

CHAPITRE 1.1.6.

VALIDATION DES EPREUVES DIAGNOSTIQUES POUR LES MALADIES INFECTIEUSES DES ANIMAUX TERRESTRES

INTRODUCTION

Une validation et une vérification adéquates des caractéristiques de performance des tests diagnostiques pour les maladies infectieuses sont essentielles pour garantir que les essais sont appliqués et interprétés d'une manière robuste et défendable d'un point de vue scientifique (Colling & Gardner, 2021). Depuis sa première adoption en 1996, le Processus de mise au point et de validation d'un essai (Figure 1) de l'Organisation mondiale de la santé animale (OMSA, fondée en tant que OIE) a servi de norme internationalement reconnue pour la validation des tests diagnostiques vétérinaires pour les maladies infectieuses.

La validation est un processus qui détermine l'adéquation d'un essai¹ ayant été correctement développé, optimisé et normalisé, en vue d'un objectif déterminé. Tous les essais diagnostiques, qu'ils soient utilisés au laboratoire ou hors laboratoire, doivent être validés pour l'espèce chez laquelle ils seront utilisés. La validation comprend l'estimation des caractéristiques de performance analytique et diagnostique d'un test. Dans le contexte de ce chapitre, un essai qui a accompli les trois premières étapes du processus de validation (voir Figure 1 ci-dessous), y compris la caractérisation de sa performance, peut être désigné comme «validé pour le ou les objectifs initialement prévus». Pour maintenir la validité d'un essai, sa performance devrait toutefois être soigneusement surveillée dans les conditions d'une utilisation de routine en effectuant un suivi régulier du comportement des contrôles de l'essai pour chaque cycle d'analyses ainsi que par un suivi continu durant son utilisation en routine dans la population cible sur la durée. Si ces résultats s'avéraient ne plus être cohérents avec les données de validation d'origine, le test pourrait être considéré comme inadéquat pour le ou les objectifs prévus.

Les essais appliqués aux individus ou aux populations ont de nombreux objectifs, tels que contribuer à : documenter le statut indemne de maladie d'un pays ou d'une région, prévenir la propagation d'une maladie par le biais des échanges commerciaux, contribuer à l'éradication d'une infection d'une région ou d'un pays, confirmer le diagnostic des cas cliniques, estimer la prévalence d'infection pour permettre l'analyse des risques, identifier les animaux infectés afin de mettre en œuvre des mesures de lutte et classer les animaux par rapport à la santé des troupeaux ou à leur statut immunitaire post-vaccination. Un seul essai peut être validé pour un ou plusieurs objectifs en optimisant ses caractéristiques de performance pour chacun des objectifs, p. ex. en définissant une sensibilité diagnostique (SeD) élevée associée à une spécificité diagnostique (SpD) moindre pour un essai de dépistage ou, inversement, en définissant une SpD élevée associée à une SeD moindre pour un essai de confirmation (voir Section A.1 Définition du ou des objectif(s) prévu(s) pour l'essai).

Ce chapitre est axé sur les critères à remplir lors de la mise au point de l'essai et de la validation de tous types d'essais ainsi que sur les mesures utilisées pour caractériser la performance du test. Considérer que la mise au point de l'essai fait partie intégrante du processus de validation peut sembler paradoxal, mais en réalité, trois des critères à évaluer pour la validation (définition du ou des buts poursuivis, optimisation et normalisation) afin d'aboutir à un essai validé comprennent des étapes du processus de développement de l'essai. Par conséquent, ce dernier mène naturellement

1 « Essai », « méthode de test » et « test » sont des termes synonymes aux fins de ce chapitre et sont donc utilisés indistinctement.

au processus de validation, l'un et l'autre nécessitant de satisfaire aux critères de validation. Les principes généraux décrits ici s'appliquent également aux maladies infectieuses qui ne sont pas listées par l'OMSA. Des conseils plus détaillés figurent dans une série de Recommandations pour la validation des tests diagnostiques (Section 2.2. Validation des tests diagnostiques, chapitres 2.2.1 à 2.2.8), conçues spécifiquement pour plusieurs types d'essais très différents (p. ex. détection d'anticorps, d'antigènes ou d'acides nucléiques) et apportant plus d'informations sur les questions spécifiques relatives à la validation des essais diagnostiques. Pour des informations spécifiques concernant les espèces sauvages, se référer au Chapitre 2.2.7 Validation des épreuves diagnostiques pour les maladies infectieuses applicables à la faune sauvage (Jia et al., 2020; Michel et al., 2021). Les informations qui figurent dans le chapitre 2.2.7, spécifiques aux espèces de la faune sauvage, peuvent également être utiles pour la validation des épreuves chez les animaux domestiques, par exemple lorsque le nombre ou la disponibilité des échantillons est limité.

Une compilation à jour des normes de validation pertinentes (OMSA et non-OMSA) ainsi que des documents d'orientation pour toutes les étapes de la validation des tests diagnostiques et des essais d'aptitude, depuis leur conception, leur analyse, leur interprétation jusqu'aux rapports et aux études de cas, sont fournis dans le numéro La validation scientifique des tests de diagnostic (Vol. 40, avril 2021) de la Revue scientifique et technique de l'OMSA. Des normes STARD pour la révision par des comités de lecture des études sur l'exactitude des tests diagnostiques existent et sont publiées pour les maladies infectieuses de l'homme, des animaux terrestres (paratuberculose, modèle bayésien de structure latente) (Bossuyt et al., 2015; Gardner et al., 2011; Kostoulas et al., 2017) et des animaux aquatiques (Gardner et al., 2019; Kostoulas et al., 2021). La vérification (Kirkland & Newberry, 2021) et les études de comparabilité (Reising et al., 2021) sont brièvement décrites à la fin de ce chapitre pour les essais qui ont accompli au moins l'étape 2 du processus de l'OMSA. Il est urgent de développer des lignes directrices et des normes pour la validation des examens de biologie médicale délocalisée (EBMD) dont l'utilisation augmente rapidement. Généralement, les EBMDs sont utilisés sur le terrain dans des conditions environnementales variables, sur une large gamme de types d'échantillons récoltés en conditions non stériles par des opérateurs dont le niveau d'expérience, de formation et de compétence varie. Les conditions de test sur le terrain, parmi lesquelles figurent les variations extrêmes de température et d'humidité ainsi que d'autres variables comme la qualité de l'eau ou des réactifs, l'inadéquation de la chaîne du froid, les compétences de l'opérateur et l'absence ou la piètre qualité de systèmes d'assurance qualité, peuvent toutes contribuer à une exactitude des tests inférieure à celle qui est rapportée par les fabricants de ces tests ou à celle qui est obtenue dans un laboratoire accrédité (voir Figure 1, Stade 5). Les conséquences d'un résultat de test positif et le besoin de tests de confirmation et de rapports par un laboratoire accrédité, notamment lors de la réalisation d'un test pour une maladie exotique, doivent être intégrées dans les politiques et les lignes directrices existantes pour les tests. Des normes et des recommandations spécifiques aux EBMDs, telles la check-list POKET (point-of-care key evidence tool) pour les rapports d'évidence pluridimensionnels, les cartes de score et les lignes directrices pour l'évaluation des EBMDs, les lignes directrices sur la qualité des pratiques dans les EBMDs non-instrumentés (c-à-d. ceux qui ne nécessitent aucun appareil spécifique) ou encore la norme ISO/TS 22583:2019 Orientation pour les opérateurs et les superviseurs de dispositifs d'EBMDs, fournissent des orientations pour le personnel de santé. Dans certains pays, en Allemagne par exemple, les EBMDs qui permettent le dépistage de maladies animales à notification ou à annonce obligatoire requièrent une autorisation officielle des autorités nationales d'enregistrement. Dans d'autres pays, l'accréditation des EBMDs par des organisations comme l'OMSA ou les autorités nationales de tests est encouragée, mais pas obligatoire (Halpin et al., 2021; Hobbs et al., 2020; et Section 2.5 Robustesse et solidité ci-dessous).

CONSIDERATIONS PRELIMINAIRES POUR LA MISE AU POINT ET LA VALIDATION DES ESSAIS

Tous les laboratoires doivent se conformer aux exigences du Chapitre 1.1.5 Gestion de la qualité dans les laboratoires de diagnostic vétérinaire (Newberry & Colling, 2021), afin de minimiser l'influence de facteurs qui ne dépendent pas du test lui-même, tels que l'instrumentation, les erreurs de l'opérateur, le choix des réactifs (chimiques ou biologiques), l'étalonnage, les cuves et les plateformes de réaction, la qualité de l'eau, le pH et l'ionocité des tampons et des diluants, les températures et les temps d'incubation ainsi que les erreurs dans la performance technique de l'essai. Des expériences complètes et bien conçues sont nécessaires pour développer et optimiser les essais dotés

de caractéristiques analytiques prometteuses. Les principes fondamentaux peuvent s'appliquer à tous les types d'essais et fournissent, s'ils sont conduits avec la rigueur appropriée, les fondements à des tests diagnostiques de qualité, adaptés à ou aux objectif(s) prévu(s) (Bowden *et al.* (2021). Dans leur revue des tests diagnostiques recommandés par l'OMSA, Cullinane & Garvey (2021) sont parvenus à la conclusion que les essais d'immuno-absorption enzymatique (ELISA) et les essais moléculaires sont les tests les plus couramment utilisés parmi les tests recommandés par l'OMSA, raison pour laquelle les exemples proposés dans les sections suivantes se focalisent sur ces méthodes (Mayo *et al.*, 2021).

La première étape du développement d'un essai consiste à définir l'objectif de l'essai puisque cela oriente toutes les étapes ultérieures du processus de validation. Les critères de validation sont les traits caractéristiques d'un essai qui représentent des mesures, normes ou facteurs décisifs sur lesquels un jugement ou une décision peut se baser. Si l'on considère les variables qui peuvent influencer la performance d'un essai, les critères entrant dans sa validation deviennent plus clairs. Ces variables peuvent être regroupées selon les catégories suivantes : (a) l'échantillon – individuel ou groupé, la composition de la matrice et les interactions hôte/organisme ayant une influence quantitative ou qualitative sur l'analyte cible ; (b) le système d'essai – physique, chimique, biologique ainsi que les facteurs liés à l'opérateur intervenant dans l'aptitude de l'essai à détecter un analyte spécifique dans l'échantillon ; et (c) l'interprétation du résultat de l'épreuve – l'aptitude du résultat d'une épreuve, dérivée du système d'essai, à prédire de manière précise le statut de l'individu ou de la population par rapport à l'objectif pour lequel le test est utilisé.

Le choix, la collecte, la préparation, la conservation et la gestion des échantillons sont des variables importantes pour la conception et la mise au point de l'essai afin de garantir la validité des résultats des épreuves. D'autres variables comme le transport, la chaîne de contrôle, le traçage des échantillons et le système de gestion des informations du laboratoire sont également des sources majeures de variations/d'erreurs qui prennent toute leur importance lorsque l'essai est utilisé pour les tests de routine. L'intégrité des résultats expérimentaux obtenus en laboratoire durant le développement et la validation de l'essai se mesure à la qualité de l'échantillon utilisé. La prévision des facteurs susceptibles d'avoir un effet négatif sur la qualité de l'échantillon doit précéder l'effort de validation de l'essai. Les échantillons de référence utilisés pour la mise au point et la validation de l'essai doivent être constitués de la même matrice que ceux qui seront utilisés pour l'essai (p. ex. sérum, tissu, sang total) et représentatifs de l'espèce à tester. Le matériel de référence doit représenter de manière appropriée la plage de concentration des analytes à détecter avec l'essai. Les biobanques virtuelles sont devenues des ressources importantes en matière de réactifs et d'échantillons durant le développement et la validation d'un test. Par exemple, une fois que la séquence génomique du coronavirus du syndrome respiratoire aigu sévère 2 (SARS-CoV-2) est devenue disponible, les membres de l'European Virus Archive – Global (EVAg) ont rapidement développé des stocks de virus, des tests diagnostiques et des contrôles positifs, pour parvenir, en 10 semaines, à plus de 1500 produits distribués dans le monde entier à des fins de diagnostic ou de recherche (Watson *et al.*, 2021). Le Chapitre 2.2.6 *Sélection et utilisation des échantillons et panels de référence* fournit un aperçu de la sélection et de l'utilisation des échantillons et des panels de référence pour évaluer des paramètres de validation pertinents tels que répétabilité, reproductibilité, SeD et SpD, etc. Les informations sur la collecte, la préparation, la conservation, la gestion et le transport des échantillons figurent aux Chapitres 1.1.2 *Prélèvement, expédition et stockage des échantillons pour le diagnostic* et 1.1.3 *Transport de matériel biologique*.

La matrice dans laquelle l'analyte cible est détecté (sérum, fèces, tissus, etc.) peut contenir des inhibiteurs endogènes ou exogènes susceptibles d'entraver la performance de l'essai. Cela concerne particulièrement les tests qui dépendent d'enzymes comme les réactions de polymérisation en chaîne (PCR) ou les méthodes immuno-enzymatiques (ELISA). D'autres facteurs influençant la concentration et la composition de l'analyte cible (notamment les anticorps) dans l'échantillon peuvent être principalement attribués à l'hôte et sont soit inhérents (p. ex. âge, sexe, race, état nutritionnel, gravidité, réactivité immunologique), soit acquis (p. ex. anticorps acquis de manière passive, immunité active induite par vaccination ou infection). Des facteurs extérieurs à l'hôte, comme la contamination ou la détérioration de l'échantillon, peuvent également influencer l'aptitude de l'essai à détecter l'analyte cible spécifique dans l'échantillon. Il est également important que les réactifs biologiques soient exempts d'agents étrangers susceptibles de mener à des résultats erronés.

LES CRITERES DE MISE AU POINT ET DE VALIDATION DES ESSAIS

L'exactitude et la précision sont deux paramètres indépendants qui définissent fondamentalement la performance d'un test diagnostique. La sensibilité, la spécificité et des fonctions de ces variables (p. ex. rapports de vraisemblance) changent avec les valeurs limites et représentent l'exactitude tandis que la répétabilité et la reproductibilité sont des mesures de la précision. Un test fiable est à la fois exact et précis. La performance d'un essai est influencée par plusieurs facteurs, à commencer par l'optimisation de l'essai. Après une optimisation initiale pour l'objectif prévu, les caractéristiques de performance de l'essai sont testées. L'essai pourra nécessiter une optimisation supplémentaire ou être jugé adapté à l'objectif prévu sur la base des résultats des travaux de validation. Le Secrétariat de l'OMSA pour l'enregistrement des kits diagnostiques (SRDK) a un processus de validation et de certification selon lequel les nouveaux tests jugés adéquats pour le ou les objectif(s) prévu(s) sont enregistrés².

Critères pour la mise au point et la validation des essais

- i) Définition du ou des objectif(s) prévu(s)
- ii) Optimisation
- iii) Normalisation
- iv) Répétabilité
- v) Sensibilité analytique
- vi) Spécificité analytique
- vii) Seuils (valeurs limites)
- viii) Sensibilité diagnostique
- ix) Spécificité diagnostique
- x) Reproductibilité
- xi) Adéquation avec le ou les objectif(s) prévu(s)

A. PROCESSUS DE MISE AU POINT DES ESSAIS

1. Définition du ou des objectif(s) prévu(s) pour l'essai

La norme ISO/IEC 17025 (2017) pour les laboratoires qui effectuent des tests et des calibrations stipule que les laboratoires doivent utiliser des méthodes et des procédures appropriées pour toutes les activités de laboratoire (Newberry & Colling, 2021). En d'autres termes, l'essai doit être adapté à l'objectif prévu. Si l'objectif du test n'a pas été défini *a priori*, cela peut engendrer des erreurs, tant dans la conception de l'essai que dans la détermination des paramètres essentiels du test, et peut potentiellement invalider la totalité du processus de développement et de validation de l'essai pour finalement aboutir à un test qui ne répond pas aux besoins des utilisateurs. Ce type d'erreur peut survenir lorsque, par exemple, la population de référence sélectionnée est inadéquate, car hétérologue par rapport à la population pour laquelle le test a été développé.

Critères pour l'interprétation des résultats

- Valeur prédictive positive (VPP)
- Valeur prédictive négative (VPN)
- Rapport de vraisemblance positif (RVP)
- Rapport de vraisemblance négatif (RVN)

L'évaluation qualitative et quantitative de l'aptitude d'un résultat de test positif ou négatif (c'est-à-dire sa valeur prédictive et son rapport de vraisemblance) à prédire avec une confiance élevée l'infection ou l'éventuelle exposition de l'animal ou de la population d'animaux constitue la préoccupation ultime pour la validation de l'essai (Section B.4.2). Cette aptitude est dépendante du développement d'un essai soigneusement optimisé et normalisé (Section A.2.3.) et le cumul de données de validation garantira l'aptitude à l'emploi de l'essai selon l'objectif prévu (Tableau 1).

