé pour $POD = 0.5$. D'après Wehling *et al.* [270], l'écart-type de reproductibilité sera très proche de $\frac{1}{2}$.

Les formules de calculs pour les intervalles de confiance à 95 % sont différentes en fonction du nombre de résultats positifs. Trois cas sont présentés dans la publication de Wehling *et al.* :

$x=0$, $x=N$ et $0 < x < N$ où N est le nombre total d'échantillons testés par niveau.

Dans cette approche, les intervalles de confiance sont représentés sous la forme de barres d'erreur, représentées sur le graphique de POD, pour chaque niveau de concentration (**Figure 33**).

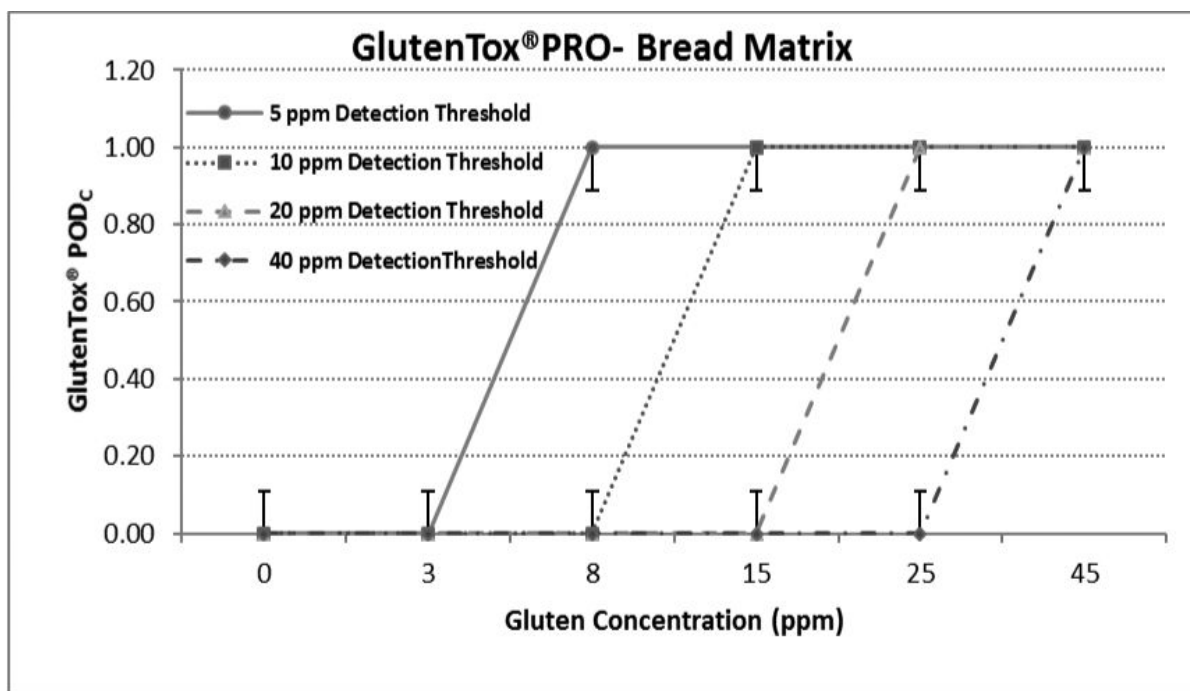
De plus, d'autres paramètres de validation peuvent être estimés à partir de la POD, par exemple une limite de détection (LOD) associée à une probabilité de détection. Par exemple, la LOD50 correspond à la concentration pour laquelle la POD est égale à 50 %.

b. Le protocole AOAC

Selon le protocole AOAC, il faut plus de réplicats pour valider une méthode qualitative que pour valider une méthode quantitative, mais l'approche reste la même [279]. En outre, valider une méthode qualitative, en utilisant des statistiques quantitatives, permettrait d'estimer les performances de la méthode avec moins de réplicats. Toutefois, aucune de ses deux affirmations n'est expliquée ou justifiée par une référence à un autre document.

Le protocole AOAC propose d'utiliser la POD, que ce soit pour une comparaison de méthodes ou pour valider une méthode dans l'absolu. Le choix du nombre de niveaux de concentration et du nombre de réplicats par niveau est crucial pour estimer au mieux le modèle de la courbe POD-concentration. Dans le protocole AOAC, il est conseillé de tester au moins 3 concentrations, avec un niveau très bas (ou absence d'analyte : un blanc) qui va donner une POD très basse, un niveau très élevé qui va donner une POD proche de 1, et enfin une concentration intermédiaire (entre 0.25 et 0.75 POD) [279]. Le nombre de niveaux peut être augmenté, pour accroître la confiance dans la détermination de la LOD. Dans certains cas, le nombre de niveaux peut être réduit à deux. De plus, il est possible d'analyser plus de réplicats à une concentration d'intérêt qu'aux autres concentrations, pour focaliser l'intervalle de confiance sur celle-ci.

Figure 33. Courbe POD pour un test immunochromatographique pour la détection du gluten dans le pain [280].



Enfin, pour comparer deux méthodes (MA/MR), il vaut mieux se focaliser sur la région marginale (0.25-0.75), car c'est là que la probabilité d'observer des différences entre les deux méthodes est la plus élevée. Pour comparer deux méthodes ou une méthode avec une valeur de référence (eg. matériau de référence certifié), la différence entre les deux POD (dPOD) est calculée, ce qui s'apparente à calculer le biais entre deux méthodes quantitatives. Si l'intervalle de confiance de la dPOD ne contient pas le zéro, alors la différence entre les deux méthodes est significative au niveau 5%.

Un exemple d'application à la validation d'un test immunochromatographique commercial pour la détection du gluten dans le pain est présenté dans la **Figure 33**, validé selon le protocole AOAC.

c. La norme NF EN FDIS 16140-2

Dans le protocole de la norme NF EN FDIS 16140-2, le principe de la comparaison de méthodes n'a pas changé par rapport à l'ancienne version de la norme et il n'est pas possible de valider la méthode alternative seule [266]. Il est recommandé de tester au moins 3 niveaux de concentration par type d'aliment, au moins un négatif, un niveau bas (niveau de détection théorique) et un niveau supérieur (légèrement supérieur au niveau de détection théorique). Le niveau bas doit pouvoir donner des résultats positifs et négatifs avec au moins une des deux méthodes. Au moins 5 réplicats doivent être analysés pour le témoin négatif et au niveau supérieur et au moins 20 répétitions au niveau bas (niveau de détection théorique).

Si la limite de détection attendue de la méthode alternative est beaucoup plus basse que celle de la méthode de référence, le niveau bas doit cibler la LOD de la méthode de référence, ce qui signifie que la LOD de la méthode alternative est sous-estimée.

La limite de détection relative la RLOD va être calculée comme suit :

$$RLOD = \frac{LOD_{alt}}{LOD_{ref}} \quad \text{Equation n°18}$$

Les RLOD peuvent être déterminés à partir d'un modèle statistique logarithme-logarithme complémentaire (LLC). Pour ce calcul, les valeurs de contaminations des échantillons ne sont pas nécessaires. Le graphique de la POD va représenter la probabilité de détection en fonction du logarithme de la concentration (niveau de contamination).

3.2.1.2.3. Conclusions sur l'approche globale

Les avantages majeurs de l'approche globale, communs au profil d'exactitude et à la POD, sont :

- l'harmonisation des concepts statistiques entre méthodes quantitatives et méthodes qualitatives,
- un outil de représentation graphique des courbes de réponse des méthodes quantitatives et méthodes qualitatives,
- un outil de prise de décision graphique pour statuer sur la validité de la méthode,

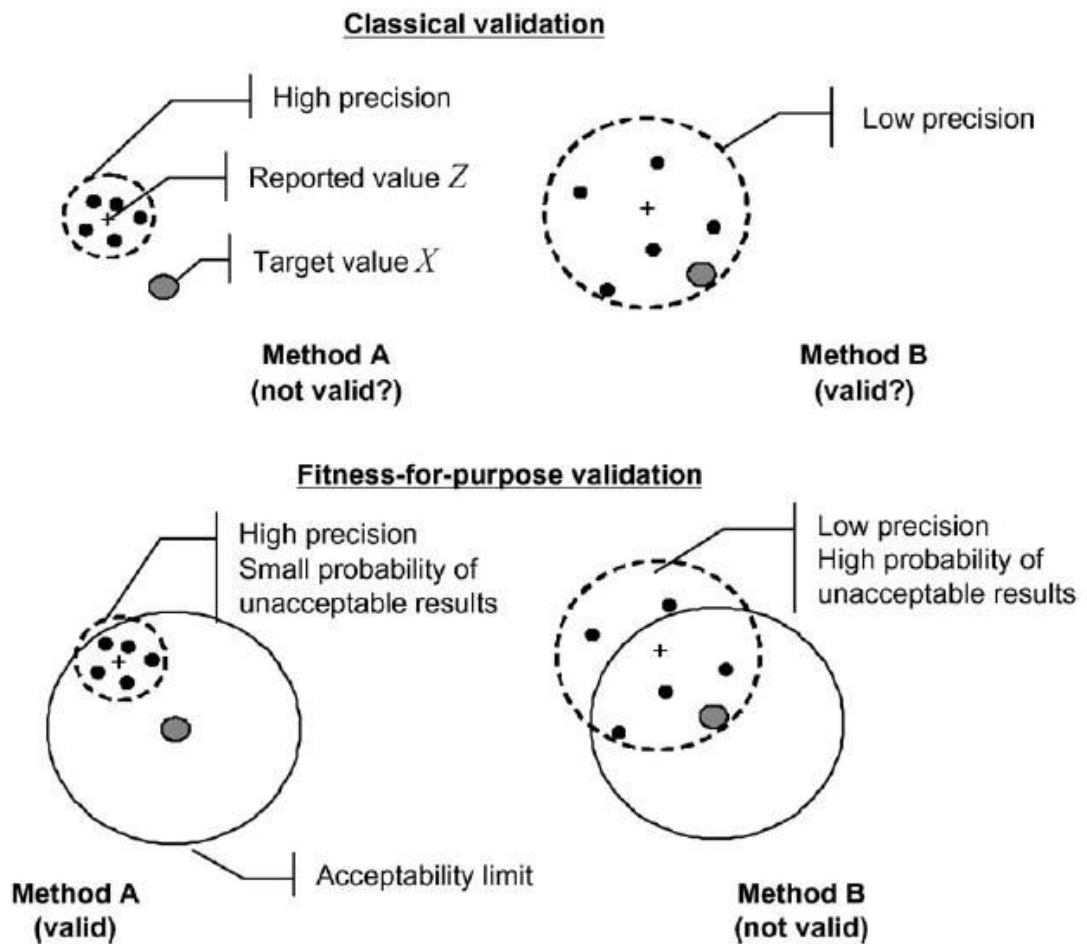
- la prise en compte la totalité des processus de la méthode, comme cela se passera en routine,
- la combinaison de la détermination des caractéristiques de performance qui facilite la prise de décision sur la validité de la méthode. Il n'y a pas de risque d'avoir des conclusions opposées puisque tout est regroupé en un seul paramètre.
- des outils de calcul sous forme de classeurs Excel, libres d'emploi, proposés pour le traitement des données de validation selon le profil d'exactitude et selon la POD, dans les normes V03-110 [246] et NF EN FDIS 16140-2 [266], et des logiciels commerciaux (eg. e-noval®, ProLab®, Statgraphics centurion® (Sigma Plus), SOS Stat VPAQ®, etc).
- des méthodes simples et qui ne demandent pas de connaissances poussées en statistiques pour utiliser les outils,
- l'application possible, aussi bien à une méthode seule (approche absolue), qu'à une comparaison de méthodes (approche relative) (norme V03-110 [246], NF EN FDIS 16140-2 [266]).
- l'application possible en intra-laboratoire, mais aussi en inter-laboratoires.

Les avantages spécifiques du profil d'exactitude sont les suivants :

- combiner deux exigences de la norme NF EN ISO/CEI 17025, c'est-à-dire valider une méthode selon des critères de performance, mais aussi déterminer l'incertitude de la mesure,
- mettre en évidence les effets matrices et possibilité d'appliquer un facteur de correction pour les analyses en routine,
- identifier les points critiques de la méthode,
- sélectionner les QC pour la routine,
- La fonction d'étalonnage peut être non linéaire (eg. logistique, quadratique, etc), comme dans la décision 2002/657/CE.
- Le profil d'exactitude est un outil de prédiction pour la réalisation des mesures en routine, mais également un outil de diagnostic des sources d'erreurs à contrôler en routine.
- L'approche globale du profil d'exactitude peut s'appliquer aux méthodes de comptage des bactéries, mais aussi aux méthodes de PCR quantitative.
- Transfert de la méthode en routine : carte de contrôle ou revalider partiellement avec un plan d'expérience à une concentration.
- dériver le profil d'incertitude à partir du profil d'exactitude [281].

Feinberg *et al.* ont illustré l'intérêt d'utiliser l'approche globale par rapport à l'approche individuelle classique, en représentant des résultats identiques, sur lesquels les critères de chaque approche sont appliqués (**Figure 34**) [271]. Les conclusions de la méthode classique peuvent être opposées car les caractéristiques sont validées individuellement. En effet, les deux méthodes présentées possèdent la même justesse, mais une fidélité différente. La méthode de gauche sur la figure présente un biais, mais une bonne fidélité, alors que la méthode de droite est juste, mais présente une fidélité faible.

Figure 34. Comparaison de l'approche classique de validation et de l'approche globale du profil d'exactitude (Feinberg *et al.* 2007) [271].



Cette figure met en évidence le fait que les conclusions de l'approche globale, grâce aux intervalles d'acceptabilité, sont simples à établir. Seule la méthode de gauche est valide, en fonction des limites d'acceptabilité fixées. Avec l'approche classique, comme la fidélité est bonne à gauche, il semble que la méthode est valide. Toutefois, les critères de justesse doivent être appliqués pour conclure sur la validité de la méthode. Pour la méthode de droite, la justesse est bonne, mais la fidélité est faible. Il est plus difficile de conclure sur la validité de la méthode, en prenant les caractéristiques individuellement. Chacune des caractéristiques doit être comparée à un critère de performance.

Les avantages spécifiques de la POD sont :

- La POD permet de déterminer plusieurs caractéristiques : les taux de faux-positifs et faux-négatifs, la sensibilité et la spécificité, la LOD et le cut-off.
- La LOD peut être déterminée pour différentes probabilités de résultats positifs, par exemple la LOD_{50} ou la LOD_{95} .
- L'estimation de la fonction POD intègre des intervalles de confiance pour les caractéristiques estimées.
- Le niveau de confiance de l'intervalle de confiance de la $LOD_{50\%}$ est égal au niveau de confiance choisi de 95%.
- Un effet matrice peut être visualisé grâce à la POD.

