

Rejane Silva Diniz

**VALIDAÇÃO DE MÉTODOS LABORATORIAIS APLICADOS AO
DIAGNÓSTICO DE DOENÇAS DOS ANIMAIS.**

Tese apresentada à Universidade Federal de Minas Gerais, Escola de Veterinária, como requisito parcial para a obtenção do grau de doutora em Ciência Animal.

Área de concentração: Epidemiologia

Orientador: João Paulo Amaral Haddad

Belo Horizonte
Escola de Veterinária da UFMG
2013

D585v Diniz, Rejane Silva, 1967-
Validação de métodos laboratoriais aplicados ao diagnóstico de doenças dos animais /
Rejane Silva Diniz. – 2013.
83 p. : il.

Orientador: João Paulo Amaral Haddad
Tese (doutorado) – Universidade Federal de Minas Gerais, Escola de Veterinária
Inclui bibliografia

1. Animais – Doenças – Diagnóstico – Teses. 2. Diagnóstico de laboratório – Técnica –
Teses. I. Haddad, João Paulo Amaral. II. Universidade Federal de Minas Gerais. Escola de
Veterinária. III. Título.

CDD – 636.089 607 5

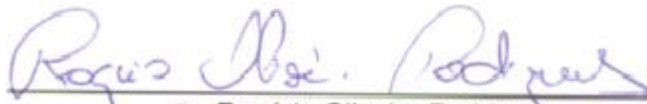
Tese defendida e aprovada em 13 de dezembro de 2013, pela Comissão Examinadora constituída por:



Prof. João Paulo Amaral Haddad
Presidente - Orientador



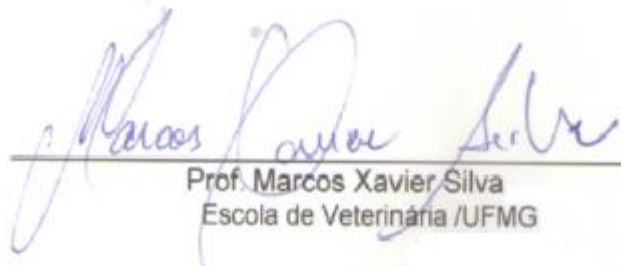
Dr. Jorge Caetano Júnior
Ministério da Agricultura, Pecuária e Abastecimento



Dr. Rogério Oliveira Rodrigues
Fundação Estadual de Pesquisa Agropecuária/RS



Prof. Cristiano Barros de Melo
Universidade de Brasília



Prof. Marcos Xavier Silva
Escola de Veterinária /UFMG

A minha querida filha, Ana Luiza.

“A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original”.

Albert Einstein

AGRADECIMENTOS

Um enorme carinho aos meus pais, Roberto e Rita, cuja dedicação me fizeram seguir em frente.

A minha filha Ana Luiza pela constante parceria, entusiasmo e tanta alegria no dia a dia.

Ao comitê de orientação Prof. Dr. João Paulo Amaral Haddad, Prof. Dr. Jenner Karlisson Pimenta dos Reis e Prof. Dr. Rômulo Cerqueira Leite pela orientação, pelo meu crescimento científico e pelas perspectivas dessa pesquisa.

Meu reconhecimento e agradecimento a Universidade Federal de Minas Gerais pelo ambiente criativo e apoio acadêmico.

A Escola de Veterinária pelo compromisso com a minha formação na Pós-Graduação.

A minha banca de exame de qualificação, Prof. Dr. Roberto Maurício Carvalho Guedes, Dr. Jorge Caetano Junior e Prof. Dr. Marcos Bryan Heinemam pelas importantes avaliações e contribuições para a completude dessa tese.

Muito obrigada aos colaboradores, Dr. Guilherme Nunes de Souza, aos diretores e coordenadores das unidades dos cursos de medicina veterinária, que autorizaram o envio dos questionários e, aos pesquisadores que participaram dessa pesquisa.

Ao Colégio Técnico da Universidade Federal de Minas Gerais e ao setor de Patologia Clínica pelas ações no sentido de possibilidades de crescimento.

Aos colegas do laboratório de Epidemiologia pela excelente convivência, diálogo, cooperação e desenvolvimento mútuo.

Aos Professores, pela dedicação na formação de qualidade do aluno de Pós-Graduação.

Aos funcionários da escola de veterinária, pelo compromisso com o bom andamento das atividades do meu doutoramento.

Aos colegas do Laboratório de Retrovírus da Escola de Veterinária (Retrolab) pela integração, incentivo, e pela amplitude das reuniões científicas.

À Pró- Reitoria de Pesquisa da Universidade Federal de Minas Gerais, na gestão do Prof. Dr. Renato de Lima Santos pelo apoio para participação de evento científico, e aos órgãos de fomento CAPES, INCT e CNPq que direta ou indiretamente vieram a colaborar com o desenvolvimento dessa pesquisa.

Um agradecimento especial à minha extraordinária banca de defesa de tese pelo envolvimento profissional, reconhecimento, e “lapidação” da tese.

SUMÁRIO

	LISTA DE ÓRGÃOS	10
	RESUMO	11
	ABSTRACT	12
1.	INTRODUÇÃO	13
2.	LITERATURA CONSULTADA	13
2.1	VALIDAÇÃO DE MÉTODOS	13
2.1.1	Aspectos gerais.....	13
2.1.2	Definição de validação	15
2.1.3	Critérios para validação.....	15
2.1.4	O que cada critério de validação expressa.....	16
2.1.4.1	Sensibilidade analítica.....	16
2.1.4.2	Especificidade analítica.....	17
2.1.4.3	Repetibilidade.....	17
2.1.4.4	Reprodutibilidade	18
2.1.4.5	Robustez.....	18
2.1.4.6	Ponto de corte.....	19
2.1.4.7	Sensibilidade e Especificidade diagnóstica.....	19
2.1.4.8	Valor preditivo positivo e negativo	21
2.1.5	Estágios da validação	21
2.2	TESTES ESTATÍSTICOS.....	24
2.2.1	Aspectos gerais.....	24
2.2.2	Desvio-padrão e Coeficiente de variação	25
2.2.3	Coeficiente de correlação	26
2.2.4	Teste Qui-Quadrado	27
2.2.5	Teste Kappa.....	27
2.2.6	Curva ROC.....	28
2.2.7	Modelo Bayesiano.....	30
2.3	QUALIDADE E METROLOGIA	31
2.3.1	Aspectos gerais.....	31
2.3.2	Normas ISO.....	32
2.3.3	Metrologia	32
2.3.4	Acreditação, certificação e credenciamento laboratorial.....	33
2.3.5	Qualidade e manutenção dos critérios de validação.....	34
2.3.5.1	Controle de qualidade interno	34
2.3.5.2	Controle de qualidade externo, teste de proficiência.....	34
2.3.6	STARD e validação de ensaios diagnósticos veterinários.....	35
3.	MATERIAL E MÉTODOS	36
3.1	Aspectos gerais.....	36
3.2	Metodologia.....	36

3.2.1	Estudo dos documentos	37
3.2.2	Estudo do processo de desenvolvimento e validação de um método diagnóstico	38
3.2.3	Estudo da presença dos critérios de validação de métodos do documento da OIE (OIE..., 2010) nos artigos científicos publicados nos anos de 2011 e 2012.....	40
4.0	RESULTADOS E DISCUSSÃO	41
4.1	Estudo dos documentos de validação	41
4.2	Estudo do processo de desenvolvimento e validação de um método diagnóstico	45
4.3	Estudo da presença dos critérios de validação de métodos nos artigos científicos publicados nos anos de 2011 e 2012.....	48
5.	CONCLUSÃO	51
6.	REFERÊNCIAS BIBLIOGRÁFICAS	52
7.	ANEXOS	57
7.1	Anexo 1- Termo de consentimento	57
7.2	Anexo 2- Questionário Validação	59
7.3	Anexo 3- Modelo de correspondência.....	65
7.4	Anexo 4- Manual de Testes de Diagnóstico e Vacinas para Animais Terrestres (OIE..., 2010).....	66

LISTA DE TABELAS

Tabela 1-	Tabela de contingência para avaliar o desempenho do teste	20
Tabela 2-	Classificação qualitativa de valores de Kappa como grau de concordância além do acaso.....	27
Tabela 3-	Equilíbrio entre sensibilidade e especificidade	30
Tabela 4-	Distribuição das frequências absolutas e percentuais dos critérios de desenvolvimento e validação de métodos segundo a OIE (OIE..., 2010) nos documentos estudados.....	43
Tabela 5-	Distribuição das frequências absolutas e percentuais das ferramentas estatísticas recomendadas nos documentos estudados.....	44
Tabela 6-	Distribuição das frequências absolutas e percentuais da completude dos documentos estudados em relação à presença de conceitos, estágios, métodos estatísticos e planejamento de avaliação dos testes.....	45
Tabela 7-	Sumário dos resultados para avaliação da importância das etapas de validação de métodos.....	47
Tabela 8-	Sumário dos resultados do estudo do processo de validação	50

Tabela 9-	Sumário dos resultados do estudo da presença dos critérios de validação na amostra dos artigos científicos publicados nos anos de 2011 e 2012.....	51
-----------	--	----

LISTA DE FIGURAS

Figura 1.	Os cinco estágios no processo de validação do ensaio segundo Jacobson (1998).	23
Figura 2.	Kappa	28
Figura 3.	Curva ROC.....	29
Figura 4.	Distribuição dos conceitos dos cursos na amostra.....	39
Figura 5.	Distribuição de Pesquisadores das Instituições na amostra.....	40
Figura 6.	Fluxograma da pesquisa da literatura de validação de métodos publicada em 2011 e 2012.....	42
Figura 7.	Distribuição das respostas dos pesquisadores em relação aos documentos utilizados na avaliação da acurácia diagnóstica	46
Figura 8.	Distribuição das respostas dos pesquisadores em relação ao uso das ferramentas estatísticas na avaliação do desempenho de testes diagnósticos...	47
Figura 9.	Intervalos do grau de discordância e concordância do estudo do processo de validação.....	48

LISTA DE ÓRGÃOS

OIE- World Organisation for Animal Health

CFIA- Canadian Food Inspection Agency

SCAHLS- Subcommittee on Animal Health Laboratory Standards

AAVLD- American Association of Veterinary Laboratory Diagnosticians

VICH- International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Products

WHO- World Health Organisation

MAPA- Ministério da Agricultura, Pecuária e Abastecimento

FAO-CODEX- Food and Agriculture Organisation of the United Nations - Codex *Alimentarius*

OECD- Organisation for Economic Co-operation and Development

RESUMO

Métodos diagnósticos aplicados à veterinária são detentores de notória relevância para diagnósticos de doenças dos animais, tanto no desenvolvimento de atividades de pesquisa quanto para a realização de análises laboratoriais de rotina. A credibilidade atribuída aos resultados obtidos a partir dos ensaios laboratoriais é dependente, em grande medida, da validação do método utilizado. Um método validado deve haver comprovado por avaliação e fornecimento de evidências que os requisitos para um determinado uso pretendido tenham sido cumpridos. Atualmente, existem organizações e agências que editam diretrizes e, recomendações com esse propósito, entretanto, nem sempre completas. Além disso, tais recomendações podem diferir em alguns requisitos e ainda, nos parâmetros de validação mesmo em se tratando de um mesmo tipo de método diagnóstico. Como resultado, observa-se uma busca constante pela harmonização de normas objetivando assegurar a equivalência de ensaios laboratoriais. O objetivo dessa pesquisa foi construir um modelo teórico-metodológico norteador da validação de métodos sorológicos para diagnósticos laboratoriais de doenças dos animais. Nessa pesquisa, em um primeiro momento, a partir do ano 2000, foram analisadas as recomendações, para o desenvolvimento de processos de validação de métodos para diagnóstico veterinário, utilizados por 40 países, entre aqueles, 34 membros e seis parceiros da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Em um segundo momento, foi realizado um inquérito sobre o processo de desenvolvimento e validação de ensaios laboratoriais veterinários em uma amostra de 38 pesquisadores, de 10 universidades brasileiras. Num terceiro momento foi verificada a frequência dos critérios de validação em artigos publicados em periódicos internacionais. Os resultados das análises realizadas apontaram que os critérios mais frequentemente empregados em processos de desenvolvimento e validação de métodos diagnósticos foram: proposta de uso e reprodutibilidade (88,9%) e sensibilidades e especificidades analíticas e diagnósticas (72,2%). A referência mais utilizada em pesquisa e validações de ensaios aplicados à veterinária no nível nacional foi o Manual de Provas de Diagnóstico e de Vacinas para os Animais Terrestres da Organização Mundial da Saúde Animal (OIE). Os artigos científicos pesquisados que referenciaram as normas e literatura de validação de métodos apresentaram maior frequência (46,3%) dos critérios descritos nesta tese. Em conclusão, o manual de Provas da OIE (OIE..., 2010) apresentou o maior número de critérios de validação de métodos, o que geraria uma maior aceitação internacional do ensaio a ser validado. Dessa forma, a consulta de tal documento, associado à consulta de artigos científicos especializados poderiam potencializar a validação de ensaios veterinários.

Palavras-chave: Validação de métodos, validação de ensaios laboratoriais veterinários, critérios de validação.

ABSTRACT

Diagnostic methods applied to veterinary have notorious relevance for diagnosis of diseases of animals, both in the development of research activities and for conducting routine laboratory analysis. The credibility assigned to the results obtained from the laboratory assays depends largely on the validation of the method. A validated method should be confirmed by evaluating and providing evidence that the requirements for a specific intended purpose have been fulfilled. Currently, there are organizations and agencies who publish guidelines and recommendations for this purpose, however, these are not always complete. In addition, such recommendations may differ in some requirements and also the validation parameters even in the case of the same type of diagnostic method. As a result, there is a constant search for the harmonization of standards in order to assure the equivalence of laboratory assays. The objective of this research was to construct a theoretical -methodological validation model of serological methods for laboratory diagnosis of infections of animals. In this research, at first, from the year 2000, were analyzed the recommendations for the development of validation processes for veterinary diagnostic methods, used by 40 countries, among those, 34 current membership and six partners of the Organization for Economic Cooperation and Development (OECD). In a second step, a survey was conducted on the process of development and validation of veterinary laboratory tests in a sample of 38 researchers from 10 Brazilian universities. In the third step was investigated the frequency of validation criteria in articles published in international journals. The results of analyzes indicated that the criteria most often employed in the development and validation of diagnostic methods process were: purpose of the assay and reproducibility (88.9%) and analytical and diagnostic sensitivities and specificities (72.2%). The reference most used in research and validation of tests applied to the veterinarian at the national level was the Manual of Diagnostic Tests and Vaccines for Terrestrial Animals of the World Organisation for Animal Health (OIE). The papers surveyed that were referenced with standards and method validation literature showed a higher frequency (46.3%) of criteria described in this thesis. In conclusion, the OIE Manual of Diagnostic Tests (OIE..., 2010) had the highest number of validation criteria, which would generate greater international acceptance of test to be validated. Thus, the use of such document associated with the consultation of specialized scientific articles could potentiate the validation of veterinary tests.

Key words: Validation of methods, validation of veterinary laboratory tests, validation criteria.

1. INTRODUÇÃO

As informações contidas nos sistemas de informações sobre doenças no Brasil são, tanto na área humana quanto na área animal, imprescindíveis para a realização de avaliações e intervenções relevantes. A qualidade dessas informações, em grande medida dependente da utilização de métodos laboratoriais adequados e validados, é indispensável para as tomadas de decisões acertadas em saúde pública, agronegócios, tramitação de comércio internacional e no desenvolvimento de novas tecnologias e formas de tratamentos.

A globalização tem aumentado o risco de epidemias de doenças infecciosas e, conseqüentemente, reforçado a imprescindibilidade da qualidade diagnóstica em ensaios laboratoriais. Encontram-se disponíveis diversos métodos diagnósticos com complexidade variada, podendo ser utilizados para demonstrar a presença ou ausência de infecção. Dentro dessa variedade, há uma busca constante pelo desenvolvimento de testes rápidos e capazes de detectar a presença de agentes etiológicos com alto grau de confiança e produzindo consideráveis efeitos positivos para o seu controle.

A validação dos ensaios é um estudo necessário para obtenção de resultados válidos e da estimativa da acurácia. Para alcançar a amplitude requerida para a avaliação de testes diagnósticos, imprescindível à sua aplicação, nas áreas de diagnóstico humano e animal, várias propostas de documentos de validação têm sido apresentadas por órgãos oficiais e certificadores de qualidade. A Organização Mundial do Comércio (OMC) reconheceu três organizações internacionais como órgãos competentes para a definição de normas: a Organização Mundial de Saúde Animal (OIE), reconhecida como referência em padrões de saúde animal; a Comissão do Codex Alimentarius (Codex), órgão da

Organização das Nações Unidas para Agricultura e Alimentação (FAO); e a Organização Mundial da Saúde (OMS), que elabora normas para a segurança alimentar e saúde pública (Wiegers, 2002).

Ensaio laboratoriais voltados ao diagnóstico de infecção por agentes etiológicos de doenças dos mamíferos, aves e abelhas e, de animais aquáticos têm sido reconhecidos internacionalmente pela OIE com base em provas de sua utilidade regional, nacional ou internacional. A validação adequada do método é inerente à sua transferência, seja do meio acadêmico para laboratórios oficiais ou credenciados pelo Governo ou privados ou, ainda, para entidades do setor produtivo.

Considerando a crescente preocupação com a confiabilidade de testes diagnósticos e, da demanda nacional de geração de produtos agropecuários estratégicos e, visando o alcance de novos patamares de competitividade, de acordo com a política nacional de desenvolvimento de biotecnologia, se propõe no presente trabalho uma avaliação das diferentes normas para validação dos ensaios. Assim, esta pesquisa visa, como resultado, o preenchimento de lacuna no que diz respeito às recomendações de validação de ensaios laboratoriais de uso veterinário, destinados ao diagnóstico de doenças dos animais no Brasil e, contribuir para com a pesquisa e o desenvolvimento de ensaios que atendam às normas e diretrizes nacionais e internacionais, favorecendo sua aceitação em diferentes instâncias.

2. LITERATURA CONSULTADA

2.1 VALIDAÇÃO DE MÉTODOS

2.1.1 Aspectos gerais

A pesquisa e o desenvolvimento de ensaios laboratoriais são importantes para a melhoria do diagnóstico das infecções e para auxiliar no entendimento da sua patogênese. Em

consulta da Organização Mundial da Saúde (OMS) sobre "as tendências futuras na saúde pública veterinária", realizada em Téramo, Itália, em 1999, a Saúde Pública Veterinária foi definida como "a soma de todas as contribuições para o bem-estar físico, mental e social dos seres humanos através da compreensão e aplicação da ciência veterinária". É reconhecido que a saúde humana está fortemente ligada à saúde e à produção animal. Segundo a OMS aproximadamente 75% das novas doenças que têm afetado os seres humanos ao longo dos últimos 10 anos têm sido causadas por agentes de origem animal (Future..., 2002).

O desenvolvimento ou aprimoramento de novos métodos diagnósticos potencializa o controle e a prevenção das doenças. Porém, segundo o documento final versão 2005 da Plataforma Tecnológica Europeia para a Saúde Animal Global, visão 2015, não há legislação comunitária para licenciamento de testes de diagnósticos em saúde animal. No mencionado documento é relatado que, em 2005, a OIE desenvolveu um novo processo de validação de testes de diagnóstico em conformidade com o Manual de Testes de Diagnóstico e Vacinas para Animais Terrestres. No entanto, em relação aos aspectos regulatórios, não existe ainda nenhum mecanismo formal para a aprovação ou autorização de testes de diagnóstico, embora existam diretrizes internacionais para a sua validação e padronização (European..., 2005).

A OIE estabeleceu um "processo de reconhecimento" de testes de diagnóstico, incluindo kits comerciais, através de sua rede mundial de mais de 150 laboratórios de referência em 31 países. Tal processo foi importante para a aceitação de ensaios seletivos voltados à diferenciação entre animais vacinados de infectados, em situação nas quais o endosso internacional se revelasse essencial para que os países importadores aceitassem os testes (European..., 2005).

Nos últimos anos surtos das doenças têm reforçado a necessidade urgente de melhorar os testes de diagnóstico existentes. Para muitas infecções, o diagnóstico na fase inicial e o tratamento podem ter um papel importante na prevenção de complicações, ou na interrupção da transmissão do agente infeccioso. Embora existam ensaios de alta qualidade, o relatório do Comitê Científico Europeu da Saúde e Bem-Estar, identificou, em 2003, limitações para os testes de diagnóstico (European..., 2005). A fim de superar tais limitações, o comitê ressaltou a necessidade de desenvolver: métodos eficazes e de baixo custo para: triagem de produtos de origem animal, testes simples e rápidos para uso no campo e laboratórios regionais para apoiar a investigação de suspeitas clínicas de doenças; testes rápidos e altamente sensíveis para a detecção de animais recém- infectados; testes mais sensíveis e específicos para detectar a infecção de um animal, e métodos mais rápidos e sensíveis para o diagnóstico diferencial. Para que esses testes sejam desenvolvidos e utilizados com credibilidade é imprescindível que sejam validados (European..., 2005).

O trabalho em rede dos laboratórios de referência da OIE devem assegurar resultados congruentes em todos os laboratórios. Os resultados devem ser obtidos através da adesão às normas da OIE, com uso de material de referência dentro do grupo e, participação em testes de proficiência apropriados (OIE..., 2011b). A participação no grupo de trabalho é compulsória para os laboratórios de referência da OIE. Esses laboratórios são designados para seguir todos os problemas técnicos e científicos relativos a uma doença ou tópico específico (OIE..., 2011d).

Para a obtenção da posição de laboratório de referência da OIE os laboratórios candidatos devem apresentar informações sobre seus especialistas, estudos e experiências em diagnósticos. Dentre essas, deve ser descrito

os projetos atuais de pesquisa e desenvolvimento de métodos na doença, incluindo as publicações científicas sobre o tema. Bem como as experiências em ensaios diagnósticos de acordo com os padrões OIE e a ‘expertise’ em técnicas diagnósticas. Além disso, é requerida informação sobre experiência em padronização e validação de ensaios diagnósticos (OIE..., 2011c)

As orientações para os laboratórios candidatos são disponibilizadas no “Guideline for applicants for OIE Reference Laboratory Status” (OIE..., 2011c). A conformidade com a OIE e outros padrões internacionais para qualidade assegurada e medidas de Biossegurança, são critérios aplicados na seleção dos laboratórios de Referência (OIE..., 2011a). Dentre as cláusulas de referência encontra-se descrito que os laboratórios de referência devem coordenar estudos técnicos e científicos em colaboração com outros laboratórios, centros e organizações; fornecer, quando for o caso, conselhos técnicos e científicos sobre medidas de controle da doença (OIE..., 2011e).

2.1.2 Definição de validação

A validação é um processo que determina a capacidade de um ensaio, o qual foi devidamente desenvolvido, otimizado e padronizado, para um fim pretendido. Tem sido definido também como a confirmação por exame e fornecimento de evidência objetiva de que os requerimentos específicos para um determinado uso pretendido são cumpridos (OIE..., 2010).

Para manter um estado de ensaio validado, no entanto, é necessário acompanhar atentamente o desempenho diário do ensaio e muitas vezes, monitorar o comportamento dos controles internos ao longo do tempo. Este garante que o ensaio, tal como inicialmente validado, mantenha consistentemente suas características de desempenho (OIE..., 2010).

2.1.3 Critérios para validação

Um método será considerado validado se suas características estiverem em conformidade com os critérios e limites estabelecidos de acordo com a particularidade do ensaio e da doença em estudo.

A metodologia ideal de validação é aquela que tenha progredido através de um estudo baseado em protocolos internacionais de realização e interpretação harmonizados de estudos propostos para avaliação de desempenho dos testes (Validation..., 1997)

A OIE propõe doze critérios para validação de ensaios aplicáveis a doenças infecciosas: a proposta de uso, a otimização, a padronização, a robustez, a repetibilidade, a sensibilidade analítica, a especificidade analítica, o ponto de corte, a sensibilidade diagnóstica, a especificidade diagnóstica, a reprodutibilidade e *Ruggedness* (OIE..., 2010).

Para que um método seja aceito pelos governos dos países, o método validado deve ser prático e apropriado para o uso. As propostas de uso mais frequentes para métodos de diagnóstico de doenças dos animais são: demonstrar que uma população definida se encontra livre de infecção, com ou sem vacinação; restabelecer o estado livre de infecção após surtos; avaliar a erradicação da doença ou a eliminação da infecção de populações animais definidas; confirmação de casos suspeitos; determinar o estado imunológico de animais submetidos à vacinação; certificar possibilidade de infecção ou a presença do agente em animais ou a presença do agente em produtos de origem animal; estimar prevalência de uma doença; para uso em laboratório de pesquisa; e para certificação com finalidade de exportação (Wiegers, 2003; OIE..., 2010).

Para Wiegers (2003) também devem ser consideradas as seguintes utilizações para métodos de diagnóstico de doenças dos

animais: teste de um único animal ou de um rebanho e, verificação de lotes de produtos. Informações sobre as amostras (*pool* de amostras ou amostras individuais) e tipo de amostra (espécie, tecido, matriz).

No desenvolvimento do ensaio algumas condições são consideradas essenciais: a garantia da qualidade, a seleção de equipamentos calibrados e a seleção, coleta, preparação e gestão das amostras. Durante o desenvolvimento do teste os limites de detecção superiores e inferiores são estabelecidos. Para o conhecimento dessa faixa, uma amostra de referência elevada é selecionada (OIE..., 2010).

A otimização é uma etapa recomendada no desenvolvimento do ensaio. Otimização é o processo por meio do qual os parâmetros físicos, químicos e biológicos mais importantes de um teste são avaliados e ajustados de modo a assegurar que as características de desempenho do ensaio sejam as mais adequadas para a aplicação pretendida (OIE..., 2010). Controles, por exemplo, de equipamentos e materiais, diluições e faixa de pH do teste são parâmetros que fazem parte da otimização do ensaio. É útil selecionar, pelo menos, três amostras de referência bem definidas, que representam o que se pretende determinar, variando de negativa a forte positividade. O ideal é que estas amostras representem animais sabidamente infectados e não infectados, selecionados a partir de população na qual o ensaio será aplicado ou de população que eventualmente tornar-se-á o alvo do ensaio, quando o teste encontrar-se validado. Uma análise científica criteriosa e a utilização das melhores práticas proporcionam uma boa otimização do ensaio. Exemplos de componentes para a otimização de um imunoenensaio enzimático (ELISA) são: as diluições das variáveis antígeno e anticorpo que proporcionam melhor detecção do que se propõe detectar e a definição do melhor tempo de incubação (Jacobson, 1998).

Muitos ensaios de uso veterinário e a maioria dos equipamentos de laboratório utilizados são adaptados de métodos de ensaio humanos, mas devem ser adaptados com sucesso para a diversidade de espécies animais para que as aplicações sejam úteis aos propósitos finais (Guideline..., 2007). Para Hoffmann *et al.* (2008), a padronização do método o torna comparável, diante de resultados que foram obtidos em condições diferentes, por exemplo em dias diferentes ou em diferentes laboratórios. Sua confiabilidade é estabelecida após padronização.

Já a OIE (OIE..., 2010) denomina essa fase de “normalização” e relata que devido à variação inerente aos resultados dos testes que são frequentemente observados entre as corridas executadas do mesmo ensaio ou entre laboratórios utilizando ensaios iguais ou semelhantes, é virtualmente impossível comparar diretamente os dados quantitativos. Para melhorar sensivelmente a comparabilidade dos resultados dos testes dentro e entre laboratórios, um ou mais reagentes padrão são incluídos em cada corrida de um ensaio. Valores primários do teste para cada amostra do ensaio podem então ser convertidos em unidades de atividade em relação ao padrão de trabalho pelo processo chamado “normalização”. O valor “normalizado” pode ser expresso em várias formas, tais como, por exemplo, uma porcentagem de um controle positivo.

2.1.4 O que cada critério de validação expressa

2.1.4.1 Sensibilidade analítica

Comumente, o limite de detecção e a sensibilidade analítica são considerados sinônimos. A sensibilidade analítica representa a menor quantidade de analito numa amostra que pode ser detectada por um ensaio de detecção direta em pelo menos 50% das repetições para cada diluição, numa

série de diluições do analito na matriz (OIE..., 2010).