Afin de s'assurer que les résultats du test permettent des déductions diagnostiques utiles sur des animaux ou des populations compte tenu de l'objectif prévu, le processus de validation comprend le développement initial et la documentation concernant le fonctionnement de l'essai ainsi que l'évaluation continue du contrôle qualité et de l'assurance qualité.

La Figure 1 présente le processus de validation de l'essai, à partir de sa conception jusqu'à sa mise en œuvre, son déploiement et son maintien, en passant par les processus de mise au point et de validation.

La première étape de la mise au point d'un essai réside dans le choix du type d'essai approprié et susceptible d'être validé pour un usage donné (adéquation à l'objectif).

2 <https://www.woah.org/fr/ce-que-nous-proposons/produits-veterinaires/kits-de-diagnostic/le-registre-des-kits-de-diagnostic/>

Les objectifs les plus courants énumérés dans le *Manuel terrestre* de l'OMSA sont les suivants :

- 1) contribuer à démontrer l'absence d'infection dans une population donnée (pays/zone/compartiment/troupeau) (prévalence apparente égale à zéro) :
 - 1a) « indemne » avec et/ou sans vaccination,
 - 1b) recouvrement du statut indemne après des foyers ;
- 2) certifier l'absence d'infection ou la présence de l'agent chez des animaux pris individuellement ou dans des produits à des fins de commerce/déplacement ;
- 3) contribuer à l'éradication de la maladie ou à l'élimination de l'infection dans des populations données ;
- 4) confirmer le diagnostic de cas suspects ou de cas cliniques (y compris confirmation d'un test de dépistage positif) ;
- 5) estimer la prévalence de l'infection ou de l'exposition à l'infection afin de faciliter l'analyse de risque (études, statut sanitaire du troupeau, mesures de lutte contre la maladie) ;
- 6) déterminer le statut immunitaire d'animaux pris individuellement ou de populations (après vaccination).

Tableau 1. Objectifs du test et importance relative de la sensibilité diagnostique (SeD), de la spécificité diagnostique (SpD), d'une valeur prédictive positive (VPP), d'une valeur prédictive négative (VPN), du rapport de vraisemblance d'un résultat de test positif (RVP) et du rapport de vraisemblance d'un résultat de test négatif (RVN)

Objectifs	Exemples et mesures de l'exactitude diagnostique en fonction de l'objectif du test (Reid <i>et al.</i> , 2022)
1a) Statut indemne sur la base de données historiques (avec ou sans vaccination)	Dans une population qui est historiquement indemne d'une maladie ou d'un agent pathogène particulier, la prévalence est de zéro (ou proche de zéro). Pour être adapté à son objectif, le test ou l'algorithme de test vise à minimiser les chances de résultats faux positifs et requiert idéalement une SpD élevée , une VPP élevée et un RVP élevé . Cela peut être obtenu avec un test unique doté d'une SpD élevée ou avec des tests en série ^(*) .
1b) Recouvrement du statut indemne après un foyer	Au cours d'un programme de contrôle d'une maladie fructueux, un passage graduel d'une prévalence élevée (durant le pic du foyer) à une prévalence basse (à la toute fin du foyer) peut être escompté. Durant les stades précoces du programme de test prouvant l'absence de maladie, alors que la prévalence de la maladie reste à un niveau non négligeable, un test adapté à son objectif doit avoir une SeD élevée , une VPN élevée et un RVN élevé . Cette approche minimise les chances de résultats faux négatifs et permet la détection des individus positifs. Cela peut être obtenu avec un test unique doté d'une SeD élevée ou avec des tests en parallèle ^(*) . À la fin du programme de contrôle de la maladie, lorsque les animaux infectés restants ont été retirés de la population, la prévalence de la maladie sera très faible et l'algorithme de test prouvant l'absence de maladie devra être modifié pour augmenter la SpD (et augmenter ainsi la VPP et le RVP), comme dans 1a.
2) Certification de l'absence d'infection ou de l'agent chez des animaux pris individuellement ou dans des produits à des fins de commerce/de déplacement	Lorsque l'objectif est le commerce ou le déplacement, la probabilité de résultats faux négatifs doit être minimisée. Sinon, des animaux infectés peuvent être échangés ou déplacés, avec le risque de propager l'infection à des populations saines, non infectées. Comme le test s'applique à des individus, aucune information, ou peu, n'est disponible, avant les tests, concernant la prévalence ou la probabilité d'une infection. Pour être adapté à son objectif, le test ou l'algorithme de test vise à minimiser les chances de résultats faux négatifs et requiert idéalement une SeD élevée , une VPP élevée et un RVN élevé . Cela peut s'obtenir avec un test unique doté d'une SeD élevée ou avec des tests en parallèle ^(*) .
3) Contribution à l'éradication d'une maladie ou à l'élimination d'une infection dans des populations données	Cet objectif suit un modèle similaire à 1b où il est attendu que la prévalence passe d'élevée à basse au cours du temps dans une population donnée.
4) Confirmation du diagnostic de cas suspect ou de cas cliniques (y compris confirmation d'un test de dépistage positif)	L'objectif d'un test de confirmation est de minimiser les chances d'un résultat faux positif. <i>Confirmation de cas cliniques</i> Afin de confirmer un cas clinique, le test idéal aura une SpD élevée , une VPP élevée et un RVP élevé . En raison de la manifestation clinique de la maladie et de la charge en agents pathogènes vraisemblablement élevée, la SeD n'est pas considérée comme pertinente. <i>Confirmation de tests de dépistage positifs</i> Les tests de dépistage sont appliqués à des populations saines. Ils ont généralement une SeD élevée pour s'assurer de ne pas manquer les individus infectés. L'animal n'est considéré comme positif que si cela est confirmé par un test de confirmation doté d'une SpD élevée. Dans ce cas, le test de confirmation* doit avoir une SpD élevée , une VPP élevée et un RVP élevé . Cette approche suit l'algorithme de test en série ^(*) .
5) Estimation de la prévalence d'une infection ou d'une exposition afin de faciliter l'analyse du risque	Les épidémiologistes ont besoin d'estimations fiables de l'exactitude du test pour concevoir des plans d'échantillonnage destinés aux études de prévalence, aux sondages, au statut sanitaire des troupeaux et aux mesures de contrôle des maladies. L'utilisation d'un test de dépistage doté d'une SeD élevée suivi d'un test de confirmation doté d'une SpD élevée est une approche courante pour les estimations de prévalence.
6) Détermination du statut immunitaire a) chez des animaux pris individuellement après vaccination b) estimation de la séroprévalence après vaccination (recherche et suivi de l'efficacité d'un vaccin)	Pour ce groupe, l'objectif est d'avoir une SpD élevée , une VPP élevée et un RVP élevé . Un résultat faux positif pourrait avoir des conséquences fatales car un tel animal pourrait en fait ne pas être vacciné/protégé. Plus l'exactitude de ce test est élevée, plus précise sera l'estimation de la séroconversion après vaccin chez les individus et dans les populations. Un exemple est le test de neutralisation virale révélée par des anticorps fluorescents (<i>fluorescent antibody virus neutralisation</i> , FAVN) pour estimer le statut immunitaire des chiens et des chats après vaccination antirabique. Pour les voyages internationaux, un résultat >0.5 UI/ml est considéré comme indiquant une protection acceptable.

^(*)Combinaison de tests multiples

La combinaison de tests multiples consiste à utiliser plus d'un test pour déterminer le statut infectieux d'un animal. Les algorithmes les plus courants proposent d'effectuer des tests en série ou en parallèle. Par exemple, si deux tests sont utilisés en série, un échantillon est considéré comme positif seulement si le premier et le second test sont positifs. Les tests en série augmentent la SpD mais diminuent la SeD, tout en augmentant la VPP et le RVP. Les tests de confirmation appliquent l'approche des tests en série puisque le résultat positif d'un test de dépistage (SeD élevée) doit être confirmé par un second test à la SpD élevée. Le test de confirmation doit avoir une SeD au moins aussi élevée que le test de dépistage sous peine de fournir un résultat faux négatif, qui serait considéré comme un vrai négatif selon cet algorithme. Les tests de confirmation augmentent la SpD mais diminuent la SeD, tout en augmentant la VPP et le RVP. Si deux tests sont utilisés en parallèle, un échantillon est considéré comme positif si l'un des tests ou les deux sont positifs. Les tests en parallèle augmentent la SeD, mais diminuent la SpD, tout en augmentant la VPN et le RVN. De plus, pour que les algorithmes de la combinaison de tests multiples soient efficaces, les tests de dépistage et de confirmation ne devraient pas être conditionnellement dépendants (par exemple mesure du même analyte, comme pour deux tests ELISA utilisant le même antigène). Une dépendance positive de la sensibilité du test réduit la sensibilité de l'interprétation des tests en parallèle alors qu'une dépendance positive de la spécificité du test réduit la spécificité de l'interprétation des tests en série (Gardner et al., 2000).

2. Mise au point d'un essai – Études expérimentales

2.1. Modèle de méthode de test et démonstration de faisabilité

Des connaissances, une réflexion et une planification préalables doivent entrer dans la conception de toutes les étapes d'un nouvel essai destiné à la validation ou dans la modification d'un essai existant. Une aide est fournie par les recommandations pour la validation des tests diagnostiques³ qui décrivent les bonnes pratiques de développement et de validation des essais pour différents analytes (p. ex. détection des anticorps, des antigènes ou de l'acide nucléique), ainsi que par les Chapitres 2.2.1 *Mise au point et optimisation des méthodes de détection d'anticorps*, 2.2.2 *Mise au point et optimisation des méthodes de détection des antigènes* et 2.2.3 *Mise au point et optimisation des méthodes de détection de l'acide nucléique*, respectivement.

La mise au point de tout essai dépend des échantillons de référence, reflétant l'analyte cible, la matrice dans laquelle celui-ci se trouve et la population dans laquelle il est prévu d'utiliser l'essai. Les échantillons de référence peuvent être des sérums, des liquides (y compris jus de viande) ou des tissus contenant l'analyte concerné ou une construction génomique correspondant à l'analyte cible. Ce matériel de référence est utilisé dans les expériences menées tout au long du processus de développement jusqu'à la validation de l'essai.

Les facteurs susceptibles d'affecter les caractéristiques analytiques des essais diagnostiques sont nombreux et varient selon le type d'essai ; les principaux facteurs affectant les caractéristiques analytiques des essais sérologiques et moléculaires sont décrits dans Bowden et al. (2021).

Pour les essais moléculaires, la SeD est davantage dépendante de l'aptitude à obtenir l'analyte cible dans un échantillon traité provenant d'un animal porteur de la maladie que de l'aptitude intrinsèque de l'essai à détecter des concentrations très faibles de l'analyte. Même si un essai, comme la PCR en temps réel, peut être extrêmement sensible d'un point de vue analytique, cela ne se traduira pas systématiquement par une SeD élevée, au vu des problèmes d'échantillonnage éventuels (faible volume, substances inhibitrices, variations du spectre clinique de la maladie chez l'animal pris individuellement et efficacité de l'extraction). L'inhibition de la Taq polymérase en raison des caractéristiques de l'échantillon peut être à l'origine de résultats faux négatifs dans un test doté par ailleurs d'une sensibilité analytique élevée (SeA). De plus, l'évaluation de deux essais PCR en temps réel distincts pour le virus de la fièvre porcine africaine (PPA) a montré que tous deux avaient une SeA comparable (équivalente à 14 copies du génome du virus de la PPA) en utilisant un plasmide artificiel de qualité contenant la protéine virale capsidiale VD72 du virus de la PPA comme matrice. Toutefois, lors de l'évaluation de différents échantillons prélevés sur des porcs cliniquement malades, la SeD s'est avérée être de 83% pour l'un des essais et de 92% pour l'autre (données non incluses). La taille de l'amplicon de l'essai doté de la SeD la plus faible (250 paires de

3 https://www.woah.org/fileadmin/Home/fr/Health_standards/tahm/2.02.00_INTRODUCTION_F.pdf

bases ou bp) était significativement plus grande que celle de l'autre essai (75 bp). De manière générale, l'efficacité de l'amplification est sensée diminuer lorsque la taille de l'amplicon augmente. Cette caractéristique est d'autant plus prononcée lorsqu'il s'agit de tester des acides nucléiques viraux extraits d'échantillons cliniques, leur qualité pouvant être altérée selon le degré de dégradation. Si la détermination de la SeA de l'essai s'était faite avec le plasmide pour matrice, le résultat obtenu n'aurait pas été de cet ordre. En outre, puisqu'un essai comme la PCR en temps réel est d'une sensibilité analytique si élevée, il faut prendre le plus grand soin pour éviter une contamination croisée avec une matrice précédemment amplifiée. Une contamination de ce type serait à l'origine d'un résultat faux positif dans un essai considéré par ailleurs comme doté d'une spécificité (analytique) très élevée (Bowden & Wang 2021).

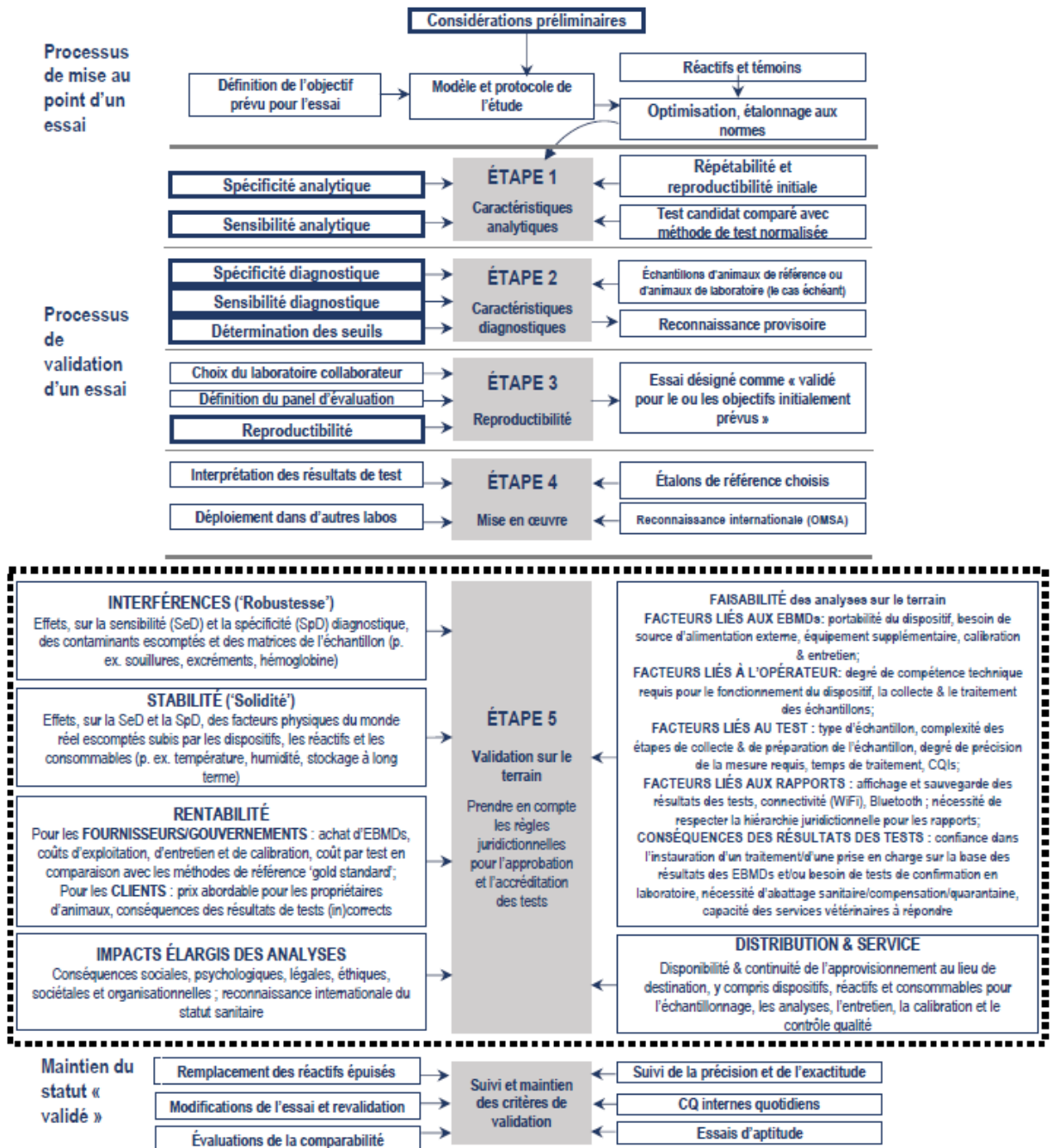


Fig. 1. Processus généraux modifiés de mise au point de validation d'un essai (Étapes 1-4 et maintien du statut validé), les critères de validation figurant en gras dans les encadrés grisés. Critères de validation spécifiques pour les EBMDs ajoutés à l'étape 5 dans l'encadré pointillé (Halpin et al., 2021); (CQIs = échantillons pour le contrôle interne de qualité).