En raison de ces nombreux avantages, le modèle statistique de la POD est à présent intégré à la norme NF EN FDIS 16140-2-2 par exemple pour l'analyse des données des méthodes qualitatives en microbiologie alimentaire lors de l'étude préliminaire et lors de l'étude collaborative. La LOD_{50} est déterminée dans le cadre de la norme NF EN FDIS 16140-2-2 (2013). Des liens avec modèles de feuilles de calcul Excel sont intégrées dans la nouvelle norme NF EN FDIS 16140-2 [266]. Pourtant, la norme NF EN FDIS 16140-2 conserve l'approche classique des caractéristiques de performance relatives (sensibilité et spécificité relative).

En outre, cette approche est aussi intégrée depuis 2014 aux lignes directrices de l'AOAC pour la validation des méthodes chimiques qualitatives [279]. Dans ce cadre, une validation d'une méthode alternative de microbiologie alimentaire (méthode PCR qualitative) [282] et une méthode alternative pour la détection du gluten, un test immunochromatographique en bandelettes, qui est très similaire sur le principe aux tests en bandelettes pour les antibiotiques [280] ont été publiées. A ce jour, aucune publication n'a été trouvée mettant en place l'approche POD pour la validation de méthodes de dépistage des antibiotiques dans les aliments, sans doute car depuis il n'y a pas eu de nouvelle demande de certification ou de demande de modification d'un kit déjà certifié.

Toutefois, l'intérêt de cette approche explique son développement dans cette thèse. Ce modèle pourrait potentiellement s'appliquer aux méthodes microbiologiques ou immunologiques (eg. ELISA) de dépistage des résidus d'antibiotiques. Dans ce cas, l'exigence serait que la POD soit au minimum égale à 95 % pour une concentration cible inférieure ou égale à la limite réglementaire (eg. LMR). En effet, la décision 2002/657/CE exige que l'erreur beta soit inférieure ou égale à 5 % à la LMR ou en dessous.

Le premier inconvénient du profil d'exactitude est que les conclusions sur la validité de la méthode dépendent des critères fixés β et λ . Or le choix de ces deux paramètres peut se révéler compliqué. Dans certains domaines, il n'existe pas de critères réglementaires pour fixer β et λ . Dans le domaine des résidus d'antibiotiques, λ est un critère réglementaire en relation avec la décision européenne 2002/657/CE, mais β n'est pas fixé réglementairement. L'expérience de l'analyste en fonction du type de méthode doit permettre de fixer le critère β . Même si λ est un critère réglementaire, des valeurs minimales et maximales sont fixées pour la justesse et la fidélité dans la décision 2002/657/CE, mais aucune valeur pour l'exactitude. La pratique courante est de sommer les valeurs de justesse et de fidélité pour calculer λ , puisque l'exactitude est la combinaison de la justesse et de la fidélité.

Le deuxième inconvénient de cette approche est que l'optimisation de la fonction de réponse, c'est-à-dire le choix du modèle d'étalonnage le plus adapté ne peut se faire qu'avec un logiciel de traitement des données. Donc, sans logiciel adapté, la fonction de réponse utilisée ne sera peut-être pas optimale et cela aura une influence sur le profil d'exactitude et la validité de la méthode.

L'inconvénient majeur de la POD est que la charge de travail est importante, en regard du nombre de niveaux de concentrations par antibiotique à valider et du nombre de réplicats par niveau, nécessaires pour avoir une confiance raisonnable dans le modèle. Quand la méthode recherche un germe par exemple, même si les matrices alimentaires doivent être variées, cela équivaut à valider un seul antibiotique. Alors que les méthodes de dépistage microbiologique et multiplex de type biocapteurs sont capables de détecter des dizaines d'antibiotiques différents.

Uhlig *et al.* ont critiqué l'approche proposée par Wilrich *et al.* [269] et par Wehling *et al.* [270] dans le cadre d'une étude inter-laboratoires [283]. D'après eux, les variances ne sont pas homogènes entre les laboratoires et donc l'analyse de variance (ANOVA) ne peut pas être réalisée puisque l'homogénéité des variances est un pré-requis pour l'ANOVA. Wehling ne parle de ce problème et Wilrich a proposé une approche pour éliminer ce problème. D'après Uhlig *et al.*, l'écart-type de reproductibilité ne peut permettre d'évaluer la performance d'une méthode qualitative dans le cadre d'une étude collaborative. Quant à l'écart-type inter-laboratoires, il semble présenter des limites. Cette équipe propose une autre approche.

3.2.1.3. Les approches critère par critère :

Les différentes étapes de la validation suivant l'approche critère par critère sont présentées dans la **Figure 35**. L'approche appelée critère par critère est une approche qui nécessite la détermination d'une caractéristique de performance à la fois. De plus, l'interprétation des résultats est individualisée, caractéristique par caractéristique. Des critères de validation différents sont appliqués sur chacune des caractéristiques calculées. Le but de la validation est de démontrer que la méthode analytique est conforme aux critères applicables pour les caractéristiques de performance appropriées.

3.2.1.3.3. Les approches critère par critère en fonction de la réglementation :

Deux grands types d'approche vont être discutés : la validation AOAC, qui représente l'approche des Etats-Unis et la validation suivant la réglementation européenne (**Figure 36**). Ensuite, les caractéristiques de performance à déterminer ainsi que les critères d'acceptabilité seront présentés. Certaines caractéristiques et critères peuvent varier en fonction de l'approche.

a. La validation AOAC

La validation selon les principes et procédures de l'AOAC permet d'obtenir la certification AOAC pour une méthode. Ces études suivent les exigences du AOAC Research Institute et de l'administration américaine (U.S. Food and Drug Administration (FDA)). Les études de validation AOAC sont basées sur une approche par comparaison entre une méthode alternative et une méthode de référence dans le domaine de la microbiologie alimentaire [284, 285]. Par contre, dans le domaine des kits pour le dépistage de résidus d'antibiotiques, toutes les validations sont réalisées selon une approche absolue, critère par critère, basée uniquement sur la méthode alternative. Seule l'étude de l'effet matrice peut faire intervenir une méthode de référence, si elle est disponible.

Trois cas de figures ont été trouvés dans la littérature concernant l'organisation de la validation AOAC pour des méthodes de dépistage des résidus d'antibiotiques :

- Soit toute l'étude de validation a été réalisée entièrement par un laboratoire indépendant [23],
- Soit la majeure partie de l'étude a été réalisée par un laboratoire indépendant, mais l'étude de robustesse a été réalisée par le laboratoire du fabricant [286, 287],
- Soit une partie de la validation a été réalisée par le laboratoire d'origine (ou développeur) et le reste de la validation par un laboratoire indépendant [288, 289]. Le protocole de l'étude de validation indépendante, dans ce cas, est destiné à vérifier les performances déclarées de la méthode, dans des conditions contrôlées en laboratoire et à caractériser la méthode dans les conditions d'utilisation prévues.

Au final, le laboratoire développeur rédige un rapport synthétisant les résultats des deux études de validation. Ce rapport de validation est évalué par des pairs (un expert et deux rapporteurs), qui décident ou non d'accorder la certification AOAC. La certification est accordée si :

- les résultats de l'étude de validation réalisée par le laboratoire développeur sont en accord avec les performances revendiquées dans la notice du kit,
- les résultats de l'étude de validation indépendante confirment les résultats de l'étude de validation du laboratoire développeur, dans les limites statistiques fixées dans le protocole d'étude de validation.

Figure 35. Les différentes étapes de la validation suivant l'approche critère par critère.

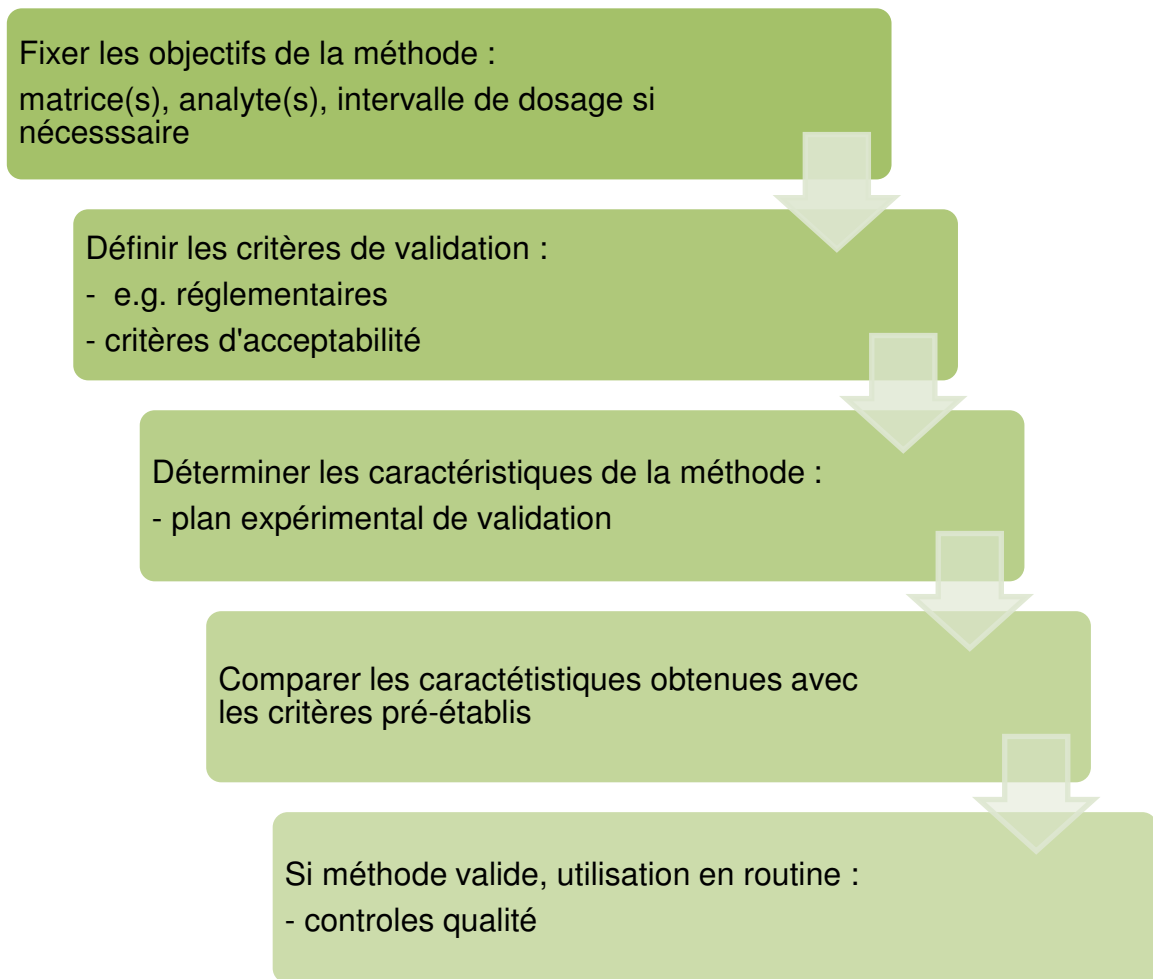
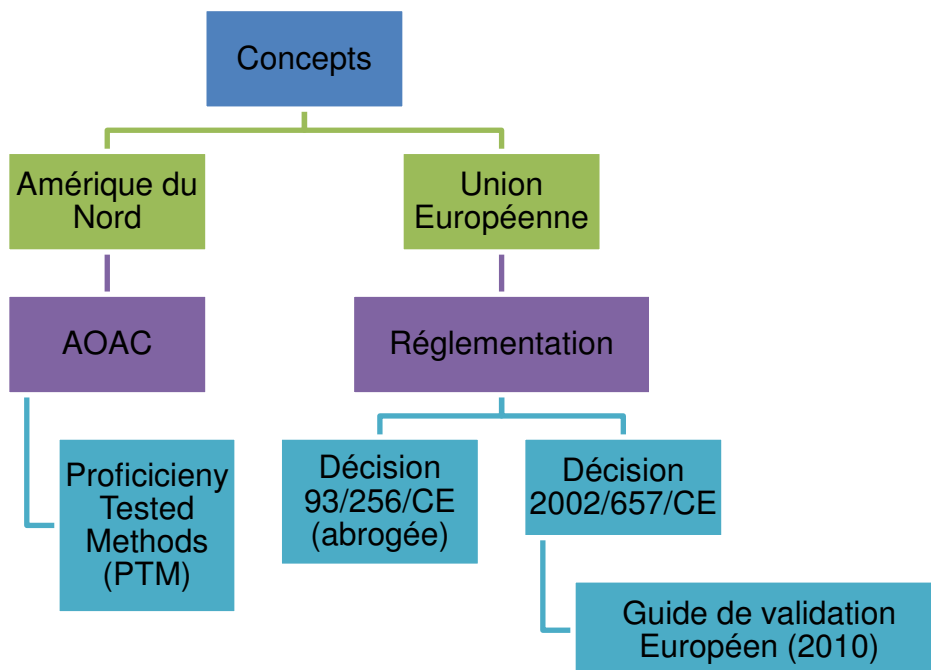


Figure 36. Les deux approches critère par critère en fonction de la réglementation.



- les résultats permettent de conclure que le kit a des performances similaires ou meilleures que la méthode de référence appropriée (le cas échéant),
- les résultats sont conformes aux critères d'acceptation figurant dans les protocoles d'étude de validation,
- les résultats répondent aux exigences minimales de performance (si elles existent).

Le rapport d'étude de validation approuvé par les pairs doit être soumis à l'AOAC pour publication dans le Journal l'AOAC INTERNATIONAL. La méthode obtient alors la certification « Performance Tested Method » (PTM). Cette méthode peut ensuite être utilisée aux Etats-Unis dans le cadre des contrôles officiels.

b. La décision européenne 93/256/CE

En Europe, la réglementation concernant les exigences de validation est différente. La validation des méthodes d'analyse pour le dépistage et la confirmation des médicaments vétérinaires et des contaminants dans les aliments d'origine animale était décrite dans la décision 93/256/CE de la Commission du 14 avril 1993 arrêtant les méthodes à utiliser pour la recherche de résidus de substances à effet hormonal et de substances à effet thyrostatique [290].