Para ensaios de detecção direta, a sensibilidade analítica pode ser expressa como dose infecciosa ou, por exemplo, como o número de cópias do genoma do agente que pode ser detectado e distinguido do resultado da matriz. Para os ensaios de detecção indireta, a sensibilidade analítica é a quantidade mínima de anticorpo detectado, geralmente, na penúltima diluição da amostra. Na referida diluição o analito seria indistinguível da atividade de uma matriz da amostra controle (OIE..., 2010).

2.1.4.2 Especificidade analítica

Especificidade analítica é o grau em que o ensaio distingue entre o analito alvo e outros componentes que podem ser detectados. Por exemplo, interferentes, produtos de degradação ou a ligação não específica de reagentes a uma fase sólida. Essas variáveis podem causar respostas falsamente reduzidas ou elevadas no ensaio afetando negativamente a sua especificidade analítica (OIE..., 2010).

Os procedimentos utilizados para demonstrar a especificidade dependerão do objetivo do procedimento analítico. Nem sempre é possível demonstrar que um procedimento analítico é específico para o que se pretende medir em particular, ou seja, sua discriminação completa. Neste caso, uma combinação de dois ou mais processos de análise é recomendada para atingir o nível necessário de discriminação (Validation..., 1998).

O nível de discriminação de um procedimento pode ser confirmado através da obtenção de resultados positivos. Por exemplo, através da comparação com um material de referência conhecido ou, a partir de amostras que contêm o que se quer detectar, em conjunto com resultados negativos a partir de amostras que não contêm o que se quer detectar. Além disso, o

ensaio utilizado para essa medição pode ser aplicado à materiais estruturalmente semelhantes ou estreitamente relacionados com o analito para confirmar que, nestes casos, um resultado positivo não é obtido. A escolha de tais materiais potencialmente interferentes deve ser baseada em julgamento científico considerando as interferências que poderiam ocorrer na metodologia que está sendo utilizada (Validation..., 1998).

Segundo publicação do grupo de trabalho da comissão de representantes de laboratório veterinário (SCAHLs), estabelecido pelo governo da Austrália, revisado em agosto de 2012, a especificidade analítica de um método é a sua capacidade de diferenciação do analito alvo de analitos não-alvo, mas relacionados. O objetivo é descrever o grau em que o ensaio não reage de forma cruzada com outro analito, isto é, número de falsos positivos. Inclui uma comparação dos resultados com um padrão-ouro. O teste deve ser baseado em um painel de soros de animais com estado conhecido, o qual deve incluir soros de animais expostos a agentes encontrados em ambientes semelhantes, que produzem sintomas similares e / ou que tenham taxonomia semelhante (Measurement..., 2010).

2.1.4.3 Repetibilidade

A repetibilidade expressa precisão sob as mesmas condições de funcionamento durante um intervalo de tempo curto e é também denominada estudo intra-ensaio. A precisão intermediária expressa variações dentro de laboratórios em dias diferentes, analistas diferentes, equipamentos diferentes, etc (Validation..., 1999).

Existe uma consistência na apresentação da precisão de um método analítico que é geralmente expressa como a variância, desvio padrão ou o coeficiente de variação de uma série de medições. Porém para considerações sobre repetibilidade observamos diferenças. A OIE, por

exemplo, conceitua a repetibilidade como o nível de concordância entre os resultados de repetições de uma amostra tanto no âmbito do próprio ensaio quanto entre as execuções do mesmo ensaio num dado laboratório (OIE..., 2010).

A repetibilidade é estimada avaliando a variação nos resultados de repetições com o número de replicatas preferencialmente determinado em consulta com um estatístico, e sugerida a partir de um mínimo de três amostras representativas do analito dentro da faixa de operação do ensaio (OIE..., 2013). É recomendado o uso de amostras com medidas próximas do nível de decisão ou ponto de corte, e também aquelas que apresentam maior intensidade de reação (Banoo *et al.*, 2010; OIE..., 2010).

2.1.4.4 Reprodutibilidade

A reprodutibilidade é a capacidade de um método proporcionar resultados consistentes, tal como determinado pelas estimativas de precisão, quando aplicado a alíquotas de uma mesma amostra testadas em diferentes laboratórios. Expressa a precisão entre laboratórios (Validation..., 1999).

Para os especialistas em avaliação de diagnósticos (Banoo *et al.*, 2010) a reprodutibilidade de um teste é uma medida do grau de concordância entre os seus resultados quando as condições para testar ou medir se alteram. Por exemplo, a reprodutibilidade pode ser medida comparando-se os resultados de distintos operadores, de diferentes locais de teste, utilizando diferentes instrumentos, entre diferentes lotes de reagentes preparados, ou os resultados obtidos em dias diferentes. Esses especialistas consideram o teste estatístico Kappa uma medida útil de concordância entre os tipos de teste ou lotes, e entre os usuários. Quando os resultados dos testes são dicotômicos, ou seja, positivo ou negativo, reprodutibilidade é normalmente avaliada com base nos

aspectos que dependem do operador. Essa avaliação é considerada especialmente importante para os ensaios nos quais a interpretação dos resultados é subjetiva, como por exemplo, leituras visuais realizadas nos testes de imunodifusão em gel de ágar.

Fatores limitantes para os estudos de validação multilaboratório incluem custos elevados, a falta de especialistas de laboratórios disponíveis e dispostos a participar em tais estudos, e restrições de tempo em geral (Validation..., 1997).

2.1.4.5 Robustez

Segundo Wieggers (2003) *Robustness* destina-se a ser uma medida da capacidade de um método permanecer inalterado por pequenas variações analisadas nos parâmetros do método, funções como uma medida de confiabilidade de um método, durante a utilização rotineira. *Robustness* também é referido como a capacidade de um ensaio permanecer inalterado em situações que podem ocorrer ao longo do teste em um único laboratório (OIE..., 2013). *Rudgedness* foi conceituado como a capacidade de um ensaio permanecer inalterado quando pequenas variações nas condições de funcionamento ou de ambiente são feitos (Wieggers, 2003).

Consistentemente OIE (OIE..., 2010) e VICH (Validation..., 1999) definiram robustez como uma medida da capacidade do ensaio permanecer não afetado pelas variações dos parâmetros do método, indicando a sua confiabilidade durante o uso normal. Todos os reagentes críticos são identificados e submetidos a uma avaliação fatorial que compara todas as possíveis combinações de reagentes. Por exemplo, em detecção de anticorpos por ELISA, os fatores podem incluir a concentração de antígeno ligado à fase sólida, diluição do conjugado e vários testes de soros que representam o intervalo de operação do ensaio. A resposta do ensaio com relação a

essas pequenas mudanças não pode resultar em variabilidade inaceitável.

Wieggers (2003) relata que algumas referências igualam *Ruggedness* com *Robustness*. Neste caso, segundo essa autora, *Robustness* pode ser visto como uma medida de variabilidade não aleatória (isto é, quando se conhece as alterações) e *Ruggedness*, de variabilidade aleatória.

Burns *et al.* (2009) publicaram uma discussão sobre as características associadas a *Robustness e Ruggedness*, usadas nas descrições de procedimentos analíticos. Os autores identificaram a existência de uma considerável confusão na literatura no que diz respeito à utilização dos termos robustos e ‘resistentes’ e características associadas da robustez e ‘resistência’. Muitos autores usam os dois termos e suas características associadas, como se fossem sinônimos. Outros usam apenas um termo e / ou característica, ou seja, robusto / robustez. Além disso, ainda segundo Burns *et al.* (2009), o uso do termo robusto em ligação com os testes estatísticos de dados é suportado por muitos químicos e estatísticos.

Os autores (Burns *et al.*, 2009) recomendaram as seguintes definições para a *Robustness e Ruggedness*. Robustez de um método analítico é a propriedade que indica a insensibilidade contra mudanças de parâmetros operacionais sobre os resultados do método e, conseqüentemente, a sua adequabilidade para a sua função definida. Esta é apurada durante os processos de desenvolvimento de método e determina os limites aceitáveis para todos os parâmetros críticos que afetam os valores medidos. Além disso, fornece informações sobre a praticabilidade de um método, ou seja, traz conhecimento a respeito da simplicidade na operação, em termos de amostra e custo, para atingir o critério de desempenho.

2.1.4.6 Ponto de corte

Segundo a OIE (OIE..., 2010) o ponto de corte é o valor selecionado para distinguir entre resultados positivos e negativos em uma escala contínua de valores e deve refletir a finalidade do ensaio e sua aplicação. É apontado também como limiar de decisão. Intermediário ou inconclusivo são termos usados como sinônimos para uma zona de valores entre aqueles correspondentes a resultados positivos e negativos.

De acordo com o documento de exemplos trabalhados da Subcomissão de Laboratório Padrão em Saúde Animal (Measurement..., 2010), limites de decisão são necessários para a interpretação dos valores dos testes laboratoriais e podem ser pontos de corte ou intervalos de referência. Pelo menos um tipo de limite de decisão é necessário para responder a perguntas. Pontos de corte são usados para diferenciar doença de saúde ou para diferenciar doenças apresentando problemas semelhantes. A sensibilidade e especificidade do diagnóstico podem ser determinadas em pontos de corte diferentes e a curva ROC (*Receiver Operating Characteristics*) pode ser usada para identificar pontos de corte ideais para diferentes necessidades clínicas (Measurement..., 2010). Além da curva ROC outras formas para se determinar o ponto de corte são utilizadas: análise da distribuição das amostras com resultados positivos e negativos, e a tradicional média mais ou menos dois ou três desvios padrão dos resultados das amostras negativas.

2.1.4.7 Sensibilidade e especificidade diagnóstica

Segundo a OIE (OIE..., 2010) sensibilidade diagnóstica é a proporção das amostras conhecidas de animais referência infectados com resultado positivo num ensaio; enquanto que a especificidade diagnóstica é a proporção das amostras conhecidas de animais referência não infectados com resultado negativo num ensaio.

Uma medição é válida se é apropriada para a questão que se está pesquisando ou se mede corretamente o que se propõe. A medida dos verdadeiros positivos e negativos de um teste de diagnóstico reflete a sua sensibilidade e especificidade, respectivamente. No estudo da sensibilidade e especificidade diagnóstica comumente é construída uma tabela de contingência para avaliar o desempenho do teste (Tabela 1). Essa tabela também facilita o entendimento do conceito de sensibilidade e especificidade, mostrando os resultados dos testes referência e teste pesquisado quando avaliado as amostras dos animais (Jacobson, 1998).

A acurácia do teste relata a sua habilidade de reportar a verdadeira medida do que está sendo testado. OIE (OIE..., 2010) define a acurácia como a proximidade de um valor de teste para o esperado, verdadeiro, ou valor de referência com concentração ou título conhecido. O estudo da acurácia envolve corrida do teste com amostras apresentando quantidade conhecida do que está sendo medido, utilizando, por exemplo, amostras que representem animais conhecidamente infectados no caso de testes com resultados positivos e negativos.

Para avaliar um novo teste precisamos de um padrão- ouro que é a designação de um método padrão, utilizado em estudo comparativo com o novo teste. Um método padrão é aquele reconhecido nacionalmente e internacionalmente como melhor, ou na sua ausência o melhor método disponível preferencialmente publicado em revistas científicas de elevado conceito. Após a submissão das amostras utilizando o teste em estudo e o método referência o desempenho diagnóstico ou validade é avaliado (OIE..., 2010).

Observando a Tabela 1 podemos dizer que a sensibilidade é a probabilidade de um teste ser positivo em um infectado ou doente, isto é, traduz a percentagem de infectados ou doentes corretamente diagnosticados por um teste positivo. A sensibilidade corresponde à proporção de verdadeiros positivos do teste: **$a/(a+c)$** .

Já a especificidade é a probabilidade de um teste ser negativo em um não infectado ou não doente, isto é, traduz a percentagem de não doentes corretamente identificados por um teste negativo. A especificidade corresponde à proporção de verdadeiros negativos do teste: **$d/(b+d)$** .

Tabela 1- Tabela de contingência para avaliar o desempenho do teste

Teste sob avaliação	Teste referência		Total
	Positivo	Negativo	
Positivo	a	b	a+b
Negativo	c	d	c+d
Total	a+c	b+d	

Sensibilidade = $a/(a+c)$; especificidade = $d/(b+d)$

VPP = $a/(a+b)$; VPN = $d/(c+d)$

a= verdadeiro positivo;

b= falso positivo;

c= falso negativo;

d= verdadeiro negativo

2.1.4.8 Valor preditivo positivo e negativo

A probabilidade de infecção ou doença, dados os resultados de um teste diagnóstico é denominada valor preditivo do teste. No caso de se fazer uma triagem na população que proporção de indivíduos infectados ou doentes seria corretamente identificada? Isso tem importância evidente em saúde animal e em saúde pública. O valor preditivo pode ser dividido em positivo e negativo. Seu valor é determinado pela sensibilidade, especificidade e pela prevalência da doença na população testada. Quanto mais sensível o teste, melhor seu valor preditivo negativo. Quanto mais específico o teste, melhor seu valor preditivo positivo (Pagano, 2004; Florkowski, 2008).

Então, além da sensibilidade e da especificidade, o desempenho de um teste de classificação binária pode ser medido com seu valor preditivo positivo (VPP) e com seu valor preditivo negativo (VPN). O valor preditivo positivo responde a questão: - Se o resultado do teste é positivo, como prever uma presença real de, por exemplo, uma doença?, e pode ser calculado da seguinte forma: verdadeiros positivos/(verdadeiros positivos + falsos positivos). Ou seja, o valor preditivo positivo, é dado pela proporção de verdadeiros positivos a partir de todos os resultados positivos do ensaio. Utilizando os resultados negativos do ensaio, o mesmo se aplica para o cálculo do valor de predição negativo (Pagano, 2004; Florkowski, 2008).

É essencial notar uma importante diferença entre os dois conceitos. Isto é, a sensibilidade e especificidade são independentes da população, no sentido de que eles não alteram dependendo da proporção de positivos e negativos testados. Na verdade, podemos determinar a sensibilidade do teste, testando apenas casos positivos. Já, os valores de previsão são dependentes da população (Guidelines..., 2012).

Na Tabela 1 pode-se identificar o Valor Preditivo Positivo (VPP) – probabilidade de um indivíduo ter a doença quando o teste é positivo: $a/(a+b)$. E o Valor Preditivo Negativo (VPN) – probabilidade de um indivíduo não ter a doença quando o teste é negativo: $d/(c+d)$.

2.1.5 Estágios da validação

Segundo Greiner e Gardner (2000), a validação é um processo contínuo e a análise do método deve ser planejada de acordo com a estrutura dos dados. São recomendados por Jacobson (1998) cinco diferentes estágios de validação. O método de validação proposto (Jacobson, 1998) considerou a validação de ensaios como processos inter-relacionados (Figura 1). Os processos foram divididos em duas partes (I e II).

Na primeira parte, foi estabelecido parâmetros e características do ensaio através de 3 estágios. O estágio 1 consistiu em estudos para determinar se os reagentes selecionados e o protocolo, possuem a capacidade para distinguir entre uma faixa de concentração de anticorpo, para o agente infeccioso, provendo o mínimo de atividade de “background”. Para os estudos de viabilidade foi descrito a necessidade de amostras que devem representar animais conhecidamente infectados e não-infectados originados da população que eventualmente será alvo do ensaio validado.

No estágio 2, citado como desenvolvimento e padronização, foram evidenciados a determinação ótima dos reagentes e os parâmetros do protocolo. Nessa etapa, foi fornecida uma estimativa inicial da repetibilidade. Ainda no estágio 2, a determinação da sensibilidade e especificidade analíticas foram recomendadas. Essas foram conceituadas respectivamente, como o mínimo detectável do analito em questão, e a avaliação para que o ensaio não apresente reação cruzada com outro analito.

A caracterização do desempenho do ensaio foi relatada no estágio 3. Para essa caracterização foram descritas as estimativas da sensibilidade e especificidade diagnósticas.

A parte II compreendeu o estágio de monitoramento do desempenho do ensaio para validade (estágio 4) e a manutenção e extensão dos critérios de validação (estágio 5). No caso do monitoramento da precisão, uma vez que o ensaio foi estabelecido na rotina, foi apresentado o uso do controle interno da qualidade, tendo os resultados plotados em um gráfico, o de Levey-Jennings (Westgard, 2002), para acessar a repetibilidade e a acurácia do ensaio. Foi recomendado também a avaliação da reprodutibilidade, que segundo o autor, deve ser verificada pelo menos 2 vezes a cada ano (Jacobson, 1998).

O teste de proficiência foi apresentado como forma de controle externo da qualidade. Usualmente administrado por um laboratório de referência que distribui amostras, recebe os resultados dos laboratórios, analisa os dados e reporta os resultados para os laboratórios (Jacobson, 1998).

No estágio 5, considerando a extensão das variáveis que tem impacto no desempenho do ensaio, foi mostrado vantajoso a expansão do número de amostras referências, quando possível, devido ao princípio que, o erro é reduzido com o aumento do tamanho da amostra. Essa expansão também deve ser utilizada para atualizar as estimativas das sensibilidades e especificidades diagnósticas para a população alvo do ensaio. Além disso, essa expansão foi considerada essencial para revalidar o ensaio para quando for transferido para região geográfica diferente.

O resumo da informação essencial necessária para avaliar o estado de validação

de um ensaio sorológico segundo a SCAHLS (SCAHLs..., 2012) apresentou quatro níveis de validação. O primeiro nível consiste no desenvolvimento e viabilidade; o segundo é o desempenho do diagnóstico e implementação; o terceiro é a transferência de tecnologia; e o quarto a reprodutibilidade. Os níveis de validação representaram uma progressão de desenvolvimento através da aplicação. Nos níveis são incluídos controle e limites de aceitabilidade.

Para a OECD (Guidance..., 2005), a decisão sobre se uma avaliação retrospectiva do estado de validação é adequada ou se estudos de validação prospectivos são necessários depende da existência ou não de informação confiável para substanciar a validade de um método. A OECD relatou que na prática, é necessário muitas vezes adotar uma combinação de abordagens, retrospectiva e prospectiva, a fim de gerar dados e informações adequadas.

Propõe um quadro conceitual geral para documentar os estudos necessários para avaliar o estado de validação de um método, chamado de "abordagem modular" para a validação. Nessa abordagem, necessária para apoiar a validade de um método de ensaio, os módulos estão organizados para fornecer as seguintes informações: a definição de teste; repetibilidade intralaboratorial e reprodutibilidade; interlaboratorial de transferência; a reprodutibilidade Interlaboratorial e; a acurácia, com domínio de aplicabilidade e padrões de desempenho.

De acordo com esta proposta cada um desses "módulos" de informação deve estar disponível, a fim de avaliar o estado de validação do método de ensaio para uma finalidade específica, mas eles não precisam ser concluídos na forma sequencial.

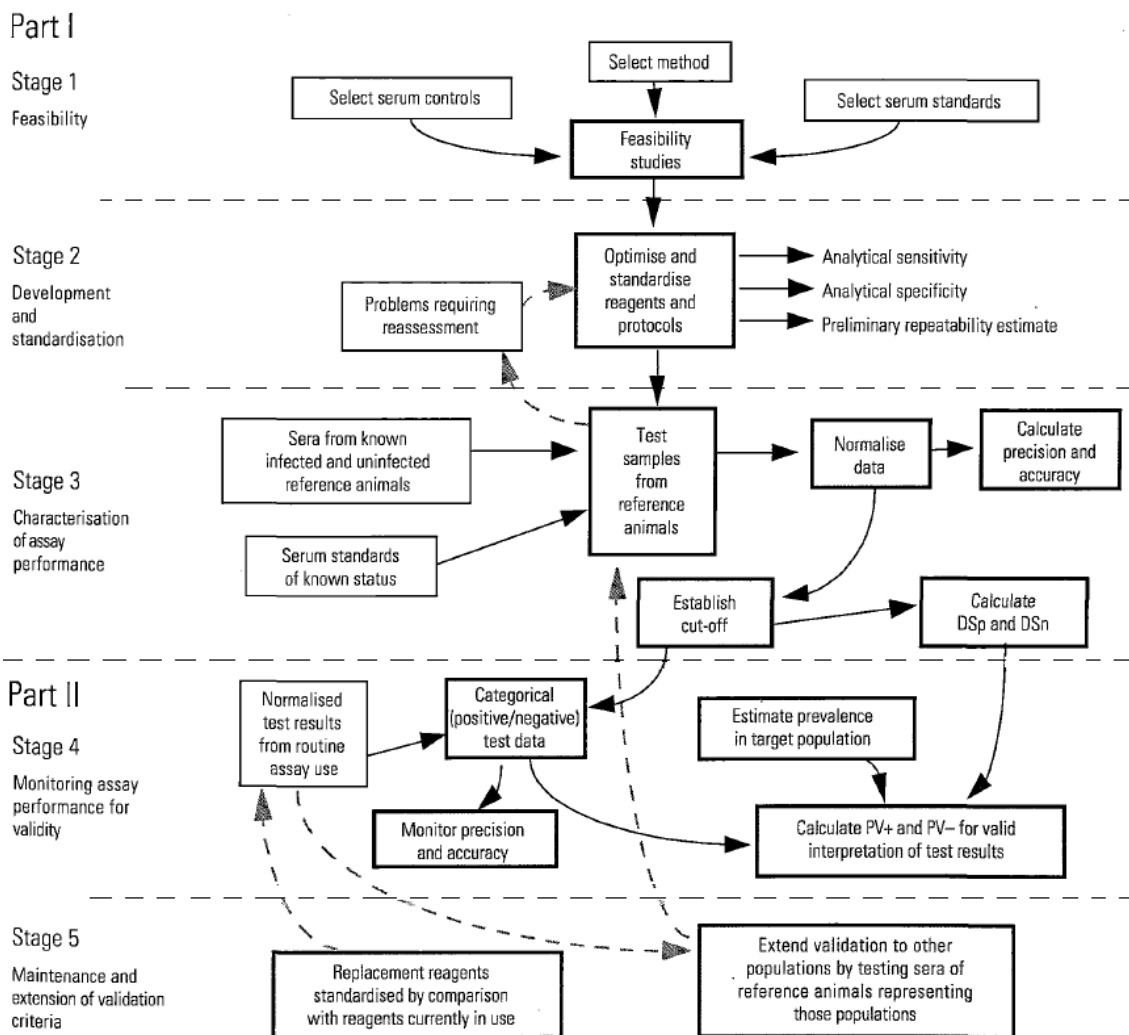


Figura 1. Os cinco estágios no processo de validação do ensaio diagnóstico segundo Jacobson (1998). Na primeira parte, são estabelecidos parâmetros e características do ensaio através da análise de viabilidade. Cinco estágios no processo de validação do ensaio (Jacobson, 1998), desenvolvimento e padronização, e caracterização do ensaio. Na segunda parte, com o objetivo de assegurar a validade constante do teste, é contemplado o monitoramento do seu desempenho e a manutenção e extensão dos critérios de validação. No decorrer dos níveis de avaliação do ensaio, o autor indica pontos de ação, dentro de cada estágio do processo, subsidiando a análise crítica e permitindo a reavaliação do ensaio diagnóstico.

O documento da SCAHLS reúne propostas de diretrizes a serem seguidas para reunir os dados essenciais utilizados para a avaliação de um novo teste sorológico, kit de ensaio sorológico ou reagente de ensaio essencial sorológico.

2.2 TESTES ESTATÍSTICOS

2.2.1 Aspectos gerais

Os cientistas fazem muito uso da estatística em busca de inferências. Segundo Deming e Mahalanobis (1972), quanto melhor entendermos as limitações de uma inferência sobre um conjunto de resultados, mais útil se torna a inferência.

Conhecimentos suficientes de estatística devem estar disponíveis para garantir o planejamento adequado de estudos de validação e avaliação dos dados resultantes. O responsável pela análise estatística deve estar familiarizado com a base biológica e com as limitações práticas do teste proposto e do teste de referência na espécie de interesse (Guidance..., 2005).

Este conhecimento irá ajudar na seleção de procedimentos estatísticos adequados, bem como no desenvolvimento de critérios de decisão apropriados, e pode auxiliar na publicação dos resultados do estudo. O bioestatístico deve estar envolvido no desenho de todas as fases dos estudos de validação e no estabelecimento de procedimentos adequados para a manutenção de registros, coleta de dados e envio de dados (Guidance..., 2005).

Métodos estatísticos a serem utilizados devem ser definidos antes do início dos estudos e devem apoiar o propósito do estudo. No entanto, detalhes dos dados, como por exemplo, sua variação analítica e a presença de *outliers*, nem sempre podem ser previstos. Nesses casos, pode ser necessário rever os procedimentos estatísticos, identificar ou desenvolver novos

procedimentos ou modelos após avaliação dos dados do teste (Guidance..., 2005).

Em vários estágios no processo de validação, avaliações estatísticas devem ser feitas para verificar se os resultados, de um novo teste ou de um teste em melhoria, estão de acordo com os dados de referência utilizando os critérios de decisão. Além disso, as condições do teste devem ser avaliadas estatisticamente para fornecer recomendações para melhorar a eficiência do protocolo. Uma análise deste tipo pode também servir para reduzir e otimizar a utilização de animais, por exemplo, mostrando a potência do teste como uma função do número de animais por grupo em relação à reprodutibilidade / variabilidade e à resposta do ponto de extremidade que está sendo medido (Guidance..., 2005).

Para aplicar a estatística na avaliação dos métodos é importante estar claro o que se quer medir para que possa ser feita a melhor escolha do teste estatístico que será utilizado. Os testes estatísticos paramétricos são utilizados no estudo de populações cujos dados amostrados são assumidos como normalmente distribuídos ou com distribuição normal aproximada. Se os dados não satisfazem às essas suposições, métodos estatísticos não-paramétricos devem ser usados.

No procedimento operacional padrão (POP) para validação e certificação de ensaios diagnósticos (Validation..., 2009) é requerido que os dados incluam: o desvio padrão e o coeficiente de variação para análise de repetibilidade e reprodutibilidade; a média e desvio padrão ou curva ROC para determinação do ponto de corte; a tabela 2x2 para estimativas de sensibilidade e especificidade com animais de referência definidos; a inferência Bayesiana para estimativas de sensibilidade e especificidade sem animais de referência definidos; e testes de concordância para comparação de ensaios.

2.2.2 Desvio padrão e coeficiente de variação

A média aritmética é a medida de tendência central mais frequentemente usada. É calculada mediante a soma de todas as observações de um conjunto de dados dividida pelo número total de medidas. A média pode ser usada como uma medida resumo tanto para as medidas discretas quanto para as contínuas. Medidas discretas são dados restritos a ter somente valores específicos frequentemente inteiros ou contagens que diferem por quantidades fixadas; nenhum valor intermediário é possível. Exemplo de medida discreta poderia ser, o número de equídeos infectados pelo vírus da anemia infecciosa equina em um determinado período do ano. Dados contínuos são aqueles que representam quantidades mensuráveis, mas que não estão restritos a assumir certos valores especificados (tais como inteiros). O único fator que limita uma observação contínua é o grau de precisão com o qual pode ser medida.

Quando temos dados dicotômicos e os resultados possíveis são representados pelos valores 0 e 1 também podemos utilizar a média. Nessa situação, a média das observações é igual à proporção de número 1 no conjunto de dados (Pagano, 2004).

Independentemente da medida de tendência central usada em uma situação específica, pode ser equivocado assumir que esse valor seja representativo de todas as observações do grupo. É útil saber onde o centro de um conjunto de dados se encontra. No entanto, usualmente essa informação não é suficiente para caracterizar uma distribuição inteira de medidas. As observações podem ser próximas do centro ou dispersas por um amplo intervalo de valores.

Medidas de dispersão comumente utilizadas em avaliação de testes diagnósticos são variância, desvio-padrão e coeficiente de

variação (CV). Essas medidas são concordantemente recomendadas na avaliação da imprecisão do teste.

A variância quantifica a variabilidade ao redor da média das medidas. A variância é calculada ao se subtrair a média de um conjunto de valores de cada uma das observações, elevar ao quadrado esses desvios, somá-los e dividir a soma pelo número de observações do conjunto de dados menos 1. Representando a variância por s^2 , temos que:

$$s^2 = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2$$

O desvio-padrão de um conjunto de dados é o módulo da raiz quadrada da variância: $s = \sqrt{s^2}$

Na prática, o desvio padrão é usado mais frequentemente do que a variância. A razão disso é que o desvio-padrão tem a mesma unidade de medida que a média, em vez da unidade elevada ao quadrado. Em uma comparação de dois grupos de dados, o grupo com menor desvio-padrão tem as observações mais homogêneas; o grupo com maior desvio padrão exibe maior variabilidade. Como o desvio-padrão tem unidade de medida, não tem sentido comparar desvios-padrão para duas quantidades não relacionadas. A variabilidade entre dois ou mais conjuntos de dados que representam quantidades variadas com diferentes unidades de medida, pode ser comparada utilizando o coeficiente de variação (Pagano, 2004).