2.2. Normalisation et optimisation

L'optimisation est le processus par lequel les paramètres physiques, chimiques et biologiques les plus importants d'un essai sont évalués et ajustés pour garantir que les caractéristiques de performance de l'essai soient les mieux adaptées à l'utilisation prévue. Il est utile de sélectionner au moins trois échantillons de référence bien définis, représentant la substance à analyser et allant de fortement positif à négatif (soit fortement positif, faiblement positif et négatif). Idéalement, ces échantillons doivent représenter des animaux connus pour être infectés ou non infectés dans la population qui deviendra la cible de l'essai. Il n'est toutefois pas toujours possible d'obtenir de tels échantillons de référence, notamment pour les tests de détection d'acide nucléique et d'antigènes. L'alternative qui consiste à préparer des échantillons de référence enrichis avec des agents de culture ou avec des sérums positifs est de moindre valeur, ces échantillons ne représentant pas réellement l'interaction naturelle entre la matrice et l'agent (voir également Chapitre 2.2.6 *Sélection et utilisation des échantillons et panels de référence*). Lorsqu'il n'existe pas d'autre solution, enrichir un échantillon avec une quantité connue d'analyte ou d'agent issu d'une culture ou diluer un sérum hautement positif dans un sérum négatif de la même espèce peut constituer la seule option possible. Dans chacun de ces cas, il est impératif que la matrice dans laquelle l'analyte est placé ou dilué soit identique ou ressemble autant que possible aux échantillons qui seront finalement analysés. Idéalement, les échantillons de référence auront été bien caractérisés par une ou, de préférence, deux autres méthodes au moins. Ces échantillons peuvent être utilisés dans des expériences pour déterminer si l'essai est à même de distinguer des quantités variables d'analytes, de différencier l'analyte cible et des analytes étroitement apparentés ou encore d'optimiser les concentrations de réactif et de perfectionner le protocole. En principe, pour tous les types d'essais, il est fortement souhaitable de préparer et de stocker une quantité suffisante de chaque échantillon de référence, réparti en aliquotes, afin de pouvoir s'en servir lors de chaque cycle de l'essai candidat, dans la mesure où celui-ci est évalué durant tout le développement et le processus de validation. Changer d'échantillons de référence pendant le processus de validation introduit une variable irréductible qui peut fortement compromettre l'interprétation des données expérimentales et, par conséquent, l'intégrité du développement et du processus de validation.

Le processus laborieux d'optimisation d'un essai est fondamental et essentiel pour obtenir une performance fiable et prévisible. Un jugement scientifique et l'utilisation des bonnes pratiques scientifiques telles que proposées par Bowden *et al.* (2021) sont recommandés pour guider l'optimisation de tous les éléments qui entrent dans la mise au point et la validation de l'essai. L'approche décrite fournit de solides fondements pour le développement d'un essai fiable. Souvent, des prototypes d'essais sont développés en utilisant des réactifs et des équipements disponibles dans le laboratoire. Cependant, si l'essai est destiné à une utilisation diagnostique de routine dans plusieurs laboratoires, sa normalisation devient essentielle. Chaque formulation chimique, chaque solution tampon normalisée doit être entièrement décrite. Tous les réactifs doivent être définis du point de vue de leur pureté et de leur acidité (y compris l'eau). Les plages de fonctionnement admissibles doivent être établies et documentées pour les paramètres tels que le pH ou la molarité. Les normes de qualité, de pureté, de concentration et de réactivité des produits biologiques doivent être définies. Les durées de conservation et les conditions de stockage doivent également être prises en compte pour les produits chimiques tout comme pour les produits biologiques. Les plages admissibles pour les temps et les températures de réaction doivent aussi être établies. Les équipements essentiels requis pour la performance de l'essai doivent être décrits en détail, y compris les spécifications opérationnelles ainsi que l'étalonnage. Un contrôle (qualité) des processus doit faire partie intégrante de l'optimisation et être pris en compte dès le tout début plutôt qu'au terme du développement d'un essai, comme c'est souvent le cas. De plus, les aspects situés en aval tels que la saisie, le traitement et l'interprétation des données peuvent également nécessiter une normalisation et une optimisation. Enfin, tous ces paramètres, une fois optimisés, doivent être intégralement décrits dans le protocole de la méthode de test.

Au cours de l'optimisation d'un essai, il est important de prendre note des étapes de la procédure et des paramètres de l'essai qui ont une plage de fonctionnement optimal étroite, puisque ceux-ci constituent les points critiques qui, au final, se répercuteront sur la fiabilité de l'essai (voir la section A.2.7). Pour certains types d'essai, des étapes spécifiques de la procédure peuvent jouer un rôle plus important que d'autres sur le fonctionnement final de l'essai (pour des informations complémentaires sur la démonstration de la comparabilité lorsque des réactifs ou des processus sont modifiés, voir la section B.5 ci-après ainsi que le Chapitre 2.2.8 *Comparabilité des épreuves suite à des changements introduits dans une méthode de test validée* ainsi que Reising *et al.*, 2021).

Un éventail d'échantillons de référence de substances à analyser et d'autres contrôles des processus, automatiquement inclus dans tout système d'essais, est déterminé dans les sections qui suivent. Celles-ci décrivent les principales fonctions de surveillance nécessitant une attention spéciale durant l'optimisation des essais. De plus, une attention particulière doit être apportée à la préparation et au stockage de tous les réactifs biologiques et matériels de référence afin de garantir leur stabilité (voir Chapitre 1.1.2 ; Watson *et al.*, 2021).

2.3. Plage de fonctionnement de l'essai

La plage de fonctionnement d'un essai équivaut à l'intervalle des concentrations ou des titres d'analyte sur lequel la méthode offre une exactitude et une précision adéquate. L'exactitude équivaut à la proximité de la valeur testée avec la (vraie) valeur attendue (moyenne ou médiane) pour un réactif étalon dont la concentration ou le titre sont connus. La précision est le degré de dispersion (variance, écart type [ET] ou coefficient de variation [Cv]) au sein d'une série de mesures du même échantillon testé dans des conditions spécifiques. Les sources de variation du laboratoire qui influencent la précision de l'essai se situent : 1) au sein de chaque cycle d'analyse, 2) entre deux cycles concurrents, 3d) entre des cycles effectués à des moments différents d'un même jour ou à des jours différents dans des conditions similaires, 3b) entre des cycles effectués à des jours différents par des opérateurs différents, 4) entre des laboratoires différents. Dans ce chapitre, les catégories 1-3 sont des estimations de la répétabilité et la catégorie 4 est synonyme de reproductibilité. La répétabilité des résultats d'opérateurs travaillant dans des laboratoires différents sera vraisemblablement inférieure (ET plus élevée, Cv plus élevé) que celle d'opérateurs travaillant dans le même laboratoire. Au cours du développement de l'essai, les limites inférieures et supérieures de la plage de fonctionnement sont déterminées. Pour déterminer formellement cette plage, un échantillon de référence fortement positif est sélectionné (idéalement, cet échantillon sera l'un des trois échantillons décrits à la Section A.2.3 ci-après). Cet échantillon fortement positif est dilué en série jusqu'à extinction de la réponse de l'essai dans une matrice négative pour cet analyte, mais de la même composition que la matrice des échantillons prélevés sur les animaux de la population ciblée par l'essai. Les résultats sont représentés sous forme d'une courbe de réponse, la réponse (densité optique, cycle seuil, comptages des agents pathogènes, etc.) étant fonction de la concentration d'analyte (quantité). La courbe établit la plage de fonctionnement de l'essai. Si cette plage est jugée inacceptable pour l'objectif prévu, une optimisation supplémentaire pourra s'avérer nécessaire. La courbe typique d'étalonnage pour la plupart des essais a une forme sigmoïde. Les données sont transformées pour s'approcher d'une relation linéaire entre la réponse et la concentration au moyen d'un algorithme adapté (Findlay et Dillard, 2007).

2.4. Facteurs inhibiteurs dans la matrice de l'échantillon

Chaque matrice destinée à un essai doit être utilisée dans le processus de validation. Certaines matrices d'échantillons contiennent des facteurs inhibiteurs qui interfèrent avec le fonctionnement de certains types d'essais. Le sérum, notamment s'il est hémolytique, peut contenir des facteurs toxiques pour les cellules utilisées dans les épreuves de neutralisation virale, tandis que des substances endogènes présentes dans certains tissus et fluides peuvent interférer avec ou inhiber les essais par fixation de ligand ou les dosages enzymatiques du type ELISA. Les fèces, les tissus autolysés et les échantillons de sperme ont tendance à contenir plus de substances interférentes et sont donc plus problématiques pour le fonctionnement des essais que ne le sont le sérum, le sang ou les tissus frais. Pour ce qui est des essais moléculaires, des substances interférentes, dégradantes ou inhibitrices des enzymes dans mélange réactionnel peuvent être présentes dans la matrice. C'est essentiellement la sélectivité qui constitue le test permettant de détecter la présence d'inhibiteurs. Sans doute parmi les moins reconnues, les variations des matrices d'échantillons constituent pourtant l'une des plus importantes sources d'erreur dans les mesures analytiques. L'évaluation de la SpA devrait se faire avec des matrices pertinentes au regard de l'objectif visé, comme des tissus solides, du sang complet ou des écouvillons. Les substances interférentes (inhibitrices) peuvent provenir des échantillons, p. ex. l'hémoglobine (Schrader *et al.*, 2012; Wilson, 1997), de l'environnement ou du processus de collecte et de transport de l'échantillon (p. ex. milieu ou anticoagulants) (Druce *et al.*, 2012; Garcia *et al.*, 2002; Gibb *et al.*, 1998; Miyachi *et al.*, 1998; Yokota *et al.*, 1999). Le sang complet est par exemple l'un des principaux échantillons couramment envoyés pour la détection du hendravirus (HeV) par RT-PCR en temps réel. Or, il a été montré que l'héparine de lithium, un anticoagulant courant, a un impact important sur les tests RT-PCR en temps réel du HeV, produisant des résultats faux négatifs. En comparaison, l'effet inhibiteur de l'anticoagulant acide éthylène diamine tétra-acétique (EDTA) est moindre. Par ailleurs, pour la détection du virus de la langue bleue (*bluetongue virus*, BTV) par RT-PCR en temps réel, la dilution 1/10 (volume par volume ou v/v) du

sang dans une solution saline tamponnée au phosphate améliore la sensibilité du test, ce qui est attesté par des valeurs Ct inférieures en comparaison avec du sang non dilué. Une revue récente des différents types de contrôles internes disponibles pour le monitoring de l'inhibition des essais PCR en temps réel (Yan *et al.*, 2020) fournit des informations sur les stratégies de contrôle interne utilisées comme élément de gestion systématique de la qualité des épreuves moléculaires vétérinaires. Les données récoltées pendant le processus de validation concernant la performance de l'essai et utilisant les matrices d'échantillon à cibler devraient permettre de décider, sur la base du risque, si un contrôle de l'inhibition doit être inclus pour chaque échantillon ou s'il est peu probable que le système de test soit affecté par l'inhibition. Si les substances inhibitrices constituent un problème important, un contrôle de l'inhibition doit être inclus pour chaque échantillon à tester (Section 1.2; Bowden *et al.*, 2021).

2.5. Robustesse

La robustesse fait référence à la capacité d'un essai à ne pas être influencé par les variations mineures des conditions de test, susceptibles de se produire en cours d'analyse. L'évaluation de la robustesse doit débuter durant la mise au point de l'essai et se poursuivre pendant les étapes d'optimisation. Des variations délibérées des paramètres de la méthode peuvent être investiguées dans des expériences une fois que les conditions optimales pour l'essai ont été établies. Néanmoins, lorsque des titrages multifactoriels des réactifs sont utilisés pour optimiser l'essai, des signes indiquant que la robustesse est compromise peuvent apparaître. Si des différences minimales des conditions d'expérience ou des concentrations de réactifs causent une variabilité inacceptable, c'est que l'essai n'est probablement pas robuste. L'identification précoce de ce genre de situations constitue une étape clé du processus décisionnel pour déterminer si cela vaut la peine de continuer la validation de l'essai puisque, si l'essai n'est pas robuste dans un laboratoire aux conditions relativement idéales, il est peu susceptible d'être reproductible une fois transféré dans d'autres laboratoires.

Les facteurs les plus susceptibles d'influencer la robustesse d'un essai comprennent le pH, la température, le lot de réactifs, la marque des plaques de microtitrage ainsi que des facteurs liés aux matrices aqueuses ou organiques (Bowden *et al.*, 2021; Dejaegher et Vander Heyden, 2006). Une fois l'optimisation terminée, la robustesse de l'essai devient partie intégrante de l'évaluation de la répétabilité. Pour les EBMDs (voir Figure 1 Étape 5), des évaluations de la solidité (exprimant l'absence d'influence, sur les résultats du test, des variables opérationnelles et environnementales de la méthode d'analyse) sont aussi nécessaires car la performance des EBMDs ne devrait pas être sensible à la compétence de l'opérateur, aux fluctuations de température, d'humidité, de lumière, ni à d'autres facteurs environnementaux.

2.6. Étalonnage de l'essai avec des réactifs étalons

2.6.1. Étalons internationaux et nationaux

Idéalement, les étalons approuvés par l'OMSA ou d'autres étalons internationaux contenant une concentration ou un titre connus de la substance à analyser sont utilisés pour normaliser tous les essais (voir les lignes directrices⁴ de l' OMSA ainsi que le Chapitre 2.2.6). Ces étalons sont préparés et distribués par certains Laboratoires de référence de l' OMSA ou par d'autres laboratoires de référence internationaux. Les étalons nationaux sont déterminés par comparaison avec un étalon international lorsque cela est possible ; ils sont préparés et distribués par un laboratoire de référence national. En l'absence d'étalon international, l'étalon national sert à comparer l'essai candidat. Ces étalons sont largement caractérisés selon des analyses approfondies et les méthodes privilégiées pour leur caractérisation, leur préparation et leur stockage sont publiées dans des revues à comité de lecture (Watson *et al.*, 2021).

2.6.2. Étalons internes

En règle générale, un étalon interne devrait être étalonné par rapport à un étalon international ou national. En l'absence de l'un et l'autre de ces étalons et dans la mesure du possible, l'étalon interne sera caractérisé en détail de la même manière que les étalons internationaux ou nationaux (voir

4 Accessible sous : <http://www.oie.int/en/scientific-expertise/veterinary-products/reference-reagents/>

Chapitre 2.2.6). Cet étalon interne sera donc le meilleur disponible et sera conservé en quantités aliquotes pour être utilisé régulièrement dans l'étalonnage des réactifs de travail.

2.6.3. Étalons de travail

Un ou plusieurs étalons de travail, communément appelés contrôles d'analytes ou de processus, sont déterminés par rapport à un étalon international, national ou interne et sont préparés en grandes quantités, aliquotés et stockés pour être systématiquement utilisés dans chaque cycle d'essais diagnostiques.

2.7. « Normaliser » les résultats par rapport à un étalon de travail

En raison des variations intrinsèques dans les résultats de test bruts souvent observées entre les cycles de tests d'un même essai ou entre les laboratoires utilisant des essais identiques ou similaires, il est quasiment impossible de comparer directement des données (semi-)quantitatives. Pour améliorer notablement la comparabilité des résultats des tests au sein d'un laboratoire et entre différents laboratoires, un ou plusieurs étalons de travail sont inclus dans chacun des cycles d'un essai. Les valeurs brutes des tests pour chaque échantillon analysé peuvent ainsi être converties en unités d'activité par rapport aux étalons de travail selon un processus appelé normalisation. Les valeurs normalisées peuvent être exprimées de différentes manières, par exemple sous forme de pourcentage du contrôle positif (ex. : dans un ELISA) ou sous forme d'une estimation de la concentration, par exemple des copies du génome ou du titre d'un analyte à partir d'une courbe type telle que la valeur du cycle de seuil (Ct) dans un essai utilisant une sonde d'hydrolyse. Les bonnes pratiques exigent d'inclure des étalons de travail dans tous les cycles de l'essai au cours de son développement et sa validation, car cela permet une normalisation des données et constitue un bon moyen de comparer directement les résultats entre les cycles d'un essai. Il est impératif de contrôler la variation (absolue) des étalons pour la normalisation, faute de quoi cette dernière peut introduire un biais. Pour plus d'informations, voir les Chapitres 2.2.1., 2.2.2. et 2.2.3.; Bowden *et al.* (2021).