Les exigences concernant les méthodes de dépistage étaient très faibles dans cette décision. En effet, il était écrit : « Il n'est pas possible de fixer des exigences pour les méthodes de dépistage. Leur principale qualité réside dans le fait que l'incidence des faux résultats négatifs doit être la plus faible possible ». Cette décision contenait très peu d'information sur la façon de déterminer les caractéristiques de performance, telles que la spécificité. Il y avait très peu d'exigences à respecter dans le protocole de validation, en dehors d'un nombre minimal de blancs à analyser pour calculer la limite de détection.

Les premières validations de méthodes immunologiques suivant la décision 93/256/CE ont été réalisées en considérant la méthode développée comme une méthode quantitative [290]. Par la suite, au laboratoire, les méthodes immunologiques seront validées uniquement en qualitatif. En effet, les caractéristiques de performance quantitatives (fidélité, exactitude) ne respectent pas généralement les critères de performance attendus pour une méthode quantitative. De plus, dans le contrôle réglementaire, les résultats positifs du dépistage sont nécessairement confirmés par une méthode physico-chimique qui identifie et quantifie la (ou les) substance(s) présente(s).

c. La décision européenne 2002/657/CE

Suite aux évolutions très importantes des méthodes analytiques, la décision 93/256/CE a été abrogée et remplacée par la décision 2002/657/CE [13]. La décision 2002/657/CE est liée à la directive 96/23/CE [2], qui prévoit des mesures de contrôle, dont la liste des substances à l'annexe I (c'est-à-dire, des substances ayant un effet anabolisant ou substances non autorisées (groupe A) et des médicaments vétérinaires et de contaminants (par exemple, des substances antibactériennes,

d'autres médicaments vétérinaires (vermifuges, coccidiostatiques, sédatifs), d'autres substances et contaminants environnementaux (pesticides, éléments chimiques, les mycotoxines, colorants) (groupe B)). Ainsi, en Europe, chaque méthode analytique développée pour détecter ou confirmer les contaminants alimentaires ci-dessus devrait être validée selon les exigences de la décision de la 2002/657/CE [13].

La décision européenne 2002/657/CE établit des caractéristiques et des critères de performance pour caractériser, puis conclure sur la validité de la méthode [13]. Une caractéristique de performance y est définie comme une qualité fonctionnelle qui peut être attribuée à une méthode d'analyse. Un critère de performance est défini comme une exigence en matière de caractéristique de performances à partir desquelles il est possible de juger qu'une méthode d'analyse convient pour l'objectif poursuivi et donne des résultats fiables. Cette décision explique comment déterminer les caractéristiques de performance des méthodes d'analyse et définit les critères de performance. Toutefois, cette décision s'applique particulièrement pour les méthodes de confirmation, mais donne très peu d'information quant aux méthodes de dépistage.

d. Le guide de validation Européen pour les méthodes de dépistage :

De 2005 à 2009, les Laboratoires de Référence de l'Union Européenne (LRUE) avec la DGSANTE ont formé un groupe de travail auquel j'ai participé. Ce groupe de travail a rédigé un guide pour la validation des méthodes de dépistage qui a été publié sur le site de la DGSANTE en Janvier 2010 [256] ([http://ec.europa.eu/food/food/chemicalsafety/residues/lab_analysis_en .htm](http://ec.europa.eu/food/food/chemicalsafety/residues/lab_analysis_en.htm)). Ce guide est applicable à tout type de méthode de dépistage : microbiologique (méthode interne, test commercial), immunologique (eg. ELISA, test récepteur, biocapteur) et physico-chimique (de la Chromatographie Couche Mince (CCM) à la Chromatographie Liquide Haute Performance couplée à la Spectrométrie de masse en tandem (CL-SM/SM)). De plus, il s'applique aux méthodes qualitatives et quantitatives, ainsi que pour toutes les denrées alimentaires d'origine animale (eg. lait, muscle, miel, etc).

Ce guide est un outil pour aider les laboratoires à mettre en place la validation des méthodes de dépistage, en accord avec les exigences de la décision 2002/657/CE. Le guide donne des recommandations pour construire le plan de validation afin de déterminer les différentes caractéristiques de performance exigées par la décision 2002/657/CE et indique aussi les critères d'acceptabilité quand c'est nécessaire.

Le guide de validation [256] pose les bases pour deux types de validation :

- La validation initiale concerne une nouvelle méthode ou une méthode dont les données de validation ne sont pas publiques. Elle peut être réalisée par le laboratoire qui a développé la méthode ou par un autre laboratoire. Elle consiste en une validation complète et approfondie.
- La validation de transfert concerne une méthode dont les données de validation sont publiques (eg. kit commercial) ou une méthode transférée du laboratoire développeur vers un laboratoire d'application, qui souhaite mettre en place la méthode dans son propre laboratoire, sans modifications notables. Elle est réalisée par le laboratoire d'application et consiste en une validation réduite (volume d'analyses réduit), qui permet de vérifier les performances du test, en comparaison avec la validation initiale et des critères de performance.

La première différence au niveau de ces deux types de validation est que toutes les caractéristiques de performance ne sont pas à déterminer à nouveau lors de la validation de transfert. Seules la spécificité et les capacités de détection doivent être déterminées en transfert, afin de les comparer à celles de la validation initiale. La deuxième différence est que le nombre d'échantillons à analyser pour déterminer le(s) CC β est réduit dans la validation de transfert : 20 échantillons quelle que soit la concentration cible de dépistage et son rapport avec la limite réglementaire.

Nous allons maintenant présenter les différentes caractéristiques de performance à déterminer, en fonction de l'approche utilisée (décision 93/256/CE, AOAC et 2002/657/CE).

3.2.1.3.4. Caractéristiques de performance à déterminer:

Les différentes caractéristiques de performance à déterminer pour une méthode de dépistage sont présentées dans la **Figure 37**. Quelle que soit l'approche, pour une méthode de dépistage qualitative, il faut déterminer la spécificité/sélectivité, la limite de détection, l'applicabilité, la robustesse et la stabilité. Pour une méthode de dépistage quantitative, il faut ajouter à ces caractéristiques la fidélité. Pour déterminer ces caractéristiques de performance, un plan de validation doit être construit.

Ce plan va définir plusieurs paramètres, permettant cette détermination, comme par exemple le nombre d'échantillons à analyser ou la répartition des analyses en intra-jour ou en inter-jours. Le plan de validation présenté dans la **Figure 38** illustre les expériences à réaliser pour déterminer les caractéristiques de performance, suivant les recommandations du guide Européen de validation des méthodes de dépistage et donc suivant les exigences de la décision 2002/657/CE.

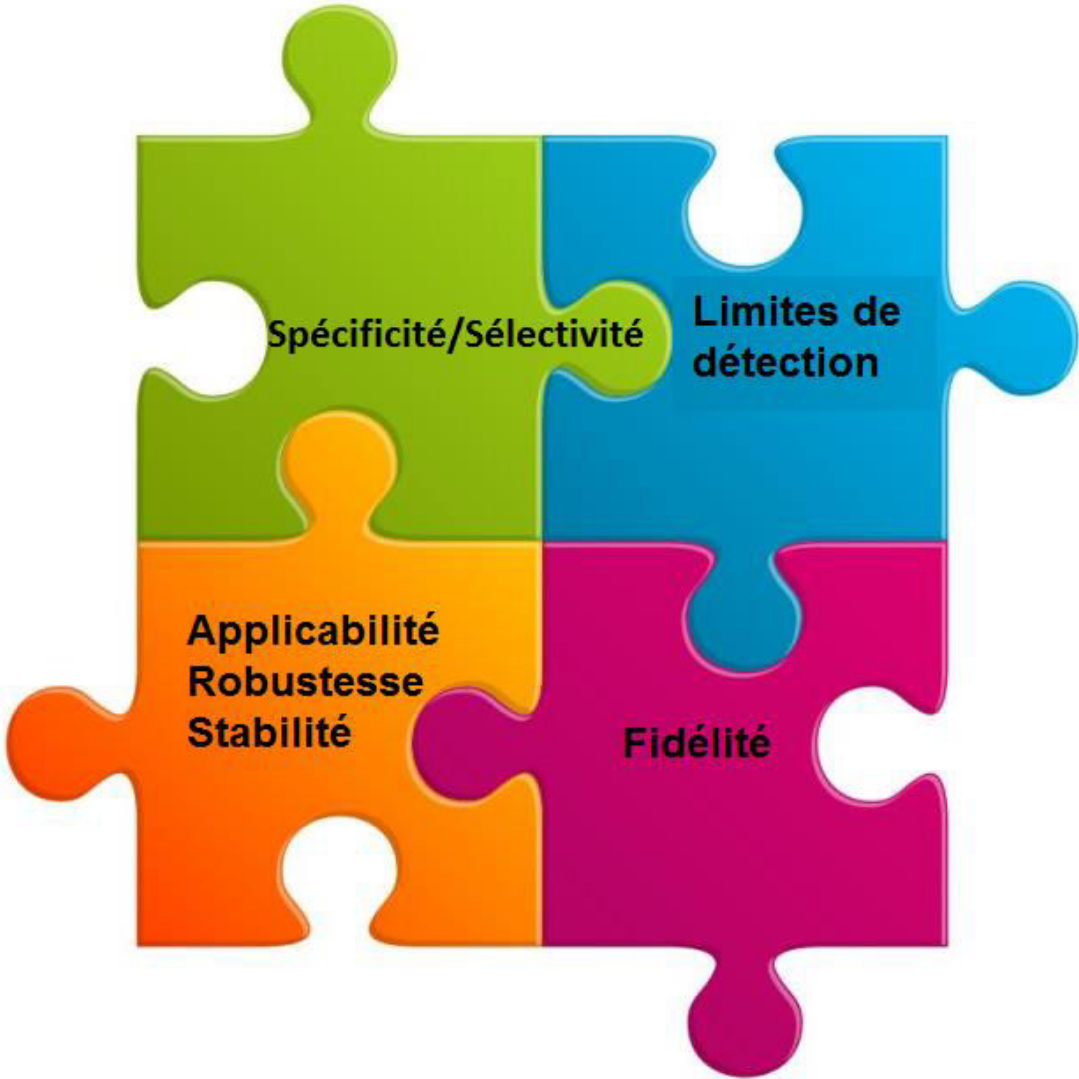
a. Limites de détection (LOD) :

Le terme limite de détection est défini comme la plus petite concentration qui peut être détectée. En dessous de cette limite, l'analyte ne peut être détecté. En fonction de la réglementation ou du référentiel de validation, la LOD peut être estimée de différentes façons.

Limite de détection (Décision 93/256/CE)

La limite de détection (LOD) était définie comme « la teneur minimale mesurée à partir de laquelle il est possible de déduire la présence de l'analyte avec une certitude statistique raisonnable (au moins 95% pour les substances non autorisées) ». Vingt échantillons blancs représentatifs devaient être analysés. La LOD se calculait, soit à partir de la moyenne des teneurs des échantillons blancs plus 3 fois l'écart-type, soit à partir de trois fois la hauteur du bruit de fond entre pics. Ce mode de calcul était ciblé sur les méthodes physico-chimiques et peu adapté pour les méthodes biologiques.

Figure 37. Caractéristiques de performance à déterminer pour une méthode de dépistage,.



Sensibilité 90/95% (AOAC)

La sensibilité correspond au pourcentage de résultats positifs trouvés parmi les résultats positifs attendus.

Selon le référentiel de l'AOAC, le niveau de sensibilité à 90/95% correspond à la concentration la plus basse estimée qui donne un résultat positif pour au moins 90 % des échantillons supplémentés testés, contenant cette concentration en analyte, avec un intervalle de confiance de 95 %. Ce niveau de sensibilité est déterminé à partir de la courbe dose/réponse, qui représente le pourcentage de résultats positifs obtenus en fonction de la concentration.

Le principe de la validation est le même pour une méthode immunologique ou pour une méthode microbiologique, avec seulement quelques différences dans le protocole. Pour un test microbiologique à large spectre pour le dépistage des antibiotiques dans le lait, le niveau de sensibilité à 90/95% a été calculé à partir d'environ 1900 échantillons pour 4 antibiotiques [23]. Pour des tests immunologiques spécifiques des bêta-lactamines, la détermination de la sensibilité 90/95% a nécessité environ 1400 échantillons pour 5 antibiotiques [286, 288].

Le niveau de sensibilité 90/95%, ainsi que la sélectivité sont déterminés soit par le laboratoire du fabricant, puis vérifiés par le laboratoire indépendant, soit directement déterminés par le laboratoire indépendant.