O coeficiente de variação relaciona o desvio-padrão de um conjunto de valores à sua média; ele é o desvio-padrão dividido pela média e multiplicado por 100 e é, portanto, uma medida de variabilidade relativa. Como o desvio padrão e a média partilham as mesmas unidades de medida, as

unidades cancelam-se e deixam o coeficiente de variação adimensional (Pagano, 2004).

O desempenho de um teste de diagnóstico é definido por duas medidas independentes: precisão e acurácia. Cada medição feita tem um erro associado à ela. Uma vez que o processo de medição não é totalmente reproduzível, não existe um valor exato, que pode ser associado com o que está sendo medido. Assim, o resultado é expresso de forma mais precisa como uma estimativa, juntamente com um nível de imprecisão associada. Essa imprecisão é a incerteza de medição (Measurement..., 2010).

Incerteza de medição é propriedade de um resultado de medição. É definida como um parâmetro associado com o resultado de uma medição, que caracteriza a dispersão dos valores os quais podem ser razoavelmente atribuídos ao mensurando (International..., 2008). O conhecimento da incerteza de medição é necessário para a comparação das medições e para a comparação eficaz das medições com limites de especificação. Portanto, a estimativa de incerteza de medição pode ser considerada um elemento essencial de validação do método (Guidelines..., 2012).

Para o SCHALS (Measurement..., 2010), a incerteza de medições fornece estimativas quantitativas do nível de confiança que um laboratório tem na precisão analítica dos resultados dos testes, e é, portanto, componente essencial de um sistema de qualidade para laboratórios de ensaio de diagnóstico veterinário.

2.2.3 Coeficiente de correlação

Correlação é definida como a quantificação do grau em que duas variáveis aleatórias estão relacionadas, desde que a relação seja linear. É aplicada para investigar as relações que podem existir entre variáveis contínuas (Pagano, 2004).

Antes de conduzir qualquer tipo de análise de correlação devemos sempre criar um gráfico de dispersão bidimensional. Se colocarmos os resultados da variável “x” ao longo do eixo horizontal e os resultados da variável “y” ao longo do eixo vertical, cada ponto no gráfico representa uma combinação de valores (x,y). Frequentemente podemos determinar se existe uma relação entre x e y (Pagano, 2004).

A correlação quantifica o grau da relação linear entre os resultados x e y. O coeficiente de correlação (r) é um número adimensional, não tem unidade de medida. O valor máximo que r pode assumir é 1 e seu valor mínimo é -1. Portanto, para qualquer conjunto de observações, $-1 \leq r \leq 1$. Os valores $r=1$ e $r=-1$ ocorrem quando há uma exata relação linear entre x e y; se representássemos todos os pares de resultados (x,y), os pontos se encontrariam sobre uma linha reta (Pagano, 2004).

Tal como acontece com um coeficiente de correlação de Pearson, o coeficiente de correlação de concordância (CCC) pode ser utilizado para comparar dois conjuntos de resultados e reflete melhor o nível de concordância entre esses dois conjuntos do que a Coeficiente de correlação de Pearson. Se houver dois conjuntos de resultados de ensaios de escala contínua concordando perfeitamente, uma parcela de um conjunto contra o outro iria produzir uma linha reta em um ângulo de 45° (a linha de igualdade). O CCC é calculado a partir de três parâmetros. A localização do parâmetro, mede o quanto ele está acima ou abaixo da linha de igualdade. O parâmetro de mudança de escala, mede a diferença entre a inclinação para os dados da amostra e a linha de igualdade. O produto dos parâmetros de localização e inclinação é referido como parâmetro da precisão. O CCC é indicado para trabalhar com dois conjuntos de resultados de ensaios e trabalhar com dados categóricos (Dohoo, 2009).

2.2.4 Teste qui-quadrado

Ao se trabalhar com dados nominais agrupados em categorias, frequentemente dispõe as contagens em um formato tabular conhecido como tabela de contingência. No caso mais simples, envolvidas duas variáveis aleatórias dicotômicas; as linhas da tabela representam os resultados de uma variável, e as colunas, os resultados da outra. As entradas da tabela são as contagens que correspondem a uma combinação particular de categorias (Pagano, 2004).

O teste qui-quadrado compara as frequências observadas em cada categoria da tabela de contingência com as frequências esperadas, sendo a hipótese nula verdadeira. É usado para determinar se os desvios entre as contagens observadas e esperadas são muito grandes para serem atribuídas ao acaso (Pagano, 2004).

A distribuição qui-quadrado não é simétrica. Uma variável aleatória qui-quadrado não pode ser negativa, pois assume valores de zero até o infinito e é assimétrica à direita. Há uma distribuição qui-quadrado diferente para cada valor possível de graus de liberdade (Pagano, 2004).

Alonzo e Pepe (1999) usando uma combinação de testes de referência para verificar a acurácia de um novo teste diagnóstico, resumiram os dados de comparação em tabelas de contingência e calcularam a sensibilidade e a especificidade dos testes.

2.2.5 Teste Kappa

Especialistas fazendo uma leitura de resultado de um teste diagnóstico qualitativo às vezes concordam simplesmente por acaso. A estatística kappa corrige esta chance de concordância e informa o quanto de concordância além do acaso os leitores conseguiram.

A estatística kappa nos permite medir a concordância além do esperado por acaso. Na tabela 2 podemos verificar a interpretação comumente utilizada para o teste Kappa (McGinn *et al.*, 2004; Dohoo, 2009).

Em um exemplo citado por McGinn *et al.* (2004) adaptado, apresentado na figura 2, a barra horizontal superior representa 100% de concordância entre dois observadores. Para a situação hipotética representada na figura 2, a probabilidade estimada da concordância entre os dois observadores é de 50%. Isto pode ocorrer se, por exemplo, cada um dos dois observadores aleatoriamente encontrar metade das avaliações positivas. Dada esta informação, o que é a concordância possível para além do acaso? A linha vertical na figura 2 cruza as barras horizontais no ponto de 50%, que foi identificada como a concordância esperada por acaso. Todas as concordâncias para a direita desta linha correspondem à concordância além do acaso.

Tabela 2- Classificação qualitativa de valores de Kappa como grau de concordância além do acaso

Valor de Kappa	Grau de concordância além do acaso
0	Nenhum
0-0,2	Leve
0,2-0,4	Razoável
0,4-0,6	Moderada
0,6-0,8	Substancial
0,8-1,0	Quase perfeita

Fonte: McGinn *et al.* (2004)

Por isso, o máximo de concordância além do acaso é de 50% (100% - 50%). O outro dado necessário para calcular a pontuação kappa é o grau de concordância além do acaso. A concordância observada, como mostrado pela barra horizontal inferior na figura 2, é de 75%, de modo que o grau de

concordância além do acaso é de 25% (75% - 50%).

Kappa é calculado como a concordância observada além do acaso (25%) dividido pelo nível máximo de concordância além do acaso (50%); então, kappa é de 0,50.

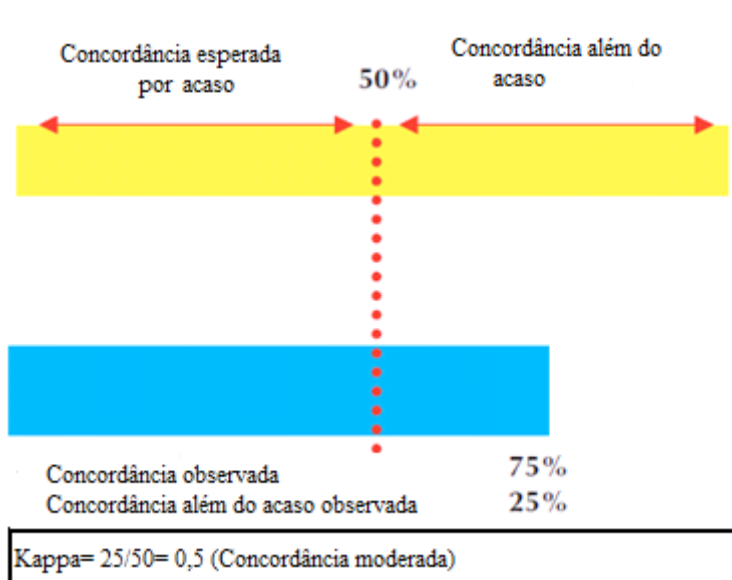


Figura 2. Dois observadores independentes avaliam um resultado. Cada observador determina que a constatação está presente em exatamente 50% dos sujeitos. Suas avaliações concordam em 75% dos casos. A barra horizontal amarela representa a concordância potencial (100%), e a barra azul representa concordância real. A porção de cada barra que se encontra à esquerda da linha vertical pontilhada representa a concordância esperada por acaso (50%). A concordância observada além do acaso é metade da concordância por acaso. A razão entre estes dois números é a pontuação kappa (McGinn *et al.*, 2004).

2.2.6 Curva ROC

A curva ROC (*Receiver Operating Characteristics*) é uma descrição gráfica do desempenho de um teste representando a relação entre a fração dos verdadeiros positivos (sensibilidade) e os falsos positivos (1- especificidade). Normalmente, a proporção de verdadeiros positivos é plotada no eixo vertical e o número proporcional de falso positivo é plotado no eixo horizontal (Assessment..., 1995). A análise da curva ROC quantifica a acurácia de um teste diagnóstico. Outra modalidade de avaliação utilizando a curva ROC é a discriminação de

dois estados ou condições que são referidas como doente e não doente. A acurácia discriminatória de um teste diagnóstico é mensurada pela habilidade da classificação correta do animal como infectado ou não infectado (Gardner e Greiner, 2006).

Segundo Gardner e Greiner (2006) a análise da curva ROC tem sido largamente usada em estudos na patologia clínica veterinária embora deva ser considerada um complemento eficiente para estimar a sensibilidade e especificidade em estudos de avaliação de testes diagnósticos. Os autores observaram que comparada com a medicina

humana, a medicina veterinária tem adotado essa técnica lentamente.

O uso de uma curva ROC tem vantagem sobre o uso de um valor de ponto de corte para determinar sensibilidade e especificidade em que se descreve a capacidade total do ensaio para distinguir animais doentes de não-doentes devido a dispor de um alcance maior de pontos de corte (Dohoo, 2009).

A área sob a curva ROC (AUC) pode ser interpretada como a probabilidade de que um indivíduo positivo, selecionado

aleatoriamente, apresente um valor de teste (densidade óptica), maior que a de um indivíduo negativo selecionado aleatoriamente, assumindo que a distribuição estatística do teste no grupo positivo é maior que no grupo negativo (Pepe, 2003).

O desempenho global do teste diagnóstico é comumente sumarizado pela área sob a curva ROC (Figura 3). Essa área expressa a probabilidade do resultado do teste diagnóstico. Quanto maior a área sob a curva ROC melhor é o desempenho global do teste diagnóstico.

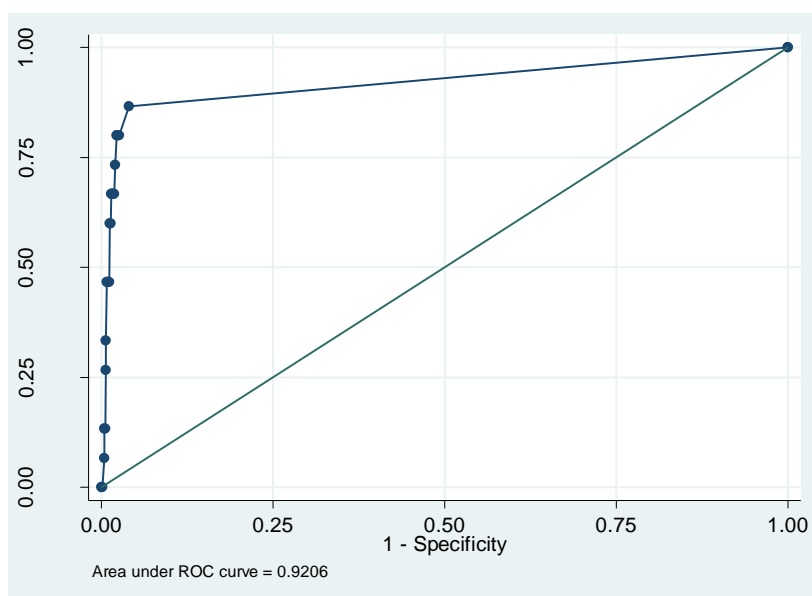


Figura 3- Curva ROC

A faixa de valores para área (AUC) varia entre 0,1 e 1. Um teste perfeito tem uma área de 1 e testes inadequados apresentam área de 0,5. *Guidelines* tem sugerido interpretações para valores intermediários para área: baixo ($0,5 < AUC \leq 0,7$), moderada ($0,7 < AUC \leq 0,9$) e acurácia elevada ($0,9 < AUC \leq 1$) (Gardner e Greiner, 2006). Então, considerando essa interpretação o teste representado na figura 3 apresenta acurácia elevada.

Além da análise da área, a curva ROC mostra graficamente o equilíbrio entre sensibilidade e especificidade permitindo a análise e escolha, atribuindo o melhor ponto de corte para o propósito de uso (Dohoo, 2009). Como exemplo, observa-se na tabela 3 cada ponto de corte com o respectivo equilíbrio entre sensibilidade e especificidade e o valor corretamente classificados.

Tabela 3. Equilíbrio entre Sensibilidade e Especificidade

Ponto de corte	Sensibilidade	Especificidade	Corretamente classificados
(≥ 0)	100,00%	0,00%	1,64%
($\geq ,211$)	80,00%	96,89%	96,62%
($\geq ,222$)	80,00%	97,45%	97,16%
($\geq ,233$)	66,67%	98,00%	97,49%
($\geq ,244$)	60,00%	98,56%	97,93%
($\geq ,255$)	26,67%	99,11%	97,93%
($\geq ,266$)	0,00%	99,67%	98,03%
(> ,266)	0,00%	100,00%	98,36%

2.2.7 Modelo Bayesiano

A base para a inferência estatística é a teoria da probabilidade. Evento é o elemento básico para o qual a probabilidade pode ser aplicada. No estudo da probabilidade, os eventos são representados por letras maiúsculas, tais como A, B e C.

Na definição frequentista, a probabilidade de um evento A é sua frequência relativa de ocorrência, ou a proporção de vezes que o evento ocorre, em um número grande de repetidas tentativas, sob condições virtualmente idênticas.

Para Paulino (2003) um aspecto importante da inferência estatística clássica consiste em reconhecer a variabilidade que se verifica de amostra para amostra. Para estabelecer inferências, considera indispensável ter em mente que os dados observados formam apenas um dos muitos, possivelmente infinitos, conjuntos que poderiam ter sido obtidos nas mesmas circunstâncias. Segundo o princípio da amostragem repetida, o comportamento dos métodos estatísticos deve ser apreciado em um número indefinido de repetições efetuadas nas mesmas condições.

Uma das características das probabilidades subjetivas é poder aplicar-se a situações não repetitivas. Para os bayesianos cada problema é único e tem um contexto real próprio onde o parâmetro θ é uma

quantidade significativa acerca da qual existem, em geral, graus de conhecimento que variam de problema para problema. A filosofia bayesiana é que, no caso em questão, o parâmetro θ é incerto e toda a incerteza deve ser quantificada em termos de probabilidade. A quantidade desconhecida é estimada utilizando informações além da amostra em estudo, informações *a priori*, que por expressarem a incerteza do pesquisador sobre θ antes da observação dos dados são consideradas subjetivas. Essas informações podem ser originadas de teste piloto, dados históricos, a partir de opiniões de especialistas ou por conhecimento prévio. Exemplificando: Brascum *et al.*(2005) estudando a sensibilidade e especificidade de um teste ELISA e um teste de imunofluorescência indireta, para o diagnóstico de Neospora canina em gado, construíram a distribuição *a priori* para os parâmetros do modelo usando informações provenientes do estudo realizado por Frössling *et al.*(2003).

A estimativa de parâmetros em estudo pode ser obtida de modo iterativo através da combinação da distribuição dos dados e das distribuições *a priori* por meio do Teorema de Bayes e da utilização da técnica de simulação *Gibbs Sampling*.

Segundo Gardner (2002), o Teorema de Bayes pode ser formalmente escrito como:

$Pr(H|Dados) = Pr(Dados|H) * Pr(H) / Pr(Dados)$

A equação indica que a probabilidade de uma dada hipótese de dados é igual à probabilidade dos dados, dado que a hipótese original era correta, multiplicada pela probabilidade da hipótese, antes da obtenção dos dados, dividida pela probabilidade média dos dados.

A distribuição Beta é apropriada para o modelo de probabilidade binomial, com valores de dois parâmetros (α , β). As densidades *a priori* Beta (α_θ , β_θ) para cada parâmetro são: α_θ a combinação do centro do intervalo do parâmetro θ de interesse com a média da distribuição Beta; β_θ a combinação de $\frac{1}{4}$ do intervalo do parâmetro θ de interesse, com o desvio padrão da distribuição Beta.

Para Gardner (2002), a primeira vantagem prática do método bayesiano em relação ao método tradicional da estatística frequentista é permitir uma ótima flexibilidade em lidar com problemas complexos incluindo, aqueles que são em camadas múltiplas com muitos parâmetros, como as inferências sobre doenças de animais envolvendo países, rebanho dentro dos países, e animais dentro dos rebanhos.

Em estudos de testes diagnósticos o modelo bayesiano é recomendado e utilizado principalmente quando não se tem um padrão-ouro definido (Joseph *et al.*, 1995; Dendukuri, 2001; Brascum *et al.*, 2005).

2.3 QUALIDADE E METROLOGIA

2.3.1 Aspectos gerais

Para se qualificarem para o licenciamento e ficarem disponíveis para o comércio, os ensaios diagnósticos devem ter características de desempenho constantes (Morgan, 1998). As autoridades regulatórias como Animal and Plant Health Inspection Service (APHIS) nos Estados Unidos,

Canadian Food Inspection Agency (CFIA) no Canadá, Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e Agência Nacional de Vigilância Sanitária (ANVISA) no Brasil estabelecem os requisitos gerais para o licenciamento dos kits diagnósticos. Esses requisitos consistem do preparo dos reagentes, pureza dos produtos (Morgan, 1998), condições de medidas, critérios de validação de métodos e da manutenção desses critérios com monitoramento (OIE..., 2010).

Um teste validado necessita de monitoramento constante, o que pode ser realizado por meio de estudos de reprodutibilidade entre laboratórios. Nesse sentido, as boas práticas de laboratório, incluindo a implementação de programa de qualidade assegurada tornaram-se essenciais para os laboratórios que visam à certificação nacional e internacional (Jacobson, 1998). Muitos fatores influenciaram esta tendência, em particular o comércio nacional e internacional. Devido à necessidade de aprovação global, estudos têm sido feitos para estabelecer princípios baseados na ciência com o objetivo de promover um sistema transparente e seguro no comércio internacional (Wieggers, 2002).

No desenvolvimento de ensaios em laboratório ou na realização de análises de material clínico, o objetivo é a produção de dados de alta qualidade. Isso exige que os requisitos fundamentais sejam cumpridos. O estabelecimento de qualidade assegurada (QA) e do sistema de controle de qualidade (CQ) é essencial. Isso compreende um conjunto de protocolos de qualidade, incluindo a utilização de amostras de ensaio de controle que asseguram que o sistema está funcionando corretamente e confirmam a reprodutibilidade e a qualidade dos dados. O controle de qualidade e os sistemas de CQ, com pessoal treinado e competente, já foram estabelecidas em muitos laboratórios em todo o mundo (OIE..., 2010).

Farr e Freeman (2008) relataram que a definição de requisitos para vários ensaios de laboratório é um aspecto importante para o planejamento de qualidade veterinário como os requerimentos aplicados para laboratórios humanos, mas, não existem para laboratórios veterinários. Nesse mesmo estudo consideram que o planejamento da qualidade em laboratórios veterinários inclui definição de padrões de qualidade como base para processos de qualidade para laboratório, controle de qualidade, qualidade assegurada e melhoria da qualidade.

Para Gardner (2010) manuscritos de revisões de teste de acurácia para doenças dos animais avaliando a qualidade são raros e, estudos reportando a qualidade têm sido informais antes e depois do endosso do *Standards for Reporting of Diagnostic Accuracy (STARD)* em revistas veterinárias. Sua situação na medicina veterinária é, entretanto, provavelmente, semelhante àquela observada para a medicina em humanos.

2.3.2 Normas ISO

A *International Organization for Standardization (ISO)* desenvolve normas internacionais. O grupo foi fundado em 1947, e desde então publicou mais de 19.500 normas internacionais que abrangem quase todos os aspectos da tecnologia e de negócios. Essas normas apresentam as especificações para atingir o estado da arte de produtos, serviços e boas práticas. São desenvolvidas através de consenso global, e ajudam a remover barreiras ao comércio internacional (International..., 2013).

Uma norma ISO é desenvolvida por um grupo de especialistas pertencentes a uma comissão técnica. Diante da necessidade de criação de uma norma ISO, estes especialistas se reúnem para discutir e negociar seu projeto de norma. Uma vez desenvolvido, o projeto é compartilhado com os membros da ISO que são convidados a comentar e votar. Se o consenso for

alcançado o projeto se torna um padrão ISO. A norma é um documento que fornece requisitos, especificações, diretrizes ou características que podem ser usados de forma consistente para assegurar que materiais, produtos, processos e serviços são adequados para o seu propósito (International..., 2013).

A ISO e a Comissão Eletrotécnica Internacional (IEC) prepararam a ISO/IEC 17025. A primeira edição (1999) desta Norma foi produzida como resultado de extensa experiência da implementação da ISO / IEC Guia 25 e EN 45001. Essa edição continha todos os requisitos que os laboratórios de ensaio e calibração devem atender se desejam demonstrar que dispõem de um sistema de gestão, são tecnicamente competentes, e são capazes de produzir resultados tecnicamente válidos (Calibration..., 2005).

2.3.3 Metrologia

A metrologia legal no Brasil precede a Lei 5966 de 12 de dezembro de 1973 que criou o Sistema Nacional de Metrologia, Normalização e Qualidade Industrial, do qual o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) é o órgão executivo central. Assim, o INMETRO através da Diretoria de Metrologia Legal, observando a competência que lhe é atribuída pelas leis 5966/73, 9933/99, 10829/03 e pela Resolução 11, de 12 de outubro de 1988, do Conselho Nacional de Metrologia, Normalização e Qualidade Industrial, organiza e executa as atividades de metrologia legal no Brasil. Estão sujeitos à regulamentação e ao controle metrológico, os instrumentos de medição e medidas materializadas utilizados nas atividades econômicas e nas medições que interessem à isenção de perigo das pessoas nas áreas da saúde, da segurança e do meio ambiente (BRASIL, 2012).

O controle metrológico compreende: o controle dos instrumentos de medição relativo à verificação e inspeção; a supervisão metrológica constituída pelos procedimentos realizados na fabricação, na utilização, na manutenção e no conserto de um instrumento de medição, e a perícia metrológica, que é composta por um conjunto de operações que tem por fim examinar e certificar as condições em que se encontram um instrumento de medição e determinar suas qualidades metrológicas de acordo com as exigências regulamentares específicas (BRASIL, 2012).

Segundo a OIE (OIE..., 2010) a seleção de equipamentos é um dos fatores que impacta a validação de ensaios. Os equipamentos que não passam por manutenção e não são calibrados, podem impedir a realização de um ensaio de qualidade. Cita que os aparelhos como freezers, blocos de aquecimento, incubadoras, geladeiras, espectrofotômetros, termocicladores, pipetas, devem ser calibrados de acordo com os protocolos de garantia de qualidade do laboratório (OIE..., 2010).

O INMETRO estabeleceu os critérios de validação de ensaios através da Norma NIE-DIMEL-064 aprovada em janeiro de 2005. Nesse documento foram utilizados como referência a NBR ISO/IEC 17025:2001 que trata dos requisitos gerais para competência de laboratórios de ensaio e calibração e as Portarias INMETRO nºs 29 de 10/03/95 e 102 de 10/06/1988- Vocabulário internacional de Termos Fundamentais e Gerais de Metrologia e Vocabulário de Metrologia Legal (BRASIL, 2005).

Outra calibração necessária é a dos ensaios para padronizar os reagentes. Existem padrões de referência nacionais e internacionais, contendo uma concentração conhecida do analito. Esses padrões são preparados e distribuídos por laboratórios de referência. Os padrões de referência nacionais são calibrados por comparação

com os padrões internacionais sempre que possível. As amostras de referência utilizadas na validação devem estar na mesma matriz utilizada no ensaio e serem representantes da espécie a serem testadas. Além disso, os materiais de referência devem representar adequadamente a faixa de concentração do analito a ser detectado pelo ensaio (OIE..., 2010).

2.3.4 Acreditação, certificação e credenciamento laboratorial

A competência técnica e operacional do laboratório para conduzir testes é determinada pela acreditação. A acreditação é um procedimento pelo qual órgãos oficiais reconhecem formalmente que um órgão é competente para realizar tarefas específicas (Wieggers, 2002).

De acordo com o INMETRO (2013) a acreditação de laboratórios de ensaio é concedida por ensaio para um determinado produto, segundo uma norma, regulamento, resolução ou procedimento desenvolvido pelo laboratório em que é estabelecida a metodologia utilizada. A norma NIT-DICLA- 016 estabelece as diretrizes para a elaboração do escopo (BRASIL..., 2013c).

Para Wieggers (2002), uma discussão sobre a competência do laboratório deve começar com a descrição da cooperação para acreditação internacional de laboratório (ILAC) porque a harmonização de regimes deriva dessa cooperação. Formalizada como uma cooperação internacional em 1996, a ILAC a referência internacional para regimes de acreditação para uso da ISO/IEC padrão internacional 17025 - Requisitos gerais para competência de ensaios e calibração (Wieggers, 2002).

A acreditação de laboratórios, segundo os requisitos estabelecidos na norma ABNT NBR ISO/IEC 17025:2005, é aplicável a laboratórios de calibração e de ensaio. A acreditação de laboratórios de calibração é concedida para um escopo, constituído por

grupos de serviços de calibração estabelecidos na norma NIT-DICLA-012, incluindo serviços, faixas e melhor capacidade de medição (BRASIL..., 2013b).

A certificação é uma garantia por escrito, concedida por organismo de terceira parte, de um produto, processo ou serviço está em conformidade com o requisito especificado. Esse termo é usado internacionalmente para incluir o sistema da qualidade ou gerenciamento do sistema de certificação ou registro. Por exemplo, um sistema de qualidade pode ser certificado (registrado) como satisfazendo os requisitos da ISO 9001 (Quality..., 1998).

No credenciamento laboratorial o escopo de credenciamento é definido como o conjunto de um ou mais ensaios realizados por laboratório credenciado e publicados no sítio eletrônico do MAPA (BRASIL, 2013).

De acordo com o Art. 5º da INSTRUÇÃO NORMATIVA Nº 57 do Ministério da Agricultura, Pecuária e Abastecimento (BRASIL, 2013d), para a obtenção do credenciamento o laboratório interessado deverá apresentar certificação de acreditação e do escopo de acreditação na ABNT NBR ISO/IEC 17.025 - Requisitos Gerais para a Competência de Laboratórios de Ensaio e Calibração, emitidos pela Coordenação-Geral de Acreditação do Instituto Nacional de Metrologia, Normalização e Qualidade Industrial- CGCRE/Inmetro, contemplando os ensaios constantes na solicitação de credenciamento. Laboratório credenciado é definido pela IN 57/2013 como laboratório público ou privado, legalmente constituído como laboratório, homologado pelo MAPA para realizar ensaios e emitir resultados em atendimento aos programas e controles oficiais do MAPA (BRASIL, 2013d).