2.8. Étude préliminaire de la répétabilité

L'évaluation de la répétabilité devrait débuter lors de la mise au point et se poursuivre pendant les phases d'optimisation de l'essai. L'identification précoce d'une situation de ce genre constitue une étape critique du processus décisionnel en vue de déterminer si cela vaut la peine de poursuivre la validation de l'essai.

La répétabilité continue à être vérifiée durant l'étape 1 de la validation de l'essai (section B.1.1). Lorsque le test optimisé est effectué dans des conditions ordinaires en laboratoire ou sur le terrain (étape 4 de la validation d'un essai), la répétabilité est continuellement surveillée et fait partie des procédures de contrôle des procédés pour toute la durée de vie de l'essai (voir section B.5.1).

B. PROCESSUS DE VALIDATION D'UN ESSAI

La validation est un processus qui détermine l'aptitude à l'emploi d'un essai qui a été correctement développé, optimisé et normalisé pour un ou des objectifs définis. La validation inclut l'estimation des caractéristiques de performance analytique et diagnostique d'un test. Dans le contexte de ce document, un essai qui a accompli les trois premières étapes du processus de validation (Figure 1), dont la caractérisation de sa performance, peut être désigné comme « validé pour le ou les objectifs initiaux prévus ». Pour les maladies zoonotiques nouvelles ou émergentes, on observe souvent que le nombre d'échantillons provenant d'animaux infectés et non infectés n'est pas suffisant pour être statistiquement robuste, ce qui peut constituer un obstacle majeur pour obtenir des estimations fiables de l'exactitude (Colling *et al.*, 2018; Stevenson *et al.*, 2021). Dans ces circonstances, des tests dotés d'une sensibilité analytique (SeA), d'une spécificité analytique (SpA) et d'une répétabilité acceptables, avec des résultats prometteurs quoique préliminaires en termes de SeD, de SpD et de reproductibilité, basés sur un nombre limité d'échantillons, peuvent être provisoirement reconnus par les autorités nationales et les partenaires commerciaux jusqu'à ce que les résultats de tests complémentaires viennent confirmer leur adéquation globale à l'objectif. Des échantillons bien caractérisés ont été utilisés avec succès à différentes fins dans des études comparatives inter-laboratoires afin d'évaluer leur aptitude à l'objectif visé (Chapitre 2.2.6 ; Gardner *et al.*, 2021). Les nouvelles plateformes, capables de détecter simultanément plusieurs pathogènes, comme les technologies de multiplexage, le séquençage à haut débit, les essais basés sur des biomarqueurs et les EBMDs constituent de

nouveaux défis pour les études de validation orientées vers un but précis (Bath *et al.*, 2020; Halpin *et al.*, 2021; Reid *et al.*, 2021; van Borm *et al.*, 2016).

1. Étape 1 – Caractéristiques de performance analytique

Idéalement, les études décrites dans les sections qui suivent doivent être conçues avec l'aide d'un statisticien et d'un spécialiste des maladies infectieuses pour garantir que la taille de l'échantillon et que l'approche expérimentale sont valides. Il est possible de concevoir des expériences qui fournissent des informations utiles sur les sources potentielles de variation de la précision de l'essai dans un même laboratoire et entre plusieurs laboratoires⁵, informations qui définiront les caractéristiques de performance de l'essai. Le choix des organismes, des souches ou des sérotypes destinés à évaluer la sensibilité et la spécificité analytiques doit refléter les connaissances récentes, ce qui permettra d'avoir le meilleur plan expérimental possible pour cibler des analytes spécifiques.

1.1. Répétabilité

La répétabilité est le degré de concordance entre les résultats des copies d'un échantillon dans un même cycle et entre les cycles de la même méthode de test dans un laboratoire donné. La répétabilité est estimée en évaluant la variation des résultats des copies. Le nombre de copies devrait de préférence être déterminé de concert avec un statisticien, selon un minimum suggéré de trois échantillons représentatifs de l'activité de l'analyte à l'intérieur de la plage opérationnelle de l'essai. Chacun de ces échantillons sera aliquoté dans un nombre approprié de récipients individuels et constituera une copie identique de l'échantillon initial contenant la concentration d'analyte et de matrice initiale (voir Chapitre 2.2.6.). Chaque copie passera ensuite par toutes les étapes de l'essai, y compris par la création de la dilution de travail, comme s'il s'agissait d'un échantillon de test issu de la population ciblée par l'essai. Il n'est pas acceptable de préparer une dilution de travail finale d'un échantillon dans un seul tube à partir duquel des aliquotes dilués seront pipetés dans des tubes de réaction, ni de créer des copies à partir d'une extraction d'acide nucléique plutôt que d'extraire chaque copie avant dilution dans le tube de réaction. De tels échantillons ne constituent pas des copies valables pour les études de répétabilité. La variation intercycles est déterminée en utilisant les mêmes échantillons lors de plusieurs cycles qui ne sont pas effectués le même jour et qui font intervenir deux opérateurs ou plus. Bowden & Wang (2021) fournissent un exemple de répétabilité inter-essais pour trois tests du virus de la PPA par PCR en temps réel qui ont été évalués en utilisant un échantillon faiblement positif du virus de la PPA. Les analyses, extraction de l'ADN viral comprise, ont été effectués sur dix jours distincts par différents opérateurs dans un même laboratoire. La variation entre les résultats des copies peut être exprimée sous forme d'écart type, de coefficients de variation (écart type + moyenne des copies) ou autrement (voir Chapitre 2.2.4. *Incertitude des mesures* pour l'évaluation de la répétabilité).

1.2. Spécificité analytique

La SpA est la capacité de l'essai à distinguer l'analyte cible (p. ex. anticorps, organisme ou séquence génomique) des analytes non cibles, y compris les composants de la matrice. Cette évaluation est qualitative et le choix, comme la provenance des types d'échantillons, des organismes et des séquences pour l'évaluation de la SpA doit refléter les objectifs du test ainsi que le type d'essai. Voir Chapitres 2.2.1., 2.2.2. et 2.2.3. pour les instructions sur les tests de détection des anticorps, des antigènes et de l'acide nucléique respectivement. Par exemple, l'évaluation de la spécificité analytique d'un test PCR pour la fièvre aphteuse pourrait être réalisée en testant des échantillons de référence connus pour être positifs aux virus de la stomatite vésiculaire, de la maladie vésiculeuse du porc et/ou du coryza gangréneux. Les évaluations de la spécificité analytique ne peuvent pas définir tout l'éventail des analytes ayant potentiellement une réactivité croisée et présents dans la population, ni prendre en compte la variabilité des échantillons au niveau de la population ; c'est pourquoi, déterminer la spécificité analytique ne peut pas remplacer les évaluations de la spécificité diagnostique. Durant l'étape 1 de la validation, la SpA est documentée et les réactions croisées identifiées (Ludi *et al.*, 2021). La réactivité croisée (SpA inférieure à 100 %) peut être acceptable selon l'usage prévu pour le test. Les effets de la réactivité croisée sont

5 La précision peut être évaluée de plusieurs manières en testant le même échantillon répliqué : 1) sur une ou plusieurs plaques dans un seul cycle d'analyse, 2) entre des plaques analysées simultanément dans un cycle d'analyse, 3a) entre des cycles différents effectués à des moments différents d'un même jour ou à des jours différents selon des conditions similaires, 3b) entre des cycles effectués à des jours différents par des opérateurs différents, 4) entre des laboratoires différents. Dans ce chapitre, les catégories de précision 1-3 sont des estimations de la répétabilité, la catégorie de précision 4 étant synonyme de reproductibilité. Les degrés 3a et 3b sont aussi connus pour décrire une précision intermédiaire.

documentés plus en détail au cours de l'étape 2 (établissement de la SpD) et évalués lors de la mise en œuvre de l'étape 4.

1.2.1. Sélectivité

La sélectivité définit la mesure dans laquelle une méthode peut quantifier de manière exacte l'analyte ciblé en présence des éléments suivants : 1) substances interférentes comme des composantes de la matrice (p. ex. inhibiteurs ou enzymes dans le mélange réactionnel) ; 2) agents dégradants (p. ex. facteurs toxiques) ; 3) liaison non spécifique de réactifs sur une phase solide (p. ex. conjugué d'un ELISA adsorbé dans le puits d'une plaque de microtitrage) ; et 4) anticorps en réponse à la vaccination susceptibles d'être confondus avec des anticorps en réponse à une infection active. Ces substances interférentes peuvent être source de réponses au test faussement basses ou élevées, compromettant la spécificité analytique de l'essai. Vessman *et al.* (2001) apportent une bonne vue d'ensemble de la sélectivité définie pour la chimie analytique, à partir de laquelle une modification a été décrite pour être appliquée aux tests diagnostiques.

1.2.2. Exclusivité

L'exclusivité est la capacité d'un essai à détecter un analyte ou une séquence génomique unique à l'organisme ciblé et à exclure tous les autres organismes connus ayant potentiellement une réactivité croisée. Cette définition s'applique également à un essai de confirmation. Par exemple, un essai destiné à détecter les sous-types H5 du virus de la grippe aviaire devrait être évalué au regard des réactions croisées avec des sous-types non-H5 de la grippe aviaire. Les tests de spécificité devraient aussi inclure d'autres organismes responsables de signes cliniques similaires afin de démontrer l'utilité de l'essai pour la détection différentielle de l'organisme cible.

1.2.3. Inclusivité

L'inclusivité est la capacité d'un essai à détecter différentes souches ou sérovars d'une espèce, différentes espèces d'un genre ou un regroupement analogue d'organismes ou d'anticorps étroitement apparentés. Elle caractérise le spectre d'un essai de dépistage, comme dans le cas d'un ELISA ciblant le virus de la langue bleue (*bluetongue virus*, BTV) qui détecterait les anticorps contre tous les sérotypes du BTV ou un ELISA ciblant les protéines non-structurelles du virus de la fièvre aphteuse qui détecterait les anticorps contre les sept sérotypes du virus de la fièvre aphteuse.

1.3. Sensibilité analytique

La limite de détection (LD) est un indicateur de la SeA d'un essai. La limite de détection est la quantité estimée d'analytes dans une matrice donnée produisant au moins un résultat positif à un certain pourcentage. On procède généralement à l'estimation de la limite de détection par un enrichissement de l'analyte dans la matrice cible. Le choix du ou des analytes (p. ex. espèces, souches) fait partie de la définition de la SeA et doit être documenté correctement. Ces expériences peuvent être conçues pour une estimation précise et exacte du point de la probabilité (p. ex. 50 % ou 100 %) mais, dans certaines circonstances, une estimation prudente de la LD (p. ex. 100 %) peut être acceptable. Par exemple, dans un titrage utilisant des dilutions décimales, toutes les copies à toutes les dilutions peuvent montrer des réponses soit de 100 %, soit de 0 %. À ce stade, deux choix se présentent. La dernière dilution montrant une réponse de 100 % peut être acceptée comme une estimation prudente de la limite inférieure de détection. Une estimation plus exacte peut être obtenue lors d'une seconde étape de l'expérience en utilisant des intervalles plus étroits du schéma de dilution, focalisés sur la zone entre 100 % et 0 %. Les méthodes d'évaluation statistique des données de LD figurent au Chapitre 2.2.5 *Méthodes statistiques de validation*.

1.4. Exactitude analytique des tests ou des procédures complémentaires

Certaines méthodes ou procédures de test peuvent servir d'outils analytiques dans les laboratoires de diagnostic. Ceux-ci utilisent généralement des tests ou des procédures secondaires qui sont appliqués à un analyte détecté dans un essai primaire. La finalité de ce genre d'outils analytiques est de caractériser plus en détail l'analyte détecté dans l'essai primaire. Parmi les tests complémentaires, on peut citer la neutralisation virale servant à typiser un virus préalablement isolé, le séquençage moléculaire ou la

spectrométrie de masse MALDI-TOF (*matrix assisted laser desorption ionisation - time of flight*) pour les bactéries (Ricchi et al., 2016).

Ce genre de tests complémentaires doit être validé pour ses caractéristiques de performance analytique (Sections A.2 à B.1.3 ci-dessus). Toutefois, ils diffèrent des tests diagnostiques puisqu'ils ne requièrent pas de validation pour leurs caractéristiques de performance diagnostique (Section B.2 à B.4 ci-après) si leurs résultats ne sont pas utilisés pour établir un diagnostic final par rapport à l'objectif prévu. L'exactitude analytique de ces outils peut être définie par rapport à un étalon ou d'après les caractéristiques inhérentes à l'outil lui-même (tel un titrage au point final). Dans tous ces exemples, l'outil analytique permet d'affiner la caractérisation quantitative et qualitative de l'analyte cible.

2. Étape 2 – Performance diagnostique de l'essai

Les échantillons d'origine animale utilisés pour évaluer la SeD et la SpD sont principalement de quatre provenances: (1) échantillons stockés dans des banques de référence dont le statut infectieux est connu, (2) échantillons provenant d'un foyer ou de mesures de surveillance pour lesquels le statut de l'animal est inconnu, mais celui de la population (infectée ou non infectée) connu, (3) échantillons pour lesquels ni le statut de l'animal, ni celui de la population n'est connu et (4) échantillons issus d'études expérimentales (voir Tableau 3 pour les avantages et les limites de chaque méthode – dans le Tableau, les groupes 2 et 3 sont combinés par souci de concision). Les estimations de la SeD (proportion d'échantillons d'animaux de référence connus pour être infectés et testés positifs dans l'essai) et de la SpD (proportion d'échantillons d'animaux de référence connus pour être non infectés et testés négatifs dans l'essai) sont les premiers indicateurs de performance établis au cours de la validation d'un essai (voir chapitres 2.2.1., 2.2.2., 2.2.3). Ces estimations servent de base de calcul pour d'autres paramètres qui sont extrapolés à propos des résultats de test (p. ex. valeurs prédictives et rapports de vraisemblance de résultats de test positifs et négatifs). C'est pourquoi il est impératif que les estimations de la SeD et de la SpD soient aussi exactes que possible. Idéalement, elles sont dérivées de tests faits avec un panel d'échantillons d'animaux, dont l'histoire et le statut infectieux relatif à la maladie/infection concernée sont connus et représentatifs du pays ou de la région où le test sera utilisé. Une estimation de l'aire sous la courbe d'efficacité (Ceff., *receiver operating characteristic, ROC*) est un complément utile aux estimations de la SeD et de la SpD pour un test diagnostique quantitatif en cela qu'elle fournit une évaluation de son exactitude globale sur l'ensemble des valeurs possibles de l'essai (Greiner et al., 2000; Zweig & Campbell, 1993). Cette approche est décrite dans le Chapitre 2.2.5.

Sensibilité diagnostique

Pourcentage d'animaux connus pour être infectés qui sont testés positifs avec l'essai; les animaux infectés qui sont négatifs au test sont considérés comme des faux négatifs.

Spécificité diagnostique

Pourcentage d'animaux connus pour être non-infectés qui sont testés négatifs avec l'essai; les animaux non-infectés qui sont testés positifs sont considérés comme des faux positifs.

Échantillons de référence : Le nombre prescrit d'échantillons positifs et négatifs connus dépendra des valeurs probables de SeD et de SpD de l'essai candidat ainsi que du degré de confiance désiré pour leur estimation (Tableau 2 et Jacobson, 1998). Le Tableau 2 fournit deux ensembles de nombres théoriques d'échantillons requis, en admettant soit 5%, soit 2% d'erreur dans l'estimation de la SeD ou de la SpD. Pour obtenir une confiance élevée (typiquement 95%) dans les estimations de la SeD et de la SpD lorsqu'une petite marge d'erreur est souhaitée, le nombre d'échantillons requis est élevé. Par exemple, la comparaison d'une erreur de 2 % contre 5 % pour une SeD ou une SpD probable avec une confiance à 90 % ou à 95 % présente une différence considérable quant au nombre d'échantillons requis (864 contre 138). Pour les maladies listées les plus importantes d'un point de vue économique, une confiance de 99% pourrait être privilégiée. Des limitations logistiques et financières peuvent contraindre à diminuer le nombre d'échantillons testés par rapport au nombre statistiquement requis, auquel cas l'intervalle de confiance calculé pour la SeD et la SpD indiquera une confiance diagnostique moindre dans les résultats. La taille de l'échantillonnage peut également être limitée si les populations de référence ou les réactifs de référence de l'OMSA font défaut (voir chapitre 2.2.5 pour plus de détails). Il peut donc s'avérer nécessaire, au départ, de se contenter de moins d'échantillons. Il est toutefois hautement souhaitable d'accroître la confiance et de réduire la marge d'erreur de l'estimation de la SeD et de la SpD en ajoutant davantage d'échantillons (d'un statut équivalent à celui du panel initial) dès que possible.