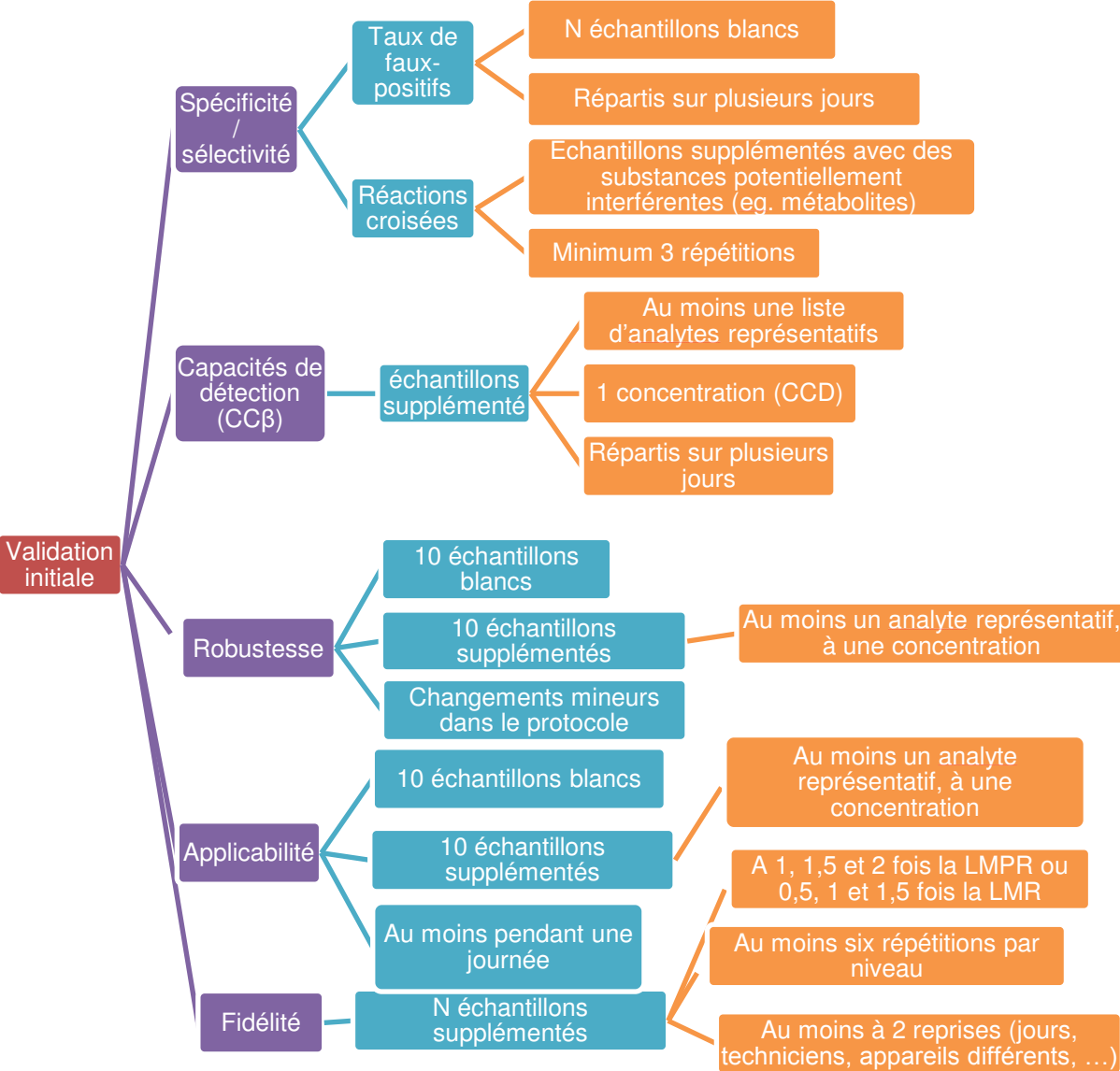
Capacité de détection (Décision 2002/657/CE et guide de validation Européen)

La caractéristique de performance la plus importante pour les méthodes de dépistage est la capacité de détection, qui est définie au point 1.12 de l'annexe de la décision 2002/657/CE [13] comme le plus petit contenu de l'analyte qui peut être détecté, identifié et / ou quantifié dans un échantillon avec une probabilité d'erreur de β . Cette définition est reprise dans le guide de validation [256]. L'erreur β est la probabilité que l'échantillon testé soit véritablement non conforme, même si une mesure conforme a été obtenue. Le $CC\beta$, contrairement à la LOD déterminée dans le cadre des validations selon la décision 93/256/CE, est déterminé à partir d'échantillons supplémentés avec l'antibiotique d'intérêt et non à partir d'échantillons blancs, qui ne prenaient en compte que le bruit de fond de la méthode.

La construction du **plan d'expérience** pour déterminer les $CC\beta$, qui est décrite dans le guide de validation Européen, se fait à partir de plusieurs informations concernant la méthode et les limites réglementaires des analytes cibles, si elles existent (**Figure 39**).

En fonction du type de la méthode, le **choix des analytes** et surtout de leur nombre va différer. En effet, le spectre de détection peut être plus ou moins large. Dans l'hypothèse maximale, une méthode peut détecter par exemple une soixantaine de substances à LMR dans le lait, ce qui représente une charge de validation très importante et un coût élevé. De plus, quand les méthodes ne sont pas spécifiques, il est impossible de mélanger plusieurs antibiotiques dans un même échantillon, ce qui multiplie le nombre d'essais. C'est pourquoi le guide de validation Européen donne des recommandations pour le choix des analytes à valider, en fonction du type de méthodes (**Tableau 16**).

Figure 38. Diagramme du plan de validation initiale pour une méthode de dépistage.



n : nombre d'échantillons requis pour la validation (20 ≤ *n* ≤ 60)

Le (ou les) analyte (s) sélectionné(s) pour l'étude de validation devrai(en)t idéalement être les moins sensibles de leur famille, à savoir la concentration cible de dépistage la plus proche (inférieure) de la limite réglementaire. Cependant, cet analyte doit être détecté au niveau ou en dessous de la limite réglementaire correspondante.

Dans le cas des méthodes microbiologiques, il est recommandé, avant la validation, de déterminer les profils d'activités pour tous les analytes concernés dans chaque famille afin de choisir au moins une ou deux substances représentatives par famille de composés, les moins bien détectées dans la famille par exemple. Les profils d'antimicrobiens sont déterminés en testant des solutions standards à différentes concentrations autour de la LMR. Lorsque la méthode de dépistage détecte plus d'une substance, la supplémentation doit être répétée pour chaque substance représentative.

Les protocoles proposés dans la décision 2002/657/CE pour déterminer le $CC\beta$ pour une méthode de dépistage ne sont pas adaptés. En effet, le premier protocole est basé sur l'utilisation de la courbe d'étalonnage, suivant les exigences de la norme ISO 11843 [291]. Or toutes les méthodes de dépistage ne sont pas utilisées avec une courbe d'étalonnage. De plus, dans cette approche, la courbe d'étalonnage devait être linéaire. Le second protocole nécessite de compléter 20 échantillons blancs au $CC\alpha$, ce qui nécessite la détermination préalable du $CC\alpha$. Le $CC\alpha$ correspond à la limite de décision de la méthode, limite au-dessus de laquelle l'échantillon va être déclaré non-conforme. Or le $CC\alpha$ n'est pas demandé pour une méthode de dépistage.

En 2008, le guide SANCO/2004/2726-rev a été publié afin de donner des lignes directrices pour compléter la décision 2002/657/CE et aider à la mettre en place [292]. Ce guide a proposé de déterminer le $CC\beta$ du dépistage en analysant 20 échantillons blancs supplémentés à deux concentrations, incluant le niveau d'intérêt (*i.e.* la limite réglementaire).

Ce n'est pas l'approche qui a été retenue dans le guide Européen. Ce guide définit la notion de **concentration cible de dépistage (CCD)** qui est la concentration à laquelle un test de dépistage donne un résultat positif pour un échantillon donné, ce qui entraîne obligatoirement une analyse de confirmation. La CCD doit être assez basse par rapport à la limite réglementaire (*eg.* LMR, LPMR, RPA) pour garantir que le $CC\beta$ sera bien inférieur ou égal à la limite réglementaire (LR). Le **nombre d'échantillons à analyser** pour déterminer le $CC\beta$ est fonction du rapport entre la CCD et la LR. En effet, plus on s'éloigne de la LR, plus la confiance statistique dans le résultat obtenu augmente et moins cela nécessite d'échantillons. En d'autres termes, plus la concentration cible est basse par rapport à la LR, moins il y a de risques d'obtenir un résultat faux-négatif pour les échantillons contenant un résidu à la LR. Le nombre d'échantillons supplémentés à la CCD à analyser sera fixé comme représenté dans la **Figure 39**.

Figure 39. Construction du plan d'expérience pour la détermination du CC β .

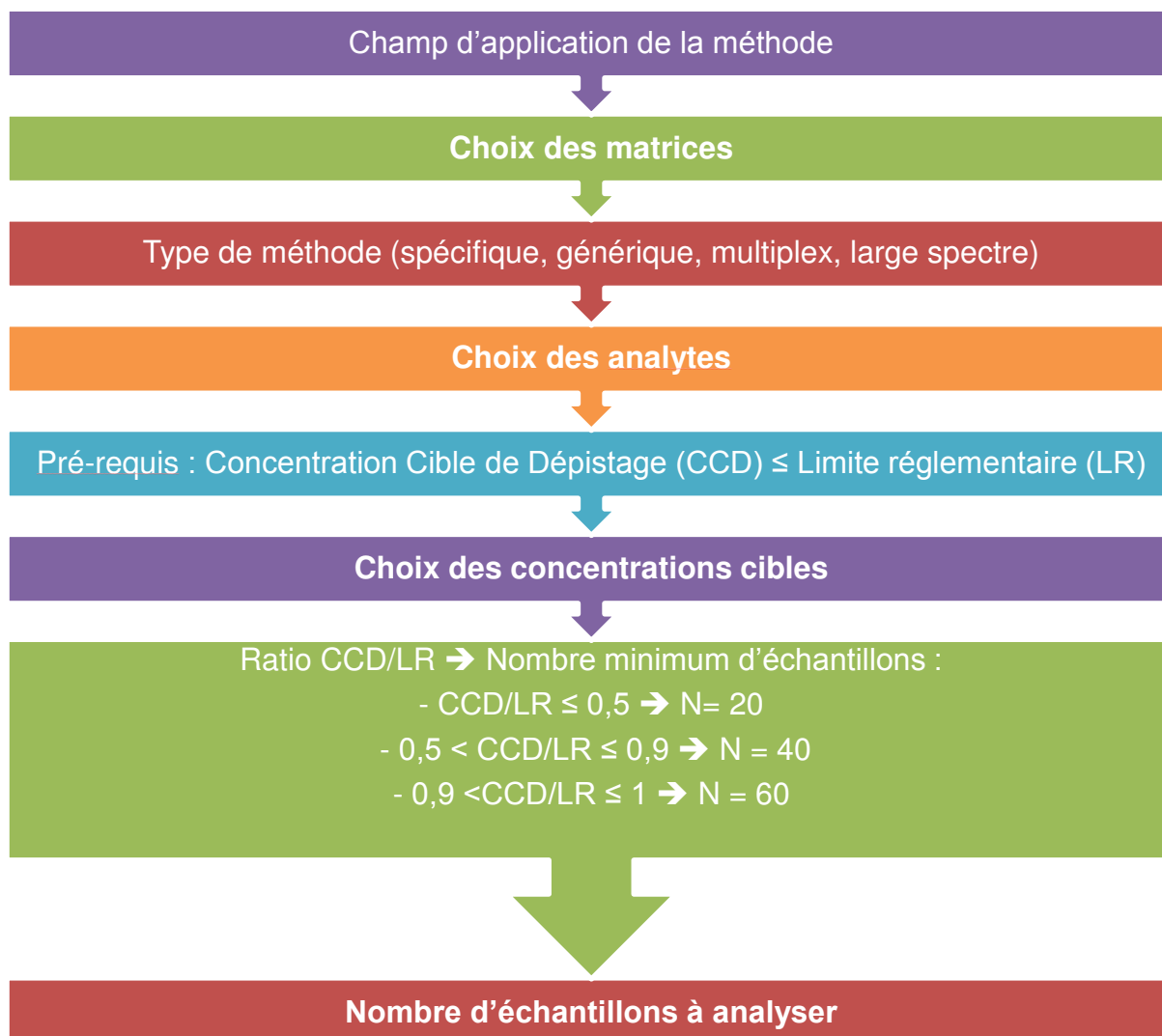


Tableau 16. Choix des analytes en fonction du type de méthode.

Méthodes	Spectre de détection	Analytes
Microbiologique	Large	Liste de substances représentatives
Immunologique	Spécifique d'un analyte	Analyte principal
	Générique (une famille)	Analyte avec la plus faible réaction croisée
	Multiplex (plusieurs familles)	Analyte avec la plus faible réaction croisée, dans chaque famille
Physico-chimique	Fonction de la méthode	Tous les composés détectables par la méthode ou au moins un composé par classe ou sous-classe chimique connue

Pour les matrices solides (eg. muscle), le guide de validation propose l'utilisation de tissus simulés, c'est-à-dire des tissus broyés et supplémentés, pour être le plus proche possible de vrais échantillons ou tissus. En effet, les méthodes microbiologiques (type boîtes) utilisent le plus souvent le muscle intact pour l'analyse. Il est donc impossible de supplémenter un muscle intact, de façon homogène. Il est aussi inconcevable de pouvoir produire des matériaux naturellement chargés (issus d'animaux traités), pour tous les antibiotiques d'intérêt, à la concentration cible précisément. Soit des morceaux de tissus congelés supplémentés sont carottés et placés directement sur les boîtes, soit des disques de papier sont dopés avec des solutions standard d'antibiotiques, les disques placés sur les boîtes, puis des morceaux de tissus entiers sont placés sur les disques dans un "sandwich". Ce dernier protocole permet de respecter l'intégrité du tissu.

En pratique, les N échantillons blancs (spécificité, voir point b.) et les N échantillons supplémentés à la CCD sont répartis et analysés sur plusieurs jours, en aveugle. Dans l'idéal, les N échantillons doivent être d'origine différente. En outre, l'utilisation de plusieurs lots (au moins deux) de kits, s'il s'agit d'un kit commercial ou de souche bactérienne, pour une méthode microbiologique par exemple, est recommandée pour la détermination du CCβ.

Pour une méthode microbiologique, le nombre d'échantillons supplémentés donnant un résultat négatif (inférieur au seuil de positivité de la méthode) va être déterminé.

Pour une méthode immunologique, les réponses analytiques pour les échantillons blancs B_i et pour les échantillons supplémentés Y_i sont mesurées. Puis, la réponse moyenne des blancs B et l'écart type « SD_b » de la réponse sont calculés. Une « valeur- seuil » ou seuil de positivité T est calculée à partir des blancs. De même, la réponse moyenne des supplémentés M et l'écart type « SD » de la réponse sont calculés. Le facteur « cut-off » F_m , est calculé à partir des supplémentés. Les facteurs T et F_m sont matrice-spécifiques.

Pour l'interprétation des résultats, deux cas se présentent :

- Le signal est proportionnel à la concentration d'antibiotiques dans l'échantillon (tests microbiologiques ou physico-chimiques) :

$$T = B + 1.64 * SD_b \quad \text{Equation n°19}$$

$$F_m = Y - 1.64 * SD \quad \text{Equation n°20}$$

Ce calcul est rarement utilisé pour les méthodes biologiques, il est plus couramment utilisé pour les méthodes physico-chimiques utilisées pour le dépistage, type CL-SM/SM. Cette possibilité ne va donc pas être développée ici.