2.3.5 Qualidade e manutenção dos critérios de validação

2.3.5.1 Controle de qualidade interno

A validade do ensaio na rotina necessita ser consistentemente monitorada pela repetibilidade, por meio do processo de controle para avaliar possíveis alterações na precisão e acurácia do teste. Essas alterações podem ser monitoradas graficamente pela plotagem de valores de controle no gráfico de controle. Desvios advindos do desempenho do ensaio devem ser investigados e ações corretivas devem ser tomadas caso seja necessário (OIE..., 2010).

Segundo Westgard (2002), por comparação, um procedimento de regra única de controle utiliza um único critério ou um único par de limites de controle, assim como um gráfico de Levey-Jennings, com limites de controle calculados como $\bar{x} \pm 2DP$ (média mais ou menos dois desvios -padrão) ou $\bar{x} \pm 3DP$ (média mais ou menos 3 desvios-padrão). Westgard propõe outra forma de controle de qualidade que consiste na utilização de cinco regras de controle diferentes para a aceitabilidade de uma corrida analítica.

No Manual de Provas da OIE, versão 2013, o gerenciamento da qualidade em laboratório de ensaio veterinário foi incluído como ferramenta para diminuir a influência dos fatores que não são dependentes do teste. Esses fatores foram citados como instrumentação, calibrações, erro do operador e qualidade da água dentre outros (OIE..., 2013).

2.3.5.2 Controle de qualidade externo, teste de proficiência

A reprodutibilidade é acessada através de programas de controle de qualidade externos como o teste de proficiência (OIE..., 2010).

O teste de proficiência é uma importante atividade para assegurar a qualidade de ensaios que contribui para que sejam válidos. Com a acreditação e o uso de métodos válidos, os testes de proficiência são a terceira parte essencial de dados do

sistema de credibilidade do laboratório (Wiegers, 2004).

No Brasil, os laboratórios que pretendem obter a acreditação através do INMETRO devem demonstrar a sua competência para realizar os ensaios e calibrações para os quais buscam a acreditação, por meio da participação satisfatória em atividades de ensaios de proficiência. A norma NIT-DICLA-026 detalha os requisitos de participação em atividades de ensaio de proficiência (BRASIL, 2011).

O teste de proficiência em ensaios aplicados a laboratórios de diagnóstico de doença animal (veterinário) é essencial para facilitar e documentar a transferência do ensaio da fase de desenvolvimento para o uso na bancada, transferi-lo de um laboratório para outro, para validar, para avaliar e comparar métodos, para verificar sua precisão, e facilitar o comércio e aprovações regulamentares e decisões. (Wiegers, 2004).

Além disso, o teste de proficiência tem se apresentado como um componente essencial para verificação do resultado do ensaio, validação do método e harmonização, aprovação do laboratório pelos órgãos reguladores, acreditação de laboratórios e gestão de questões comerciais (Wiegers, 2004).

A participação dos laboratórios em atividades de ensaio de proficiência é um dos mecanismos de controle da qualidade dos resultados previstos na NBR ISO/IEC 17025. Os benefícios advindos desta participação em ensaios de proficiência incluem ter uma avaliação externa regular e independente da qualidade de seus resultados de ensaios e calibrações e poder comparar o seu desempenho com o de outros laboratórios semelhantes (BRASIL, 2013a).

2.3.6 STARD e validação de ensaios diagnósticos veterinários

O objetivo geral do *Standards for Reporting of Diagnostic Accuracy* (STARD) é melhorar a acurácia, a integridade e transparência de relatórios de estudos de precisão diagnóstica, a fim de permitir que os pesquisadores detectem possíveis fontes de vieses no estudo e avaliar a possibilidade de generalização e utilidade dos resultados para o propósito de uso (Bossuyt *et al.*, 2003; Gardner, 2010).

Para atingir esse objetivo, o grupo de projeto do STARD desenvolveu uma lista de verificação com 25 itens. A lista pode ser utilizada para verificar se todos os elementos essenciais foram incluídos no relatório de um estudo. A ordem dos itens corresponde à sequência utilizada em muitas publicações de estudos de acurácia diagnóstica. A decisão da inclusão dos itens na lista foi sempre que possível, baseado em evidências ligando esses itens ao viés, variabilidade nos resultados, ou limitações da aplicabilidade dos resultados para outros cenários.

Segundo Bossuyt *et al.* (2003) o documento é explicativo para facilitar a utilização, compreensão e divulgação da lista de verificação. O documento contém um esclarecimento sobre o significado, justificativa e uma utilização otimizada de cada item na lista de verificação, bem como um breve resumo das evidências disponíveis sobre viés e aplicabilidade. (Bossuyt *et al.*, 2003)

O item sete, por exemplo, refere-se à descrição do padrão ouro e da justificativa para sua utilização. Em estudos de acurácia do diagnóstico, o padrão de referência é usado para distinguir indivíduos doentes com a condição alvo, daqueles sem essa condição. Algumas condições alvo não podem ser definidas inequivocamente. Dependendo do estudo em questão, a relevância clínica, prognóstico ou diagnóstico patológico pode definir a condição alvo (Knottnerus e Muris, 2002).

No STARD o teste que está sendo avaliado é referido como o ensaio índice. Os resultados obtidos com o teste índice são comparados com os resultados do padrão de referência, obtido nas mesmas amostras. Nessa situação, o padrão de referência é o melhor método disponível para o estabelecimento da presença ou a ausência da condição alvo. O padrão de referência pode ser um único teste, ou uma combinação de métodos e técnicas, incluindo o acompanhamento clínico de indivíduos testados (Bossuyt *et al.*, 2003).

Segundo Bossuyt *et al.* (2003) a metodologia para a concepção e realização de estudos de acurácia do diagnóstico ainda está amadurecendo. O entendimento das fontes de variabilidade e o potencial de polarização estão crescendo. Como resultado, espera-se uma atualização periódica da lista de verificação do STARD.

3. MATERIAL E MÉTODOS

3.1 Aspectos gerais

Os órgãos oficiais da área veterinária possuem regulamentos e orientações sobre validação de métodos diagnósticos em forma de documentos diferenciados em normas, manuais, e instruções normativas. Existe também, uma tendência de elaboração de documentos gerados a partir de consulta a expertises e documentos elaborados a partir de encontros realizados com a finalidade de harmonização dos requisitos. Esses encontros proporcionam a revisão de propostas de validação, bem como novas propostas de orientações e limites aceitáveis para a realização dos estudos de desempenho do teste diagnóstico novo ou em estado de melhoria. Faltam, no entanto, documentos completos como instrumento para avaliação de tais estudos.

Os principais objetivos dos mencionados documentos são facilitar a realização de procedimentos de avaliação e assegurar a coerência entre a avaliação do desempenho

dos testes em busca de disponibilização de testes diagnósticos com elevada qualidade. Além disso, regulamentar a comercialização e utilização de métodos destinados a diagnósticos, definir diretrizes e promover a aceitação científica de novas metodologias que cada vez mais vêm aumentando.

As diretrizes evitam e removem dificuldades no procedimento e interpretação dos resultados bem como, facilitam o processo da implantação da validação pelos órgãos governamentais e outros usuários, como laboratórios de rotina, pesquisa e laboratórios fabricantes. Como novas tecnologias vão surgindo em paralelo, os conteúdos desses documentos vão sendo, modificados ou ampliados. Essas modificações vão sendo indicadas por datas de revisões informadas nos documentos.

Na presente pesquisa foi realizado um estudo exploratório dos documentos de orientação para validação de métodos diagnósticos aplicados à viroses dos animais. Após a análise exploratória foi identificado a frequência dos itens recomendados para desenvolvimento e validação dos testes utilizando os critérios presentes no documento da OIE (OIE..., 2010) como referência.

3.2 Metodologia

Há muitas maneiras de obter informações. Os métodos mais comuns para tanto são: pesquisas bibliográficas, grupos focais, entrevistas, pesquisas por telefone, inquéritos por *e-mail*, *e-mail*, pesquisas e levantamentos pela internet. No presente estudo as informações foram obtidas através de pesquisas bibliográficas, inquérito por *e-mail* e pesquisas e levantamentos pela internet.

Levantamentos por *e-mail* e internet são relativamente novos e pouco se sabe sobre o efeito do viés de amostragem em pesquisas de internet. Embora seja claramente o método mais eficaz e mais rápido de

distribuição de uma pesquisa, o perfil demográfico do usuário de internet não representa a população em geral, embora isso se encontre em processo de mudança.

3.2.1 Estudo dos documentos

O material do estudo consistiu em documentações de validação de métodos direcionadas à diagnóstico sorológico aplicado em veterinária, sendo principais as normas internacionais e diretrizes de órgãos governamentais. A metodologia empregada foi pesquisa documental proposta por Gil (2006) aplicando a análise de conteúdo desenvolvida em três fases: pré-análise que consiste na fase de organização, escolha dos documentos e preparação do material para análise; a segunda fase denominada exploração do material é a mais longa e, compreende a codificação dos documentos envolvendo a escolha das regras de contagem; e a terceira fase, que consiste no tratamento dos dados e, sua inferência e interpretação com o objetivo de torná-los válidos e significativos.

O estudo das normas foi realizado nos 34 países membros, e nos 6 países parceiros da Organização para a Cooperação e Desenvolvimento Econômico (OECD). A OECD é uma organização internacional criada em 30 de setembro de 1961 que, agrupa os países mais industrializados da economia de mercado no mundo. Os membros da OECD são economias de alta renda com um alto índice de Desenvolvimento humano (IDH). Os representantes dos países membros se reúnem para definir políticas com o objetivo de maximizar o crescimento econômico e o desenvolvimento dos mesmos. Os países membros são: Austrália, Áustria, Bélgica, Canadá, Chile, República Checa, Dinamarca, Estônia, Finlândia, França, Alemanha, Grécia, Hungria, Islândia, Irlanda, Israel, Itália, Japão, Coreia, Luxemburgo, México, Holanda, Nova Zelândia, Noruega, Polônia, Portugal,

República Eslovaca, Eslovênia, Espanha, Suécia, Suíça, Turquia, Reino Unido e Estados Unidos, e os parceiros são: Brasil, China, Índia, Indonésia, África do Sul e Rússia.

Na pré-análise, caracterizada pela fase de organização, foram realizadas buscas, a partir do ano 2000, em redes eletrônicas utilizando o buscador PubMed incluindo normas internacionais citadas pela OIE, revistas científicas, anais de eventos científicos, periódicos de indexação, teses e dissertações, livros, manuais e obras de referência especializadas com o objetivo de obtenção de informações mais completas em relação ao significado da validação de testes diagnósticos para veterinária. Posteriormente, foram realizadas buscas nos sites dos Ministérios da Agricultura dos países membros do OECD em busca de fonte primária de documentos. Os documentos primários são desprovidos de comentários e traduções e, portanto, os mais indicados segundo Luna (1993). Em seguida, foram realizadas a seleção e a separação dos documentos de acordo com o país e instituição de origem para análise em busca do fornecimento das propostas de validação de método.

Na fase de exploração, os materiais foram codificados envolvendo a escolha dos documentos incluindo critérios de inclusão e exclusão de *guidelines* e artigos. Os critérios de inclusão foram: documentos originados de organismos reconhecidos em nível regional, nacional e internacional, com função de estudo, elaboração e aprovação de normas de validação disponíveis ao público; e, documentos nos quais apresentaram critérios de validação incluindo avaliação estatística ou não. Nessa fase, também, foi realizada a caracterização da fonte selecionada e classificação dos mesmos. A avaliação dessas informações determinou as categorias de análise e fundamentou as etapas seguintes de investigação do tema em função das lacunas existentes localizando os

principais entraves teóricos e/ou metodológicos. Os 18 documentos foram reunidos em nove grupos conforme o órgão oficial de origem.

Na fase de tratamento dos dados foi realizada uma análise descritiva das abordagens de validação. Nessa etapa foram caracterizados os estágios de validação, testes analíticos e ferramentas estatísticas disponíveis nas documentações oficiais considerando a área de aplicação do ensaio diagnóstico.

Na análise de completude do documento em estudo foram utilizados os seguintes critérios: “0” para ausência do item; “0,5” para presença do item de forma incompleta, com a presença apenas do relato do teste; e “1,0” para presença do item de forma completa, incluindo informações além do relato da necessidade de realização do teste.

Os critérios de desenvolvimento, validação de métodos e ferramentas estatísticas recomendadas foram contabilizados considerando a presença do item no documento independentemente de informações a respeito da aplicação dos mesmos. Em sequência foram construídas tabelas com os dados coletados possibilitando destacar os percentuais de presença de cada item em relação aos totais de itens presentes nos documentos estudados.

Na fase de inferência e interpretação, as propostas de validação categorizadas segundo a aderência dos estágios de validação, em relação ao documento para validação e certificação de ensaios para diagnóstico (OIE..., 2010) foram analisadas para estabelecer possíveis relações.

A interpretação dos dados foi realizada mediante análise da relação entre os dados empíricos e a teoria, suficientemente confirmada, dos aspectos epidemiológicos analisados no estudo proposto enfatizando a importância da teoria para o estabelecimento

de generalizações e sistemas de relações entre as proposições (Gil, 2006).

3.2.2 Estudo do processo de desenvolvimento e validação de um método diagnóstico

Em uma segunda fase foi aplicado um questionário sobre o processo de desenvolvimento e validação de método diagnóstico. O objetivo do estudo foi conhecer a aderência dos pesquisadores ao processo de validação de métodos na pesquisa e desenvolvimento de um ensaio diagnóstico sorológico aplicado em veterinária. Além disso, identificar pontos de estudos.

Pesquisadores colhem informações por meio de emprego de uma grande variedade de métodos e muitos dos métodos de coleta de informações envolvem uma escolha do sujeito experimental. Por exemplo, querer escolher os respondentes a serem entrevistados em uma pesquisa. Essa escolha pode ser feita usando métodos baseados em probabilidade. Alternativamente, a escolha pode ser feita utilizando de algum elemento de julgamento. Métodos que envolvem julgamento são, por vezes, conhecidos como seleção intencional, ou seleção não-probabilística. Em vários países europeus, os principais inquiridos por amostragem oficiais de empresas usam seleção proposital, por causa de problemas graves na obtenção de cooperação do entrevistado. Na Nova Zelândia, algumas partes do Índice de Preços ao Consumidor (IPC) usa amostragem não-probabilística. Nesse caso, o uso da amostragem não-probabilística tem a aprovação do Comitê Consultivo, um grupo de representantes de consumidores e utilizadores e especialistas que prestam aconselhamento aos estatísticos do Governo sobre o IPC (Doherty, 1994).

Devido a característica do presente estudo, ou seja, levantar as informações de especialistas, a amostragem foi não-

probabilística. Amostragem não-probabilística é aquela que possibilita ao pesquisador a escolha de determinado elemento do universo, por meio de uma grande variedade de técnicas (Gil, 2006). Amostragem por conveniência é usada em pesquisa exploratória na qual o pesquisador está interessado em obter uma aproximação da verdade com baixo custo (Doherty, 1994; Gil, 2006).

A amostra deste estudo foi composta por 38 indivíduos, de ambos os sexos e com idade superior a 18 anos. Na área de avaliação "Medicina Veterinária" foram identificadas as universidades com conceito 7, 6, 5 ou 4 conforme divulgação pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- CAPES/MEC no ano de 2010. Os

conceitos referem-se a produção bibliográfica distribuída segundo a estratificação Qualis. São eles: teses e dissertações defendidas e número de docentes permanentes no, triênio 2007-09; e nota final da avaliação 2010. Nessas universidades identificadas, foram selecionados os professores pesquisadores dos Cursos de Pós-Graduação, com modalidade acadêmica, contemplando mestrado e doutorado (orientação plena), com atuação ou publicação nas áreas de desenvolvimento, avaliação e /ou validação de métodos sorológicos de diagnóstico ou virologia. O "n" amostral de 38 representou o universo de acordo com os critérios estabelecidos para inclusão do pesquisador no estudo. A distribuição dos conceitos dos cursos pode ser observada na Figura 4.

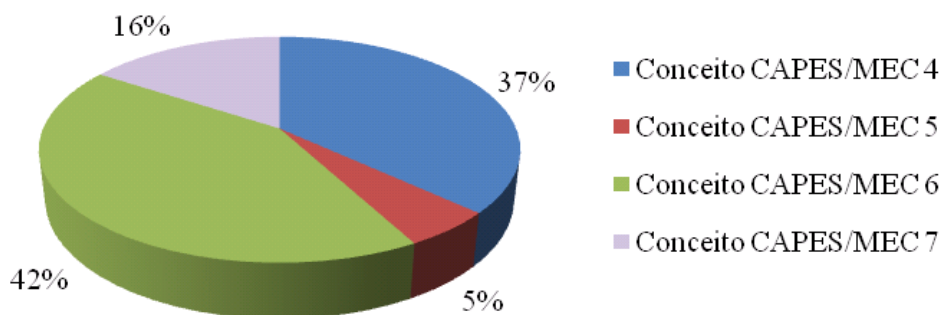


Figura 4 – Distribuição dos conceitos dos cursos na amostra

O professor/pesquisador foi convidado a participar da pesquisa através de termo de consentimento livre e esclarecido (Anexo 1). A pesquisa foi realizada através do envio do questionário eletrônico (Anexo 2) para Universidades do Brasil conforme a Figura 5, após a autorização da direção da unidade acadêmica na qual o participante estava inserido, e da aprovação do projeto no Comitê de Ética em Pesquisa (COEP) – UFMG. Foram enviadas correspondências

(Anexo 3), oficiais numeradas, para os diretores ou coordenadores das unidades dos cursos de medicina veterinária, formalizando a solicitação da realização da pesquisa e esclarecendo sobre os objetivos e compromisso de realização do estudo cumprindo os requisitos da Resolução do Conselho Nacional de Saúde 196/96 que trata da Pesquisa envolvendo Seres Humanos.

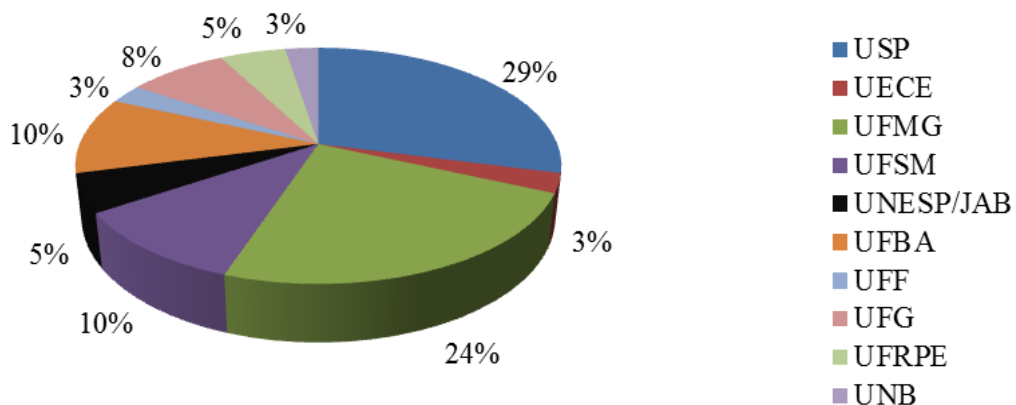


Figura 5 – Distribuição de Pesquisadores das Instituições na amostra

O instrumento proposto, o questionário, foi composto por 28 itens, sendo 17 itens segundo o objetivo 4, desenvolver uma nova ferramenta de avaliação para a avaliação da qualidade dos estudos de diagnóstico baseada em evidências, do estudo de Whiting *et al.* (2004). Os itens foram avaliados em uma escala tipo Likert (1932) de cinco pontos cujos descritores são: “Discordo totalmente”, “Discordo moderadamente”, “indiferente”, “Concordo moderadamente” e “Concordo plenamente”. Após a elaboração do questionário foi realizado um estudo piloto aplicado a 10% da amostragem com o objetivo de verificar a necessidade de ajustes antes da realização do pré-teste. Foi calculado o alfa de Cronbach (1951) para a amostra definida para verificar a consistência do questionário.

O questionário foi enviado, por *e-mail*, também para 12 pesquisadores de duas instituições públicas federais com o objetivo de fazer um levantamento sobre a utilização das normas de validação.

O sucesso da escala de Likert (1932) reside na sua sensibilidade para reconhecer a oposição entre contrários; reconhecer gradiente; e reconhecer situação intermediária. A representação aritmética de um evento qualitativo é uma estratégia para

o processamento e análise, mas a interpretação dos resultados requer um retorno ao significado original de suas medidas. Assim, para analisar as expressões do grau de concordância/discordância, nesse trabalho foram utilizados os valores -100%, -50%, 0, 50%, e 100%, respectivamente, para as categorias “discordo totalmente”, “discordo moderadamente”, “indiferente”, “concordo moderadamente” e “concordo plenamente”, utilizadas no questionário (Pereira, 2004). Após a atribuição dos valores, a média e o intervalo de confiança a 95% foram calculados e representados graficamente. Os intervalos negativos foram considerados como discordantes, os positivos como concordantes e os intervalos contendo o zero representaram a neutralidade.

3.2.3 Estudo da presença dos critérios de validação de métodos do documento da OIE (OIE..., 2010) nos artigos científicos publicados nos anos de 2011 e 2012

A literatura foi sistematicamente pesquisada. Foram incluídos os artigos com relatos de desenvolvimento, validação de métodos e avaliação de acurácia. Os artigos foram identificados como elegíveis para inclusão

utilizando as interfaces PubMed, OVID e Google Scholar utilizando o idioma inglês (Figura 6). As datas de publicação restringiram-se ao período entre 1º de janeiro de 2011 e 31 de dezembro de 2012. Os descritores para o banco de dados foram: “*veterinary diagnostic test*”, “*validation of method and veterinary*” e “*development and validation*”. Posteriormente o título e o resumo de cada artigo potencial foram examinados para determinar a potencial elegibilidade, que foi confirmada por uma análise do texto completa. As referências dos artigos foram examinadas para verificar a presença de normas de validação, artigos com temas de validação de métodos, desenvolvimento de métodos e estudos de acurácia. Em sequência, foi avaliada a aderência dos 12 critérios de validação de métodos citados no documento “Princípios e métodos de validação de ensaios diagnósticos para doenças infecciosas” (OIE..., 2010). Os dados foram sumarizados em uma tabela, onde foram tabuladas as frequências dos critérios em relação à presença das normas e artigos. A separação em colunas foi realizada com o objetivo de comparar a completude em relação à presença ou ausência das normas e artigos.

4.0 RESULTADOS E DISCUSSÃO

4.1 Estudo dos documentos de validação

Avaliando os critérios de desenvolvimento e validação de métodos segundo OIE (OIE..., 2010) nos documentos pesquisados, os critérios proposta de uso e reprodutibilidade apresentaram maior frequência percentual (ambas com 88,9%). As sensibilidades e especificidades analíticas e diagnósticas apresentaram uma homogeneidade com frequência percentual (72,2%). Os critérios com menor citação foram ‘otimização’ (33,3%), “*Ruggedness*” (33,3%) e ponto de corte (27,8%). Enquanto a robustez e

repetibilidade ficaram com valores próximos de frequência percentual, 77,8% e 83,3% respectivamente (Tabela 4).

As frequências encontradas para sensibilidade e especificidade justificam-se devido à sua utilização nas avaliações da acurácia dos ensaios e também quando é preciso realizar modificações nos ensaios. Segundo a OIE (OIE..., 2010) essas modificações são necessárias para ajustar o desempenho do diagnóstico ou para melhorar a eficiência do ensaio ou custo-efetividade. Esses critérios são avaliados também na revalidação que é recomendada quando ocorre a mudança da aplicação do ensaio em relação à região geográfica e / ou população. Nessa situação se encontram as estirpes de vírus, derivadas de animais em diferentes localizações geográficas por apresentarem sequências alvo diferentes. As mutações de agentes infecciosos podem afetar a eficiência do ensaio e invalidar as características de desempenho estabelecidas. Assim, a OIE (OIE..., 2010) recomenda que seja confirmada regularmente a sequência alvo nas regiões genômicas para os isolados (nacional ou regional) dos agentes infecciosos.

Ruggedness foi um dos critérios menos citados, possivelmente porque foi considerado um estudo da sequência da análise da imprecisão interlaboratorial e, às vezes, foi citado como robustez. Burns *et al.* (2009) definiram claramente os termos *Robust* e *Rugged* e seus usos distinguidos, juntamente com as características associadas de robustez e resistência. Mostraram que as características de robustez e resistência expressam resistência contra as condições e influências. Essas características diminuem tanto a precisão como a exatidão dos resultados analíticos, obtidos por um procedimento em particular, em um dado laboratório ou em laboratórios diferentes.

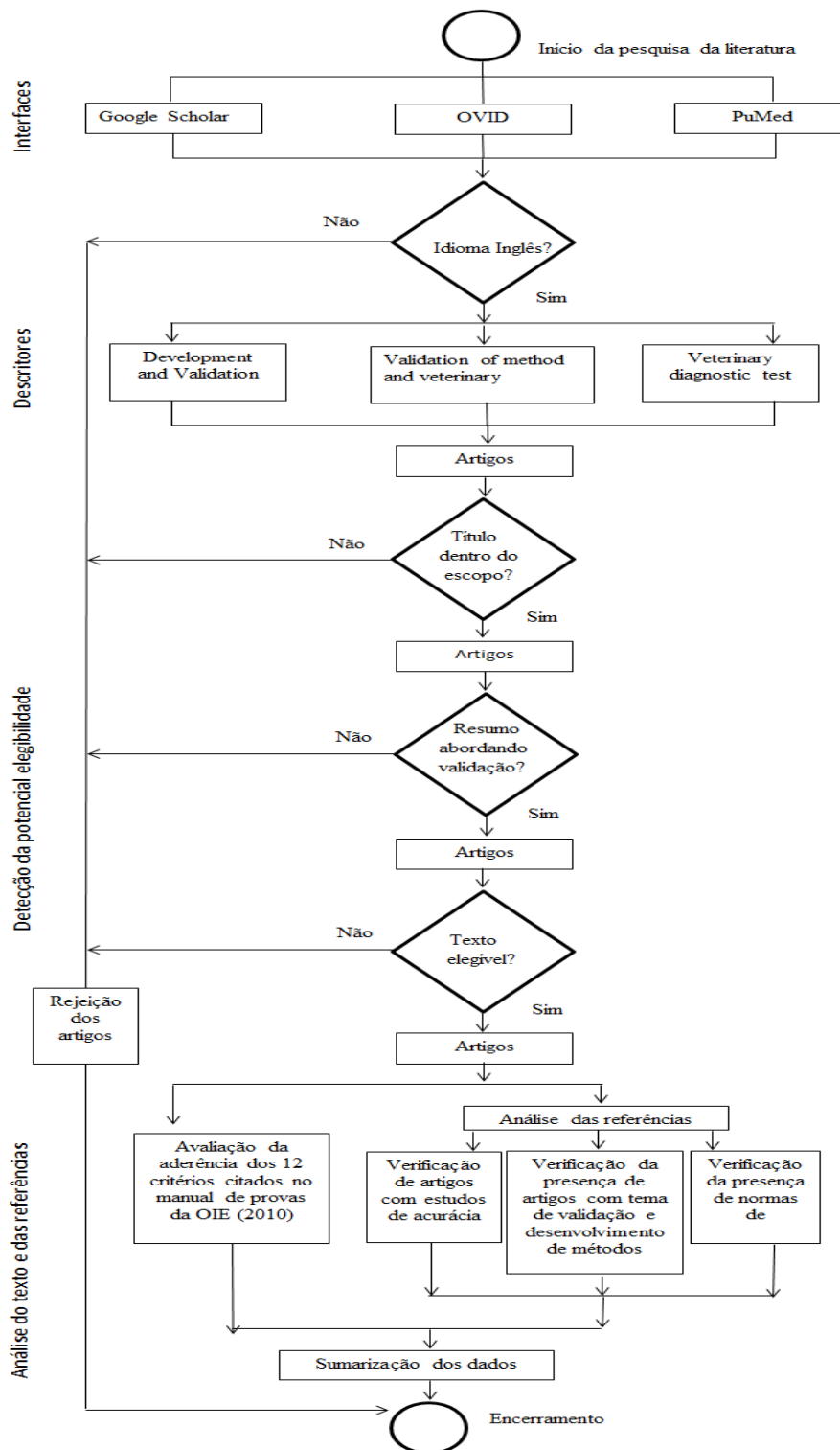


Figura 6 – Fluxograma da pesquisa da literatura de validação de métodos publicada em 2011 e 2012.