Définition du cas, c'est-à-dire en quoi consiste un animal infecté et en quoi consiste un animal non-infecté, respectivement, par ex. signes cliniques, fièvre, résultats d'un test de référence, etc.

Tableau 2. Nombre théorique d'échantillons provenant d'animaux au statut infectieux connu nécessaires pour établir les estimations de la sensibilité (SeD) et de la spécificité (SpD) diagnostiques en fonction des valeurs de SeD ou de SpD probables et de la marge d'erreur et de la confiance souhaitées

SeD ou SpD estimatives	2 % d'erreur admis pour les valeurs estimatives de SeD et de SpD			5 % d'erreur admis pour les valeurs estimatives de SeD et de SpD		
	Confiance			Confiance		
	90 %	95 %	99 %	90 %	95 %	99 %
90 %	610	864	1493	98	138	239
92 %	466	707	1221	75	113	195
94 %	382	542	935	61	87	150
95 %	372	456	788	60	73	126
96 %	260	369	637	42	59	102
97 %	197	279	483	32	45	77
98 %	133	188	325	21	30	52
99 %	67	95	164	11	15	26

Les exemples ci-dessous de populations et de méthodes de référence sont susceptibles d'aider à estimer les caractéristiques de performance du test à valider.

2.1. Populations animales de référence

Idéalement, la sélection d'animaux de référence nécessite que les principales variables de la population cible soit représentées dans les animaux choisis parce qu'ils ont été infectés par ou exposés à l'agent ciblé, ou inversement (Tableau 4). Les variables à prendre en compte comprennent, sans s'y limiter, l'espèce, l'âge, le sexe, la race, le stade d'infection, les commémoratifs de vaccination et les antécédents concernant les maladies du troupeau (voir Chapitre 2.2.6 pour plus de détails). Après la détection initiale d'une nouvelle maladie, il se peut que les échantillons de référence n'existent pas ou seulement en quantités limitées. Dans ces situations, les échantillons provenant d'études expérimentales peuvent être les seuls échantillons disponibles pour la validation initiale des essais.

2.1.1. Échantillons de référence négatifs

Des échantillons vrais négatifs, issus d'animaux ne pouvant pas avoir été infectés ou exposés à l'agent concerné, peuvent être difficiles à trouver. Il est souvent possible d'en obtenir dans des pays ou des zones qui ont éradiqué ou n'ont jamais connu la maladie en question. De tels échantillons peuvent être utiles tant que la population ciblée pour l'essai est suffisamment similaire à la population source de l'échantillon.

2.1.2. Échantillons de référence positifs

Il est généralement problématique de trouver suffisamment d'animaux de référence vrais positifs, cela étant défini par l'isolement de l'agent pathogène. Il peut s'avérer nécessaire de recourir à des échantillons d'animaux qui ont été identifiés par un autre test d'une exactitude suffisante, tel un test validé de détection de l'acide nucléique. Le test candidat sera appliqué à ces échantillons de référence et les résultats (positifs et négatifs) seront recoupés dans un tableau à double entrée. Cela est appelé « modèle de référence absolue » (*gold standard model*) dans la mesure où l'étalon y est supposé parfait. Un exemple de calcul figure dans le Tableau 4 de la section B.2.5. Les situations où une référence parfaite existe pour les animaux positifs ou négatifs et celles où la référence est parfaite pour tous deux sont décrites dans la validation des tests diagnostiques par Heuer & Stevenson (2021).

2.2. Échantillons d'animaux au statut inconnu

Il arrive, dans certaines situations, que le statut infectieux d'une population soit connu. Si la population est connue pour être non infectée, tous les animaux de cette population seront également présumés non infectés. Dans une population infectée, tous les animaux ne seront pas infectés, en particulier si la maladie n'est pas hautement contagieuse. Déduire le statut infectieux d'une population est plus difficile si l'infection est subclinique ou latente, ou si les animaux sont protégés par la vaccination ou par une exposition antérieure à un pathogène. Lorsque l'étalon dit de référence (statut infectieux réel) est imparfait, ce qui est la règle pour tous les tests diagnostiques utilisés pour analyser des échantillons du terrain, les estimations de la SeD et de SpD pour l'essai candidat reposant sur cette référence sont faussées. Une manière de surmonter ce problème est d'effectuer une analyse de structure latente (ASL) des résultats conjugués de deux ou de plusieurs tests en partant de l'hypothèse qu'aucun de ces tests n'est parfait (voir Johnson *et al.*, 2019).

Les modèles de structure latente ne se basent pas sur le postulat d'une référence parfaite mais estiment plutôt l'exactitude du test candidat et d'un test de référence en utilisant les résultats conjugués des deux tests (Branscum *et al.*, 2005 ; Enøe *et al.*, 2000 ; Georgiadis *et al.*, 2003 ; Hui et Walter, 1980). L'utilisation d'un cadre bayésien pour l'ASL permet d'incorporer dans l'analyse les connaissances déjà disponibles quant à la SeD et à la SpD du test de référence et du test candidat. Le statut infectieux des populations source peut aussi être spécifié dans une ASL bayésienne et l'inclusion d'échantillons d'animaux provenant de populations connues pour ne pas être infectées peut renforcer l'aptitude de ces modèles à estimer la SpD ou d'autres paramètres du modèle (p. ex. SeD ou prévalence) de manière plus précise que ne le permettraient l'utilisation des seules données de deux populations infectées ou plus. Dans une ASL, trois hypothèses fondamentales doivent normalement être remplies : (1) lorsque des données provenant de plusieurs populations sont utilisées, la prévalence dans chaque population devrait être différente ; (2) la SeD et la SpD du test sont constantes dans toutes les populations testées ; et (3) les tests sont conditionnellement indépendants. Cependant, s'il existe une dépendance conditionnelle entre le test-indice et le test de référence (c-à-d. que tous deux mesurent le même analyte), des modèles de structure latente avec une structure de dépendance différente ont été développés pour modéliser la dépendance conditionnelle entre les tests.

Comme ces modèles bayésiens de structure latente sont complexes et nécessitent l'adhésion à des hypothèses fondamentales, il convient de rechercher une aide statistique pour guider l'analyse et déterminer l'échantillon de la ou des populations cibles; les caractéristiques des autres tests inclus dans l'analyse, le choix du modèle approprié et des méthodes d'estimation devraient se faire sur la base de publications à comité de lecture (voir Chapitre 2.2.5. pour plus de détails et Cheung *et al.*, 2021). Le calcul des tailles d'échantillon pour une ASL requiert l'aide d'un statisticien ou d'un épidémiologiste disposant d'expérience avec ces techniques.

2.3. Animaux de référence infectés expérimentalement ou vaccinés

Les échantillons successifs prélevés sur des animaux infectés expérimentalement ou vaccinés sont utiles pour déterminer la cinétique des réponses humorales ainsi que la présence/l'absence d'antigènes ou d'organismes dans les échantillons provenant de ces animaux (Tableau 3). Cependant, les résultats successifs multiples obtenus avant ou après exposition d'animaux individuels ne sont pas acceptables pour procéder à l'estimation de la SeD et de la SpD, l'exigence statistique d'observations indépendantes n'étant pas respectée. L'échantillonnage à un moment précis d'animaux de laboratoire pris individuellement peut être acceptable (p. ex. un échantillon prélevé aléatoirement sur chaque animal). Il faut cependant noter que, pour les méthodes de détection indirecte d'analytes, l'exposition à des micro-organismes dans des conditions expérimentales ou la vaccination peuvent déclencher une réponse humorale qui n'est ni quantitativement, ni qualitativement typique de l'infection naturelle dans la population cible (Jacobson, 1998). La souche du microorganisme, la dose, la voie d'administration aux animaux de laboratoire sont autant d'exemples des variables qui peuvent introduire une erreur lors de l'extrapolation des estimations de la SeD et de la SpD à la population cible. Dans les cas où la quasi-impossibilité d'obtenir des échantillons de référence appropriés provenant d'animaux naturellement exposés nécessite de recourir à des échantillons provenant d'animaux de laboratoire pour des études de validation, les mesures de la SeD et de la SpD qui en résultent seront considérées comme des résultats de type "preuve de concept", inférieures aux estimations idéales des vraies SpD et SeD.

Tableau 3. Populations source pour les échantillons destinés à la validation des tests

Banques d'échantillons de référence (p.ex. provenant de laboratoires de référence)	Populations sur le terrain avec des animaux au statut infectieux inconnu	Infections d'épreuve expérimentales
<p style="text-align: center;">Avantages</p> <ul style="list-style-type: none"> • Statut infectieux connu • Analyse statistique simple <p style="text-align: center;">Limites</p> <ul style="list-style-type: none"> • Souvent pas disponible pour les maladies rares ou nouvelles • Potentiellement non représentatif des animaux à tester dans le futur • Statut clinique, résultats d'autres tests et informations démographiques potentiellement manquantes 	<p style="text-align: center;">Avantages</p> <ul style="list-style-type: none"> • Représentatif des animaux à tester dans le futur • Modèles de structure latente utilisés pour l'analyse des données ; possibilité d'ajouter le statut infectieux de la population et les covariables • Possibilité d'évaluer l'exactitude des essais multiples <p style="text-align: center;">Limites</p> <ul style="list-style-type: none"> • Les modèles de structure latente requièrent formation et compétences ainsi que l'adhésion aux hypothèses du modèle 	<p style="text-align: center;">Avantages</p> <ul style="list-style-type: none"> • Statut infectieux basé sur un groupe d'épreuve • Évaluation de la "fenêtre diagnostique" d'un essai <p style="text-align: center;">Limites</p> <ul style="list-style-type: none"> • Échantillons des infections d'épreuve potentiellement non représentatifs des échantillons collectés sur le terrain • Résultat souvent secondaire dans une étude de pathogenèse • Impact potentiel de la voie d'exposition, de la dose infectieuse et d'autres conditions expérimentales • Plus indiqué pour les infections aiguës • Considérations éthiques

2.4. Sélection de valeurs seuils (points limites) pour la classification des résultats de test

Pour obtenir des estimations de la SeD et de la SpD du test candidat mesurées sur une échelle continue, les résultats du test doivent d'abord être regroupés en deux (positifs ou négatifs) ou trois (positifs, intermédiaires [douteux] ou négatifs) catégories de résultats. Cela s'effectue en insérant un ou deux points limites (seuils ou limites de décision) sur l'échelle des résultats. Le choix de ces points limites doit refléter l'objectif prévu de l'essai ainsi que son utilisation et venir appuyer la SeD et la SpD requises pour l'essai. Différentes options et méthodes descriptives pour déterminer la meilleure manière d'exprimer la SeD et la SpD existent (Branscum *et al.*, 2005 ; Georgiadis *et al.*, 2003 ; Greiner *et al.*, 1995 ; Greiner *et al.*, 2000 ; Jacobson, 1998 ; Zweig et Campbell, 1993 ; et Chapitre 2.2.5). Si un chevauchement notable apparaît dans la distribution des valeurs de test d'animaux dont le statut infectieux ou non infectieux est connu, il sera impossible de choisir un point limite permettant de classer avec exactitude ces animaux en fonction de leur statut. Plutôt qu'un seul point limite, deux seuils peuvent être choisis pour obtenir une SeD élevée (p. ex. inclusion de 99 % des valeurs provenant d'animaux infectés) et une SpD élevée (p. ex. inclusion de 99 % des valeurs provenant d'animaux non infectés) (Greiner *et al.*, 1995).

La principale difficulté pour établir des points limites basés sur les caractéristiques de performance diagnostiques réside dans le manque d'échantillons (de référence) bien caractérisés en nombre suffisant. Des solutions sont présentées dans la section B.2.6 pour l'acceptation provisoire d'un essai durant la phase d'accumulation des données afin de consolider les estimations de la SeD et de la SpD.

2.5. Calcul de la SeD et de la SpD sur la base des résultats de test des échantillons de référence

Une méthode typique pour déterminer les valeurs estimatives de SeD et de SpD est de tester des échantillons de référence dans le nouvel essai et de recouper les résultats nominaux dans un tableau croisé à double entrée. Dans un exemple hypothétique, supposons que le concepteur du test a choisi une taille d'échantillons pour la SeD et la SpD du nouvel essai en estimant que les valeurs les plus probables étaient de 97 % (SeD) et de 99 % (SpD) respectivement, avec une confiance souhaitée de 95 % pour ces deux valeurs estimatives. La marge d'erreur souhaitée a été fixée à 2 % pour ces estimations. Le Tableau 4 montre que 279 échantillons provenant d'animaux connus pour être infectés sont nécessaires à l'estimation de la SeD et que 95 échantillons connus pour être négatifs sont nécessaires pour établir la valeur estimative de la SpD. Prêsumons que les échantillons sont ensuite analysés avec le nouveau test.

Le Tableau 4 présente un ensemble hypothétique de résultats à partir desquels les estimations de la SeD et de la SpD sont obtenues.

Tableau 4. Estimations de la sensibilité et de la spécificité diagnostiques calculées à partir d'un ensemble hypothétique de résultats d'échantillons testés provenant de populations connues pour être infectées ou non infectées

		Nombre d'échantillons de référence requis *	
		Positifs connus (279)	Négatifs connus (95)
Résultats du test	Positif	270	7
	Négatif	9	88
		VP	FP
		FN	VN
		Sensibilité diagnostique*	
		VP/(VP + FN)	
		96.8% (94.0 – 98.5%)**	
		Spécificité diagnostique*	
		VN/(VN + FP)	
		92.6% (85.4 – 97.0%)**	

*basé sur le Tableau 2 pour un essai dont les paramètres sont les suivants :

- 1) Avant le test, estimation de la SeD à 97 % et de la SpD à 99 %
 - 2) 95 % = confiance requise pour les estimations de la SeD et de la SpD
 - 3) 2 % = marge d'erreur pour les estimations de la SeD et de la SpD
- VP et FP = vrai positif et faux positif, respectivement
VN et FN = vrai négatif et faux négatif, respectivement

**limites de confiance exactes par distribution binomiale à 95 % pour les valeurs calculées de SeD et de SpD (voir Chapitre 2.2.5 pour les informations concernant les limites de confiance)

Dans cet exemple, l'estimation de la SeD est telle qu'attendue, mais la SpD est beaucoup plus basse (92 %) que la valeur attendue de 99 %. En conséquence, l'intervalle de confiance pour la SpD est plus important que prévu. Un nouvel examen du Tableau 1 montre que 707 échantillons sont nécessaires pour atteindre une marge d'erreur de $\pm 2\%$ pour une SpD de 92 %. Or, un tel accroissement de la taille de l'échantillonnage peut s'avérer irréalisable (voir Chapitre 2.2.5 pour plus de détails).

2.6. Reconnaissance provisoire d'un essai⁶

Dans certaines situations, il n'est pas possible ou souhaitable d'accomplir l'étape 2 du processus de validation, car les échantillons appropriés provenant de la population source sont rares ou que les animaux sont difficiles d'accès (dans le cas, par exemple, de maladies infectieuses transfrontalières ou de maladies de la faune sauvage).

L'expérience a montré que le plus grand obstacle pour passer à l'étape 2 du processus de validation est le nombre défini d'échantillons requis pour calculer la SeD et la SpD. La formule est bien connue et des tableaux existent, qui servent à déterminer le nombre d'échantillons nécessaires pour estimer différentes valeurs de SeD et de SpD selon la marge d'erreur souhaitée et le niveau de confiance des estimations (Tableau 2 et Jacobson, 1998). La formule part du principe que la multitude de facteurs liés à l'hôte ou au microorganisme susceptibles d'affecter le résultat du test est prise en compte. Comme cette hypothèse est discutable, les tailles d'échantillons estimées sont, au mieux, minimales. Pour une maladie qui n'est ni endémique ni généralisée, il peut s'avérer impossible, au départ, d'obtenir le nombre d'échantillons requis mais, avec le temps, l'accumulation de données supplémentaires peut permettre un ajustement des seuils ou, si aucun ajustement n'est nécessaire, renforcer la confiance dans les estimations.

6 La reconnaissance provisoire n'implique pas l'acceptation par l'OMSA. Celle-ci reconnaît toutefois les décisions informelles des autorités au niveau local, étatique, national ou international d'approbation conditionnelle d'un essai partiellement validé.