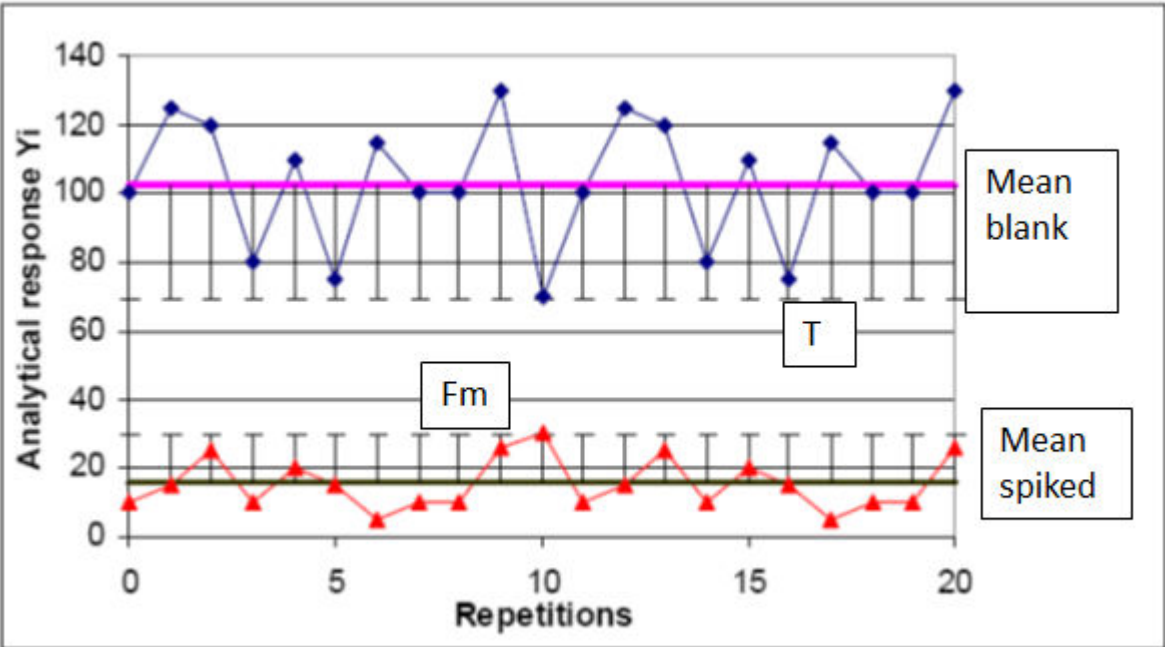
- Le signal est inversement proportionnel à la concentration d'antibiotiques dans l'échantillon (tests immunologiques type ELISA) :

Ce calcul est très souvent utilisé pour les méthodes immunologiques, type ELISA ou biocapteurs (**Figure 40**). Nous n'allons traiter que cette façon d'interpréter les résultats.

$$T = B - 1.64 * SD_b \quad \text{Equation n°21}$$

$$F_m = Y + 1.64 * SD \quad \text{Equation n°22}$$

Figure 40. Représentation graphique des résultats des échantillons blancs et des supplémentés, pour une méthode immunologique, en lien avec les valeurs seuils T et Fm.



b. Spécificité/Sélectivité

Dans la décision européenne 2002/657/CE, la spécificité est définie comme la capacité d'une méthode à distinguer l'analyte mesuré d'autres substances présentes dans l'échantillon (eg. métabolites, produits de dégradation, constituants de la matrice, etc.) [13]. Cette caractéristique est principalement fonction du type de méthode utilisée, mais dépend aussi de la classe de l'analyte ou de la matrice. Cette définition n'a pas évolué depuis la décision 93/256/CE et convient aussi pour la caractéristique appelée sélectivité, déterminée dans le cadre des validations AOAC.

La sélectivité est associée à la spécificité dans la décision 2002/657/CE, avec un protocole de détermination commun. Toutefois, seule la spécificité est clairement définie dans la décision. Les termes spécificité et sélectivité sont employés pour caractériser les méthodes de dépistage des résidus d'antibiotiques. La différence entre les deux termes, qui est communément admise dans le domaine des méthodes de dépistage des résidus d'antibiotiques, se situe au niveau de l'interprétation du calcul de spécificité.

- Le terme « spécificité » est habituellement utilisé pour les tests ciblés sur un seul analyte. Une méthode est dite spécifique quand elle donne un signal analytique exclusivement pour cet analyte. Par exemple, une méthode peut être spécifique de la pénicilline G.

Le terme « sélectivité » est généralement utilisé pour les tests de dépistage à large spectre (ou méthodes multi-résidus). La méthode est dite sélective quand elle produit des résultats analytiques pour une (ou plusieurs) classe(s) de substances. Par exemple, une méthode peut être sélective de la famille des bêta-lactamines. La spécificité correspond au pourcentage de résultats négatifs trouvés parmi les négatifs attendus (échantillons exempts de résidus).

Selon le référentiel de l'AOAC, le taux de sélectivité est déterminé comme le pourcentage de résultats faux-positifs parmi les 60 échantillons blancs testés.

Le guide de validation européen recommande, quel que soit le type de méthode (spécifique ou large spectre), la spécificité/sélectivité est déterminée après l'analyse de 20 à 60 échantillons de matrice blanche (exempte de résidus d'antibiotiques) (**Figure 36**). A partir des résultats, le taux de résultats faux-positifs peut être calculé :

$$\text{Taux Faux positifs} = \frac{\text{nb résultats positifs}}{\text{nb échantillons blancs}} * 100 \quad \text{Equation n°16}$$

Le nombre de résultats faux-positifs pour une méthode immunologique, type ELISA, sera déterminé globalement pour tous les jours de validation. Les échantillons blancs qui présenteront un résultat (signal) inférieur à la valeur seuil (eg. Fm) seront considérés comme des résultats faux-positifs.

Pour un test de dépistage ciblé, conçu pour répondre sélectivement à un composé unique ou à une famille d'antimicrobiens (eg. kit ELISA, test récepteur), la spécificité/sélectivité va en plus être constituée de la détermination des réactions croisées (CR), entre la substance principale et une liste de substances (**Figure 37**). Ces substances potentiellement interférentes peuvent être des substances chimiquement proches (métabolites, dérivés, etc.) ou d'autres substances susceptibles de se trouver en présence du composé d'intérêt dans les échantillons (médicament vétérinaire, en association avec d'autres (médicaments)). La présence éventuelle d'interférences sur les résultats du test va être analysée.

Les réactions croisées sont déterminées (*eg.* autres antibiotiques de la même famille, autres familles). Les concentrations en analytes sont calculées par rapport à une gamme d'étalonnage. La réponse de chaque analyte testé est comparée à la réponse de l'analyte principal pour déterminer la concentration équivalente à « l'analyte principal ». Cette concentration est exprimée en pourcentage de la réponse « de l'analyte principal ». Dans ce calcul, la spécificité de l'anticorps pour « l'analyte principal » est fixée à 100 %.

Le pourcentage de CR permettra d'estimer les CC β pour les molécules testées ici, à partir du CC β de l'analyte principal :

$$CC\beta \text{ analyte PI} = \frac{CC\beta \text{ analyte principal}}{CR \text{ a nalyte PI}} * 100 \quad \text{Equation n°24}$$

De plus, dans le protocole AOAC, les effets masquants potentiellement dûs à la présence d'autres substances dans l'échantillon sont mesurés.

c. Fidélité

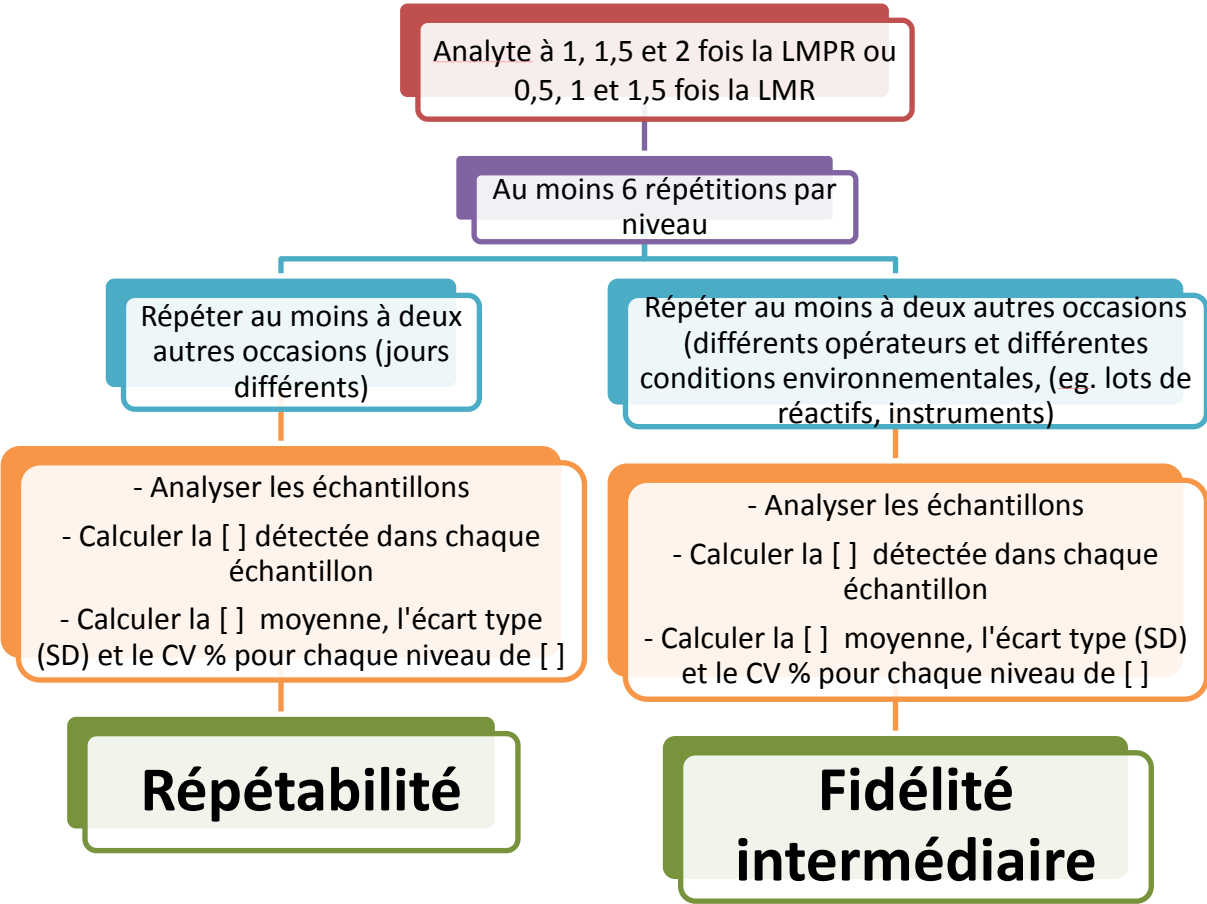
Pour une méthode quantitative de dépistage, la détermination de la justesse n'est pas demandée dans la décision européenne 2002/657/CE [13], seule la fidélité doit être déterminée. Dans la décision européenne 93/256/CE [290], deux caractéristiques devaient être déterminées pour une méthode de dépistage quantitative, même si, suite à la traduction de la décision en anglais, des abus de langage peuvent être signalés. Le terme précision était utilisé à la place du terme fidélité et le terme exactitude à la place de justesse.

De plus, aucun critère de décision n'était fixé pour ces deux caractéristiques. La seule exigence formulée était : « Les faux résultats positifs peuvent être acceptés mais les faux résultats négatifs doivent être réduits au minimum au niveau recherché ». Cette exigence ne représente pas vraiment une exigence pour une méthode qui produit des résultats quantitatifs. Elle est bien loin des exigences formulées plus tard par la décision 2002/657/CE [13].

Dans la décision 2002/657/CE, la fidélité est définie comme « l'étroitesse d'accord entre des résultats de tests indépendants obtenus dans des conditions stipulées (prédéterminées) » [13]. La fidélité traduit uniquement la distribution des erreurs aléatoires et n'a aucune relation avec la valeur vraie ou spécifiée.

La détermination de la fidélité est exigée seulement pour les méthodes quantitatives. Même si des tests microbiologiques (comme les méthodes boîtes) peuvent donner des résultats quantitatifs (*eg.* zone d'inhibition), la détermination de la fidélité n'est pas obligatoire pour ce type d'essais. De même, si un test ELISA a pour but d'être utilisé seulement en tant que méthode qualitative, l'évaluation de la fidélité pendant la validation n'est pas nécessaire.

Figure 41. Protocole de détermination de la répétabilité et de la fidélité intermédiaire.



Coefficient de variation (CV %)

La fidélité est calculée à partir de l'écart type des résultats d'essais, réalisés dans des conditions de répétabilité (même laboratoire, même opérateur, même équipement) et de reproductibilité (différents laboratoires, différents opérateurs, différents équipements), pour les échantillons supplémentés pour chaque niveau de concentration (**Figure 41**). Le guide SANCO/2004/2726-rev 4 complète la décision pour les substances interdites qui n'ont pas de LPMR [292]. Les concentrations à tester sont alors 1, 1.5 et 2 fois le CC β .

Dans le cadre de la certification AOAC, toutes les méthodes de dépistage ont été validées comme des méthodes qualitatives. Ainsi, la fidélité n'a pas été déterminée.

d. Applicabilité

L'applicabilité n'est pas abordée comme caractéristique spécifique dans les validations AOAC. De plus, toutes les méthodes validées pour les résidus d'antibiotiques ont été validées pour une matrice unique (*i.e.* le lait de vache).

Dans la décision 93/256/CE, l'applicabilité était définie comme « la liste des types d'échantillons et/ou d'analytes auxquels la méthode peut s'appliquer telle quelle ou moyennant certaines modifications mineures » [290]. Aucun mode de détermination n'était proposé.

Dans la décision européenne 2002/657/CE, applicabilité et robustesse sont traitées simultanément [13]. Dans le guide de validation Européen, l'applicabilité est traitée à part [256]. L'applicabilité est la possibilité d'appliquer une méthode à plusieurs matrices, avec des performances similaires. En effet, les matrices peuvent avoir un impact sur les résultats d'un essai. La plupart des méthodes sont conçues pour une matrice principale. En raison du grand nombre de combinaison analyte/matrice possibles, il est nécessaire de définir la portée de la méthode en termes de matrices et d'espèces animales. Les nouvelles matrices peuvent être des tissus différents (muscle, rein, foie,...), d'origine animale différente (œufs, miel, crevettes,...) ou venant de différentes espèces (bovin, ovin, volaille, lapins,...). De plus, les LMR peuvent différer entre les matrices, mais aussi entre différentes espèces animales pour une même matrice.

L'applicabilité peut être réalisée à deux moments distincts :

- soit dès la validation initiale si on a déjà fixé la (ou les) matrice(s) ou espèce(s) animale(s) pour lesquelles on veut évaluer l'applicabilité,
- soit à posteriori si on veut étendre la méthode déjà validée à une nouvelle matrice ou espèce animale.