Não foi encontrado um modelo que contemplasse todos os critérios e ferramentas estatísticas disponíveis para a validação de métodos. Uma das razões para isso poderia estar relacionada ao fato de que o estudo de validação de métodos na área veterinária é mais recente em relação à área humana (Guidance..., 2005). A maioria dos critérios e ferramentas estatísticas apresentadas para avaliação de métodos na área humana, como nos documentos do *National Committee For Clinical Laboratory Standards* (NCCLS), são aplicáveis na avaliação de ensaios veterinários. No entanto, embora os critérios pareçam ser os mesmos, Greiner e Gardner (2000) ressaltaram que na área veterinária algumas distinções devem ser observadas como, por exemplo, no que se refere à aplicação do ensaio a um *pool* de amostras.

É importante ressaltar nesse ponto que as diretrizes de validação estão ligadas as relações econômicas e aos interesses políticos de comercialização de produtos entre os países, com reflexos na proposta de uso do produto e no que deve ser avaliado através do ensaio diagnóstico.

Algumas iniciativas como a elaboração do documento (Validation..., 1998) atualizado em 2010 demonstram relações de interesses de comercialização. Esse documento foi elaborado com o objetivo de fornecer recomendações sobre como considerar características de validação como parte dos pedidos de registros para a aprovação de medicamentos veterinários apresentados à União Europeia, Japão e os Estados Unidos (Validation..., 1998).

Tabela 4. Distribuição das frequências absolutas e percentuais dos critérios de desenvolvimento e validação de métodos segundo OIE (OIE..., 2010) nos documentos estudados.

Critérios de validação de ensaios	Frequência absoluta	Frequência percentual
Proposta de uso	8,0	88,9%
Otimização	3,0	33,3%
Padronização	4,5	50,0%
Robustez	7,0	77,8%
Repetibilidade	7,5	83,3%
Sensibilidade analítica	6,5	72,2%
Especificidade analítica	6,5	72,2%
Ponto de corte	2,5	27,8%
Sensibilidade diagnóstica	6,5	72,2%
Especificidade diagnóstica	6,5	72,2%
Reprodutibilidade	8,0	88,9%
<i>Ruggedness</i>	3,0	33,3%

Na avaliação das ferramentas estatísticas recomendadas, os resultados encontrados foram: desvio-padrão e coeficiente de variação com 55,6%, curva ROC, teste Kappa e outros com 22,2% e o Modelo Bayesiano apresentou menor frequência (11,1%). O Qui-Quadrado representou 33,3% das ferramentas estatísticas recomendadas nos documentos estudados (Tabela 5).

A frequência percentual do desvio padrão e coeficiente de variação foram os mesmos, o que pode ser explicado pelo fato de que para obter-se o coeficiente de variação é necessário o valor do desvio padrão.

O modelo Bayesiano foi o menos frequente tendo em vista que se trata de uma ferramenta estatística avançada, que requer conhecimento mais específico. Necessita de pressupostos críticos para definição dos

dados *a priori* e, investimento intelectual para utilização dos softwares para sua aplicação (Gardner, 2002). Atualmente, na avaliação de ensaios diagnósticos o uso desse modelo tem sido recomendado e utilizado quando não existe um padrão-ouro ou quando os ensaios que estão sendo comparados são considerados como imperfeitos (OIE..., 2010).

Gardner e Greiner (2006) observaram que a medicina veterinária vem adotando a curva ROC lentamente, mas de forma crescente. Piza et al. (2007) utilizou a curva ROC para

determinar o ponto de corte de um ELISA para diagnóstico da anemia infecciosa equina. No exemplo discutido no estudo para expressar a acurácia de teste diagnóstico para identificar insuficiência cardíaca congestiva, Florkowski (2008) concluiu que a análise da curva ROC permitiu definir o melhor ponto de corte para a finalidade clínica. Yang *et al.* (2012) obtiveram melhores resultados de sensibilidade e especificidade na avaliação de múltiplos ELISAs para detecção de vírus da peste suína clássica quando utilizaram a curva ROC.

Tabela 5. Distribuição das frequências absolutas e percentuais das ferramentas estatísticas recomendadas nos documentos estudados.

Ferramentas estatísticas recomendadas nos documentos	Frequência absoluta	Frequência percentual
Coefficiente de variação	5	55,6%
Curva ROC	2	22,2%
Desvio Padrão	5	55,6%
Kappa	2	22,2%
Modelo Bayesiano	1	11,1%
Qui- Quadrado	3	33,3%
Outros	2	22,2%

Na avaliação da completude dos documentos estudados, “conceitos” e “estágios” apresentaram a mesma percentagem (55,6%) (Tabela 6). A percentagem maior para conceitos e estágios relaciona-se a iniciativas de alguns países que formam grupos para estudar, discutir e harmonizar documentos de recomendação de validação de métodos que facilitarão o registro e a comercialização de produtos entre esses países. Austrália e Nova Zelândia através do SCAHLS procuram sustentar e melhorar a qualidade dos produtos animais e garantir o acesso ao mercado aplicando as melhores práticas para serviços de laboratórios veterinários. O objetivo primário é assegurar a coerência entre os laboratórios usando métodos selecionados por sua ótima precisão, sensibilidade, especificidade e robustez (ANZSDP..., 2011). Os objetivos

secundários são fornecer diretrizes para os métodos a serem utilizados em programas de ensaios de proficiência externos e para auxiliar no desenvolvimento de documentação para sistema de qualidade. Para ser aprovado, o novo ensaio, deve passar por quatro estágios de avaliação. Após esses estágios, deve ser elaborado o documento com as informações, realizado a submissão do documento com a solicitação da aprovação junto ao grupo SCAHLS (SCAHL..., 2012).

Percebe-se a necessidade da harmonização dos documentos internacionais de validação de ensaio com finalidade de minimizar as diferenças. Melhorando, assim, o conhecimento sobre a sua qualidade. Mas, por outro lado, as generalizações podem comprometer a validação no que tange as

particularidades das doenças em estudo ou dos métodos. Aumentar a aceitação dos produtos que dependem de liberação através da avaliação, utilizando ensaios diagnósticos ou mesmo da comercialização dos próprios ensaios, envolve também questões políticas relacionadas ao estabelecimento de restrições ou limites de aceitação de acordo com as condições de investimento e competição do país.

Os métodos estatísticos e o planejamento de avaliação dos testes estavam presentes em 33,3% dos documentos estudados (Tabela 6). A não recomendação dos testes estatísticos em todos os documentos pode ser explicada pelo fato de serem mais complexos e se encontrarem em situação de discussão técnica em relação às aplicações dos mesmos nas avaliações de desempenho dos ensaios diagnósticos.

Tabela 6. Distribuição das frequências absolutas e percentuais da completude dos documentos estudados em relação à presença de conceitos, estágios, métodos estatísticos e planejamento de avaliação dos testes.

Critérios de completude	Frequência absoluta	Frequência percentual
Conceitos	5	55,6%
Estágios	5	55,6%
Métodos estatísticos	3	33,3%
Planejamento de avaliação dos testes	3	33,3%

4.2 Estudo do processo de desenvolvimento e validação de um método diagnóstico

Após a coleta de dados realizada por meio do envio de questionários para o *e-mail* dos entrevistados a taxa de retorno obtida foi de 26,0%, retorno que se encontra dentro do esperado para envio de questionário eletrônico, posto que, segundo Marconi e Lakatos (2005), questionários que são enviados para entrevistados alcançam em média 25% de devolução. No estudo quanto ao uso de questionários via *e-mail* realizado por Vieira *et al.* (2010), foram selecionados 218 participantes para a realização de coleta de dados. A seleção foi realizada utilizando o cálculo da amostragem aleatória simples com erro amostral de 5%, intervalo de confiança de 95%, e ao final da coleta foram obtidos 55 questionários respondidos, representando um retorno de 25,23%.

Caracterizando os respondentes do presente estudo, pode-se observar que, segundo o tempo de pesquisa veterinária, 77% dos

pesquisadores trabalham com pesquisa em medicina veterinária há mais de 6 anos e, 23%, entre 3 e 6 anos. Considerando a formação acadêmica, as análises epidemiológicas de testes diagnósticos se encontram presentes na formação de 69% dos pesquisadores, mas, apenas 31% estão inseridos em programas de Pós-Graduação com disciplina relacionada à qualidade assegurada. Esse estudo descritivo sugere a realização de formação em qualidade e validação de métodos, com a finalidade de ajustar-se com a competência às novas demandas geradas pelo progresso científico e tecnológico.

Quanto ao item “qual método diagnóstico é utilizado com maior frequência?”, foi possível identificar que o método ELISA é o teste mais utilizado (62%), outros testes apresentaram percentagem considerável (31%) e o teste de soroneutralização apresentou a menor frequência (8%) na amostra em estudo. A técnica de soroneutralização é uma técnica complexa e demorada. Apresenta maior dificuldade de

realização, de infraestrutura e habilidade de operador, o que não é adequado para o ensaio em larga escala em sistemas de vigilância. Esse fato, possivelmente levou o pesquisador a optar pelo uso mais frequente do ELISA. A justificativa pelo uso do ELISA tem sido reportada por ser uma técnica que pode ser aplicada em larga escala em estudos epidemiológicos, e ser um ensaio simples e rápido.

O documento de validação com leitura mais frequente foi o Manual de Validação da Organização Mundial de Saúde Animal (69%) e a Instrução Normativa do Ministério da Agricultura, Pecuária e Abastecimento (38%). A leitura do documento de validação da Cooperação Internacional para a Harmonização de Requisitos Técnicos para o Registro de Produtos Veterinários (VICH) apresentou uma frequência de leitura menor (8%). Não

foi identificada a leitura de outras fontes de recomendações além das opções constantes do item “Documento de validação”. Embora todas essas normas estejam disponíveis *on line* o Manual de validação da Organização Mundial de Saúde Animal (OIE..., 2010) é na maioria das vezes recomendado nos sites dos ministérios da agricultura dos países, nas agências de regulação de alguns países e, nos cursos de medicina veterinária das universidades que realizam pesquisa de ensaios diagnósticos.

Observa-se na Figura 7 que os artigos (92%) foram a forma mais prevalente para orientação da avaliação da acurácia diagnóstica, seguidos pelas recomendações das normas internacionais (46%) e normas nacionais (31%). Nesse item, foi solicitado ao entrevistado que marcasse todas as opções de recomendações por ele utilizadas.

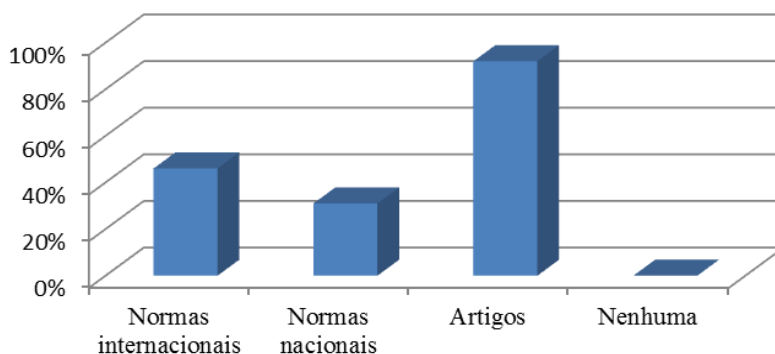


Figura 7- Distribuição das respostas dos pesquisadores em relação aos documentos utilizados na avaliação da acurácia diagnóstica

Entre as etapas de validação de métodos, a amostragem e os estudos de acurácia apresentaram o maior grau de importância, seguidos pela escolha do método de referência e por estudos de imprecisão, segundo dados expostos na Tabela 7. A maioria dos entrevistados é da área acadêmica, portanto, o resultado das pesquisas por eles realizadas são artigos

direcionados para publicação em periódicos científicos. Assim, existindo um envolvimento maior com a amostragem. A amostra de estudo é um item de avaliação por parte dos relatores dos periódicos por se tratar de generalizações que possam vir a serem feitas em relação aos estudos de doenças em populações definidas. Quanto ao estudo de acurácia, o que se pretende

demonstrar é que a pesquisa resultou em um ensaio aplicável ao diagnóstico laboratorial

das infecções com acurácia necessária para tal proposta.

Tabela 7. Sumário dos resultados para avaliação da importância das etapas de validação de métodos

Etapas	Escala Likert ^a				
	1	2	3	4	5
Amostragem	0	0	1	0	11
Estudos de imprecisão	2	0	1	3	5
Estudos de acurácia	0	1	1	0	10
Escolha do método de referência	0	0	1	3	8
Estudos primários	0	0	4	5	2

^a descritores são: (1) Discordo totalmente, (2) Discordo moderadamente, (3) indiferente, (4) Concordo moderadamente e (5) Concordo plenamente.

Na Figura 8 observamos que o Teste Qui-Quadrado (85%), o coeficiente de variação (69%) e o teste Kappa (77%) foram ferramentas estatísticas mais utilizadas na avaliação do desempenho de ensaios. O

menor percentual encontrado foi para utilização do modelo bayesiano (23%). Nesse item, foi solicitado ao entrevistado a marcação de todas as ferramentas utilizadas.

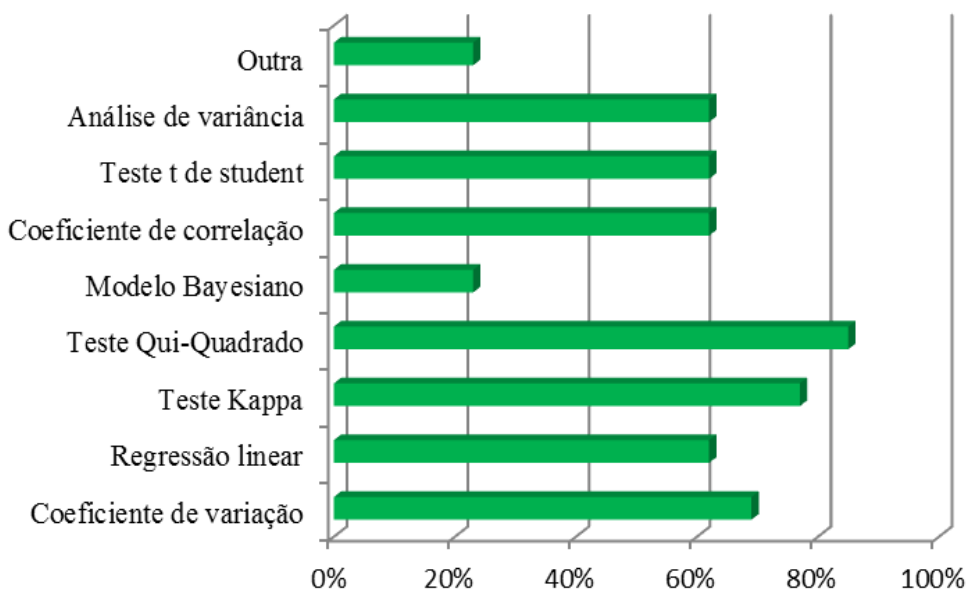


Figura 8- Distribuição das respostas dos pesquisadores em relação ao uso das ferramentas estatísticas na avaliação do desempenho de testes diagnósticos.

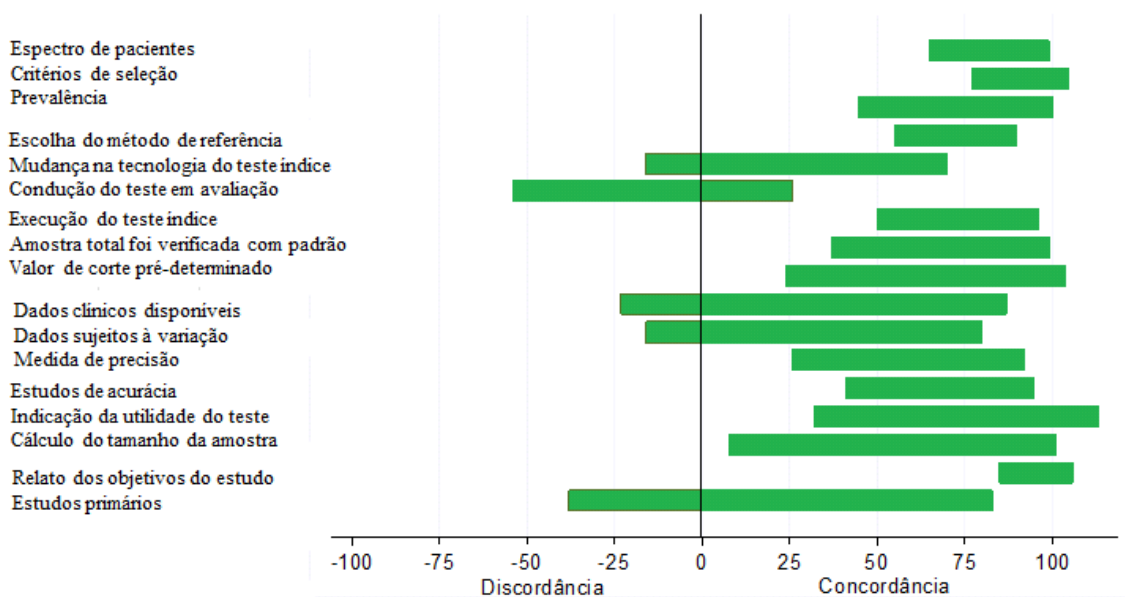


Figura 9- Intervalos do grau de discordância e concordância do estudo do processo de validação.

Na avaliação dos itens ao desenvolver e validar um método utilizando a escala Likert (1932) de cinco pontos, os resultados obtidos foram expostos na tabela 8. Na Figura 9 observa-se o intervalo de confiança a 95% dos resultados (da tabela 8) codificados conforme a manifestação de Concordância ou Discordância como sugerido por Pereira (2004).

Os itens relacionados com a composição do espectro (espectro de pacientes, critérios de seleção e prevalência) apresentaram concordância com menores intervalos de confiança. Esse resultado pode estar indicando uma maior atenção por parte dos pesquisadores com as características das amostras na realização dos estudos de validação dos ensaios diagnósticos. Esses, foram itens que, também apresentaram escore elevado (5) na primeira rodada da aplicação do Delphi, no objetivo 4, do estudo realizado por Whiting *et al.* (2004).

O objetivo do estudo de Whiting *et al.* (2004) foi combinar a evidência empírica e a opinião de especialistas (nove) na área de investigação diagnóstica em um método

formal de consenso para desenvolver uma ferramenta (QUADAS) para ser utilizada em avaliações sistemáticas da qualidade de estudos preliminares da acurácia diagnóstica.

O grau de concordância aferido para os itens relacionados com o teste índice padrão de referência: dados clínicos disponíveis e dados sujeitos a variação indica neutralidade, posto que, os intervalos incluíram o zero. Porém, diferentemente apresentaram escore elevado (5) no estudo de Whiting *et al.* (2004).

4.3 Estudo da presença dos critérios de validação de métodos nos artigos científicos publicados nos anos de 2011 e 2012

Foram identificados sessenta e seis artigos. Após examinados os títulos e o resumo, seis artigos foram excluídos e sessenta artigos foram mantidos. Os resultados estão sumarizados na Tabela 9. Quando avaliado a presença de normas e artigos na literatura, nove artigos apresentaram na referência bibliográfica, artigos e normas relacionados

com o tema validação de métodos. Vinte e três artigos apresentaram apenas normas. Treze artigos apresentaram apenas artigos e, quinze não continham na referência qualquer norma ou artigo que fizesse menção direta à validação de métodos.

Quando avaliado os critérios, verificou-se que os mais frequentes foram sensibilidade diagnóstica (51) e especificidade diagnóstica (49) seguidos da proposta de uso (43) e ponto de corte (37).

Segundo OIE (OIE..., 2010) as estimativas da sensibilidade e especificidade são indicadores preliminares do desempenho estabelecido durante a validação de um ensaio. Essas estimativas são a base para o cálculo de outros parâmetros a partir do qual são feitas inferências sobre os resultados dos testes, como exemplo, os valores preditivos (positivos e negativos). Banoo *et al.* (2010) confirmaram essa ideia considerando a sensibilidade e a especificidade diagnósticas como as características fundamentais de desempenho de um ensaio, e as indicam como destinadas à discernir entre infectados de não infectados.

Para Nielsen *et al.* (2011), proposta de uso significa que os resultados do ensaio devem ser interpretados em relação a algum objetivo e finalidade de aplicação. Este objetivo deve ser definido *a priori* pelos tomadores de decisão, que devem especificar antes da avaliação do ensaio diagnóstico o que eles querem que o ensaio detecte. No caso de infecções crônicas, o momento do diagnóstico, o que corresponde ao estágio da doença, precisa ser incluído na descrição da proposta de uso.

Gardner (2010), explorando a proposta do STARD e a sua possível modificação para estudos de doenças dos animais, considera que o projeto de pesquisa selecionado depende da finalidade e utilização prevista

do ensaio. Além disso, o autor relata que diferentes modelos e estratégias analíticas podem ser apropriados para a mesma finalidade do ensaio.

Nielsen *et al.* (2011) afirmaram que frequentemente, a finalidade específica do ensaio não é considerada formalmente quando um ensaio é avaliado. Assim, muitas vezes resulta em uma avaliação onde as estimativas de sensibilidade e especificidade são tendenciosas, seja por causa de problemas com o estabelecimento do verdadeiro estado da infecção, ou porque o ensaio detecta outro aspecto da infecção que não foi o inicialmente previsto.

O critério de menor frequência foi o *Robustness* provavelmente por ser critério recomendado para após transferência para outros laboratórios (OIE..., 2010). Não foi encontrado citação do critério *Ruggedness* o que pode ser explicado pela junção de conceitos com o critério *Robustness*, como relatado por Burns *et al.* (2009). Além disso, apesar de ser conceituado na versão 2013 do Manual de Provas OIE, *Robustness* não foi contemplado na lista de critérios validação de ensaio nesse mesmo documento (OIE..., 2013). Os critérios otimização (28) e padronização (33) apresentaram frequências próximas. Os critérios relacionados com a precisão do ensaio, a repetibilidade e reprodutibilidade apresentaram as mesmas frequências.

A sensibilidade analítica (13) e especificidade analítica (10) apresentaram frequências próximas, sendo que, a sensibilidade diferiu para um valor acima.

Para os artigos que apresentaram somente as normas na referência bibliográfica foi encontrado frequência dos critérios de 43% no total de critérios presentes.

Tabela 8. Sumário dos resultados do estudo do processo de validação

	Escala Likert ^a				
	1	2	3	4	5
Composição do espectro					
O espectro de pacientes foi descrito no artigo e foi escolhido adequadamente.	0	0	0	1	11
Os critérios de seleção foram descritos com clareza.	0	0	0	2	10
A prevalência e a gravidade da doença foi reportada.	0	0	2	3	7
Teste índice e padrão de referência: seleção e execução					
No estágio tecnológico da época, o padrão de referência escolhido era adequado para a verificação dos resultados do teste.	0	0	0	6	6
É possível que uma mudança na tecnologia do teste índice tenha ocorrido desde a publicação deste artigo.	1	2	0	6	3
Houve um período anormalmente longo entre a condução do teste em avaliação e a confirmação do diagnóstico segundo o padrão de referência.	2	3	2	4	1
A execução do teste índice foi descrita em detalhe suficiente para permitir sua replicação.	0	0	1	4	7
A amostra total ou uma fração aleatória da amostra foi verificada usando-se um padrão referencial de diagnóstico.	0	1	0	4	7
O valor de corte foi predeterminado ou era aceitável à luz de trabalhos anteriores.	1	0	0	4	7
Teste índice e padrão de referência: interpretação					
Havia dados clínicos disponíveis quando os resultados do teste foram interpretados.	2	1	1	2	6
Os dados estavam sujeitos a variação do observador ou instrumental que pudesse ter afetado as estimativas de desempenho do teste.	2	0	1	5	3
Análise					
Apresentou-se uma medida de precisão dos resultados.	0	1	1	4	6
Os resultados foram apresentados numa tabela 2.2.	0	0	2	3	7
Ofereceu-se alguma indicação da utilidade do teste.	1	0	0	2	9
Planejamento da pesquisa					
Foi feito um cálculo de tamanho adequado da amostra.	1	1	0	3	7
Os objetivos de estudo foram relatados claramente.	0	0	0	1	11
Havia alguma evidência de que um protocolo de estudo teria sido desenvolvido antes de o estudo ser iniciado.	3	1	0	2	6

^a descritores são: (1) Discordo totalmente, (2) Discordo moderadamente, (3) indiferente, (4) Concordo moderadamente e (5) Concordo plenamente.

Para os artigos que apresentaram somente artigos de validação como referência bibliográfica a frequência dos critérios foi de 42%. Aqueles artigos que apresentaram na referência bibliográfica normas e artigos juntos apresentaram maior valor percentual

(46%) de critérios presentes. Enquanto os artigos que não utilizaram as normas de validação nem os artigos relacionados com o tema validação de métodos apresentaram 38% dos critérios de validação.

Tabela 9. Sumário dos resultados do estudo da presença dos critérios de validação na amostra dos artigos científicos publicados nos anos de 2011 e 2012.

	Apenas artigo (13)	Apenas normas (23)	Artigos e normas (9)	Ausentes artigos e normas (15)	Total de critérios presentes
Critérios	156	276	108	180	720
Proposta de uso	6	20	7	10	43
Otimização	7	11	3	7	28
Padronização	8	13	5	7	33
<i>Robustness</i>	0	5	0	0	5
Repetibilidade	5	6	4	3	18
Sensibilidade analítica	2	5	3	3	13
Especificidade analítica	1	6	1	2	10
Ponto de corte	9	14	6	8	37
Sensibilidade diagnóstica	12	17	9	13	51
Especificidade diagnóstica	12	16	8	13	49
Reprodutibilidade	4	6	4	4	18
<i>Ruggedness</i>	0	0	0	0	0
Percentual da presença dos critérios	(66) 42,3%	(119) 43,1%	(50) 46,3%	(70) 38,9%	(305) 42,4%
IC *	34-50%	37-49%	37-56%	32-46%	39-46%

IC* Intervalo de confiança a 95%

5. CONCLUSÃO

Em conclusão, o Manual de Testes Diagnósticos e Vacinas para Animais Terrestres da OIE (OIE..., 2010) apresentou o maior número de critérios de validação de métodos, o que geraria uma maior aceitação internacional do ensaio a ser validado. Dessa forma, a consulta de tal documento, associado à consulta de artigos científicos

especializados poderiam potencializar a validação de ensaios veterinários.

Os artigos especializados apresentaram pesquisas com diferentes populações, diferentes propósitos de uso, ferramentas estatísticas, e formas de realizar os estudos dos critérios de avaliação de ensaios diagnósticos, indicando a validação como um processo contínuo.

Segundo a amostra da pesquisa da literatura, quando as normas e a literatura especializada em validação foram utilizadas nos estudos de ensaios veterinários, a frequência dos critérios descritos nesta tese foi maior, favorecendo a completude da validação.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALONZO, T. A.; PEPE, M. S. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statist. Med.*, v. 18, p. 2987-3003, 1999.

ANZSDP standard operating procedure, 2011. Disponível em: <<http://www.scahls.org.au/Procedures/authors/Pages/ANZSDP-Standard-Operating-Procedure.aspx>>. Acesso em: 31 de agosto de 2011.

Assessment of the clinical accuracy of laboratory tests using Receiver Operating Characteristic (ROC) Plots; Approved Guideline. *NCCLS*, v.15, n.19, 1995.

BANOO, S.; BELL, D.; BOSSUYT, P. *et al.* Evaluation of diagnostic tests for infectious diseases: general principles. *Nat. Rev. Micro.*, v. 8, n. 12, p. s18-s29, 2010.

BOSSUYT, P. M.; REITSMA, J. B.; BRUNS, D. E. *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clinical chemistry*, v. 49, n. 1, p. 7-18, 2003.

BRASCUM, A. J.; GARDNER, I. A.; JOHNSON, W. O. estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev. Vet. Med.*, v. 68, n.2-4, p.145-163, 2005.

BRASIL. INMETRO. Instituto Nacional de Metrologia. Norma n. NIE-DIMEL-064 Rev.00 Aprovada em Jan/2005. Critérios para validação de métodos de ensaio. Disponível em: <<http://www.inmetro.gov.br/>

ftp_hp/kits/nie_dimel_064_rev00.pdf>. Acesso em: 8 de outubro de 2013.