La reconnaissance provisoire établit qu'un essai a été évalué selon l'étape 1 pour les principaux paramètres de référence de l'essai (SeA, SpA et répétabilité) avec, en plus, une estimation préliminaire de la SpD et de la SeD basée sur un panel restreint d'échantillons choisis bien caractérisés contenant l'analyte cible ainsi qu'une estimation préliminaire de sa reproductibilité. Cela représente la réalisation partielle de l'étape 2. Des estimations préliminaires de la reproductibilité de l'essai candidat peuvent être faites en utilisant ce panel choisi d'échantillons bien caractérisés afin d'encourager l'acceptation provisoire de l'essai. La méthode de test candidate serait ensuite reproduite dans les laboratoires d'au moins deux instituts différents et le panel d'échantillons évalué en utilisant l'essai candidat dans chacun de ces laboratoires, selon le même protocole, avec les réactifs spécifiés dans le protocole et un équipement comparable. Il s'agit là d'une version réduite de l'étude de reproductibilité de l'étape 3 de la validation des essais. Le protocole de test ne doit pas varier lorsque l'on suit cette procédure de reconnaissance provisoire.

Reconnaissance provisoire

Test avec SeA, SpA et répétabilité acceptables ; SeD et SpD préliminaires prometteuses basées sur un petit ensemble sélectionné d'échantillons bien caractérisés contenant l'analyte cible & estimation préliminaire de la reproductibilité. Les résultats peuvent être provisoirement reconnus jusqu'à ce que des tailles d'échantillon adéquates tant pour la SeD que pour la SpD soient atteintes (voir Tableau 2) et qu'une étude de reproductibilité plus étendue soit faite pour confirmer l'adéquation globale à l'objectif visé.

La reconnaissance provisoire d'un essai par un État ou par les autorités nationales signifie que l'essai n'a pas été évalué pour ses caractéristiques de performance diagnostique. À ce titre, le laboratoire devrait développer et suivre un protocole pour ajouter et évaluer des échantillons, lorsqu'ils sont disponibles, pour remplir cette exigence. Idéalement, ce processus sera limité à une période spécifique, pendant laquelle un cumul de ce type sera utilisé pour accomplir les étapes 2 et 3 du processus de validation, ou à des situations particulières (urgences, espèces mineures, absence d'autres tests, etc.).

3. Étape 3 – Reproductibilité et estimation élargie de la répétabilité

3.1. Reproductibilité

La reproductibilité est la capacité d'une méthode de test à fournir des résultats constants, tels que déterminés par les estimations de sa précision, lorsqu'elle est appliquée à des aliquotes d'échantillons identiques testés dans des laboratoires différents, de préférence situés dans des régions ou pays distincts en utilisant le même essai (protocole, réactifs et contrôles). Pour évaluer la reproductibilité d'un essai, chacun des laboratoires, qui seront au minimum trois, doit tester le même panel d'échantillons (en aveugle) contenant un minimum suggéré de 20 échantillons en aliquotes identiques attribués à chacun des laboratoires (voir Chapitre 2.2.6). L'exercice génère également des données préliminaires sur les effets non aléatoires imputables au déploiement de l'essai à d'autres laboratoires. Aux estimations de la répétabilité au sein du laboratoire viennent s'ajouter les copies utilisées dans les études de reproductibilité. Les mesures de la précision peuvent être estimées pour les données de reproductibilité et de répétabilité (voir Chapitre 2.2.4 pour plus d'informations à ce propos et sur son application). Les facteurs affectant la reproductibilité des tests d'un laboratoire à l'autre ainsi que des exemples pratiques d'essais d'aptitude et de comparaisons inter-laboratoires sont fournis par Johnson & Cabuang (2021) et Waugh & Clark (2021). Une étude de cas, effectuée avec la fièvre aphteuse, pour la sélection et l'utilisation de panels de référence est présentée par Ludi *et al.* (2021) et l'intérêt des bio-banques virtuelles à des fins de transparence quant aux réactifs et aux échantillons utilisés pendant la mise au point et la validation des tests est décrite par Watson *et al.*, 2021.

Pour les EBMDs, il convient d'évaluer la reproductibilité selon les conditions d'usage prévu.

3.2. Désignation d'un essai validé

Au terme de l'étape 3 de la validation, partant du principe que les étapes précédentes ont été accomplies entièrement et de manière satisfaisante, l'essai pourra être désigné comme « validé pour l'objectif initialement prévu ». Le maintien de cette désignation dépend d'une surveillance continue de la performance de l'essai, comme décrit à la section 5.1.

4. Étape 4 – Mise en œuvre des programmes

La réussite du déploiement d'un essai apporte la preuve supplémentaire et précieuse de sa performance, conformément aux attentes. Par ailleurs, la prévalence (réelle) des caractéristiques diagnostiques dans la population cible est un facteur important qui nécessite d'être pris en compte, comme décrit plus loin.

4.1. Aptitude à l'emploi

Même si ce chapitre concerne la validation et l'aptitude à l'emploi d'un point de vue scientifique, il convient de noter que des facteurs pratiques peuvent avoir des répercussions sur l'utilité d'un essai du point de vue de l'utilisation à laquelle il est destiné. Ces facteurs ne comprennent pas seulement l'adéquation diagnostique de l'essai mais également son acceptabilité par les communautés scientifiques et réglementaires, son acceptabilité par le client et sa faisabilité compte tenu des ressources disponibles au laboratoire. Pour certaines maladies, de nombreux essais sont disponibles et utilisés en combinaison dans les programmes de surveillance et de lutte contre les maladies, l'utilité d'un essai pouvant, de ce fait, mériter d'être estimée en évaluant les changements graduels de la SeD, de la SpD et les valeurs prédictives des tests combinés.

L'incapacité d'un essai à remplir les exigences opérationnelles peut le rendre inadapté à l'usage auquel il était destiné. Ces exigences peuvent comprendre les coûts de performance, la disponibilité de l'équipement, le niveau de sophistication technique et de compétences requises pour son interprétation, la disponibilité des kits ou des réactifs, sa durée de conservation, les exigences de transport, la sécurité, la biosécurité, le débit d'échantillons, les temps d'exécution jusqu'aux résultats du test ainsi que des aspects de contrôle qualité et d'assurance qualité ou encore la possibilité ou non que le test soit déployé en pratique dans d'autres laboratoires. Les kits de tests utilisés sur le terrain sont grandement souhaitables au vu de leur facilité d'utilisation mais, du fait de leur exécution à l'extérieur d'un environnement contrôlé de laboratoire, ils nécessitent des précautions supplémentaires pour conserver leur adéquation à l'objectif prévu (Crowther *et al.*, 2006; Halpin *et al.*, 2021). Des exemples à l'appui des six objectifs sont fournis dans le Tableau 1 *Objectifs du test et importance relative de la sensibilité diagnostique (SeD), de la spécificité diagnostique (SpD), d'une valeur prédictive positive (VPP), d'une valeur prédictive négative (VPN), du rapport de vraisemblance d'un résultat de test positif (RVP) et du rapport de vraisemblance d'un résultat de test négatif (RVN).*

4.2. Interprétation des résultats de test

Valeurs prédictives des résultats de test : la valeur prédictive positive (VPP) est la probabilité qu'un animal ayant été testé positif soit effectivement positif du point de vue diagnostique. La valeur prédictive négative (VPN) est la probabilité qu'un animal ayant été testé négatif soit effectivement négatif du point de vue diagnostique.

Les valeurs prédictives des résultats de test sont une application du théorème de Bayes et se calculent ainsi :

$$VPP = \frac{P \times SeD}{P \times SeD + (1 - P) \times (1 - SpD)} \quad \text{et} \quad VPN = \frac{(1 - P) \times SpD}{P \times (1 - SeD) + (1 - P) \times SpD}$$

où :

VPP = valeur prédictive d'un résultat de test positif

VPN = valeur prédictive d'un résultat de test négatif

P = prévalence de l'infection

SeD = sensibilité diagnostique

SpD = spécificité diagnostique

Contrairement à la SeD et à la SpD, les valeurs prédictives sont influencées par la prévalence réelle de l'infection dans la population cible. En d'autres termes, les valeurs prédictives ne sont pas des caractéristiques inhérentes à un test diagnostique spécifique, mais sont fonction de sa SeD et de sa SpD ainsi que de la prévalence locale de l'infection dans une population donnée à un moment donné.

Les valeurs prédictives ont une grande importance pour les services vétérinaires chargés de la gestion de programmes de lutte ou d'éradication. Si l'on considère l'inverse de la VPP (soit 1/VPP), cela fournit une indication sur les montants à investir pour éliminer les vrais et les faux positifs pour chaque animal vrai positif détecté par les activités de surveillance. En d'autres termes, si la VPP est de 0,67, cela signifie

que deux animaux sur trois sont de vrais positifs et que le troisième est un faux positif. Comme la prévalence d'une infection évolue continuellement au cours de la mise en œuvre d'un programme de lutte, le suivi de la VPP constitue une manière d'évaluer les coûts du programme.

Par ailleurs, au cours de la mise en œuvre d'un programme de lutte, il est habituellement conseillé de modifier la sensibilité des tests employés sur la base de la variation de la prévalence de l'infection dans la population cible et des objectifs du programme. La VPP peut être utilisée pour faire ces modifications de la SeD et de la SpD sur la base de considérations économiques. Autrement dit, s'il devient nécessaire de modifier la SeD et la SpD du test, un certain nombre de points limites virtuels peuvent être fixés le long de la courbe ROC de la validation du test et les valeurs pertinentes de SeD et de SpD pour chaque seuil peuvent être utilisées pour évaluer les coûts attendus pour l'élimination de chaque animal infecté.

Si l'objectif est d'établir la preuve de l'absence de maladie, la VPN est la mesure la plus importante. La VPN dépend essentiellement de la SeD. Si les valeurs prédictives peuvent être un outil utile pour l'interprétation des tests diagnostiques, elles sont fortement dépendantes de la prévalence de la maladie dans la population. Les valeurs prédictives calculées dans des populations où la prévalence est élevée, ou lors du pic d'un foyer alors que la prévalence de la maladie est élevée, ne sont pas applicables aux populations où la prévalence est basse, ni à la toute fin d'un foyer lorsque la prévalence de la maladie a nettement diminué.

Le rapport de vraisemblance (RV) indique la puissance diagnostique d'un résultat de test donné et peut être utilisé pour aider à l'interprétation du test. Le rapport de vraisemblance est une caractéristique intrinsèque du test ; il dépend uniquement de la sensibilité et de la spécificité diagnostiques combinées et, de ce fait, ne varie pas avec la prévalence. Les rapports de vraisemblance sont extrêmement puissants, dans la mesure où ils peuvent être utilisés pour calculer la probabilité de maladie 'après analyse', sur la base du résultat quantitatif du test observé et de l'estimation par le diagnosticien de la probabilité d'infection avant que le test n'ait été effectué (Caraguel & Colling, 2021).

Le RV est calculé en divisant la probabilité d'un résultat de test donné obtenu chez des individus infectés par la probabilité du même résultat de test obtenu chez des individus non infectés. Si le RV est inférieur à un, le résultat du test corrobore l'absence d'infection, c'est-à-dire qu'il est moins probable d'obtenir ce résultat de test chez des animaux infectés que chez des animaux non infectés. Un RV égal à '1' signifie que le résultat du test n'a pas de puissance diagnostique, c'est-à-dire que la probabilité de l'obtenir est identique chez les animaux infectés et non infectés. Plus le RV est éloigné de 1, soit en direction de zéro, soit vers l'infini, plus forte est l'évidence fournie par le résultat du test. Dans un contexte clinique, les résultats de test avec un $RV > 10$ ou < 0.1 sont considérés comme de bonnes évidences diagnostiques qu'une infection est présente ou absente, respectivement.

$$RV^+ = \frac{SeD}{1 - SpD} \quad \text{et} \quad RV^- = \frac{1 - DSe}{DSp}$$

où :

RV^+ = Rapport de vraisemblance d'un résultat de test positif

SeD = Sensibilité diagnostique

RV^- = Rapport de vraisemblance d'un résultat de test négatif

SpD = Spécificité diagnostique

Le RV varie de zéro à l'infini. Si le RV d'un résultat de test donné est supérieur à 1, ce résultat corrobore la présence de l'infection. Les rapports de vraisemblance peuvent s'appliquer à la valeur seuil du test ou à différentes plages de résultats.

4.3. Reconnaissance internationale

Traditionnellement, les essais sont internationalement reconnus par l'OMSA lorsqu'ils sont désignés comme des épreuves prescrites ou comme leur remplaçant à des fins commerciales. Cette reconnaissance se base souvent sur la preuve de leur utilité à un niveau national, régional ou international. Pour les kits diagnostiques commerciaux qui ont été soumis à la procédure de l'OMSA pour la validation et la certification des essais diagnostiques, l'étape finale consiste à rejoindre la liste de tests dans le Registre de l'OMSA. Les tests listés dans ce Registre sont certifiés comme étant adaptés à un objectif spécifique s'ils ont franchi les étapes 1, 2 et 3 de la validation et qu'une révision favorable par un panel d'experts indépendants s'en est suivie. Ce Registre est destiné à fournir aux utilisateurs potentiels de kits une source d'information éclairée et objective à propos du kit et de ses caractéristiques de

performance pour un objectif donné (Gifford *et al.*, 2021). Ce Registre est disponible sur le site web de l'OMSA (voir note de bas de page 2).

4.4. Déploiement de l'essai

La dernière preuve de l'utilité d'un essai est apportée par le succès de son utilisation dans d'autres laboratoires et par son incorporation dans les programmes de lutte ou de surveillance nationaux, régionaux et/ou internationaux. Les laboratoires de référence jouent un rôle essentiel dans ce processus (Brown *et al.*, 2021). Avec l'évolution naturelle des améliorations diagnostiques et/ou technologiques, de essais récemment validés sont susceptibles de devenir les nouvelles méthodes de référence auxquelles d'autres essais seront comparés. À ce titre, ils peuvent progressivement accéder à la reconnaissance nationale, régionale et internationale. En leur qualité de norme reconnue, ces essais seront également utilisés pour développer des étalons pour les contrôles qualité, pour les essais d'aptitude ou à des fins d'harmonisation. Ces étalons sont aussi susceptibles de devenir des étalons internationaux.

Il convient de réévaluer la reproductibilité lorsque le test est transféré du laboratoire où il a été mis au point vers le terrain, qu'il soit utilisé dans des laboratoires locaux ou sur le terrain. Les modifications prévisibles, comme les températures extrêmes ou le niveau d'expérience de l'opérateur, doivent être évaluées puisqu'elles constituent des sources supplémentaires de variation des résultats de l'essai, susceptibles d'influencer les estimations de sa reproductibilité.

5. Suivi de la performance de l'essai après sa validation initiale

5.1. Suivi de l'essai

Pour maintenir le statut d'un essai validé, il est nécessaire de s'assurer que l'essai, tel qu'il a été initialement validé, conserve les caractéristiques de performance définies lors de sa validation. Cela peut être déterminé dans un programme d'assurance qualité caractérisé par un suivi attentif de la performance quotidienne de l'essai, principalement des estimations de sa précision et de son exactitude pour les contrôles internes ainsi que des valeurs éventuellement aberrantes (1.1 Répétabilité). La performance peut être suivie graphiquement en consignnant les mesures de contrôle de l'essai dans des graphiques de contrôle⁷ (Crowther *et al.*, 2006). Les écarts par rapport à la performance attendue sont investigués de manière à prendre des mesures correctives si nécessaire. Un suivi de ce type apporte la preuve fondamentale que l'essai conserve son caractère « validé » pendant sa phase de mise en œuvre. Sa reproductibilité sera évaluée au moyen de programmes externes de contrôle qualité, tels les essais d'aptitude (Johnson & Cabuang, 2021) (3.1 Reproductibilité). Si l'essai échouait à produire des résultats conformes aux données de validation initiales, il serait déclaré inadéquat à l'objectif prévu. Un essai validé doit être évalué de manière continue pour garantir qu'il conserve son adéquation à l'objectif (Waugh & Clark, 2021).

5.2. Modifications et perfectionnement – Réflexions à propos des changements introduits dans un essai

Au fil du temps, des modifications de l'essai seront vraisemblablement nécessaires pour répondre aux changements de l'objectif prévu, des analytes cibles (p. ex. modification de l'essai pour ajuster sa performance diagnostique) ou des modifications techniques seront apportées pour en améliorer l'efficacité ou la rentabilité économique. Lorsqu'il s'agit de modifier l'objectif prévu de l'essai, une révision de sa validation à partir de l'étape 2 est obligatoire.