A l'étape de la validation initiale, deux approches sont possibles pour déterminer l'applicabilité (**Figure 42**). Dans le cas de la première approche, l'hypothèse est faite que le(s) CC β pour les x matrices est (ou sont) équivalent(s), hypothèse qui peut être vérifiée par des essais préalables. Alors, la spécificité et le(s) CC β sont déterminés à partir d'un pool d'échantillons comprenant un mélange des différentes matrices (*eg.* 10 échantillons de muscle et 10 échantillons de rein). Si en fait les CC β étaient différents, déterminer un CC β unique pour les 3 matrices implique que le CC β de certaines matrices est surestimé ou sous-estimé. La deuxième approche consiste à déterminer spécificité et CC β pour une première matrice. Ensuite, spécificité et CC β sont déterminées pour une ou plusieurs autres matrices. Ces caractéristiques sont comparées à celles déterminées pour la première matrice validée.

Le guide de validation recommande que l'applicabilité soit démontrée pour chaque analyte que le laboratoire a inclus dans un programme d'analyse des résidus (*eg.* analyte cible (test ciblé)), ou au moins sur un nombre sélectionné d'analytes qui sont représentatifs du groupe d'analytes, quand la méthode a un spectre large de détection. L'applicabilité doit être démontrée par des analyses en aveugle (échantillons inconnus), à différents jours, avec différents opérateurs qualifiés, si possible.

e. Robustesse

La robustesse est définie dans la décision européenne 2002/657/CE comme « la sensibilité d'une méthode d'analyse aux variations des conditions expérimentales, qui peut s'exprimer par une liste des échantillons, des analytes, des conditions de stockage, des conditions d'environnement et/ou de préparation de l'échantillon pour lesquels la méthode peut être appliquée telle quelle ou moyennant certaines modifications mineures » [13]. L'objectif est d'identifier les points critiques d'une méthode, qui en pratique pourraient varier et influencer le résultat de l'analyse.

La robustesse est une étape très importante, qui n'est pas toujours intégrée dans la partie validation de la méthode, mais qui est parfois être considérée comme faisant partie du développement de la méthode. En effet, l'identification de points critiques dans la méthode qui peuvent influencer sur les résultats doit se faire avant la caractérisation de la méthode.

La robustesse se définit globalement de la même façon dans toutes les approches présentées. Seuls les facteurs étudiés peuvent varier. La première étape est de choisir des facteurs du pré-traitement de l'échantillon et de l'analyse, qui pourraient influencer les résultats de mesure (*eg.* l'analyste, la source et l'âge des réactifs, la température, le pH).

Dans le protocole AOAC, le laboratoire indépendant réalise des études de robustesse, sur des échantillons congelés (effet de la congélation), les interférences (germes, cellules). Le plus souvent elle a dû être prédéterminée par le fabricant, puisqu'elle doit précéder l'étape de caractérisation des performances proprement dite. Toutefois, elle est parfois réalisée par le laboratoire indépendant, selon les études.

Les facteurs doivent être modifiés dans un ordre de grandeur qui correspond à des déviations usuelles. Le guide de validation Européen recommande de se concentrer sur une molécule représentative de la série d'analytes considérés si la méthode a un spectre large de détection pour étudier la robustesse. Au moins 10 matériaux blancs différents et 10 matériaux différents supplémentés au niveau d'intérêt, c'est-à-dire la CCD, devront être analysés en aveugle, à différents jours, avec différents opérateurs qualifiés, si possible. La capacité de détection CC β pour cette molécule et la spécificité vont ainsi être déterminées. L'influence de ces variations peut être évaluée, par exemple, en utilisant des plans d'expérience, afin de minimiser le nombre d'expériences.

Figure 42. Détermination de l'applicabilité de la méthode : deux approches.

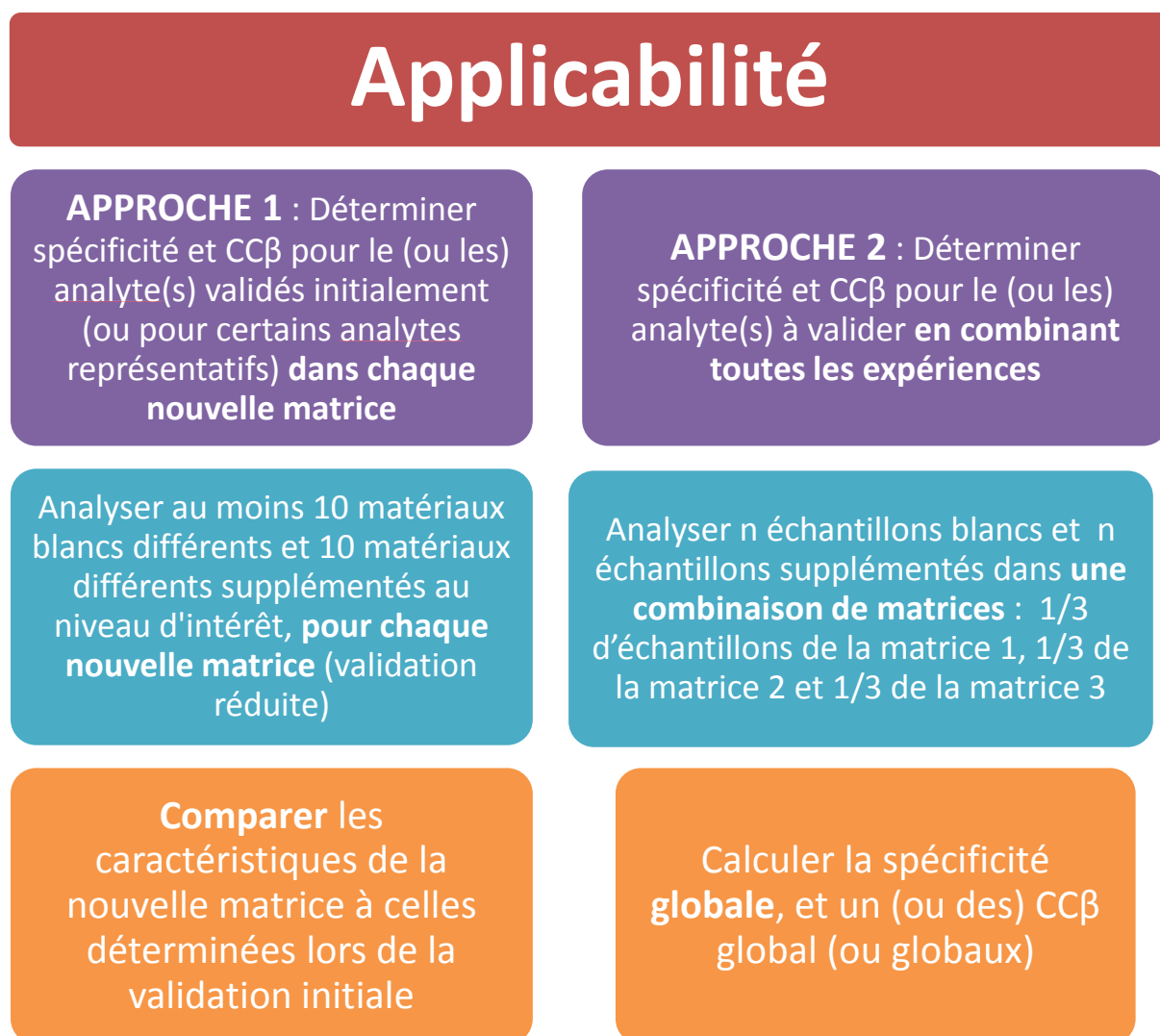


Tableau 17. Détermination de la stabilité des analytes en solution et dans la matrice [13].

	Stabilité des analytes en solution	Stabilité des analytes dans la matrice
Préparation	Solutions d'analytes	Homogénéiser une matrice blanche, partager en 5 aliquotes et supplémenter chaque aliquote avec le (ou les) analyte(s)
Mesure Cfraîche	Concentration en analyte dans la solution fraîche	Concentration en analyte dans un des aliquotes immédiatement
Différentes températures de stockage	-20°C, +4°C et +20°C, et les analyser après 1, 2, 3 et 4 semaines ou plus si nécessaire	au moins - 20°C ou à une température inférieure si nécessaire et les analyser après 1, 2, 4 et 20 semaines
Mesure Crésiduel	Concentration en analyte (C_i = concentration à l'instant considéré) : Analyte résiduel (%) = $C_i \times 100/C_{\text{fraîche}}$	

f. Stabilité

Dans le protocole AOAC, le laboratoire indépendant peut réaliser des études de stabilité des kits (ne fonction de leur date de production) et parfois des études de stabilité de l'analyte dans la matrice (eg. effet de la congélation, contrôles positifs lyophilisés à 4°C).

Dans la décision 2002/657/CE, les études de stabilité concernent les analytes [13]. La stabilité doit être vérifiée en solution et dans les matrices d'intérêt. En effet, une stabilité insuffisante de l'analyte ou des constituants de la matrice durant le stockage d l'échantillon ou pendant l'analyse peut entraîner des écarts significatifs au niveau du résultat.

Deux possibilités sont proposées dans le guide de validation européen :

- Si les données de stabilité existent déjà en interne au laboratoire ou dans la bibliographie, une étude bibliographique sera incluse dans le rapport de validation. En effet, des données de stabilité des analytes en solution ou dans les matrices peuvent être trouvées dans la littérature (données publiées) ou être déterminées par une coopération avec d'autres laboratoires si les études de stabilité ont été réalisées mais non publiées dans des journaux scientifiques. Les données de stabilité peuvent être extraites d'études réalisées avec d'autres méthodes analytiques, dans d'autres laboratoires puisqu'elles ne dépendent pas de la méthode employée pour l'analyse. Cette possibilité est aussi proposée dans le guide de validation en 2010 et dans l'autre guide SANCO/2004/2726-rev 4 [292]. Quand la stabilité n'est pas connue, des exemples sont donnés dans la décision CE/2002/657 [13] sur la façon dont la stabilité peut être déterminée (**Tableau 17**).

Dans la plupart des cas, des échantillons naturellement contaminés ne sont pas disponibles pour réaliser les études de stabilité dans la matrice. Pour les méthodes de dépistage biologiques, dans la plupart des cas, c'est l'approche bibliographique qui est retenue pour estimer la stabilité des analytes.

La caractérisation de la méthode par la détermination des caractéristiques de performance doit être suivie de la validation de la méthode, grâce à des critères d'acceptabilité prédéterminés. Ces critères vont être présentés et discutés.

3.2.1.3.5. Critères de performance :

a. Limites de détection

Limite de détection (Décision 93/256/CE)

Dans la décision 93/256/CE, le critère était, que la LOD soit « suffisamment basse pour que les résidus puissent être détectés à ce niveau » pour les substances à LMR, et qu'elle soit « aussi basse que possible » pour les substances interdites.

Sensibilité 90/95% (AOAC)

Les critères d'acceptation pour l'AOAC concernant la sensibilité 90/95% sont une sensibilité inférieure ou égale aux limites réglementaires, européennes ou américaines, en fonction du marché visé par le kit.

Capacité de détection (Guide de validation européen suivant la décision 2002/657/CE)

Les critères de validation pour cette caractéristique sont :

- 1) le cut-off Fm doit être inférieur au seuil de positivité T pour une méthode immunologique de dépistage,**
- 2) le risque d'obtenir un résultat faux-négatif à la CCD doit être inférieur ou égal à 5 % (erreur beta), ce qui permet de déterminer le CC β comme égal à la CCD,**
- 3) le CC β doit être inférieur ou égal à la limite réglementaire.**

1) Pour les méthodes immunologiques (eg. ELISA, biocapteur), la validité de la méthode va découler des calculs des facteurs T et Fm.

Pendant les essais de validation de la méthode, les facteurs T et Fm sont calculés chaque jour, à partir des résultats des échantillons blancs et des résultats des échantillons supplémentés. L'essai est valide si et seulement si le cut-off Fm est inférieur au seuil de positivité T. A la fin des essais de caractérisation, tous les résultats sont regroupés en une analyse globale. De nouveau, les facteurs T et Fm sont calculés. Si le T global est supérieur au Fm global, l'essai est valide. Dans certains cas, le T global pourra être inférieur au Fm global, mais l'essai sera quand même considéré comme valide si les T et Fm journaliers sont valides.

Le nombre de résultats faux-négatifs pour chaque jour d'essai va être déterminé, soit par rapport au T, soit par rapport au Fm, en fonction du seuil que l'on choisit pour interpréter les données. Selon toute logique, si le T est choisi comme valeur seuil pour décider de la positivité des échantillons, la capacité de détection pourra être abaissée, mais le pourcentage de résultats faux-positifs augmentera. Par contre, si le Fm est choisi comme seuil, le CC β sera un peu plus élevé, mais le risque d'obtenir des résultats faux-positifs diminuera. Dans tous les cas, au final, le nombre de résultats faux-négatifs global (inter-jours) doit être inférieur ou égal à 5% pour pouvoir conclure que la concentration cible de dépistage, qui correspond à la concentration en antibiotiques dans les échantillons supplémentés, est le CC β de cet antibiotique. Si c'est le cas, à cette concentration, le risque d'obtenir des résultats faux-négatifs sera bien inférieur ou égal à 5 %. De plus, si le cut-off Fm a été calculé par exemple avec une concentration de supplémentation égale à la ½ LMR, le risque d'avoir des résultats faux-négatifs au niveau de la LMR est bien inférieur au 5 % comme recommandé.

Au contraire, si plus de 5 % des échantillons supplémentés donnent des résultats faux-négatifs (supérieurs à la valeur seuil), la CCD choisie était trop faible. De nouveaux essais de caractérisation devront être entrepris avec une concentration plus élevée.

L'expérience nous a montré que le facteur Fm en pratique est couramment utilisé. En effet, même si un taux de faux-positifs maximum n'est pas fixé par la réglementation européenne, contrairement au taux de faux-négatifs (5%), le taux de faux positifs ne doit pas être trop élevé. En effet, chaque

échantillon positif en dépistage devra être confirmé en CL-SM/SM, ce qui est coûteux en investissement, en analyse et en temps personnel. Donc, le choix du Fm est le plus souvent un bon compromis entre un $CC\beta$ assez bas (inférieur à la limite réglementaire) et un taux de faux-positifs raisonnable.

2) Pour les tests de dépistage, **le critère de performance** est l'erreur β (taux de faux négatifs) qui **doit être inférieure à 5%**, à la concentration cible de dépistage [13].

Un résultat faux-négatif en dépistage est un résultat négatif obtenu pour un échantillon qui contient un résidu d'antibiotiques. La capacité de détection $CC\beta$ est égale à la concentration qui donne un taux de faux-négatifs de 5 % maximum. Par exemple, la survenue d'un ou aucun résultat faux-négatif après l'analyse de 20 échantillons supplémentés (*eg.* à la LMR) est suffisante pour démontrer que la capacité de détection est inférieure ou égale à la concentration testée (*eg.* à la LMR).