BRASIL. INMETRO. Instituto Nacional de Metrologia. Norma n. NIT-DICLA-026 Rev.n.08 Aprovada em Out/2011. Requisitos sobre a participação dos laboratórios de ensaio e de calibração em atividades de ensaio de proficiência. Disponível em: <http://www.inmetro.gov.br/Sidoq/Arquivos/Dicla/NIT/NIT-Dicla-26_08.pdf>. Acesso em: 8 de outubro de 2013.

BRASIL. INMETRO. Metrologia legal. A metrologia legal no Brasil, 2012. Disponível em: <<http://www.inmetro.gov.br/metlegal/metBrasil.asp>>. Acesso em: 8 de outubro de 2013.

BRASIL. INMETRO. Instituto Nacional de Metrologia. Acreditação de laboratórios (ABNT NBR ISO/IEC 17025:2005), 2013a. Disponível em: <http://www.inmetro.gov.br/credenciamento/acre_lab.asp>. Acesso em: 8 de outubro de 2013.

BRASIL. INMETRO. Instituto Nacional de Metrologia. Norma n. NIT-DICLA-012 Rev.n.15 Aprovada em Fev/2013. 2013b. Relação padronizada de serviços acreditados para laboratórios de calibração. Disponível em: <http://www.inmetro.gov.br/Sidoq/Arquivos/Dicla/NIT/NIT-Dicla-12_15.pdf>. Acesso em: 8 de outubro de 2013.

BRASIL. INMETRO. Instituto Nacional de Metrologia. Norma n. NIT-DICLA-016 Rev.n.05 Aprovada em Abr/2013. 2013c. Elaboração de escopo de ensaios. Disponível em: <http://www.inmetro.gov.br/Sidoq/Arquivos/Dicla/NIT/NIT-Dicla-16_05.pdf>. Acesso em: 8 de outubro de 2013.

BRASIL, Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa n.57 de 11 de dezembro de 2013. 2013d. Estabelece os critérios e requisitos para credenciamento e monitoramento de laboratórios pelo Ministério da Agricultura, Pecuária e Abastecimento. *Diário Oficial da*

União, Brasília, DF, 12 dez. 2013. Seção n.1, p.5-9.

BRASIL, Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa n.4 de 19 de fevereiro de 2008. Aprova as Normas Técnicas para a Fiscalização da Produção, Controle, Comercialização, Modo de Utilização de Produtos Uso Veterinário destinados a diagnosticar Doenças dos animais. *Diário Oficial da União*, Brasília, DF, 20 fev. 2008. Seção 1, p.

BURNS, D. T.; DANZER, K.; TOWNSHEND, A. A Tutorial discussion of the use of the terms "Robust" and "Rugged" and the Associated Characteristics of "Robustness" and "Ruggedness" as used in Descriptions of Analytical Procedures. *JAPA.*, v. 37, p. 40-60, 2009.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*. v. 16, n.3, p. 297-334, 1951.

DEMING, W. E.; MAHALANOBIS, P. C. *American Statistician.*, v. 26, n.4, p. 49, 1972.

DENDUKURI, N.; JOSEPH, L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics.*, v. 57, n.1, p. 158-167, 2001.

DOHERTY, M. Probability versus non-probability sampling in sample surveys. *N. Z. Stat. Rev.*, March issue, p. 21-28, 1994.

DOHOO, I.; MARTIN, W.; STRYHN, H. *Veterinary epidemiologic research*. 2 ed. Charlottetown: VER, 2009. 865 p.

EUROPEAN technology platform for global animal health. Vision 2015, 2005. Disponível em: <http://ec.europa.eu/research/agriculture/pdf/etpgah_vision2015_paper-final_en.pdf>. Acesso em: 30 setembro de 2011.

FARR, A. J.; FREEMAN, K. P. Quality control validation application of sigma metrics, and performance comparison between two biochemistry analyzer in a commercial veterinary laboratory. *J. Vet Diagn Invest.*, v. 20, n.5, p.536-544, 2008.

FLORKOWISK, C. M. Sensitivity, specificity, Receiver-Operating Characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin. Biochem.*, v. 29, n.1, p. 83-87, 2008.

FROSSLING, J., BONNETT, B., LINDBERG, A., BJORKMAN, C. Validation of *Neospora caninum* iscom ELISA without a gold standard. *Prev. Vet. Med.*, v. 57, n.3, p.141-153, 2003.

FUTURE trends in veterinary public health. Geneva: WHO, 2002. Technical Report Series.

GARDNER, I. A. The utility of Bayes' theorem and bayesian inference in veterinary clinical practice and research. *Aust. Vet. J.*, v. 80, n. 12, p. 758-761, 2002.

GARDNER, I. A.; GREINER, M. Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. *Vet. Clin. Pathol.*, v. 35, n. 1, p. 8-17, 2006.

GARDNER, I. A. Quality standards are needed for reporting of test accuracy studies for animal diseases. *Prev. Vet. Med.*, v. 97, p. 136-143, 2010.

GIL, A.C. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas, 2006. 206 p.

GREINER, M.; GARDNER, I. A. Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.*, v. 45, n.1-2, p. 3-22, 2000.

GUIDANCE document on the validation and international acceptance of or update test methods for hazard assessment. Environment, Health and Safety Publications. Series on Testing and Assessment, n.34, 2005. [ENV/JM/MONO(2005)14] Disponível em: <<http://www.oecd.org>> Acesso em: 12 dezembro de 2011.

GUIDELINE – ordering, performing and interpreting laboratory tests in veterinary clinical practice. Ontario: CVO, 2007.

GUIDELINES for the validation and verification of quantitative and qualitative test methods. NATA, 2012. Technical Note, 17. Disponível em: <<http://www.nata.com.au/component/content/article/44-june-2012/493-nata-technical-note-17-guidelines-for-the-validation-and-verification-of-quantitative-and-qualitative-test-methods>>. Acesso em: 5 setembro de 2012.

HOFFMANN, S.; EDLER, L.; GARDNER, I. A. *et al.* Points of reference in the validation process. *Atla.*, v. 36, n.3, p. 343-352, 2008.

INTERNATIONAL organization for standardization. Calibration and testing laboratory accreditation systems — General requirements for operation and recognition. ISO/IEC 17025, 2005.

INTERNATIONAL organization for standardization. How does ISO develop standards (ISO), 2013. Disponível em: <<http://www.iso.org/iso/home/standardsdevelopment.htm>>. Acesso em: 8 de outubro de 2013.

INTERNATIONAL vocabulary of metrology — basic and general concepts and associated terms. (VIM). *Vocabulaire international de métrologie*, 2008. Disponível em: <<http://www.bipm.org/utls/common/docum>

ents/jcgm/JCGM_200_2008.pdf>. Acesso em: 6 setembro de 2011.

JACOBSON, R. H. Validation of serological assays for diagnosis of infectious diseases. *Rev. Sci. Technol.*, v. 17, n.2, p. 469-526, 1998.

JOSEPH, L.; GYORKOS, T. W.; COUPAL, L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am. J. Epidemiol.*, v. 141, n. 3, p. 263-272, 1995.

KNOTTNERUS, J. A.; MURIS, J. W. Assessment of the accuracy of diagnostic tests: the cross-sectional study. In KNOTTNERUS, J. A. Ed. *The evidence base of clinical diagnosis*. London: BMJ Publishing Group, 2002. p. 39-59.

LIKERT, R.A. A technique for the measurement of attitudes. *Arch. of Psychol.*, v. 140, p. 5-55, 1932.

LUNA, S. V. Sobre o problema de pesquisa. *Chronos*, v. 26, n. 2, p. 80-92, 1993.

MARCONI, M. A.; LAKATOS, E. M. Fundamentos de metodologia científica. São Paulo: Atlas, 2005. 311p.

MCGINN, T.; WYER, P. C.; NEWMAN, T. B. *et al.* Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *Can. Med. Assoc. J.*, v. 171, n. 11, p. 1369-1373, 2004.

MEASUREMENT of uncertainty: standardising new approaches for validation of diagnostic assays and estimation of measurement uncertainty in veterinary diagnostic testing, 2010. Disponível em: <http://www.scahls.org.au/guidelines/measurement_of_uncertainty>. Acesso em: 30 setembro de 2011.

MORGAN, A. P. Regulatory control of veterinary diagnostic test kits. *Rev.Sci. Tech. Off. Int. Epiz.*, v. 17, n. 2, p.562-567, 1998.

NIELSEN, S. S.; TOFT, N.; GARDNER, I. A. Structured approach to design of diagnostic test evaluation studies for chronic progressive infections in animals. *Vet. Microbiol.*, v. 150, n.1, p.115-125, 2011.

OFFICE INTERNATIONAL DES EPIZOOTIES (OIE). World Organisation for Animal Health. Manual of diagnostic tests and vaccines for terrestrial animals. 5th ed. Paris: OIE, 2010. Disponível em: <http://www.oie.int/eng/normes/manual/A_00082.htm>. Acesso em: 20 agosto de 2011.

OFFICE INTERNATIONAL DES EPIZOOTIES (OIE). World Organisation for Animal Health. Collaborating Centres. Paris: OIE, 2011a. Disponível em: <<http://www.oie.int/our-scientific-expertise/collaborating-centres/mandate-and-internal-rules/>>. Acesso em: 16 junho de 2014.

OFFICE INTERNATIONAL DES EPIZOOTIES (OIE). World Organisation for Animal Health. Guidance for the Management of OIE Reference Centre Networks. Paris: OIE, 2011b. Disponível em: <<http://www.oie.int/our-scientific-expertise/reference-laboratories/reference-centre-networks/>>. Acesso em: 16 junho de 2014.

OFFICE INTERNATIONAL DES EPIZOOTIES (OIE). World Organisation for Animal Health. Guidelines for applicants for OIE Reference Laboratory status. Paris: OIE, 2011c. Disponível em: <<http://www.oie.int/our-scientific-expertise/collaborating-centres/guidelines-for-applicants/>>. Acesso em: 16 junho de 2014.

OFFICE INTERNATIONAL DES EPIZOOTIES (OIE). World Organisation for Animal Health. Reference Laboratories. Paris: OIE, 2011d. Disponível em: <<http://www.oie.int/our-scientific-expertise/reference-laboratories/introduction/>>. Acesso em: 16 junho de 2014.

OFFICE INTERNATIONAL DES EPIZOOTIES (OIE). World Organisation for Animal Health. Terms of Reference. Paris: OIE, 2011e. Disponível em <<http://www.oie.int/our-scientific-expertise/reference-laboratories/introduction/>>. Acesso em: 16 junho de 2014.

OFFICE INTERNATIONAL DES EPIZOOTIES (OIE). World Organisation for Animal Health. Manual of diagnostic tests and vaccines for terrestrial animals. 6th ed. Paris: OIE, 2013. Disponível em: <http://www.oie.int/eng/normes/manual/A_00082.htm>. Acesso em: 20 janeiro de 2014.

PAGANO, M.; GAUVREAU, K. *Princípios de bioestatística*. 2. ed. São Paulo: THONSON, 2004. 506 p.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística bayesiana*. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.

PEPE, M. S. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press, 2003. 302 p.

PEREIRA, J.C.R.; *Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais*. 3 ed. São Paulo: Editora da Universidade de São Paulo, 2004. 156p.

PIZA, A. S. T, PEREIRA, A. R, TERRERAN, M. T. *et al.* Serodiagnosis of equine infectious anemia by agar gel immunodiffusion and ELISA using a recombinant p26 viral protein expressed in Escherichia Coli as antigen. *Prev. Vet. Med.*, v. 78, n.3, p. 239-245, 2007.

QUALITY assurance for research and development and non-routine analysis. 2 ed. 1998. 61p. Disponível em: <<http://www.eurachem.org/images/stories/G>

- uideo/pdf/rdguide.pdf> Acesso em: 8 de outubro de 2013.
- SCAHL'S validation template for serological assays, 2012. Disponível em: <<http://www.scahls.org.au/LabTests/Pages/Validation-Templates.aspx>> Acesso em: 5 de setembro de 2012.
- VALIDATION and certification of diagnostic assays. OIE, 2009. Disponível em: <http://www.oie.int/fileadmin/Home/eng/Our_scientific_expertise/docs/pdf/A_SOP_2012_web.pdf>. Acesso em: 9 novembro de 2012.
- VALIDATION of analytical methods for food control. Vienna, AUSTRIA: FAO, 1997. A Report of a Joint FAO/IAEA Expert Consultation.
- VALIDATION of analytical procedures: definition and terminology (VICH, GL1), 1998. Brussels, Belgium: VICH, 1998.
- VALIDATION of analytical procedures: methodology (VICH, GL2), 1999. Brussels, Belgium: VICH, 1999.
- VETERINARY biologics guideline 3.19E: guideline for licensing of veterinary diagnostic tests kits in Canada. 2012. Disponível em: <<http://www.inspection.gc.ca/animals/veterinary-biologics/guidelines-forms/3-19e/eng/1328630364471/1328630723713>>. Acesso em: 4 de julho de 2012.
- VIEIRA, H. C.; CASTRO, A. E.; SCHUCH JUNIOR, V. F. O uso de questionários via e-mail em pesquisas acadêmicas sob a ótica dos respondentes. In: Seminário em Administração, XIII SEMEAD, n.13., 2010., São Paulo. *Anais...* São Paulo: FEA-SP, 2010. p. 1-13.
- WESTGARD, J. O. Multirule and "Westgard Rules" What are They? Copyright Westgard QC, Inc, 2002. Disponível em: <http://westgard.com>. Acesso em : 8 de outubro de 2013.
- WHITING, P.; RUTJES, A. W. S.; DINNES, J. et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol. Assess.*, v. 8, n. 25, p. 1-234, 2004.
- WIEGERS, A. L. The age of competence: an update on the international laboratory accreditation scene for veterinary testing laboratories. *J. Vet. Diagn. Invest.*, v. 14, n.2, p. 89-96, 2002.
- WIEGERS, A. L. Valid methods: the quality assurance of test method development, validation, approval, and transfer for veterinary testing laboratories. *J. Vet. Diagn. Invest.*, v. 15, n.4, p. 303-310, 2003.
- WIEGERS, A. L. The quality assurance of proficiency testing programs for animal disease diagnostic laboratories. *J. Vet. Diagn. Invest.*, v. 16, n.4, p. 255-263, 2004.
- YANG, Z.; LI, L.; PAN, Z. Development of multiple ELISAs for the detection of antibodies against classical swine fever virus in pig sera. *Viol. Sin.*, v. 27, n. 1, p. 48-56, 2012.

7. ANEXOS

7.1 Anexo 1- Termo de Consentimento Livre e Esclarecido

PARA MAIORES DE 18 ANOS

Caro professor (a),

Você está sendo convidado (a) para participar, como voluntário, da pesquisa “Validação de métodos sorológicos de diagnósticos laboratoriais, aplicados a estudos de viroses dos animais”. As necessidades de validação de ensaios diagnósticos e, da completude e da clareza das orientações para conduzir essa validação nos levaram a realização desse estudo. Para tal, avaliamos a metodologia e os testes utilizados para validar um teste diagnóstico. Além do estudo descritivo das normas internacionais realizado pelos pesquisadores será aplicado um questionário. Esse questionário com perguntas referentes a método geral de validação avaliando composição do espectro, teste de referência, análise e planejamento de pesquisa, será enviado de forma eletrônica. Após a análise e divulgação dos resultados os questionários serão destruídos. Foram selecionados 38 professores de dez universidades brasileiras.

O estudo proposto garante o anonimato, não serão divulgados os nomes dos professores participantes, e não há riscos ou prejuízos decorrentes da sua participação. A coleta de respostas do questionário eletrônico a ser aplicado foi elaborada de forma que não é identificado de qual participante ou universidade foi originada a resposta recebida. Dessa forma, evitando possíveis riscos, constrangimentos ou prejuízos pelo fato de seu preenchimento. Além disso, foi solicitado às universidades que fazem parte da amostragem autorização para realização do estudo.

A qualquer momento, você pode se retirar do estudo sem justificar sua decisão. Você terá acesso aos pesquisadores para esclarecimentos de eventuais dúvidas.

Sua colaboração é fundamental. Espera-se que a análise das informações sirva como subsídio para identificar impactos das diferentes abordagens de validação de testes diagnósticos, permitindo construir um modelo simples, completo e com aplicação prática em pesquisa e desenvolvimento.

Após ser esclarecido (a) sobre as informações acima, responda a primeira questão do questionário e você será automaticamente considerado como participante do estudo.

Assinatura: _____

Orientador: Prof. João Paulo Amaral Haddad

Médico Veterinário/Escola de Veterinária/UFGM

Telefone: (31)3409-2125 e-mail: jphaddad@vet.ufmg.br

Assinatura: _____

Doutoranda: Rejane Silva Diniz

Curso Ciência Animal /Epidemiologia/ Escola de Veterinária/UFMG

Tel.: 3409-2125 – e-mail: rejaned@ufmg.br

End. COEP/UFMG: Av. Pres. Antonio Carlos, 6627 – Unidade Administrativa II, 2º andar – Sala 2005 –
CEP: 31270-901 – BH – MG – Tele fax: (031) 3409-4592 -e-mail: coep@prpq.ufmg.br

Consentimento da participação:

Declaro que fui devidamente informado (a) e esclarecido (a) sobre a pesquisa, os procedimentos nela envolvidos, a não existência de riscos, e sobre garantia de confidencialidade e esclarecimentos decorrentes de minha participação. Foi-me garantido que posso desistir da minha participação a qualquer momento, sem qualquer penalidade. Concordo em participar do estudo.

7.2 Anexo 2- Questionário Validação

Consentimento de participação

Declaro que fui devidamente informado (a) e esclarecido (a) sobre a pesquisa, os procedimentos nela envolvidos, a não existência de riscos, e sobre garantia de confidencialidade e esclarecimentos decorrentes de minha participação. Foi-me garantido que posso desistir da minha participação a qualquer momento, sem qualquer penalidade. Concordo em participar do estudo.

- Aceito Não aceito

Método Diagnóstico

Marque o método que você utiliza com maior frequência.

- ELISA
- Western Blot
- Soroneutralização
- Outros

Documento de validação

Qual documento de validação abaixo você já leu pelo menos uma vez?

- Manual de validação da Organização Internacional de Saúde Animal (OIE).
- Instrução Normativa do Ministério da Agricultura.
- Cooperação Internacional para a Harmonização de Requisitos Técnicos para o Registro de Produtos Veterinários (VICH).
- Nenhum
- Outro:

Avaliação da acurácia diagnóstica

Marque a(s) recomendação(ões) que você utiliza para avaliar a acurácia de métodos diagnósticos veterinários.

- | | Marque a(s)
opção(ões) |
|-----------------------|---------------------------|
| Normas internacionais | <input type="radio"/> |
| Normas nacionais | <input type="radio"/> |
| Artigos | <input type="radio"/> |
| Nenhuma | <input type="radio"/> |

Tempo em pesquisa veterinária

Há quanto tempo você trabalha com pesquisa em medicina veterinária?

- Menos de 3 anos Entre 3 e 6 anos Acima de 6 anos

Etapas de validação de métodos

Avalie a importância das etapas de validação de métodos diagnósticos. Do (1) menos importante ao (5) mais importante.

	1	2	3	4	5
Amostragem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Estudos de Imprecisão	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Estudos de Acurácia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Escolha do método de referência	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Estudos primários	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Disciplina de Pós-Graduação

No curso de Pós-Graduação onde você está inserido existe disciplina relacionada à qualidade assegurada?

- Sim Não Não tenho certeza

Formação acadêmica

Análises epidemiológicas de testes diagnósticos faz parte da sua formação acadêmica?

- Sim Não Não tenho certeza

Ferramentas estatísticas

Marque abaixo todas as ferramentas estatísticas que você utiliza na avaliação de desempenho de testes diagnósticos.

	Marque as opções
Coeficiente de variação	<input type="radio"/>
Regressão linear	<input type="radio"/>
Kappa	<input type="radio"/>
Teste de Qui-Quadrado	<input type="radio"/>
Modelo Bayesiano	<input type="radio"/>
Coeficiente de correlação	<input type="radio"/>
Teste t de student	<input type="radio"/>
Análise de variância	<input type="radio"/>
Outra	<input type="radio"/>

Marque o grau de concordância dos itens abaixo ao você desenvolver e validar um método.

Composição do espectro

	Discordo totalmente	Discordo moderadamente	Indiferente	Concordo moderadamente	Concordo plenamente
O espectro de pacientes foi descrito no artigo e foi escolhido adequadamente.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Os critérios de seleção foram descritos com clareza.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A prevalência e a gravidade da doença foi reportada.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Teste índice e padrão de referência: seleção e execução

	Discordo totalmente	Discordo moderadamente	Indiferente	Concordo moderadamente	Concordo plenamente
No estágio tecnológico da época, o padrão de referência escolhido era adequado para a verificação dos resultados do teste.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
É possível que uma mudança na tecnologia do teste índice tenha ocorrido desde a publicação deste artigo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Houve um período anormalmente longo entre a condução do teste em avaliação e a confirmação do diagnóstico segundo o padrão de referência.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A execução do teste índice foi descrita em detalhe suficiente para permitir sua replicação.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A amostra total ou uma fração aleatória da amostra foi verificada usando-se um padrão	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Discordo totalmente	Discordo moderadamente	Indiferente	Concordo moderadamente	Concordo plenamente
referencial de diagnóstico.					
O valor de corte foi predeterminado ou era aceitável à luz de trabalhos anteriores.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Teste índice e padrão de referência: interpretação

	Discordo totalmente	Discordo moderadamente	Indiferente	Concordo moderadamente	Concordo plenamente
Havia dados clínicos disponíveis quando os resultados do teste foram interpretados.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Os dados estavam sujeitos a variação do observador ou instrumental que pudesse ter afetado as estimativas de desempenho do teste.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Análise

	Discordo totalmente	Discordo moderadamente	Indiferente	Concordo moderadamente	Concordo plenamente
Apresentou-se uma medida de precisão dos resultados.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Discordo totalmente	Discordo moderadamente	Indiferente	Concordo moderadamente	Concordo plenamente
Os resultados foram apresentados numa tabela 2.2.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ofereceu-se alguma indicação da utilidade do teste.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Planejamento da pesquisa

	Discordo totalmente	Discordo moderadamente	Indiferente	Concordo moderadamente	Concordo plenamente
Foi feito um cálculo de tamanho adequado da amostra.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Os objetivos de estudo foram relatados claramente.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Havia alguma evidência de que um protocolo de estudo teria sido desenvolvido antes de o estudo ser iniciado.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Obrigada pela sua participação!

7.3 Anexo 3- Modelo de correspondência



7.4 Anexo 4- Manual de Teste de Diagnóstico e Vacinas para Animais Terrestres (OIE..., 2010)

This thoroughly revised chapter replaces Chapter 1.1.4 Principles of validation of diagnostic assays for infectious diseases and Chapter 1.1.5 Validation and quality control of polymerase chain reaction methods used for the diagnosis of infectious diseases from the sixth edition of the OIE *Terrestrial Manual*. Adopted in May 2009.

CHAPTER 1.1.4/5.

PRINCIPLES AND METHODS OF VALIDATION OF DIAGNOSTIC ASSAYS FOR INFECTIOUS DISEASES

INTRODUCTION

Validation is a process that determines the fitness of an assay, which has been properly developed, optimised and standardised, for an intended purpose. Validation includes estimates of the analytical and diagnostic performance characteristics of a test. In the context of this chapter, an assay that has completed the first three stages of the validation pathway (see Figure 1 below), including performance characterisation, can be designated as "validated for the original intended purpose(s)"¹. To maintain a validated assay status, however, it is necessary to carefully monitor the assay's daily performance, often by tracking the behaviour of internal controls over time. This ensures that the assay, as originally validated, consistently maintains its performance characteristics. Should it no longer produce results consistent with the original validation data, the assay may be rendered unfit for its intended purpose. Thus, a validated assay is continuously assessed to assure it maintains its fitness for purpose through assessment of results of internal controls in each run of the assay.

Assay, test method, and test are synonymous terms for purposes of this chapter, and therefore are used interchangeably.

The terms **"valid"** (adjective) or **"validity"** (noun) refer to whether estimates of test performance characteristics are unbiased with respect to the true parameter values. These terms are applicable regardless of whether the measurement is quantitative or qualitative.

Assays applied to individuals or populations have many purposes, such as aiding in: documenting freedom from disease in a country or region, preventing spread of disease through trade, eradicating an infection from a region or country, confirming diagnosis of clinical cases, estimating infection prevalence to facilitate risk analysis, identifying infected animals toward implementation of control measures, and classifying animals for herd health or immune status post-vaccination. A single assay may be validated for one or several intended purposes by optimising its performance characteristics for each purpose, e.g. setting diagnostic sensitivity (DSe) high, with associated lower diagnostic specificity (DSp) for a screening assay, or conversely, setting DSp high with associated lower DSe for a confirmatory assay.

The ever-changing repertoire of new and unique diagnostic reagents coupled with many novel assay platforms and protocols has precipitated discussions about how to properly validate these assays. It is no longer sufficient to offer simple examples from serological assays, such as the enzyme-linked immunosorbent assay, to guide assay developers in validating the more complex assays, such as nucleic acid detection tests. In order to bring coherence to the validation process for all types of assays, this chapter focuses on the criteria that must be fulfilled during assay development and validation of all assay types. The inclusion of assay development as part of the assay validation process may seem counterintuitive, but in reality, three of the validation criteria that must be assessed in order to achieve a validated assay, comprise steps in the assay development process. Accordingly the assay development process seamlessly segues into an assay validation pathway, both of which contain validation criteria that must be fulfilled. This chapter also provides guidance for evaluation of each criterion through provision of best scientific practices contained in the chapter's appendices. The best practices are tailored for each of several fundamentally different types of assays (e.g. detection of nucleic acids, antibodies, or antigens).

¹ **Validation** does not necessarily imply that test performance meets any minimum value or that the test has comparable performance to any comparative test, unless this has been specifically considered in the design of the test evaluation study.

DIRECT AND INDIRECT METHODS THAT REQUIRE VALIDATION

The diagnosis of infectious diseases is performed by direct and/or indirect detection of infectious agents. By direct methods, the particles of the agents and/or their components, such as nucleic acids, structural or non-structural proteins, enzymes, etc., are detected. The indirect methods demonstrate antibodies or cell-mediated immune responses induced by exposure to infectious agents or their components. The most common indirect methods of infectious agent detection are antibody assays such as classical virus neutralisation, antibody enzyme-linked immunosorbent assay (ELISA), haemagglutination inhibition, complement fixation, and the recently appearing novel methods, such as biosensors, bioluminometry, fluorescence polarisation, and chemoluminescence,

The most common direct detection methods are isolation or *in-vitro* cultivation of viable organisms, electron microscopy, immunofluorescence, immunohistochemistry, antigen-ELISA, Western immunoblotting, and nucleic acid detection systems (NAD). The NAD systems include nucleic-acid hybridisation (NAH), macro- and microarrays and the various techniques of nucleic acid amplification, such as the polymerase chain reaction (PCR), or the isothermal amplification methods, such as nucleic acid sequence-based amplification (NASBA), and invader or loop-mediated isothermal amplification (LAMP). NAD assays are rapidly becoming commonplace and in many cases replacing virus isolation and bacteria cultivation, particularly for the detection of agents that are difficult or impossible to culture. NAD tools are also used as a secondary means for highly specific identification of strains, groups, or lineages of organisms following isolation or culture of viruses, bacteria and parasites. Molecular diagnostics, such as PCR, do not require: a) the presence of replicating organisms, b) expensive viral isolation infrastructure, c) up to several weeks to achieve a diagnosis, or d) special expertise, which is often unavailable in many laboratories – all practical advantages. These methods have become relatively inexpensive, safe and user-friendly (1–4, 6, 7, 16, 20, 21). Various real-time PCR methods, nucleic acid extraction robots, and automated workstations for NAD, antibody, antigen, and agent detection have resulted in a large repertoire of high throughput, robust, very rapid and reliable assays. Although NAD systems often have the advantage of a greater diagnostic sensitivity and analytical specificity than infectious agent recovery or antigen-capture ELISA procedures, that advantage usually carries with it a greater challenge for validation of such assays.

PRELIMINARY CONSIDERATIONS IN ASSAY DEVELOPMENT AND VALIDATION

The first consideration is to define the purpose of the assay, because this guides all subsequent steps in the validation process. By considering the variables that affect an assay's performance, the criteria that must be addressed in assay validation become clearer. The variables can be grouped into three categories: (a) the sample – individual or pooled, matrix composition, and host/organism interactions affecting the target analyte quantitatively or qualitatively; (b) the assay system – physical, chemical, biological and operator-related factors affecting the capacity of the assay to detect a specific analyte in the sample; and (c) the test result interpretation – the capacity of a test result, derived from the assay system, to predict accurately the status of the individual or population relative to the analyte in question.