Si l'essai doit être utilisé dans une autre région géographique et/ou dans une population différente, la revalidation de l'essai dans ces nouvelles conditions est recommandée. Les lignées ou sous-lignées d'un agent infectieux, issues d'animaux vivant dans des zones géographiques distinctes, sont connues pour varier et requièrent une revalidation de l'essai pour la population cible donnée. Cela est particulièrement vrai pour les systèmes de détection de l'acide nucléique, les mutations ponctuelles étant très courantes chez de nombreux agents infectieux (notamment chez les virus ARN). Les mutations susceptibles de survenir dans la zone de l'amorce ou de la sonde peuvent nuire à l'efficacité de l'essai, voire invalider les caractéristiques de performance établies. Il est également conseillé de confirmer de manière régulière

7 *Graphique de contrôle* : représentation graphique des données de mesures répétées d'un ou de plusieurs échantillons de contrôle testés dans différents cycles de l'essai au cours du temps.

la séquence cible dans les régions génomiques sélectionnées pour les isolats nationaux ou régionaux des agents infectieux. Cela est particulièrement vrai pour les zones de l'amorce ou de la sonde afin de garantir qu'elles restent stables et que la SeD et la SpD de l'essai ne sont pas compromises. Des problèmes similaires peuvent survenir avec les essais immunologiques de détection des antigènes ou des anticorps.

Une situation similaire peut se produire avec l'émergence de nouveaux sous-types d'agents pathogènes existants. Dans ces circonstances, il peut s'avérer nécessaire de modifier les essais existants.

5.2.1. Modifications techniques et évaluation de la comparabilité

Les modifications techniques apportées à un essai, comme des changements dans l'instrumentation, dans les protocoles d'extraction ou la conversion d'un essai à un système semi-automatisé ou entièrement automatisé et robotisé ne nécessiteront généralement pas une revalidation complète de l'essai. Une étude comparative des méthodes sera effectuée pour déterminer si la modification relativement mineure apportée à l'essai modifie ses caractéristiques de performance, telles que précédemment documentées. La comparabilité peut être établie en accomplissant la procédure modifiée et la procédure originale en parallèle, avec le même panel d'échantillons pour les deux et sur plusieurs cycles. Si les résultats de la procédure modifiée et de la méthode initialement validée s'avèrent comparables dans une expérience basée sur un critère préalablement spécifié, l'essai modifié restera valide pour l'objectif prévu. Voir le Chapitre 2.2.8 pour la description des expériences appropriées aux essais de comparabilité, et le Chapitre 2.2.6 sur les panels d'échantillons de référence, Bowden & Wang, 2021 et Reising *et al.* 2021.

5.2.2. Modifications biologiques et évaluation de la comparabilité

Des situations peuvent se présenter où le changement de certains des produits biologiques utilisés dans l'essai peut s'avérer nécessaire et/ou justifié. Cela peut concerner un changement dans l'échantillon à tester (p. ex. un changement des tissus à tester ou éventuellement un test à effectuer chez une espèce complètement différente). Cela peut concerner un changement dans un réactif (p. ex. la substitution d'un antigène recombinant par un antigène issu de culture cellulaire ou d'un conjugué d'anticorps par un autre d'une spécificité immunologique similaire pour un ELISA). La difficulté que comporte toute modification est de déterminer si celle-ci implique une revalidation complète de l'essai, sur la paillasse ainsi que sur le terrain. Toute modification nécessite au minimum que les « exigences analytiques » appropriées de l'étape 1 soient évaluées. La décision la plus difficile porte sur la « performance diagnostique » de l'étape 2. Pour être utile, l'essai original (de référence) doit initialement être comparé à l'essai modifié (candidat) dans une étude contrôlée en utilisant un panel défini d'échantillons diagnostiques positifs et négatifs. Voir le Chapitre 2.2.8 et Reising *et al.* (2021) pour la description de l'évaluation de la comparabilité. Si l'évaluation de la comparabilité ne laisse pas entrevoir de modification de la performance diagnostique, l'essai modifié pourra être progressivement utilisé en routine. Si, au contraire, des différences de SpD et de SeD sont observées, l'essai modifié nécessitera la répétition de l'étape 2 ou une validation complémentaire sur le terrain avant d'être adopté.

Une **étude de comparabilité** est requise lorsque qu'une modification a été apportée au protocole d'un test validé afin de garantir que la performance du test est comparable.

5.2.3. Remplacement de réactifs épuisés

Lorsqu'un réactif comme un échantillon de contrôle ou un étalon de travail arrive à épuisement, il est essentiel de préparer et de tester à plusieurs reprises son successeur, avant que ce contrôle soit épuisé. Il convient d'intégrer le nouvel échantillon de contrôle éventuel dans plusieurs cycles de l'essai, parallèlement au contrôle d'origine, afin d'établir leur rapport de proportionnalité. Il est important de changer un seul réactif à la fois afin d'éviter de cumuler les problèmes liés à l'évaluation simultanée de plusieurs variables.

5.3. Accroître la confiance dans les critères de validation

Comme de nombreuses variables ont un effet sur la performance diagnostique des essais, il est grandement souhaitable d'accroître, avec le temps, le nombre d'échantillons de référence ou d'échantillons appropriés à une analyse de structure latente. La conception de l'échantillonnage, la collecte, le transport et l'environnement de test pour les nouveaux échantillons devraient être les mêmes que ceux utilisés pour l'étude de validation initiale. L'augmentation du nombre d'échantillons améliore la précision des estimations globales de la SeD et de la SpD et peut permettre le calcul de l'estimation de la SeD en fonction de facteurs comme l'âge, le statut sanitaire et la charge en microorganismes. L'expérience pratique montre qu'il est difficile d'actualiser régulièrement les dossiers de validation avec de nouvelles données. La participation aux essais d'aptitude utilisant des panels pertinents des souches les plus récentes peut contribuer à prouver l'exactitude et la précision de l'essai.

5.3.1. Gestion des données

Le stockage à long terme des données, la révision des données de validation et la vérification continue sont des éléments importants pour rester confiant dans le fait que les essais continuent à performer avec une exactitude diagnostique acceptable. Pour atteindre cet objectif, des systèmes de gestion des informations de laboratoire (*Laboratory Information Management Systems*, LIMS) qui intègrent les données de validation et de diagnostic et qui fournissent des mises à jour régulières de la performance de l'essai sont requis. Les systèmes de gestion des données devraient au minimum faciliter : (1) le stockage des données de validation à des emplacements centraux, (2) la révision et/ou le partage des données lorsque, par exemple, des essais sont déployés auprès de laboratoires externes (Étape 4 du Processus de l'OMSA de validation d'un essai, voir Figure 1), (3) l'accumulation continue et l'intégration des résultats diagnostiques pour actualiser les caractéristiques diagnostiques d'un essai (Étape 3, voir Figure 1), et (4) le stockage des données de contrôle de qualité interne (CQI) et de contrôle de qualité externe (CQE) à utiliser pour la révision de la sensibilité de l'essai (Étape 1, voir Figure 1), en particulier lors de changements de réactifs ou d'équipement.

L'accès et l'utilisation de systèmes de gestion des informations de laboratoire (LIMS) aux fins de validation d'un test ou d'une vérification continue restent limités dans la mesure où ces systèmes nécessitent des investissements importants et une expertise en technologies de l'information pour les entretenir.

6. Vérification des essais validés existants

Si un laboratoire envisage d'utiliser un kit commercial validé ou un essai candidat sur la base d'une littérature publiée comportant des données de validation, une certaine forme de vérification sera requise pour déterminer si l'essai est conforme aux assertions du fabricant du kit ou de l'auteur, par rapport aux critères de validation de l'étape 1, dans le cadre de l'utilisation qui en est prévue. Cela peut nécessiter une vérification limitée de la SpA ainsi que de la SeA en utilisant le matériel de référence disponible, qu'il soit externe et/ou obtenu localement auprès de la population cible. Une fois que le laboratoire est convaincu que l'essai fonctionne de la manière décrite d'un point de vue analytique, la réalisation d'une validation limitée à l'étape 2 devrait alors être envisagée, dans le cadre de l'utilisation prévue et de la population cible, avant que l'essai ne soit utilisé en routine (Kirkland & Newberry, 2021).

Une **étude de vérification** est requise lorsqu'un test validé est utilisé dans un nouveau laboratoire, afin de garantir que les résultats de l'étude de validation originale se vérifient et que la performance du test est comparable.

7. Nouvelles technologies

L'utilisation du séquençage à haut débit (SHD) et les perspectives d'applications qu'il offre en termes de diagnostic sont en plein essor ; les principaux objectifs sont le séquençage non biaisé pour la découverte d'agents pathogènes et le séquençage ciblé pour la détection et la caractérisation ultérieure. Si l'essai est utilisé pour détecter des microorganismes non identifiés jusque-là, comme c'est le cas pendant l'investigation d'un foyer de maladie, le principal objectif est alors le diagnostic. Si l'essai est utilisé pour caractériser plus en détails un agent pathogène déjà identifié ou pour suivre l'épidémiologie moléculaire de l'agent pathogène pendant un foyer, son objectif général peut alors être décrit comme étant un test complémentaire. En l'absence d'analyte cible connu, il n'est pas possible de suivre

le processus traditionnel de validation. Le monitoring des critères de qualité, tels que la profondeur de la couverture, l'uniformité de la couverture, les biais GC, les indices de qualité de la transmission de base, le déclin de l'intensité du signal ou de la longueur des séquences ou la qualité de la cartographie et l'inclusion de contrôles internes sont utilisés pour évaluer la performance relative de différents essais de séquençage à haut débit (Halpin *et al.*, 2021; van Borm *et al.*, 2016).

Une autre difficulté de la validation des tests diagnostiques est le développement d'essai multiplexes tels que les essais à base de billes ou la PCR/RT-PCR multiplex en temps réel lorsque plus d'une cible a été identifiée. L'exactitude de chacune des cibles dont les concentrations et la distribution diffèrent est importante et difficile à valider.

8. Conclusions

Une validation adéquate et la vérification des caractéristiques de performance des tests diagnostiques pour les maladies infectieuses sont essentielles pour garantir que les essais sont utilisés et interprétés de manière robuste et valide (Colling & Gardner, 2021). Depuis sa première adoption en 1996, le Processus de l'OMSA de mise au point et de validation d'un essai (Figure 1) a servi de norme internationalement reconnue pour la validation des tests diagnostiques vétérinaires pour les maladies infectieuses.

Alors que la norme de l'OMSA décrit une approche globale pour la validation d'un test, l'expérience a montré que les développeurs de tests restent confrontés à des difficultés nombreuses et majeures pour se conformer à certaines étapes de ce processus. L'identification de ces obstacles a permis de définir les possibilités d'amélioration des normes en matière de validation des tests diagnostiques ainsi que des approches et des activités de vulgarisation ; elles sont résumées dans le Tableau 5 (Reid *et al.*, 2022).

Tableau 5. Résumé des difficultés et des perspectives pour la validation des tests diagnostiques

Difficultés	Perspectives
<ul style="list-style-type: none"> Quel est l'objectif du test? 	<ul style="list-style-type: none"> Définir clairement le ou les objectifs ainsi que les paramètres associés ; par exemple, un test de dépistage requiert une sensibilité diagnostique élevée alors qu'un test de confirmation requiert une spécificité diagnostique élevée
<ul style="list-style-type: none"> Définir la portée et les limites du test 	<ul style="list-style-type: none"> Que peut-on attendre du test (portée) et que ne doit-on pas attendre (limites)?
<ul style="list-style-type: none"> Définition du cas : en quoi consiste un animal infecté et en quoi consiste un animal non infecté? 	<ul style="list-style-type: none"> Par exemple : positif au(x) test(s) de référence, lésions caractéristiques, infection expérimentale, échantillon prélevé sur un individu provenant d'une population historiquement négative, etc. Si le statut infectieux n'est pas connu, un modèle bayésien de structure latente peut s'appliquer
<ul style="list-style-type: none"> Le test de référence est-il d'une exactitude imparfaite ou probablement inférieure au nouveau test en cours d'évaluation? 	<ul style="list-style-type: none"> Utiliser un modèle bayésien de structure latente pour estimer la sensibilité et la spécificité diagnostique, les rapports de vraisemblance et les autres paramètres pertinents, comme la prévalence
<ul style="list-style-type: none"> Espèces et échantillons 	<ul style="list-style-type: none"> Définir l'espèce et l'échantillon pour lesquels le test sera validé, par ex. poulet domestique, écouvillon nasopharyngé
<ul style="list-style-type: none"> Population source et population cible 	<ul style="list-style-type: none"> La population source (dans laquelle le test a été validé) est-elle similaire aux populations cibles (dans lesquelles le test sera utilisé) ?
<ul style="list-style-type: none"> Conception, analyse, interprétation et rapports des études de validation/vérification (absence de données de validation originales)? 	<ul style="list-style-type: none"> Chapitres 1.1.6, 2.2.1–2.2.8 du <i>Manuel terrestre</i>, processus de certification et d'enregistrement de l'OMSA, STARD, ParaTB, Aquatic, BLCM, ISO/IEC 17025:2017, modèles de validation d'organisations et d'ateliers nationaux and internationaux fournis par les Laboratoires de référence et les Centres collaborateurs de l'OMSA
<ul style="list-style-type: none"> Absence d'échantillons (maladies nouvelles ou émergentes, maladies rares, maladies subcliniques, maladies de la faune sauvage, etc.) 	<ul style="list-style-type: none"> Panels de référence d'échantillons bien décrits, si existants, et échantillons pour la comparaison inter-laboratoires (réseau collaboratif, "Vetlab"), reconnaissance provisoire

La rigueur avec laquelle les essais devraient être validés peut sembler un défi colossal pour les laboratoires diagnostiques et les équipes de recherche. Toutefois, des ressources existent pour aider à planifier et guider les études de validation et de vérification. Parmi elles figurent les chapitres pertinents du *Manuel terrestre* de l'OMSA, le numéro spécial récent de la *Revue scientifique et technique* de l'OMSA sur la validation scientifique des tests diagnostiques (Colling & Gardner, 2021) ainsi que les documents d'orientation publiés par les organes d'accréditation nationaux et régionaux. Les leaders internationaux dans le domaine de la validation des tests diagnostiques, dont l'OMSA ainsi que les Laboratoires de référence et les Centres collaborateurs qui lui sont associés, tout comme les organes règlementaires nationaux, ont un rôle important à jouer dans la poursuite du développement des normes et des systèmes requis pour garantir que les développeurs d'essais aient les ressources nécessaires et les incitations leur permettant d'assumer la responsabilité d'effectuer et de décrire des études de validation bien conçues et transparentes. Les utilisateurs finaux des tests diagnostiques doivent également être soutenus pour endosser la responsabilité de comprendre et de vérifier la performance de l'essai dans leur propre laboratoire ainsi que pour communiquer clairement l'incertitude associée aux résultats des tests diagnostiques à leurs partenaires.