3) Concernant une méthode de dépistage pour une substance autorisée, le $CC\beta$ doit être inférieur ou égal à la LMR, alors que pour une méthode de confirmation, le $CC\beta$ est forcément supérieur à la LMR. En effet, pour une méthode de confirmation pour une substance autorisée, le $CC\alpha$ et le $CC\beta$ sont déterminés à partir d'échantillons supplémentés à la LMR, dans une des approches. Le $CC\alpha$ qui est la limite de décision pour une méthode de confirmation est logiquement supérieur à la LMR puisque l'objectif est de conclure avec un risque de 5 % minimum que la LMR est dépassée. Par contre, l'objectif d'une méthode de dépistage est de pouvoir détecter à des niveaux inférieurs à la LMR la présence potentielle d'un analyte. Ceci explique que, pour une méthode de dépistage pour une substance autorisée, le $CC\beta$ doit forcément être inférieur à la LMR pour conclure à la validité de la méthode.

En ce qui concerne une substance interdite, les objectifs de la méthode de dépistage comme de la méthode de confirmation sont de pouvoir détecter pour le dépistage ou d'identifier et de quantifier pour la confirmation, les concentrations en analytes les plus basses possible. Il faut pouvoir discriminer des échantillons blancs, qui produisent un bruit de fond, d'échantillons supplémentés à la concentration la plus basse possible. Dans ce cas, le $CC\beta$ du dépistage et le $CC\alpha$ de la méthode de confirmation doivent être les plus bas possible et inférieurs à la LPMR ou la RPA si elles existent.

Si le $CC\beta$ est supérieur à la limite réglementaire (LR), la méthode est non valide pour cet analyte (**Figure 43**). La concentration cible de dépistage peut toutefois être augmentée et l'exercice de validation répété si le $CC\beta$ doit être déterminé plus précisément.

Figure 43. Logigramme : conclusions de la détermination du CC β .

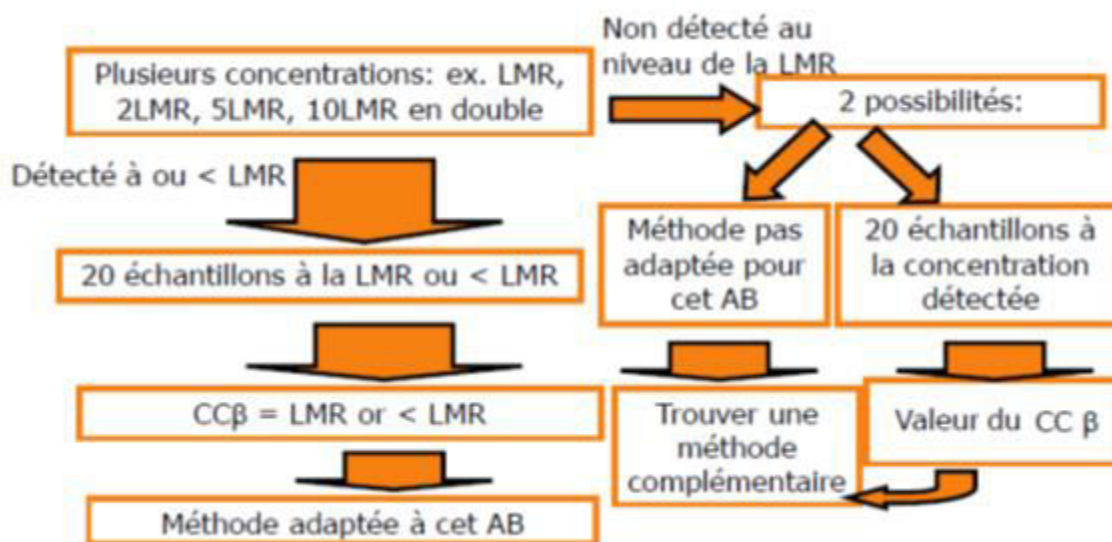


Tableau 18. Critères de validité pour estimer la fidélité de la méthode [13].

Critère sur le	Concentration < 100 µg/kg	Concentrations ≥ 100 µg/kg
CV de reproductibilité		CV calculé avec l'équation de Horwitz [13]. $CV = 2^{(1 - 0,5 \log C^*)}$
CV de répétabilité	Aussi bas que possible	la moitié et les deux tiers des CV calculés grâce à Horwitz
CV de reproductibilité intralaboratoire (fidélité intermédiaire)		CV intralaboratoire < CV de reproductibilité

*C est la fraction de masse exprimée comme une puissance (exposant) de 10 (par exemple 1 mg/g = 10⁻³), pour les concentrations supérieures à 100 µg/kg.

b. Spécificité

Le critère de l'AOAC est que les 60 échantillons blancs testés doivent donner un résultat négatif pour que le test soit déclaré valide (aucun résultat faux-positif autorisé). Par contre, au niveau européen (décision 2002/657/CE), il n'existe pas de taux de faux-positif maximum, contrairement au taux de faux-négatifs. En effet, un résultat faux-positif (un résultat positif pour un échantillon exempt de résidus) n'implique pas de risque pour la santé du consommateur. Par contre, comme tous les échantillons dépistés positifs doivent être obligatoirement confirmés par une méthode physico-chimique, un taux élevé de résultats faux-positifs entraîne des surcoûts de confirmation.

c. Fidélité

Le critère appliqué sur le coefficient de variation (CV) est lié à la concentration en analyte supplémenté. Suivant la décision 2002/657/CE, le CV inter-laboratoires pour l'analyse répétée d'un matériau de référence ou supplémenté, dans des conditions de reproductibilité, ne doit pas dépasser le niveau calculé grâce à l'équation d'Horwitz [13] (**Tableau 18**). Pour les concentrations inférieures à 100 µg/kg, l'équation d'Horwitz n'est pas adaptée car elle donne des valeurs trop élevées. Pour les concentrations supérieures à 100 µg/kg, Thompson propose une équation [293].

d. Applicabilité

Des critères ont été fixés dans le guide de validation européen des méthodes de dépistage pour conclure sur l'applicabilité de la méthode. Ces critères varient en fonction de l'approche choisie pour déterminer l'applicabilité.

- Si l'applicabilité a été évaluée dès la validation initiale, la méthode sera déclarée applicable aux matrices testées si le taux de faux-positifs est raisonnable et si le(s) $CC\beta$ du (ou des) analyte(s) validé(s) sont bien inférieur(s) ou égal (ou égaux) à la (ou aux) LR.
- Si l'applicabilité a été évaluée à posteriori pour une nouvelle matrice, la méthode sera déclarée applicable à la nouvelle matrice si le taux de faux-positif est similaire et si le(s) $CC\beta$ (s) dans la nouvelle matrice est (ou sont) proche(s) de celui (ou de ceux) déterminé(s) lors de la validation initiale. Aucune règle n'est fixée dans le guide pour limiter l'écart entre les taux de faux-positifs et/ou les $CC\beta$.

En conclusion, si des modifications mineures doivent être apportées à la méthode en fonction de la matrice, elles doivent être annoncées dans le protocole. Si l'applicabilité ne peut pas être prouvée, la méthode doit être entièrement validée pour la nouvelle matrice.

e. Robustesse

Une méthode analytique est robuste si les résultats ne sont pas (très) sensibles aux variations des conditions expérimentales. Les caractéristiques de performance (spécificité, $CC\beta$) doivent être déterminées pour tous les facteurs de variation qui ont montré un impact significatif sur le résultat de l'analyse. Lorsqu'un facteur influence notablement les résultats des mesures, des expériences

supplémentaires doivent être conduites pour déterminer les limites d'acceptabilité de ce facteur. De plus, l'impact de ce facteur sur les caractéristiques de performance doit être décrit dans le rapport de validation et ce facteur clairement identifié dans le protocole.

Il faut définir au préalable dans le plan de validation les critères applicables. Par exemple, si l'impact d'un facteur consiste en l'augmentation du CC β qui devient supérieur à la limite réglementaire, ce facteur devra être parfaitement maîtrisé en routine. Les limites d'acceptabilité de ce facteur devront être indiquées dans le protocole afin que les performances de la méthode ne soient pas affectées.

f. Stabilité

Au niveau de l'AOAC, la survenue d'au moins un résultat négatif pour un échantillon supplémenté testé, pour une méthode qualitative de dépistage des antibiotiques, entraîne la conclusion d'instabilité de l'analyte dans la matrice ou du kit [23].

Aucun critère de stabilité n'est fixé dans la décision européenne 2002/657/CE [13].

En interne, dans notre laboratoire, le guide de validation des méthodes CLHP fixe un intervalle maximum de 90 à 110 % définit les limites de stabilité de l'analyte en solution. Les critères acceptables pour la stabilité de l'analyte dans la matrice sont définis selon la méthode, avant la validation en fonction des concentrations étudiées. L'effet des cycles congélation-décongélation peut aussi être étudié.

3.2.1.3.6. Utilisation de la méthode en routine après la validation :

Après la validation initiale, l'analyse régulière de contrôles qualité (QC) pendant les analyses de routine est très importante. Ces contrôles qualité peuvent être un (ou des) échantillon(s) blanc(s) comme contrôle qualité négatif (QC-) et un (ou des) échantillon(s) supplémenté(s) au CC β comme contrôle qualité positif (QC+). Le choix des analytes à inclure dans les QC+ en routine doit s'inspirer des règles de la validation initiale, c'est-à-dire choisir l'analyte le moins bien détecté par la méthode ou l'analyte le plus pertinent dans un plan de contrôle national si c'est l'objectif de la méthode.

La préparation et l'analyse de QC en routine s'applique aussi bien aux méthodes de dépistage qualitatives, qu'aux méthodes quantitatives. Les résultats de ces QC doivent être enregistrés afin de vérifier les performances de la méthode. Il faudra avoir un taux de faux-négatifs sur le long terme inférieur ou égal à 5%.

- Pour une méthode qualitative microbiologique (*eg.* Test en tube), si le QC+ donne un résultat négatif ou si le QC- donne un résultat positif, la série d'analyses sera éliminée et il faudra refaire les analyses. Dans les deux cas, une analyse des causes sera réalisée.
- Pour une méthode immunologique de type ELISA (qualitatif ou quantitatif), les QC+ permettent de calculer un « cut-off » journalier F_m , et les QC- (blancs) un seuil de positivité T . En routine, la détermination de ces 2 seuils à partir des QC journaliers permet de conclure sur la positivité des échantillons analysés le jour même. Si la moyenne des résultats (ou le résultat) de l'échantillon inconnu est inférieure à la valeur seuil, l'échantillon est présumé contenir une concentration en antibiotique supérieure au CC β . Inversement, si la moyenne des résultats (ou le résultat) de

l'échantillon inconnu est supérieure à la valeur seuil, l'échantillon est présumé contenir une concentration en antibiotique inférieure au CC β ou aucun résidu. Dans tous les cas, il est déclaré négatif.

L'enregistrement et l'analyse des résultats des QC (ou des T et Fm journaliers) peuvent se faire sous forme de carte de contrôle, afin de suivre la dérive éventuelle des résultats d'essais [294]. La vérification continue de la méthode avec l'analyse des QC a plusieurs objectifs et avantages:

- Vérifier et tracer les performances de la méthode
- Ajouter des données au rapport de validation initiale ou de transfert,
- Vérifier la qualité de lots différents de kits commerciaux,
- Vérifier la stabilité et la qualité des réactifs,
- Vérifier l'aptitude des analystes qui mettent en place la méthode,
- Valider les essais journaliers (T>Fm),
- Conclure sur le statut des échantillons inconnus.

3.2.1.3.7. Conclusions sur l'approche critère par critère

En conclusion, l'approche actuelle de l'AOAC critère par critère pour la validation des méthodes de dépistage des résidus d'antibiotiques présente plusieurs points forts :

- La validation est réalisée, au moins en grande partie par un laboratoire indépendant. De plus, la revue des résultats par des relecteurs est un plus pour la certification AOAC.
- La notion de niveau de sensibilité à 90/95% est intéressante. Cette limite est déterminée précisément, à partir d'une courbe dose-réponse, basée sur l'analyse de plusieurs concentrations et d'un grand nombre de réplicats.
- Les effets masquants potentiels d'autres classes d'antibiotiques sur la détection des antibiotiques ciblés par le test sont étudiés.
- La robustesse est aussi un élément très important à déterminer : l'impact de facteurs primordiaux pour le lait comme le taux de cellules somatiques par exemple, sur la détection des antibiotiques est mesuré. De même, l'impact de l'utilisation de plusieurs lots de kits ayant des péremptions différentes est étudié.

Toutefois, l'approche de l'AOAC présente aussi certains points faibles :

- La charge de travail est très importante car le nombre d'échantillons à analyser pour déterminer le niveau de sensibilité 90/95% est très élevé.
- De ce fait, le nombre d'antibiotiques validés diminue. Par exemple, pour des tests spécifiques des bêta-lactamines, la sensibilité 90/95% est déterminée seulement pour 5 bêta-lactamines [287], voire les 6 bêta-lactamines autorisées aux Etats-Unis [288]. Cette validation est ciblée sur les utilisateurs américains, puisqu'aucune autre bêta-lactamine n'est étudiée. De plus, les critères de validation pour la sensibilité sont bien entendu les limites de tolérance américaines. Pour un test à large spectre, le choix des antibiotiques représentatifs est basé uniquement sur la famille des bêta-lactamines pour la détermination de la sensibilité 90/95%

[23]. 29 antibiotiques ont été testés à deux concentrations mais avec seulement trois échantillons.

- Ce protocole est impossible à réaliser sur d'autres matrices que le lait, par exemple sur le muscle. D'ailleurs, lors de la validation AOAC du Premi@Test, un test commercial microbiologique en ampoules, à large spectre, seul le niveau de sensibilité 90/95% de la pénicilline G a été déterminé [295]. Un seul échantillon a été testé pour estimer le niveau de détection de 23 antibiotiques et 10 médicaments vétérinaires.
- Contrairement à la certification AFNOR, la validation d'une méthode selon le programme « Performance Tested Methods » (PTM) de l'AOAC, qui est appliqué aux méthodes alternatives, dont les kits de dépistage des résidus d'antibiotiques, ne comporte pas d'étude collaborative.