Assay validation criterion: a characterising trait of an assay; a decisive factor, measure or standard upon which a judgment or decision may be based.

The matrix in which the targeted analyte may reside (serum, faeces, tissue, etc.) may contain endogenous or exogenous inhibitors that prevent enzyme-dependent tests such as PCRs or ELISAs from working. Other factors that affect the concentration and composition of analyte (mainly antibody) in the sample may be mainly attributable to the host and are either inherent (e.g. age, sex, breed, nutritional status, pregnancy, immunological responsiveness) or acquired (e.g. passively acquired antibody, active immunity elicited by vaccination or infection). Non-host factors, such as contamination or deterioration of the sample, also affect the ability of the assay to detect the specific targeted analyte in the sample.

Factors that interfere with the analytical performance of the assay system include instrumentation, operator error, reagent choice (both chemical and biological) and calibration, accuracy and acceptance limits of assay controls, reaction vessels and platforms, water quality, pH and ionicity of buffers and diluents, incubation temperatures and durations, and error introduced by detection of closely related analytes. It is also important that biological reagents are free of extraneous agents.

Factors that may negatively impact diagnostic performance of the assay are primarily associated with choice of reference sample panels from known infected/exposed or known uninfected animals selected for evaluating the diagnostic sensitivity (DSe) and diagnostic specificity (DSp) of the assay. This is particularly difficult because the degree to which the reference animals represent all of the host and environmental variables in the population targeted by the assay has a major impact on the confidence of test-result interpretation. For example, experienced serologists are aware that an assay, validated by using serum samples from northern European cattle, may not give valid results for the distinctive populations of cattle in Africa. Diagnostic performance of the assay is further complicated when sample panels of known infection status are not available, often because they

are impossible to obtain. In this situation, DSe and DSp may be estimated, in certain circumstances by use of latent class models (10, 14 and Appendix 1.1.4.5).

THE CRITERIA OF ASSAY DEVELOPMENT AND VALIDATION

Assay performance is affected by many factors that span from the earliest stages of assay development through the final stage of performance assessment when the test is applied to targeted populations of animals. An assay, therefore, cannot be considered validated unless a specific set of essential validation criteria (see accompanying box) have been tested and affirmed or fulfilled, either quantitatively or qualitatively (for detail on terms, see glossary in reference 25). Lack of attention to any one of these criteria will likely reduce the level of confidence that an assay is fulfilling the purpose(s) for which it is intended. The first four of these criteria typically are addressed during development of the assay (the Development Pathway), and the remaining eight are evaluated during the first three stages of assay validation (the Validation Pathway) as described below.

Assay validation criteria

1. Fitness for intended purpose(s)
2. Optimisation
3. Standardisation
4. Robustness
5. Repeatability
6. Analytical sensitivity
7. Analytical specificity
8. Thresholds (cut-offs)
9. Diagnostic sensitivity
10. Diagnostic specificity
11. Reproducibility
12. Ruggedness

A. ASSAY DEVELOPMENT PATHWAY

1. Definition of the intended purpose(s) for an assay

The OIE *Standard for Management and Technical Requirements for Laboratories Conducting Tests for Infectious Diseases* (25) states that test methods and related procedures must be appropriate for specific diagnostic applications in order for the test results to be of any relevance. In other words, the assay must be 'fit for purpose'². The capacity of a positive or negative test result to predict accurately the infection or exposure status of the animal or population of animals is the ultimate consideration of assay validation. This capacity is dependent on development of a carefully optimised (Section A.2.d), and standardised (Section A.2.g) assay that, through accrual of validation data, provides less biased and more precise estimates of DSe and DSp. These estimates, along with evidence-based data on prevalence of infection in the population being tested, are the basis for providing a high degree of confidence in the predictive values of positive and negative test results. In order to insure that test results provide useful diagnostic inferences about animal population infection/exposure status, the validation process encompasses initial development and assay performance documentation, as well as ongoing assessment of quality control and quality assurance programmes. Figure 1 shows the assay validation process, from assay design through the development and validation pathways to implementation, deployment, and maintenance of the assay

As outlined in the background information in *Certification of diagnostic assays* on the OIE website (www.oie.int), the first step is selection of an assay type that likely can be validated for a particular use.

a) Fitness for purpose

The most common purposes are to:

- 1) Demonstrate freedom from infection in a defined population (country/zone/compartments/herd) (prevalence apparently zero):
 - 1a) 'Free' with and/or without vaccination,
 - 1b) Re-establishment of freedom after outbreaks
- 2) Certify freedom from infection or presence of the agent in individual animals or products for trade/movement purposes.
- 3) Eradication of disease or elimination of infection from defined populations.
- 4) Confirmatory diagnosis of suspect or clinical cases (includes confirmation of positive screening test).

² This is a specific interpretation of the more generally stated requirements of the ISO/IEC 17025:2005 international quality standard for testing laboratories (18). The OIE Standard further states that in order for a test method to be considered appropriate, it must be properly validated and that this validation must respect the principles outlined in the validation chapters of the *Terrestrial & Aquatic Manuals*.

- 5) Estimate prevalence of infection or exposure to facilitate risk analysis (surveys, herd health status, disease control measures).
- 6) Determine immune status of individual animals or populations (post-vaccination).

These purposes are broadly inclusive of many narrower and more specific applications of assays (see Appendices for each assay type for details). Such specific applications and their unique purposes need to be clearly defined within the context of a fully validated assay.

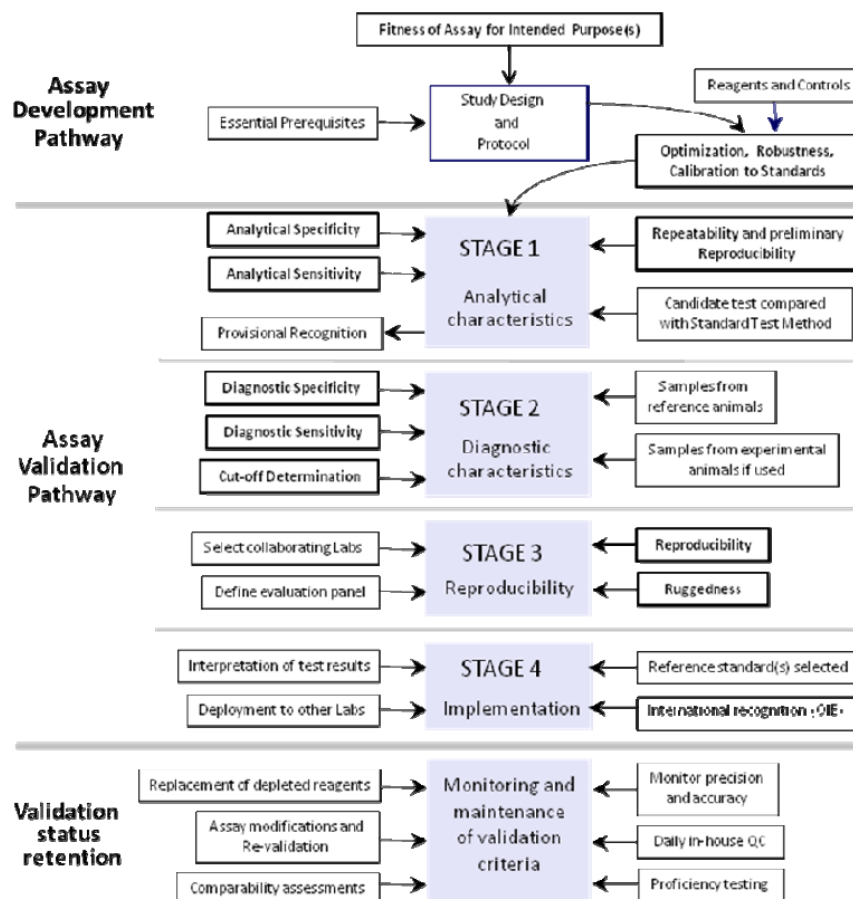


Figure 1. The assay development and validation pathways with assay validation criteria highlighted in bold typescript within shadowed boxes.

b) Fitness for use

While this chapter deals with validation and fitness for purpose from a scientific perspective, it should also be noted that other practical factors might impact the relevance of an assay with respect to its intended application. These factors include not only the diagnostic suitability of the assay, but also its acceptability by scientific and regulatory communities, acceptability to the client, and feasibility given available laboratory resources. An inability to meet operational requirements of an assay also may make it unfit for its intended use. Such requirements may include performance costs, equipment availability, level of technical sophistication and interpretation skills, kit/reagent availability, shelf life, transport requirements, safety,

biosecurity, sample throughput, turn-around times for test results, aspects of quality control and quality assurance, and whether the assay can practically be deployed to other laboratories. Test kits used in the field are highly desirable from an ease-of-use viewpoint, but because they are performed outside the confines of a controlled laboratory environment, they require added precautions to maintain fitness for purpose (8).

2. Assay development – the experimental studies

a) Essential prerequisites: factors that impact assay validation

i) **Quality assurance:** Whether developing assays in the laboratory or performing analyses of clinical material, the objective is to produce data of high quality. This requires that key requirements have to be fulfilled within the laboratory (see Chapters 1.1.3 & 1.1.1 of the *Terrestrial & Aquatic Manuals*, respectively). The establishment of quality assurance (QA) and quality control (QC) systems is essential, i.e. a set of quality protocols, including the use of assay control samples that ensure that the system is working properly and confirms data reproducibility and quality. QA and QC systems, together with trained and competent personnel, have already been established in many laboratories world-wide.

ii) **Equipment selection:** Equipment that is not maintained and calibrated can be a major impediment to achieving a quality assay. Apparatus (freezers, heating blocks, incubators, refrigerators, optical colorimeters, thermocyclers, plate washers, pipettes, etc.) must be calibrated according to the laboratory's quality assurance protocols. Examples of this need include robotics used for automation of entire assays, or parts thereof, for routine diagnostic processing. It is not sufficient to assume that robotic extraction of nucleic acid, for example, is equivalent to previously used manual extraction methods or that an automated ELISA plate washer provides uniform washing among wells of the plate. The instrument must be calibrated and the protocol validated to confirm performance efficiency and to assure cross-contamination does not occur in NAD systems or washing is adequate for all wells in a plate. (See Appendices on best practices for more details).

iii) Selection and integrity of samples

Selection, collection, preparation and management of samples are critical variables in designing, developing, and validating an assay. Other variables such as transport, chain of custody, tracking of samples, and the laboratory information management system are also critical sources of variation/error that become especially important when the assay is implemented for routine testing. Integrity of experimental outcomes during assay development and validation is only as good as the quality of the samples used in experimentation or routine diagnostic testing. Anticipating the factors that can negatively impact sample quality must precede launching an assay validation effort. Reference samples used in assay development and validation should be in the same matrix used in the assay (e.g. serum, tissue, whole blood) and representative of the species to be tested by the resulting assay. The reference materials should appropriately represent the range of analyte concentration to be detected by the assay. Details on proper sample collection, preparation, management, and transport are available in the Chapter 1.1.1 of the *Terrestrial Manual*.

b) Test method design and proof of concept

Considerable thought and planning needs to go into designing all steps of a new assay destined for validation, or an existing assay that is being modified. Assistance is offered in Appendices to this chapter, which cover best practices for development and validation of assays for detection of various analytes (e.g. antibody, antigen, and nucleic acid detection).

i) Analyte reference samples

Development of all assays is dependent on analyte reference samples that reflect the target analyte and the matrix in which the analyte is found in the population for which the assay is intended. The reference samples may be sera, fluids or tissues that contain the analyte of interest or a genomic construct consistent with the target analyte. These reference materials are used in experiments conducted throughout the development process and carried over into the validation of the assay.

Analyte reference samples, containing the analyte of interest in varying concentrations, are useful in developing and evaluating the candidate assay's validation criteria.

c) Operating range of the assay

During development of the assay, the lower and upper detection limits are established. To formally establish this range, a high positive reference sample is selected. (Ideally, this sample will be the same one from among the three samples described under "Optimisation" below). This high positive sample is serially diluted to extinction in an analyte-negative

Operating range of an assay: an interval of analyte concentrations (amounts) over which the method provides suitable accuracy and precision.

matrix of the same constitution as the sample matrix of samples from animals in the population targeted by the assay. The results are plotted as a 'response-curve', with the response (e.g., OD, Ct, etc) a function of analyte concentration (amount). The curve, establishes the range of the assay, which is the interval between the upper and lower concentration (amounts) of analyte in the sample for which a suitable level of precision³ and accuracy has been demonstrated. For most diagnostic assays, the response is the result of interaction of the analyte with an antibody or other binding reagent. These are known as ligand binding assays (LBAs). The typical calibration curve for LBAs is sigmoidal in shape, with a lower boundary (asymptote) near the background response (non-specific binding) and an upper asymptote near the maximum response. Typically, LBA data are transformed to approximate a linear relationship between response and concentration. This transformation simplifies interpolation of data using linear regression analysis, but with the disadvantage of introduced bias. Since linearization is imperfect leading to compromised estimates of accuracy and precision, numerous data-fitting algorithms have been applied to experimental calibration curve data from LBAs (11). The currently accepted reference model for calibration of LBAs is the 4-parameter logistic model, which usually optimizes accuracy and precision over the maximum usable calibration range (11). Such transformations are now more practical for the general user because of many user-friendly statistical software programs available on the internet.

Precision is the degree of dispersion among a series of measurements of the same sample tested under specified conditions (see footnote 3 for more detail)

Accuracy is the closeness of a test value to the expected (true) value for a reference standard reagent of known concentration or titer

d) Optimisation

It is useful to select at least three well-defined reference samples, representing the analyte ranging from high positive to negative (e.g., negative, low- and high-positive). These samples ideally should represent known infected and uninfected animals from the population that eventually will become the target of the assay once it is validated. Obtaining such reference samples, however, is not always possible, particularly for nucleic acid and antigen detection assays. The alternative of preparing reference samples spiked with cultured agents or positive sera is inferior because the matrix of field samples may be very different from spiked-sample matrix. But, when no other alternative exists, spiking a sample with a known amount of the analyte derived from culture, or diluting a high positive serum in negative serum of the same species may be all that is available. In either case, it is imperative that the matrix, into which analyte is placed or diluted, is identical to, or resembles as closely as possible the samples that ultimately will be tested in the assay. Ideally, reference samples have been well characterised by one or preferably at least two alternate methodologies. These samples can be used in experiments to determine if the assay is able to distinguish between varying quantities of analyte, and for optimising the reagent concentrations and perfecting the protocol. In principle, for all assay types, it is highly desirable to prepare and store a sufficient amount of each reference sample in aliquots for use in every run of the candidate assay as it is evaluated through the entire development and validation process. Switching reference samples during the validation process introduces an intractable variable that can severely undermine interpretation of experimental data and, therefore, the integrity of the development and validation process.

Optimisation is the process by which the most important physical, chemical and biological parameters of an assay are evaluated and adjusted to ensure that the performance characteristics of the assay are best suited to the intended application.

The labour-intensive process of optimising an assay is fundamental and critical to achieving a quality assay. Scientific judgment and use of best scientific practices provided in accompanying Appendices to this chapter are the best assets to guide optimisation of all elements of an assay. The approach outlined provides a solid foundation for development of an assay that fulfils the criteria of "robustness" and "ruggedness" when used over extended periods of time within a laboratory or when implemented in other laboratories. Often, prototype assays are developed using reagents and equipment at hand in the laboratory. However, if the assay is intended for routine diagnostic use in multiple laboratories, optimization becomes extremely critical. Every chemical and buffer formulation must be fully described. All reagents must be defined with respect to purity and grade (including water). Acceptable working ranges must be established for parameters such as pH, molarity, etc. Likewise for biologicals, standards for quality, purity, concentration and reactivity must be defined. Shelf lives and storage conditions must also be considered for both chemicals and biologicals. Acceptable ranges for reaction times and temperatures need also be established. Essential equipment critical to assay performance must be described in detail, including operational specifications and calibration. Process (quality) control is often an add-on at the end of assay development but it should be an integral part of optimization from the very beginning. In addition to the above, downstream aspects such as data capture,

3 Precision may be evaluated in several ways by testing the same replicated sample: 1) within a plate or plates in a run of the assay, 2) between plates run concurrently within a run of the assay, 3a) between assay runs at different times in the same day or on different days under similar conditions, 3b) between assay runs on different days with different operators, 4) between laboratories. In this chapter, precision categories 1-3 are estimates of repeatability, and precision category 4 is synonymous with reproducibility. Levels 3a and 3b are also known as intermediate precision.

manipulation and interpretation may also require standardisation and optimisation. Finally, all of these parameters, once optimised, must be fully described in the test method protocol.

In some assay types, a correct assay result is fully dependent on getting a particular step in the testing process correct, requiring special attention during optimisation. A case in point is nucleic acid extraction from the sample. Both commercial (robotic, spin columns, and magnet-based extractions, etc.) and standard chemistry-based methods are used for DNA or RNA extraction. It is crucial to determine the most reproducible and efficient extraction method through optimisation experiments. Extraction needs to be optimised for every type of tissue that may be targeted by the assay. If the method of extraction is changed, at a minimum, comparable efficiency of extraction should be demonstrated (see Section B.6 below and associated Appendix 1.1.4.6 for additional information on establishing comparability when reagents are changed).

A variety of analyte reference samples and other process controls that are routinely included in any assay system are identified in the following sections. These provide critical assay monitoring functions that require special attention during assay optimisation. In addition, proper preparation and storage of all biological reagents and reference materials must be heeded to ensure stability (see Chapter 1.1.1 of the Terrestrial Manual).

During experimentation to optimise the assay, take note of assay parameters that have a narrow range in which they perform optimally, as these are the critical points that may affect an assay's robustness (see Section A.2.f).

e) Inhibitory factors in sample matrix

Generally, for antibody detection in serum, assays are rather resistant to inhibitory factors, with the exception of certain assays, e.g. toxic factors in viral neutralisation assays, or when endogenous substances found in certain sample types inhibit enzymatic reactions in ELISAs. For nucleic acid detection, sample matrices including blood, serum, body tissues, and swab samples allow for easy extraction of target nucleic acids, while faeces, autolysed tissues and semen samples can be more difficult to handle because of the presence of factors which can inhibit downstream assays such as PCR.

f) Robustness

Robustness refers to an assay's capacity to remain unaffected by minor variations in test situations that may occur over the course of testing in a single laboratory. It is assessed by deliberate variations in method parameters (23). Assessment of robustness should begin during assay development and optimization stages. The deliberate variations in method parameters may be addressed in experiments after optimal conditions for an assay are established. However, when multi-factorial titrations of reagents are used for optimizing the assay, indications of a compromised robustness may surface. If slight differences in conditions or reagent concentrations cause unacceptable variability, the assay most likely will not be robust. Early knowledge of this situation elicits a critical decision point for determining whether to continue with validation of the assay would be worthwhile, because if an assay is not robust within one laboratory under rather ideal conditions, it is unlikely to exhibit ruggedness (reduced reproducibility) when transferred to other laboratories (ruggedness is addressed in Section 4).

Robustness is a measure of an assay's capacity to remain unaffected by small, but deliberate, variations in method parameters, and provides an indication of its reliability during normal usage.

The factors most likely to affect assay robustness are quantitative (continuous) such as pH and temperature; qualitative (categorical) such as batch of reagents or brand of microtiter plates; and mixture-related such as aqueous or organic matrix factors (9). For ligand-binding assays (LBAs), lack of robustness is not only due to less-than-optimal concentration/amount of the bio-reagent specified in the method, but may also be due to the intrinsic characteristics of the biological reagent (e.g. monoclonal antibody affinity or polyclonal antibody avidity and/or valency). Robustness, therefore, particularly of LBA-based assays, may be affected by systematic and/or random errors (22).

Robustness testing is demonstrated on a method-by-method basis. All critical reagents are identified and subjected to a factorial assessment which compares all possible combinations of reagents. For example, in antibody detection by ELISA, factors may include concentration of antigen bound to the solid phase, conjugate dilution and several test sera representing the operating range of assay. The response of the assay with respect to these small changes shall not result in unacceptable variability. Alternatively, robustness can be demonstrated through the application of factorial design experiments (9, 26)

Robustness is further verified during Stage 1 of assay validation. When the optimized test is first run under routine laboratory conditions, this practical measure of robustness is referred to as repeatability (see Section

2.a) and it is continually monitored as part of process control procedures for the duration of the life of the assay (see Section 6.a).

g) Calibration of the assay to standard reagents

i) International and national analyte reference standards

Ideally, international reference standards, containing a known concentration of analyte, are the reagents to which all assays are standardised. Such standards are prepared and distributed by international reference laboratories. National reference standards are calibrated by comparison with an international standard reagent whenever possible; they are prepared and distributed by a national reference laboratory. In the absence of an international reference standard, a national reference standard becomes the standard of comparison for the candidate assay. These standard reagents are highly characterised through extensive analysis, and preferably the methods for their characterisation, preparation, and storage have been published in peer-reviewed publications.

ii) In-house standard reagent An in-house reference standard generally has the highest metrological quality available at a given location in a given organization, and is calibrated against an International or National standard. In the absence of either of these calibrators and to the extent possible, the in-house standard is highly characterised in the same manner as international and national analyte standards. This local in-house standard therefore becomes the best available standard, and is retained in sufficient aliquotted volumes for periodic use as the standard to which working standards are calibrated.

iii) Working standard reagent

One or more working standard reagent(s), commonly known as analyte or process controls, are calibrated to an international, national, or in-house standard reagent, and are prepared in large quantities, aliquotted and stored for routine use in each diagnostic run of the assay.

h) “Normalising” test results to a working standard(s)

Due to the inherent variation in raw test results that are often observed between test runs of the same assay or among laboratories using the same or similar assays, it is virtually impossible to directly compare (semi-) quantitative data. To markedly improve the comparability of test results both within and between laboratories, one or more working standard reagent(s) are included in each run of an assay. Raw test values for each test sample can then be converted to units of activity relative to the working standard(s) by a process called ‘normalisation’ [not to be confused with transformation of data to achieve a ‘normal’ (Gaussian) distribution]. The ‘normalized’ values may be expressed in many ways, such as a percent of a positive control (e.g. in an ELISA), or as a concentration or titer of an analyte derived from a standard curve, or as a number of targeted genomic copies also derived from a standard curve of Ct (cycle threshold) values for real time PCR. It is good practice to include working standards [or at least a reasonably well-characterized sample(s) in all runs of the assay during assay development and validation because this allows ‘normalization’ of data which provides a valid means for direct comparison of results between runs of an assay. Automated assay systems may calculate and report ‘normalized’ data by, for example, a standard curve, or by reporting the cycle number at which the cycle threshold is exceeded as in real-time PCR. For more information, see Appendices 1.1.4.1 and 1.1.4.3.

B. ASSAY VALIDATION PATHWAY

1. Definition of a validated assay.

“Validation” is a process that determines the fitness of an assay that has been properly developed, optimised and standardised for an intended purpose. Validation includes estimates of the analytical and diagnostic performance characteristics of a test. In the context of this document, an assay that has completed the first three stages of the validation pathway (Figure 1), including performance characterisation, can be designated as “validated for the original intended purpose(s)”

To retain the status of a validated assay, however, it is necessary to assure that the assay as originally validated consistently maintains the performance characteristics as defined during validation of the assay (see Section 6 below). This can be determined in a quality assurance program characterized by carefully monitoring the assay’s daily performance, primarily through precision and accuracy estimates for internal controls, and by scheduled external proficiency testing. Should the assay cease to produce results consistent with the original validation data, the assay would be rendered unfit for its intended purpose. Thus, a validated assay must be continuously assessed to assure it maintains its fitness for purpose.

2. Stage 1 – Analytical performance characteristics

Ideally, the design for studies outlined in the following sections should be done with assistance of a statistician. It is possible to design experiments that efficiently provide information on likely within- and among-laboratory sources of variation in assay precision (see footnote 3 under Section A.2.c, above) which will define the performance characteristics of the assay. Stage 1 studies for repeatability, reproducibility and assessment of analytical sensitivity (limit of detection) should be performed in a blinded fashion with random selection of samples. The choice of organisms, strains or serotypes to assess analytical specificity should reflect current knowledge and therefore inform the best possible experimental design for targeting specific analytes.

a) Repeatability

Repeatability is estimated by evaluating variation in results of replicates from a minimum of three (preferably five) samples representing analyte activity within the operating range of the assay. Each of these samples is then aliquoted into individual vessels as three identical replicates of the original sample containing the original analyte and matrix concentration. Each replicate is then run through all steps of the assay, including creating the working dilution, as though it were a test sample derived from the population targeted by the assay. It is not acceptable to prepare a final working dilution of a sample in a single tube from which diluted aliquots are pipetted into reaction vessels, or to create replicates from one extraction of nucleic acid rather than to extract each replicate before dilution into the reaction vessels. Such 'samples' do not constitute valid replicates for repeatability studies. Between-run variation is determined by using the same samples in multiple runs (approximately 20) involving two or more operators, done on at least five separate days. The variation in replicate results can be expressed as standard deviations, confidence intervals, or other possible options (see Appendix 1.1.4.4 on measures of uncertainty for assessments of repeatability).

Repeatability is the level of agreement between results of replicates of a sample both within and between runs of the same test method in a given laboratory.

b) Analytical specificity (ASp)

Analytical specificity distinguishes between the target analyte and other components in the assay in at least three distinctive ways. These are described as the selectivity, exclusivity, and inclusivity of the assay.

- **Selectivity** refers to the extent to which a method can accurately quantify the targeted analyte in the presence of: 1) interferences such as matrix components (e.g. inhibitors of enzymes in the reaction mix); 2) degradants (e.g. toxic factors); 3) non-specific binding of reactants to a solid phase, (e.g. conjugate of an ELISA adsorbed to well of microtiter plate); 4) antibodies to vaccination which maybe confused with antibodies to active infection. Such interferences may cause falsely reduced or elevated responses in the assay that negatively affect its analytical specificity. Vessman, et al (24) is a useful overview of selectivity as defined for analytical chemistry from which a modification described herein was deduced for application to diagnostic tests.
- **Exclusivity** is the capacity of the assay to detect an analyte or genomic sequence that is unique to a targeted organism, and excludes all other other known organisms that are potentially cross-reactive. This would also define a confirmatory assay.
- **Inclusivity** is the capacity of an assay to detect several strains or serovars of a species, several species of a genus, or a similar grouping of closely related organisms or antibodies thereto. It characterizes the scope of action for a screening assay.

Analytical specificity is the degree to which the assay distinguishes between the target analyte and other components that may be detected in the assay.

After eliminating interferences to the extent possible, the next step is to test a panel of samples appropriate for evaluating either inclusivity, exclusivity, or both, depending on the intended purpose of the assay.

ASp applies to both direct and indirect methods of analyte detection. If exclusivity is required, field samples should be obtained from animals infected with genetically-related, non-pathogenic organisms, but this may prove difficult or even impossible. In such cases, cultivated organisms can be used in direct detection methods, or serum from animals exposed experimentally by natural routes for indirect detection methods. Acceptable cross-reactivity is largely dependent on the intended purpose of the test and the prevalence of the cross-reactive organisms/analytes in the target population samples – which must be determined for each case (see Appendix 1.1.4.5 for more detail). For PCR, it is useful to perform computer simulation studies as an adjunct to laboratory assessment of ASp; however, such studies are not sufficient by themselves to evaluate ASp.

A factor unique to viral antibody detection is the possible antibody response of animals to carrier proteins found in vaccines – another type of interferent that may negatively affect selectivity. If such proteins are also

present in the solid phase antigen on ELISA plates, they may bind antibody developed to vaccine carrier proteins and give false-positive results (lack of exclusivity in the assay). Use of vaccine preparations as antigens in ELISAs is, therefore, not recommended. See Appendix 1.1.4.1 for specific practices to determine ASp.

c) **Analytical sensitivity (ASe)**

Analytical sensitivity is synonymous with the lower limit of detection (LOD) of an analyte in an assay. For direct-detection assays, this may be expressed as the number of genome copies, infectious dose, colony-forming units, plaque forming units, etc., of the agent that can be detected and distinguished from the result of a matrix background. Most commonly, this is expressed as a number of copies, complement-fixing units or plaque-forming units giving at least 50% positive results among the replicates of a sample for a specified volume or weight (see Appendix 1.1.4.5 for detail on LOD determination). For indirect detection assays, it is the least amount of antibody detected, usually, the penultimate dilution of sample in which the analyte is indistinguishable from the activity of a sample matrix control.