REFERENCES

- BATH C., SCOTT M., SHARMA P.M., GURUNG R.B., PHUENTSHOK Y., PEFANIS S., COLLING A., SINGANALLUR N., FIRESTONE S.M., UNGVANIJBAN S., RATTHANOPHART J., ALLEN J., RAWLIN G., FEGAN M. & RODONI B. (2020). Further development of a reverse-transcription loop-mediated isothermal amplification (RT-LAMP) assay for the detection of Foot-and-Mouth Disease Virus and validation in the field with use of an internal positive control. *Transbound. Emerg. Dis.*, **67**, 2494–2506. <http://dx.doi.org/10.1111/tbed.13589>.
- BOWDEN T.R., CROWTHER J.R. & WANG J. (2021). Review of critical factors affecting analytical characteristics of serological and molecular assays. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 53–73. doi:10.20506/rst.40.1.3208.
- BOSSUYT P.M., REITSMA J.B., BRUNS D.E., GATSONIS C.A., GLASZIOU P.P., IRWIG L., LIJMER J.G. MOHER D., RENNIE D., DE VET H.C.W., KRESSEL H.Y., RIFAI N., GOLUB R.M., ALTMAN D.G., HOOFT L., KOREVAAR D.A., COHEN J.F. & STARD (Standards for Reporting Diagnostic Accuracy (STARD)) GROUP (2015). STARD. An updated list of essential items for reporting diagnostic accuracy studies. *BJM*, 351:h5527. doi: 10.1136/bmj.h5527.
- BRANSCUM A.J, GARDNER I.A. JOHNSON. W.O. (2005). Estimation of diagnostic-test sensitivity and specificity through Bayesian modelling. *Prev. Vet. Med.*, **68**, 145–163.
- BROEMELING L.D. (2011a). Bayesian Methods for Medical Test Accuracy. *Diagnostics*, **1**, 1–35; <https://doi.org/10.3390/diagnostics1010001>.
- BROEMELING L.D. (2011b). *Diagnostics*, **1**, 53–76; <https://doi.org/10.3390/diagnostics1010053>.
- BROWN I., SLOMKA M.J., CASSAR C.A., MCELHINNEY L.M. & BROUWER A. (2021). The role of national and international veterinary laboratories. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 159–172. doi:10.20506/rst.40.1.3215.
- CARAGUEL C.G.B. & COLLING A. (2021). Diagnostic likelihood ratio – the next generation of diagnostic test accuracy measurement. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 299–309. doi:10.20506/rst.40.1.3226.
- CHEUNG A., DUFOUR S., JONES G., KOSTOULAS P., STEVENSON M.A., SINGANALLUR N.B. & FIRESTONE S.M. (2021). Bayesian latent class analysis when the reference test is imperfect. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 271–286. doi:10.20506/rst.40.1.3224.
- COLLING A. & GARDNER I.A. (eds). (2021). Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (*Rev. Sci. Tech. Off. Int. Epiz.*, **40**). The Special Issue is available at: <https://doc.woaah.org/dyn/portal/index.xhtml?page=alo&alold=41245&req=21&cid=1e3ba84f-604f-47b7-bec3-927e54e04b42>
- COLLING A. & GARDNER I.A. (2021). Conclusions: Validation of tests of WOAHL-listed diseases as fit-for-purpose in a world of evolving diagnostic technologies and pathogens. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 311–317. doi: 10.20506/rst.40.1.3227

- COLLING A., LUNT R., BERGFELD J., HALPIN K., McNABB L., JUZVA S., NEWBERRY K., MORRISSY C., HLAING LOH M., CARLILE G., WAUGH C., WRIGHT L., WATSON J., EAGLES D., LOOMES C., WARNER S., DIALLO I., KIRKLAND P., BRODER C., ZUELKE K., McCULLOUGH S. & DANIELS P. (2018). A network approach for provisional assay recognition of a Hendra virus antibody ELISA: test validation with low sample numbers from infected horses. *J. Vet. Diag. Invest.*, **30**, 362–369. <https://doi.org/10.1177/1040638718760102>.
- CROWTHER J.R., UNGER H. & VILJOEN G.J. (2006). Aspects of kit validation for tests used for the diagnosis and surveillance of livestock diseases: producer and end-user responsibilities. *Rev. sci. tech. Off. int. Epiz.*, **25**, 913–935.
- CULLINANE A. & GARVEY M. (2021). A review of diagnostic tests recommended by the World Organisation for Animal Health Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 75–89. <https://doi.org/10.20506/rst.40.1.3209>.
- DEJAEGHER B. & VANDER HEYDEN Y. (2006). Robustness tests. *LCGC Europe*, **19**, online at <http://www.lcgceurope.com/lcgceurope/content/printContentPopup.jsp?id=357956>
- DRUCE J., GARCIA K., TRAN T., PAPADAKIS G. & BIRCH C. (2012). Evaluation of swabs, transport media, and specimen transport conditions for optimal detection of viruses by PCR. *J. Clin. Microbiol.*, **50**, 1064–1065. doi:10.1128/JCM.06551-11.
- ENØE C., GEORGIADIS M.P. & JOHNSON W.O. (2000). Estimating the sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.*, **45**, 61–81.
- FINDLAY J.W.A. & DILLARD R.F. (2007). Appropriate calibration curve fitting in ligand binding assays. *AAPS J.*, **9** (2): E260-E267. (Also on-line as *AAPS Journal* [2007]; **9** [2], Article 29 [<https://www.springer.com/journal/12248>]).
- FOORD A., BOYD V., WHITE J., WILLIAMS D., COLLING A. & HEINE H. (2015). Flavivirus detection and differentiation by a microsphere array assay. *J. Virol. Methods*, **203**, 65–72.
- GARCIA M.E., BLANCO J.L., CABALLERO J. & GARGALLO-VIOLA D. (2002). Anticoagulants interfere with PCR used to diagnose invasive aspergillosis. *J. Clin. Microbiol.*, **40**, 1567–1568. doi:10.1128/jcm.40.4.1567-1568.2002.
- GARDNER I.A., COLLING A., CARAGUEL C.G., CROWTHER J.R., JONES G., FIRESTONE S.M. & HEUER C. (2021). Introduction: validation of tests for OIE-listed diseases as fit-for-purpose in a world of evolving diagnostic technologies. *Rev. Sci. Tec. Off. Int. Epiz.*, **40**, 19–28. doi:10.20506/rst.40.1.3207.
- GARDNER I.A., COLLING A. & GREINER M. (2019). Design, statistical analysis and reporting standards for test accuracy studies for infectious diseases in animals: Progress, challenges and recommendations. *Prev. Vet. Med.*, **162**, 46–55. doi:10.1016/j.prevetmed.2018.10.023.
- GARDNER I.A. & GREINER M. (2006). Receiver-Operating Characteristic Curves and Likelihood Ratios: Improvements over Traditional Methods for the Evaluation and Application of Veterinary Clinical Pathology Tests. *Vet. Clin. Pathol.* **35**, 8–17.
- GARDNER I.A., NIELSEN S.S., WHITTINGTON R.J., COLLINS M.T., BAKKER D., HARRIS B., SREEVATSAN S., LOMBARD J.E., SWEENEY R., SMITH D.R., GAVALCHIN J. & EDA S. (2011). Consensus-based reporting standards for diagnostic test accuracy studies for paratuberculosis in ruminants. *Prev. Vet. Med.*, **101**, 18–34. PMID: 21601933.
- GARDNER I.A., STRYHN H., LIND P. & COLLINS M.T. (2000). Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.*, **45**, 107–122. doi: 10.1016/s0167-5877(00)00119-7.
- GEORGIADIS M., JOHNSON W., GARDNER I. & SINGH R. (2003). Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Appl. Statist.*, **52** (Part 1), 63–76.
- GIBB A.P. & WONG S. (1998). Inhibition of PCR by agar from bacteriological transport media. *J. Clin. Microbiol.*, **36**, 275–276. doi:10.1128/JCM.36.1.275-276.1998.
- GIFFORD G., SZABO M., HIBBARD R., MATEO D., COLLING A., GARDNER I. & ERLACHER-VINDEL E. (2021). Validation, certification and registration of certified tests and regulatory control of veterinary diagnostic test kits. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 173–188. doi:10.20506/rst.40.1.3216.

- GREINER M. & GARDNER I.A. (2000). Epidemiologic Issues in the Validation of Veterinary Diagnostic Tests. *Prev. Vet. Med.*, **45**, 3–22.
- GREINER M., PFEIFFER D. & SMITH R.D. (2000). Principles and practical application of the receiver operating characteristic (ROC) analysis for diagnostic tests. *Vet. Prev. Med.*, **45**, 23–41.
- GREINER M., SOHR D. & GÖBEL P. (1995). A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *J. Immunol. Methods*, **185**, 123–132.
- HALPIN K., TRIBOLET L., HOBBS E.C. & SINGANALLUR N.B. (2021). Perspectives and challenges in validating new diagnostic technologies. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 145–157. doi:10.20506/rst.40.1.3214.
- HOBBS E., COLLING A., GURUNG R. & ALLEN J. (2020). The potential of diagnostic point-of-care tests (POCTs) for infectious and zoonotic animal diseases in developing countries: technical, regulatory and sociocultural considerations. *Transbound Emerg Dis.*, **68**, 1835–1849.
- HEUER C. & STEVENSON M.A. (2021). Diagnostic test validation studies when there is a perfect reference standard. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 261–270. doi:10.20506/rst.40.1.3223.
- HUI S.L. & WALTER S.D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, **36**, 167–171.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2019). ISO/TS 22583:2019(en), Guidance for supervisors and operators of point-of-care testing (POCT) devices. <https://www.iso.org/obp/ui/#iso:std:iso:ts:22583:ed-1:v:1:en>
- JACOBSON R.H. (1998). Validation of serological assays for diagnosis of infectious diseases. *Rev. sci. tech. Off. int. Epiz.*, **17**, 469–486.
- JIA B., COLLING A., STALLKNECHT D.E., BLEHERT D., BINGHAM J., CROSSLEY B., EAGLES D. & GARDNER I.A. (2020). Validation of laboratory tests for infectious diseases in wild mammals: review and recommendations. *J. Vet. Diagn. Invest.*, **32**, 776–792. doi:10.1177/1040638720920346.
- JOHNSON P. & CABUANG L. (2021). Proficiency testing and ring trials. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 189–203. <https://doi.org/10.20506/rst.40.1.3217>
- JOHNSON W.O., JONES G. & GARDNER I. (2019). Gold standards are out and Bayes is in: implementing the cure for imperfect reference tests in diagnostic accuracy studies. *Prev. Vet. Med.*, **167**, 113–127. doi:10.1016/j.prevetmed.2019.01.010.
- KIRKLAND P.D. & NEWBERRY K.M. (2021). Your assay has changed – is it still ‘fit for purpose’? What evaluation is required? *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 205–215. <https://doi.org/10.20506/rst.40.1.3218>.
- KOSTOULAS P., GARDNER I.A., ELSCHNER M.C., DENWOOD M.J., MELETIS E. & NIELSEN S.S. (2021). Examples of proper reporting for evaluation (Stage 2 validation) of diagnostic tests for diseases listed by the World Organisation for Animal Health, *Sci. Tech. Off. Int. Epiz.*, **40**, 287–298. doi:10.20506/rst.40.1.3225.
- KOSTOULAS P., NIELSEN S.S., BRANSCUM A.J., JOHNSON W.O., DENDUKURI N., DHAND N.K., TOFT N. & GARDNER I.A. (2017). STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. *Prev. Vet. Med.*, **138**, 37–47. PMID: 28237234.
- LUDI A.B., MIOULET V., BAKKALI KASSIMI L., LEFEBVRE D.J., DE CLERCQ K., CHITSUNGO E., NWANKPA N., VOSLOO W., PATON D.J. & KING D.P. (2021). Selection and use of reference panels: a case study highlighting current gaps in the materials available for foot and mouth disease. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 239–251. doi:10.20506/rst.40.1.3221.
- MAYO C.E., WEYER C.T., CARPENTER M.J., REED K.J., RODGERS C.P., LOVETT K.M., GUTHRIE A.J., MULLENS B.A., BARKER C.M., REISEN W.K. & MACLACHLAN N.J. (2021). Diagnostic applications of molecular and serological assays for bluetongue and African horse sickness. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 91–104. doi:10.20506/rst.40.1.3210.
- MICHEL A.L., VAN HEERDEN H., PRASSE D., RUTTEN V., DAHOUK S. AL & CROSSLEY B.M. (2021). Pathogen detection and disease diagnosis in wildlife: challenges and opportunities. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 105–118. doi:10.20506/rst.40.1.3211.

- MIYACHI H., MASUKAWA A., OHSHIMA T., FUSEGAWA H., HIROSE T., IMPRAIM C. & ANDO Y. (1998). Monitoring of inhibitors of enzymatic amplification in polymerase chain reaction and evaluation of efficacy of RNA extraction for the detection of hepatitis C virus using the internal control. *Clin. Chem. Lab. Med.*, **36**, 571–575. doi:10.1515/CCLM.1998.098.
- NEWBERRY K.M. & COLLING A. (2021). Quality standards and guidelines for test validation for infectious diseases in veterinary laboratories? *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 227–237. doi:10.20506/rst.40.1.3220.
- PÉREZ L.J., LANKA S., DESHAMBO V.J., FREDRICKSON R.L. & MADDOX C.V. (2020). A validated multiplex Real Time PCR assay for the diagnosis of infectious *Leptospira* spp.: a novel assay for the detection and differentiation from both pathogenic groups I and II. *Front. Microbiol.*, **11**, 457. <https://doi.org/10.3389/fmicb.2020.00457>.
- REID T., SINGANALLUR N. B., NEWBERRY K., WAUGH C., BOWDEN T. & COLLING A. (2021). Validation of diagnostic tests for infectious diseases: challenges and opportunities. International Symposium on Sustainable Animal Production and Health Current Status and Way Forward. 28 June–2 July 2021, Joint FAO/IAEA Centre. Accepted for publication in symposium proceedings 7 April 2022.
- REISING M.M., TONG C., HARRIS B., TOOHEY-KURTH K.L., CROSSLEY B., MULROONEY D., TALLMADGE R.L., SCHUMANN K.R., LOCK, A.B. & LOIACONO C.M. (2021). A review of guidelines for evaluating a minor modification to a validated assay. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 217–226. doi:10.20506/rst.40.1.3219.
- RICCHI M., MAZARELLI A., PISCINI A., DI CARO, A, CANNAS A., LEO S., RUSSO S. & ARRIGONI N. (2016). Exploring MALDI-TOF MS approach for a rapid identification of *Mycobacterium avium* ssp. paratuberculosis field isolates. *J. Appl. Microbiol.*, **122**, 568–577.
- SAAH A.J. & HOOVER D.R. (1997). ‘Sensitivity’ and ‘Specificity’ Reconsidered: The Meaning of These Terms in Analytical and Diagnostic Settings. *Ann. Internal Med.*, **126**, 91–94.
- SCHRADER C., SCHIELKE A., ELLERBROEK L. & JOHNE R. (2012). PCR inhibitors – occurrence, properties and removal. *J. Appl. Microbiol.*, **113**, 1014–1026. doi:10.1111/j.1365-2672.2012.05384.x.
- STEVENSON M., HALPIN K. & HEUER C. (2021). Detection of emerging infectious zoonotic diseases. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 119–130. doi:10.20506/rst.40.1.3212.
- TRIBOLET, L., KERR, E. COWLED, C., BEAN, A.G., STEWART, C.R., DEARNLEY, M. AND FARR, R. (2020). MicroRNA Biomarkers for infectious diseases: from basic research to biosensing. *Front. Microbiol.*, **11**, 1197. doi: 10.3389/fmicb.2020.01197.
- VAN BORM S., WANG J., GRANBERG F. & COLLING A. (2016). Next-generation sequencing workflows in veterinary infection biology: towards validation and quality assurance. *Rev. Sci. Tech. Off. Int. Epiz.*, **35**, 67–81. doi:10.20506/rst.35.1.2418.
- VESSMAN J., STEFAN R., VAN STADEN J., DANZER K., LINDNER W., BURNS D., FAJGELJ A. & MULLER H. (2001). Selectivity in analytical chemistry. *Pure Appl. Chem.*, **73**, 1381–1386.
- WATSON J.W., CLARK G.A. & WILLIAMS D.T. (2021). The value of virtual biobanks for transparency purposes with respect to reagents and samples used during test development and validation. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 253–259. doi:10.20506/rst.40.1.3222.
- WAUGH C. & CLARK G. (2021). Factors affecting test reproducibility among laboratories. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 131–143. doi:10.20506/rst.40.1.3213.
- WILSON I.G. (1997). Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.*, **63**, 3741–3751. doi:10.1128/AEM.63.10.3741-3751.1997.
- YAN L., TOOHEY-KURTH K.L., CROSSLEY B.M., BAI J., GLASER A.L., TALLMADGE R.L. & GOODMAN L.B. (2020). Inhibition monitoring in veterinary molecular testing. *J. Vet. Diagn. Invest.*, **32**, 758–766. doi:10.1177/1040638719889315
- YOKOTA M., TATSUMI N., NATHALANG O., YAMADA T. & TSUDA I. (1999). Effects of heparin on polymerase chain reaction for blood white cells. *J. Clin. Lab. Anal.*, **13**, 133–140. doi:10.1002/(sici)1098-2825(1999)13:3<133::aid-jcla8>3.0.co;2-0.
- ZWEIG M.H. & CAMPBELL G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, **39**, 561–577.

<https://www.iaea.org/services/networks/vetlab>

<https://www.iaea.org/services/zodiac>

<https://www.woah.org/en/what-we-offer/veterinary-products/diagnostic-kits/the-register-of-diagnostic-kits/>

<https://www.awe.gov.au/agriculture-land/animal/health/laboratories/tests/test-development>

*
* *

N.B. : Il existe un Centre collaborateur de l'OMSA pour la validation scientifique des tests diagnostiques dans la région Asie-Pacifique (veuillez consulter le site web de l'OMSA: <https://www.woah.org/fr/ce-que-nous-proposons/reseau-dexpertise/centres-collaborateurs/>)

Veuillez contacter le Centre collaborateur de l'OMSA pour de plus amples informations concernant la validation.

N. B. : ADOPTE POUR LA PREMIERE FOIS EN 1996 EN TANT QUE *PRINCIPES DE VALIDATION DES EPREUVES DIAGNOSTIQUES POUR LES MALADIES INFECTIEUSES*. MISES A JOUR LES PLUS RECENTES ADOPTEES EN 2023.