Deux évolutions majeures sont à noter concernant les protocoles de validation AOAC :

- La notion de $CC\beta$ est apparue pour la première fois, récemment, dans les validations AOAC dans le cadre du programme PTM [296], ainsi que des références à la décision européenne 2002/657/CE et au guide de validation européen (2010). Le concept est le même, l'erreur beta est aussi fixée à 5 %. Le $CC\beta$ est déterminé suivant une approche proposée dans le guide européen, mais qui dans les faits n'est pas souvent utilisée. Plusieurs concentrations d'un antibiotique sont d'abord testées avec des réplicats pour estimer le $CC\beta$, à partir de la courbe dose-réponse. Ensuite, 60 échantillons sont supplémentés au $CC\beta$ estimé et le $CC\beta$ est déterminé plus précisément. C'est bien le $CC\beta$ qui est déterminé et non plus la sensibilité 90/95%.
- L'approche globale de la POD est à présent utilisée par l'AOAC, en combinaison avec une approche critère par critère, pour les kits de microbiologie alimentaire, depuis la parution de nouvelles lignes directrices en 2014 [279]. Cette approche combinée pourrait être aussi appliquée dans l'avenir aux kits de dépistage des résidus d'antibiotiques dans les aliments.

Il faudra attendre le prochain kit antibiotiques certifié PTM par l'AOAC pour voir si la détermination du $CC\beta$ se poursuit ou si l'approche de la POD est mise en place.

Dans la décision 93/256/CE, le calcul de la limite de détection était basé sur l'analyse des blancs uniquement (20 échantillons) et pas du tout sur l'analyse d'échantillons supplémentés [290]. Dans le cas de méthodes microbiologiques à très large spectre, la détermination de la LOD permettait seulement de fixer le seuil de positivité (valeur de la zone d'inhibition au-dessus de laquelle il est conclu qu'un échantillon est positif). Cette limite de détection n'apportait aucune information sur le niveau de détection associé à chacun des antibiotiques détectés par la méthode, au contraire des capacités de détection utilisées actuellement. De plus, la décision 93/256/CE précisait pour les méthodes de dépistage que l'incidence des résultats faux-négatifs devait être « la plus faible possible ». Toutefois, aucune détermination ou estimation du taux de faux-négatifs n'était prévue par cette décision.

La décision 2002/657/CE était censée apporter des améliorations par rapport à la décision 93/256/CE. Les avancées concernant les méthodes de confirmation étaient flagrantes dans cette nouvelle décision, prenant bien en compte les progrès technologiques survenus depuis la première décision

93/256/CE [290]. Concernant les méthodes de dépistage, les apports de cette nouvelle décision étaient beaucoup plus discrets. Toutefois, le concept le plus important de cette décision pour les méthodes de dépistage, la capacité de détection ou $CC\beta$ a beaucoup apporté pour la caractérisation et la validation des méthodes de dépistage. Comme nous l'avons dit précédemment, dans la décision 93/256/CE, le calcul de la limite de détection à partir du bruit de fonds produit par les échantillons blancs n'était pas très satisfaisant. Le concept plus statistique du $CC\beta$ associé à un risque d'erreur de 5 % maximum est beaucoup plus intéressant et performant. En outre, la nouvelle décision fixe un taux maximum de résultats faux-négatifs de 5%. La décision européenne 93/256/CE exigeait seulement que l'incidence des faux résultats négatifs soit la plus faible possible, sans fixer de taux maximum. Toutefois, pour la mise en pratique de ce concept, il aura fallu toutefois beaucoup de tâtonnements et finalement la rédaction du guide européen pour la validation des méthodes de dépistage (2010) [256]. Dans l'intervalle de temps entre la parution de la décision 2002/657/CE et celle du guide de validation Européen, quelques laboratoires ont publié des validations de méthodes en essayant de se conformer aux exigences de la décision 2002/657/CE et en « copiant » les procédures applicables pour les méthodes de confirmation, bien développées dans la décision. Deux exemples montrent ce « copier-coller ». Tout d'abord, la limite de décision ($CC\alpha$) devait nécessairement être déterminée pour des méthodes de dépistage dans une des approches de détermination du $CC\beta$, alors que le $CC\alpha$ n'était pas exigé. Pourtant c'était la seule possibilité pour une méthode biologique de dépistage, puisque cette approche était de calculer le $CC\beta$ à partir d'échantillons supplémentés au $CC\alpha$. Il fallait donc d'abord déterminer le $CC\alpha$ au préalable à partir d'échantillons blancs [297, 298]. L'autre approche de la décision était de déterminer les $CC\alpha$ et $CC\beta$ à partir de la courbe d'étalonnage linéaire selon la norme ISO 11843 [291], ce qui n'est pas toujours possible avec les méthodes biologiques de dépistage. La norme ISO 11843 a été étendue à des modèles non linéaires [299]. Toutefois, ces approches sont peu ou pas appliquées dans les laboratoires, pour les résidus d'antibiotiques. De plus, toutes les validations de kits (eg. ELISA) ou de biocapteurs immunologiques approchaient ce type de méthodes obligatoirement comme une méthode quantitative [26, 163]. Au contraire, à présent la tendance va vers une approche qualitative de ce type de méthodes pour la validation et pour l'utilisation en routine [61, 187]. D'une part, cela simplifie la validation et diminue la charge de travail. De plus, il n'est pas nécessaire pour une méthode de dépistage de fournir une valeur quantitative de la concentration en analyte contenue dans l'échantillon. La détermination de la valeur seuil et l'utilisation des QC négatifs et surtout positifs permet d'estimer si la concentration en antibiotiques dans l'échantillon est inférieure ou supérieure au $CC\beta$, si et seulement si toutefois une seule molécule est reconnue par l'anticorps ou le biorécepteur.

En conclusion, les avantages de l'approche de validation critère par critère sont les suivants :

- des critères objectifs pré-déterminés auxquels les caractéristiques de performance déterminées sont comparées,
- par rapport à la décision 93/256/CE, l'approche du $CC\beta$ est plus statistique que la LOD,
- un critère de risque de 5 % est fixé sur les faux-négatifs, alors qu'il n'existait pas dans la décision 93/256/CE,

- le CC β est la caractéristique majeure apportée par la décision européenne 2002/657/CE [13] qui permet d'évaluer la performance d'une méthode de dépistage et son aptitude à répondre à l'usage attendu,
- le CC β peut/doit être utilisé pour suivre les performances de la méthode en continu, pendant son application en routine, comme valeur de supplémentation du contrôle qualité positif (eg. cartes de contrôle),
- les valeurs seuils Fm et T sont un apport très important du guide de validation européen, pour vérifier la performance et la validité des essais en routine au jour le jour et pour aider à la décision sur les échantillons inconnus en routine. Ces valeurs seuils sont liées au CC β .
- grâce à l'approche proposée pour la détermination de l'applicabilité dans le guide de validation européen [256], la détermination du CC β et de la spécificité permet de vérifier, avec des essais réduits, si la méthode est valide dans le cas d'une extension de matrice (eg. denrée alimentaire, espèce animale),
- le CC β , combiné à la spécificité, va permettre de comparer deux méthodes, en déterminant ces deux caractéristiques pour chacune des méthodes.

Les inconvénients de cette approche sont les suivants :

- Le nombre d'essais à réaliser pour déterminer le(s) CC β , principalement pour les méthodes microbiologiques à large spectre représente une charge de travail lourde, malgré le choix d'une liste d'antibiotiques représentatifs et l'utilisation de tissus simulés pour les matrices solides (eg. muscle).
- Avec cette approche de la détermination du CC β , le taux de faux-négatifs à cette concentration est forcément inférieur ou égal à 5 % mais on ne connaît pas ce taux précisément comme le font remarquer Galarini *et al.* [300].
- Comme plusieurs caractéristiques de performance sont déterminées en parallèle, certaines caractéristiques peuvent être valides et pas d'autres. Dans ce cas de figure, les conclusions sont contradictoires. En général, la conclusion finale est que la méthode n'est pas valide puisqu'une des caractéristiques ne l'est pas. Toutefois, dans le cadre des méthodes biologiques, par exemple, le CC β peut ne pas être valide pour certains antibiotiques à rechercher, mais être valide pour d'autres. Si toutes les autres caractéristiques sont valides, on peut conclure à la validité de la méthode, mais seulement sur une liste de composés dont le CC β est valide.
- Le concept des CC α et CC β n'est utilisé que dans la décision 2002/657/CE [13], donc que pour les résidus concernés par la directive 96/23/CE. Donc il y a un manque d'harmonisation pour certains laboratoires qui travaillent sur d'autres contaminants chimiques, par exemple les pesticides. Toutefois, le concept de SDL dans la SANCO/12571/2013 [301] pour les pesticides pour les méthodes de dépistage qualitatives est équivalent au concept de CC β .

3.3. CONCLUSION GENERALE SUR LES APPROCHES DE VALIDATION

Les différentes caractéristiques de performance, les expériences nécessaires pour les déterminer, le mode d'évaluation et les critères de validation associés, en fonction de l'approche choisie, sont synthétisés dans le **Tableau 19** pour les méthodes de dépistage qualitatives et dans le **Tableau 20** pour les méthodes de dépistage quantitatives.

L'approche de validation par comparaison d'une méthode alternative par rapport à une méthode de référence n'est pas satisfaisante pour valider une méthode de dépistage des résidus d'antibiotiques. Le choix de la méthode de référence est difficile et entraîne des complications dans le déroulement de l'étude et dans l'interprétation des résultats. En effet, la méthode alternative est souvent basée sur des principes différents et présente le plus souvent des limites de détection plus basses que la méthode de référence.

L'approche globale pour les méthodes qualitatives (POD), comme pour les méthodes quantitatives (profil d'exactitude) sont applicables aux méthodes de dépistage des résidus d'antibiotiques. La représentation graphique permet de visualiser rapidement le biais d'une méthode et sa fidélité. De plus, le profil d'exactitude permet de choisir le modèle d'étalonnage le plus adapté, et pas forcément celui utilisé classiquement pour un type de méthodes. Enfin, la combinaison de deux caractéristiques en une, pour les deux approches permet de conclure plus simplement sur la validité d'une méthode que chaque caractéristique prise indépendamment. Le nouveau protocole AOAC a intégré l'approche POD pour déterminer la sensibilité 90/95% pour des kits de dépistage des résidus d'antibiotiques. Le profil d'exactitude a déjà été utilisé pour valider des méthodes de dépistage quantitatives, par exemple des méthodes de microbiologie alimentaire [277].

L'approche classique critère par critère recommandée par la décision européenne 93/256/CE n'était pas très performante. Le calcul de la LOD basé sur l'analyse de 20 blancs n'était pas satisfaisant, puisque la LOD n'était pas basée sur l'analyse d'échantillons supplémentés avec l'analyte d'intérêt. Aucun taux maximum n'était fixé pour le taux de résultats faux-négatifs. La décision européenne 2002/657/CE [13] qui a abrogé la décision 93/256/CE [290] a apporté une caractéristique déterminante pour la validation des méthodes de dépistage et pour le suivi de la performance de la méthode en routine, la capacité de détection ou CC β . Le taux de faux-négatifs a été fixé à 5% (erreur β) pour les méthodes de dépistage.

Tableau 19. Caractéristiques de performance pour les méthodes qualitatives de dépistage, en fonction de l'approche de validation.

Approche	Caractéristiques	Expériences	Evaluation	Critères d'acceptation	
93/256/CE	LOD	20 blancs	3*SD bruit du fond	≤ LMR	le plus bas possible
POD	LOD50 ou LOD95	Plusieurs concentrations, plusieurs répétitions	A partir de la courbe dose réponse	≤ LMR	le plus bas possible ou ≤ LPMR
	RLOD			Proche de 1 ???	
AOAC	Sensibilité 90/95%	Plusieurs concentrations, plusieurs répétitions	A partir de la courbe dose réponse	≤ limites réglementaires	/
2002/657/CE Guide de validation (2010)	CCβ	20 à 60 supplémentés à la CCD	Concentration pour laquelle le taux de faux-négatifs ≤ 5%	≤ LMR	le plus bas possible ou ≤ LPMR
Toutes	Spécificité	N échantillons blancs	Taux de faux-positifs	Taux minimal	
Guide de validation (2010)	Applicabilité	20 blancs et 20 supplémentés au CCβ de la 1ère matrice	Concentration pour laquelle le taux de faux-négatifs ≤ 5%	CCβ de la 2 ^e matrice équivalent au CCβ de la 1ère matrice	

Ecart-type (Standard deviation ou SD) ; Limite de détection (LOD) ; Limite de détection relative (RLOD) ; Concentration cible de dépistage (CCD) ; Limite Maximale de Résidus (LMR) ; Limite de Performance Minimale Requise (LPMR).

Toutefois, le guide de validation des méthodes de dépistage rédigé par les LRUE et la DGSANTE en 2010 [256] a été nécessaire pour aider les laboratoires à mettre en place cette décision. Pour les méthodes microbiologiques, les avancées principales concernent la recommandation de validation pour une liste d'antibiotiques représentatifs, faire varier le nombre d'échantillons à analyser la concentration cible en fonction du ratio CCD/LMR et les tissus simulés pour les matrices solides comme le muscle. Pour les méthodes immunologiques, le point principal concernées valeurs seuils T et Fm à partir des échantillons blancs et des échantillons supplémentés à la CC, pour déterminer le CC β et valider la méthode, pour suivre les performances de la méthode en routine (cartes de contrôle) et pour déterminer l'applicabilité de la méthode à post \acute{e} riori (*i.e.* extension de matrices).

Tableau 20. Caractéristiques de performance pour les méthodes quantitatives de dépistage, en fonction de l'approche de validation.

Approche	Caractéristiques	Expériences	Evaluation	Critères d'acceptation
93/256/CE	Exactitude <hr/> Précision	Concentration proche de la LMR Répétitions en conditions de répétabilité et de reproductibilité	Différence entre la valeur moyenne mesurée pour le matériel de référence et sa valeur réelle, exprimée en pourcentage <hr/> Calcul des CV intra-jour et inter-jours	¹ Faux-positifs acceptés Faux-négatifs au minimum
2002/657/CE Guide de validation	Fidélité	Concentrations 1, 1,5 et 2 fois la LPMR ou 0,5, 1 et 1,5 fois la LMR, plusieurs répétitions (jours)	Calcul des CV intra-jour et inter-jours	$\leq CV =$ f(concentration) (équation d'Horwitz, de Thompson)
Profil d'exactitude	Exactitude = justesse + fidélité	Plan d'étalonnage + plan de validation (plusieurs niveaux de concentration, plusieurs répétitions)	Choix du modèle d'étalonnage optimal, choix de β et λ ; construction des profils d'exactitude	Intervalle de tolérance (proportion β) le plus possible compris dans l'intervalle d'acceptabilité (λ réglementaire)

¹ Des critères en termes de coefficients de variation (CV) ne sont indiqués dans cette décision que dans le chapitre méthodes de confirmation.