Limit of detection or analytical sensitivity, is the smallest amount of analyte in a sample that can be detected by a direct detection assay in at least 50% of the replicates for each dilution, in a dilution series of analyte in matrix.

If the intended purpose is to detect low levels of analyte or subclinical infections and it is difficult to obtain the appropriate reference materials, for example samples from early stages of the infection process, it may be useful to determine comparative analytical sensitivity by running a panel of samples on the candidate assay and on another independent assay. This would provide a relative comparison of analytical sensitivity between the assays, but care must be taken in choosing the independent assay used in the comparison to ensure that the analytes being detected (if different) demonstrate the same type of pathogenic profile in terms of time of appearance after exposure to the infectious agent, and relative abundance in the test samples chosen.

Where a new, more sensitive test is developed, it may be necessary to test serial samples taken from infected animals early after infection, on through to the development clinical or fulminating disease, and to run these in parallel with previously used tests to demonstrate the increased sensitivity. This would also provide a temporal comparison of the earliest point of detection relative to the pathogenesis of the disease.

d) **Standard test method for comparison with the candidate assay test method**

There are situations where it is not possible or desirable to continue to Stage 2 of the Validation Pathway because appropriate samples from the target population are scarce and animals are difficult to access (such as for exotic diseases). However, a small but select panel of highly characterised test samples representing the range of analyte concentration should be run in parallel on the candidate assay method and by the standard method if it exists.

A **Standard test method** is the best internationally or nationally recognized method, or in their absence, the best available method preferably published in a reputable journal.

e) **Analytical accuracy of adjunct tests or procedures**

Some test methods or procedures may be qualified for use as analytical tools in the diagnostic laboratory. These usually are secondary adjunct tests or procedures that are applied to an analyte that has been detected in a primary assay. The purpose of such analytical tools is to further characterise the analyte detected in the primary assay. Examples of such adjunct tests include virus neutralisation to type an isolated virus, or molecular sequencing to confirm a real-time PCR test result. Pathogenicity indices, haemagglutination inhibition, drug resistance determinations, etc. are other examples where adjunct tests or procedures are performed, either independent of, or as part of a primary assay.

Such adjunct tests must be validated for analytical performance characteristics (Sections A.2 through B.2.c., above), but differ from routine diagnostic tests because they do not require validation for diagnostic performance characteristics (Sections B.3 through B.5, below). The analytical accuracy of these tools may be defined by comparison with a reference reagent standard, or by characteristics inherent in the tool itself (such as endpoint titration). In all of these examples, the targeted analyte is further characterised quantitatively or qualitatively by the analytical tool.

f) **Preliminary evaluation of reproducibility**

Preliminary reproducibility estimates of the candidate assay may be useful at this time in the validation process, where only a small panel of highly characterised samples is available. This panel could be used for a limited evaluation of reproducibility to enhance provisional acceptance status for the assay. The candidate test method is then duplicated in laboratories at one or more different institutes, and the panel of samples is evaluated using the candidate assay in each of these laboratories, using the same protocol, same reagents

as specified in the protocol, and comparable equipment. This is a scaled-down version of Stage 3 of assay validation.

g) **Provisional assay recognition⁴**

Experience has shown that the greatest obstacle for continuing through Stage 2 of the Validation Pathway is the number of defined samples required to calculate DSe and DSp (see requirements for Stage 2, Diagnostic Performance, below). The formula is well known and tables are available for determining the number of samples required to estimate various levels of DSe and DSp, depending on the amount of allowable error and the level of confidence in the estimates (Table 1 and reference 19). The formula assumes that the myriad of host/organism factors that may affect the test outcome are all accounted for. Since that assumption may be questionable, the estimated sample sizes are at best minimal. For a disease that is not endemic or widespread, it may be impossible, initially, to obtain the number of samples required, but over time, accrual of additional data will allow adjustment of the threshold or if no adjustment is needed, enhance confidence in the estimates.

Provisional recognition defines an assay that has been assessed through Stage 1 for critical assay benchmark parameters: ASe, ASp, repeatability, and an estimate of reproducibility.

Historical precedent would suggest that assays were generally the product of laboratory experiments with an emphasis on analytical sensitivity and analytical specificity, and evaluation of panels of field samples was nominal. Such bench validation for bovine spongiform encephalopathy (BSE) is a classical example where positive field samples were not available. Nevertheless, during extended periods of usage in diagnostic settings, such tests have undergone adjustments based on empirical evidence to reduce false-positive and false-negative results. For some of the BSE rapid tests, adjustments had to be made in the cut-off to reduce false-positive results apparent in early implementation. Bench validation provided a level of confidence in diagnostic performance that was adequate for conditional diagnostic use as determined by national authorities. But, it must never become a replacement for a full field validation. Therefore bench validation of diagnostic assays can only offer provisional recognition with anticipation that full field validation will follow.

A provisional recognition of an assay by state or national authorities recognises that the assay has not been evaluated for diagnostic performance characteristics. As such, the laboratory should develop and follow a protocol for adding and evaluating samples, as they become available, to fulfil this requirement. Ideally, this process should be limited to a specific timeframe in which such an accrual would be directed toward fulfilling Stages 2 and 3 of the validation pathway. This concept should be limited to emergency situations where rapid introduction of new tests is deemed essential by authorities. There may be other situations where bi-lateral trade agreements, based on tests (e.g. standard test methods) that have not been fully validated within a given country, are mutually accepted. In exceptional cases for rare diseases where no other assay option exists, provisional recognition may be allowed, but reporting of results should include a statement of the provisional nature of the validation of the assay. In all cases, sound evidence must exist for preliminary estimates of DSp and DSe based on a small select panel of well-characterised samples containing the targeted analyte.

3. Stage 2 – Diagnostic performance of the assay

Diagnostic Sensitivity and Diagnostic Specificity. Estimates of DSe and DSp are the primary performance indicators established during validation of an assay. These estimates are the basis for calculation of other parameters from which inferences are made about test results (e.g. predictive values of positive and negative test results). Therefore, it is imperative that estimates of DSe and DSp are as accurate as possible. Ideally, they are derived from testing a panel of samples from reference animals, of known history and infection status relative to the disease/infection in question and relevant to the country or region in which the test is to be used. Receiver operating characteristic curve analysis is a useful adjunct to estimation of DSe and DSp because it assesses the global accuracy of a quantitative diagnostic test across possible assay values (14, 15, 27). This approach is described in-depth in Appendix 1.1.4.5.

Diagnostic sensitivity is the proportion of samples from known infected reference animals that test positive in an assay.

Diagnostic specificity is the proportion of samples from known uninfected reference animals that test negative in an assay.

A sampling design must be chosen that will allow estimation of DSe and DSp. The designated number of known positive and known negative samples will depend on the likely values of DSe and DSp of the candidate assay and the desired confidence level for the estimates (Table 1 and reference 19). An abbreviated Table 1 provides two

⁴ Provisional recognition does not imply certification by the OIE. It does, however, recognise an informed decision of authorities at local, state, national or international levels of their conditional approval of a partially validated assay, usually for a time-limited use in emergency situations or as the basis for bi-lateral agreements between countries that choose to accept results from such an assay for trade purposes.

panels of the theoretical number of samples required, when either a 5% or 2% error is allowed in the estimates of DSe or DSp. Comparison of a 5% vs 2% error shows a considerable reduction in the number of samples required. A rather large number of samples is required to achieve a very high confidence for DSe and DSp when a minimal amount of error in the estimate is desired. Logistical and financial limitations may require that less than an optimal number of samples will be evaluated. However, by reducing the DSe and DSp confidence levels to less than 90% usually would not be recommended. Sample size also may be limited by the fact that reference populations and "gold standards" (perfect reference standards) may be lacking (see Appendix 1.1.4.5 for further detail). It may, therefore, be necessary to use a sub-optimal number of samples initially. It is, however, highly desirable to enhance confidence and reduce allowable error in the DSe and DSp estimates by adding more samples as they become available.

Table 1. Theoretical number of samples from animals of known infection status required for establishing diagnostic sensitivity (DSe) and specificity (DSp) estimates with known confidence

Estimated DSe or DSp	2% error allowed in estimate of DSe and DSp						5% error allowed in estimate of DSe and DSp					
	Confidence						Confidence					
	75%	80%	85%	90%	95%	99%	75%	80%	85%	90%	95%	99%
90%	257	369	475	610	864	1493	41	59	76	98	138	239
92%	210	302	389	466	707	1221	34	48	62	75	113	195
94%	161	232	298	382	542	935	26	37	48	61	87	150
95%	136	196	251	372	456	788	22	31	40	60	73	126
96%	110	158	203	260	369	637	18	25	32	42	59	102
97%	83	119	154	197	279	483	13	19	25	32	45	77
98%	56	80	103	133	188	325	9	13	16	21	30	52
99%	28	41	52	67	95	164	4	7	8	11	15	26

□ Percent error allowed in the estimate of DSe or DSp = 2% in the left panel and 5% in the right panel. For the number of samples required for 1%, 3%, and 4% allowable error in the estimate of DSe and DSp, multiply the number of samples in the left panel of the table by a factor of 4.0, 0.44, and 0.25, respectively.

The following are examples of reference populations and methodologies that may aid in determining performance characteristics of the test being validated.

a) Reference animal populations

Ideally, selection of reference animals requires that important host variables in the target population are represented in animals chosen for being infected with or exposed to the target agent, or that have never been infected or exposed. The variables to be noted include but are not limited to species, age, sex, breed, stage of infection, vaccination history, and relevant herd disease history.

- i) **Negative reference samples:** True negative samples, from animals that have had no possible infection or exposure to the agent, may be difficult to locate. It is often possible to obtain these samples from countries that have eradicated or have never had the disease in question. Such samples are useful as long as the targeted population for the assay is similar to the sample-source population.
- ii) **Positive reference samples:** It is generally problematic to find sufficient numbers of true positive reference animals, as determined by isolation of the organism. It may be necessary to resort to samples from animals that have been tested by another test such as a nucleic acid detection system.
- iii) **Samples from animals of unknown status:** See Appendix 1.1.4.5 Section 5.b.i, and Section 3.a.iii of this chapter for a discussion of latent class models.

b) Reference animal infection status

i) So-called "Gold Standard" model

The term 'gold standard' is commonly used to describe any standard of comparison, but it should be limited to methods or combination of methods that unequivocally classify animals as infected/exposed or uninfected. Some isolation methods themselves have problems of repeatability and analytical sensitivity, so are not truly gold standards particularly for purportedly negative samples. When the so-called reference standard is

imperfect, which is the common scenario for most ante-mortem tests, estimates of DSe and DSp for the candidate assay may be compromised because the error in estimates obtained by comparison to the relative standard is carried over into the estimates for the candidate assay. Indeed, when using imperfect reference assays, the DSe and DSp performance estimates of the candidate assay will be flawed and often overestimated.

NAD assays may be more sensitive and specific than existing 'gold standard' methods, which render the established 'gold standard' as not suitable for use as a comparison. If the NAD is more sensitive than the 'gold standard,' an apparent lower relative specificity will be misleading. This problem may be partially resolved by assessing sample derivation, clinical history and sequencing of any PCR products to confirm analyte identity.

ii. Latent-class models

Latent-class models (5, 10, 12 17) do not rely on the assumption of a perfect reference test but rather estimate the accuracy of the candidate test and the reference standard with the joint test results. Because these statistical models are complex and require critical assumptions, statistical assistance should be sought to help guide the analysis and describe the sampling from the target population(s), the characteristics of other tests included in the analysis, the appropriate choice of model and the estimation methods based on peer-reviewed literature (see Appendix 1.1.4.5 on statistical considerations for details).

c) Experimentally infected or vaccinated reference animals

Sera obtained sequentially from experimentally infected or vaccinated animals are useful for determining the kinetics of antibody responses or the presence/absence of antigen or organisms in samples from such animals. However, multiple serially acquired pre- and post-exposure results from individual animals are not acceptable for establishing estimates of DSe and DSp because the statistical requirement of independent observations is violated. Only single time-point sampling of individual experimental animals is acceptable. Also, for indirect methods of analyte detection, exposure to organisms under experimental conditions, or vaccination may elicit antibody responses that may not be quantitatively and qualitatively typical of natural infection in the target population (19). The strain of organism, dose, and route of administration to experimental animals are examples of variables that may introduce error when extrapolating DSe and DSp estimates to the target population. For these reasons, validation of an assay should not be based solely on samples from experimental animals.

d) Threshold (cut-off) determination

To obtain DSe and DSp estimates of the candidate assay, the test results first must be reduced to categorical (positive, negative, or intermediate) status. This is accomplished by insertion of one or two cut-off points (threshold or decision limits) on the continuous scale of test results. The selection of the cutoff(s) should reflect the purpose of the assay and its application, and must support the required DSe and DSp of the assay. Options and descriptive methods for determining the best way to express DSe and DSp are available (5, 12, 13, 15, 19, 27 and Appendix 1.1.4.5 on statistical considerations). If considerable overlap occurs in the distributions of test values from known infected and uninfected animals, it is difficult to select a single cut-off that will accurately classify these animals according to their infection status. Rather than a single cut-off, two cut-offs can be selected that define a high DSe (e.g. inclusion of 99% of the values from infected animals), and a high DSp e.g. 99% of the values from uninfected animals). The values that fall between these percentiles may be classified as intermediate (see box), and would require testing by a confirmatory assay, retesting for detection of sero-conversion, or sequencing for identity.

Threshold and cut-off are considered to be synonymous. A cut-off is the test value selected for distinguishing between negative and positive results on a continuous scale of test values.

Intermediate, inconclusive, suspicious, or equivocal are terms used synonymously for a zone of test values between the positive and negative cut-offs

The main difficulty in establishing cut-offs based on diagnostic performance characteristics is the lack of availability of the required number of well-characterised samples. Alternatives are discussed in Section B.2.g. on provisional acceptance of an assay during accrual of data to enhance estimates of DSe and DSp.

e) Calculation of DSe and DSp based on test results of reference sera

A typical method for determining DSe and DSp estimates is to test the reference samples in the new assay, and cross tabulate the categorical test results in a 2 X 2 Table. In a hypothetical example, assume the test developer decided that estimated DSe and DSp for the new assay should be 97% and 99%, respectively, with a desired confidence of 95% for both estimates. The amount of allowable error in the estimates was set at 2%. Table 1 indicates that 279 samples from known infected animals are required for the DSe assessment, and 95 known negative samples are needed for establishing the DSp estimate. The samples

were then run in the new assay. Table 2 is an hypothetical set of results and the calculated DSe and DS_p estimates based on the samples tested.

Table 2. Diagnostic sensitivity and specificity estimates calculated from hypothetical set of results for samples tested from known infected and non-infected populations

		Number of reference samples required*	
		Known positive (279)	Known negative (95)
Test results	Positive	270	7
	Negative	9	88
		TP	FP
		FN	TN
		Diagnostic sensitivity*	Diagnostic specificity*
		$TP/(TP + FN)$	$TN/(TN + FP)$
		96.7% (94.0 – 98.5%)**	92.0% (84.3 – 96.7%)**

* Based on Table 1 for an assay with the following parameters:

- 1) Prior to testing, estimated DSe of 97% and DS_p of 99%
- 2) 95% = required confidence in DSe and DS_p estimates
- 3) 2% = Allowable error in the estimates of DSe and DS_p

TP and FP = True Positive & False Positive, respectively

TN and FN = True Negative and False Negative, respectively

** 95% exact binomial confidence limits for DSe and DS_p calculated values (see Appendix 1.1.4.5 for information on confidence limits)

In this example, the DSe estimates are as anticipated, but the DS_p is much reduced from the anticipated 99%. As a consequence, the width of the confidence interval for DS_p is greater than expected. Re-inspection of Table 1 indicates that 707 samples are necessary to achieve an error margin of ± 2% at a DS_p of 92% but such an increase in sample size might not be feasible

4. Stage 3 – Reproducibility and augmented repeatability estimates

Reproducibility is an important measure of the precision of an assay when used in several laboratories located in distinct or different regions or countries using the identical assay (protocol, reagents and controls). Each of at least three laboratories test the same panel of samples (blinded) containing a minimum of 20 samples, with identical aliquots going to each laboratory (see Appendix 1.1.4.7 on panels of samples). This exercise also generates preliminary data on non-random effects attributable to deployment of the assay to other laboratories - also known as ruggedness of the assay. In addition, within-laboratory repeatability estimates are augmented by the replicates used in the reproducibility studies. Measurements of precision can be estimated for both the reproducibility and repeatability data (see Appendix 1.1.4.4 on Measurement of Uncertainty for further explanation of the topic and its application).

Reproducibility is the ability of a test method to provide consistent results, as determined by estimates of precision, when applied to aliquots of the same samples tested in different laboratories.

Ruggedness is a measure of the assay's capacity to remain unaffected by substantial changes or substitutions in test conditions anticipated in multi-laboratory utilisation.

5. Stage 4 – Programme implementation

a) Interpretation of test results.

Predictive values of test results. An assay's test results are most useful when the inferences made from them are accurate. Predictive values for test results need to be based on the true prevalence of exposure/infection in the targeted population. For screening assays used in surveillance of a 'disease-free' population, false-positive results are a significant problem. For instance, an assay may have impeccable credentials (e.g. high precision and accuracy, 99% DSe

Predictive Value (PV) of positive or negative test result. PV₊ is the probability that an animal has been exposed or infected, given that it tests positive. PV₋ is the probability that an animal is free from exposure or infection, given that it tests negative.

and 99.9% DSp). But if the prevalence of disease is close to zero, and the assay has one false positive for every 1000 animals tested, false-positive inferences are a problem (see reference 19 for PV tables for various estimates of DSe and DSp). Similarly, if the assay for a highly virulent disease has one false negative for every 100 animals, false negative inferences could have devastating consequences. This illustrates the critical importance of choosing diagnostic thresholds that are appropriate for the application at hand. Thresholds should be chosen to minimize the effect of false positives and/or false negatives on the predictive values of the test given its application and the prevalence of exposure/infection in the target population. It may also be prudent to have highly specific confirmatory assays to determine whether screening assay reactors are true or false positives.

For nucleic acid assays, it may be necessary to confirm NAD-positive results by sequence analysis of the amplified product (an example of an assay to assist in resolving errors due to non-specific target or primer binding).

b) International recognition

Traditionally, assays have been recognised internationally by the OIE when they are designated as prescribed or alternate tests for trade purposes. This has often been based on evidence of their usefulness on a national, regional or international basis. For test kits that have completed the certification process, the final step is listing of the test in the OIE Register. Tests listed in the Register are certified as fit for a specific purpose if they have completed Validation Stages 1, 2 and 3. The Register is intended to provide potential test users with an informed and unbiased source of information about the test and its performance characteristics for an intended purpose. The Register is available on the OIE website at: www.oie.int/vcda/eng/en_vcda_registre.htm

c) Deployment of the assay

Ultimate evidence of the usefulness of an assay is its successful application(s) in other laboratories and inclusion in national, regional and/or international programmes. Reference laboratories play a critical role in this process. In the natural progression of diagnostic and/or technological improvements, new assays will become the new standard method to which other assays will be compared. As such, they may progressively achieve national, regional and international recognition. As a recognised standard, these assays will also be used to develop reference reagents for quality control, proficiency and harmonisation purposes. These reference reagents may also become international standards.

An assessment of ruggedness should be repeated when the test is transferred from the development laboratory to the field, whether for use in local laboratories or in pen-side applications. Predictable changes, e.g. extremes of temperature and levels of operator experience, should be assessed as additional sources of variation in assay results that may affect estimates of ruggedness (which is mostly derived from reproducibility estimates).

6. Monitoring assay performance after initial validation

a) Monitoring the assay

A validated assay in routine use needs to be consistently monitored for repeatability through process controls to evaluate possible temporal changes in test precision and accuracy. These changes can be monitored graphically by plotting control values in control charts. Deviations from the expected performance should be investigated so corrective action can be taken if necessary. Such monitoring provides critical evidence that the assay retains its "validated" designation during the implementation phase of the assay. Subsequent ongoing evaluation of the assay's performance is also essential and is usually done through assessments of precision, accuracy, and outlier tendencies using control charts. Reproducibility is assessed through external quality control programmes such as proficiency testing.

Control chart – A graphical representation of data from the repetitive measurement of a control sample(s) tested in different runs of the assay over time.

b) Diagnostic modifications – considerations for changes in the assay

Over time, modification of the assay likely will be necessary to address changes in the analytes targeted (i.e., modification of the assay to adjust diagnostic performance) or technical modifications may be needed to improve assay efficiency or cost-effectiveness.

If the assay is to be applied in another geographical region and/or population, revalidation of the assay under the new conditions is recommended. Lineages or sub-lineages of a virus, derived from animals in different geographic locations, are known to have different target sequences or primer sites, requiring revalidation of the assay. This is especially true for NAD systems as it is very common for point mutations occur in many

infectious agents (i.e. RNA viruses). Mutations, which may occur within the primer or probe sites can affect the efficiency of the assay and even invalidate the established performance characteristics. It is also advisable to regularly confirm the target sequence at the selected genomic regions for national or regional isolates of the infectious agents. This is especially true for the primer and probe sites, to ensure that they remain stable and the estimates of DSe for the assay are not compromised.

A similar situation may occur with incursion of new viral lineages into countries or regions where that viral lineage did not previously exist. In these circumstances, existing NAD assays which did not target these novel lineages may need to be modified to include primers or probes targeting these new analytes. The same would be true for typing sera used in virus neutralisation assays.

i) Technical modifications and comparability assessments

Technical modifications to a validated assay such as changes in instrumentation, extraction protocols, and conversion of an assay to a semi-automated or fully automated system using robotics will typically not necessitate full revalidation of the assay. Rather, a methods comparison study is done to determine if the relatively minor modifications of the assay affect the test results. Comparability can be established by running the modified procedure and original procedure side-by-side, with the same panel of samples in both, over several runs. The panel chosen for this comparison should represent the entire operating range of both assays. If the results from the modified procedure and originally validated method are determined to be comparable in an experiment based on a pre-specified criterion, the modified assay remains valid for its intended purpose. See Appendix 1.1.4.6 for description of experiments that are appropriate for comparability testing and Appendix 1.1.4.7 on reference sample panels.

A comparability assessment may or may not be adequate if the test is applied to a different sample matrix, e.g. validated on blood and used on another tissue in the sample species. Revalidation may be necessary if the test is designed for use in a new species.

ii) Replacement of depleted reagents

When a reagent such as a control sample is nearing depletion, it is essential to prepare and repeatedly test a replacement before such a control is depleted. The prospective control sample should be included in multiple runs of the assay in parallel with the original control to establish their proportional relationship. Whenever possible, it is important to change only one reagent at a time to avoid the compound problem of evaluating more than one variable.

c) Enhancing confidence in validation criteria

Because many host variables have an impact on the diagnostic performance of assays, it is highly desirable over time to increase the number of reference samples from animals of known infection status. This improves the precision of the overall estimates of DSe and DSp, and may allow calculations of DSe estimates by factors such as age, stage of disease, and load of organisms. New data should be included annually in relevant test dossiers.

REFERENCES

1. BALLAGI-PORDÁNYI A. & BELÁK S. (1996). The use of mimics as internal standards to avoid false negatives in diagnostic PCR. *Mol. Cell. Probes*, **10**, 159–164.
2. BELÁK S. & THORÉN P. (2001). Molecular diagnosis of animal diseases. *Expert Rev. Mol. Diagn.*, **1**, 434–444.
3. BELÁK S. (2005). The molecular diagnosis of porcine viral diseases: a review. *Acta Vet. Hung.*, **53**, 113–124. (Review).
4. BELÁK S. (2007). Molecular diagnosis of viral diseases, present trends and future aspects. A view from the OIE Collaborating Centre for the Application of Polymerase Chain Reaction Methods for Diagnosis of Viral Diseases in Veterinary Medicine. *Vaccine*, **25**, 5444–5452.
5. BRANSCUM A.J., GARDNER I.A., JOHNSON W.O. (2005). Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Preventive Veterinary Medicine* **68**, 145–163.
6. BURNS M.J., NIXON G.J., FOY C.A. & HARRIS N. (2005). Standardisation of data from real-time quantitative PCR methods - evaluation of outliers and comparison of calibration curves. *BMC Biotechnol.*, **5**, 31–44.
7. BUSTIN S.A. (2005). Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences. *Expert Rev. Mol. Diagn.*, **5**, 493–498.

8. 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** (3), 913–935.
9. DEJAEGHER, B. & VANDER HEYDEN Y. (2006) Robustness tests. *LCGC Europe*, **19** (7) online at <http://www.lcgeurope.com/lcgeurope/content/printContentPopup.jsp?id=357956>
10. ENOE 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.
11. FINDLAY J.W.A, DILLARD R.F. (2007) . Appropriate Calibration Curve Fitting in Ligand Binding Assays. *AAPS Journal*. **9**(2): E260-E267. (Also on-line as AAPS Journal (2007); **9** (2), Article 29 (<http://www.aapsj.org>))
12. 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, pp. 63–76.
13. GREINER M., FRANKE C.R., BOHNING D. & SCHLATTMANN P. (1994). Construction of an intrinsic cut-off value for the sero-epidemiological study of *Trypanosoma evansi* infections in a canine population in Brazil: a new approach towards unbiased estimation of prevalence. *Acta Trop.*, **56**, 97–109.
14. GREINER M. & GARDNER I. (2000). Epidemiologic issues in the validation of veterinary diagnostic tests. *Vet. Prev. Med.*, **45**, 3–22.
15. 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.
16. HUGGETT J., DHEDA K., BUSTIN S. & ZUMLA A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.*, **6**, 279–284. (Review).
17. HUI S.L. & WALTER S.D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, **36**, 167–171.
18. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - INTERNATIONAL ELECTROTECHNICAL COMMISSION (ISO/IEC) (2005). ISO/IEC 17025:2005, General requirements for the competence of testing and calibration laboratories.
19. JACOBSON R.H. (1998). Validation of serological assays for diagnosis of infectious diseases. *Rev. sci. tech. Off. int. Epiz.*, **17**, 469–486.
20. LAUERMAN L.H. (2004). Advances in PCR technology. *Anim. Health Res. Rev.*, **5**, 247–248.
21. LOUIE M., LOUIE L. & SIMOR A.E. (2000). The role of DNA amplification technology in the diagnosis of infectious diseases. *CMAJ*, **163**, 301–309.22.
22. THOMPSON, M., ELLISON, S. & WOOD, R. (2002). Harmonized guidelines for single-laboratory validation of methods of analysis. *Pure Appl. Chem.*, **75** (5), 835-855.
23. VALIDATION OF ANALYTICAL PROCEDURES: TEXT AND METHODOLOGY (Q2(R1)). (2005). 1-13 Tripartite Guideline of the International Conference on Harmonisation for Technical Requirements for Registration of Pharmaceuticals for Human Use. Pp 1-13 (Available on-line at: http://bioforum.org.il/uploads/Editor/karen/q2_r1_step4.pdf)
24. 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** (8). 1381-1386.
25. WORLD ORGANISATION FOR ANIMAL HEALTH (OIE) (2008). OIE Standard for Management and Technical Requirements for Laboratories Conducting Tests for Infectious Diseases. *In: OIE Quality Standard and Guidelines for Veterinary Laboratories: Infectious Diseases*. OIE, Paris, France, 1–31.
26. YOUDEN, W. & STEINER, E. (1987). Statistical Manual of the AOAC. AOAC International. 96 Pages. 5th Printing (ISBN 0-935584-15-3).
27. ZWEIG M.H. & CAMPBELL G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, **39**, 561–577.

*
* *

APPENDICES TO VALIDATION CHAPTER – UNDER STUDY

Appendix 1.1.4.1. Development and optimisation of antibody detection assays

Appendix 1.1.4.2. Development and optimisation of antigen detection assays by immunological means

Appendix 1.1.4.3. Development and optimisation of Nucleic Acid Detection (NAD) assays

Appendix 1.1.4.4. Measurement of Uncertainty

Appendix 1.1.4.5. Statistical Approaches to Validation

Appendix 1.1.4.6. Methods Comparability

Appendix 1.1.4.7. Reference Panels – Selection and Use