Report of the

First International Meeting On The Design and Analysis Of Diagnostic Test Evaluation Studies

The Golf View Hotel, Nairn, Inverness, Scotland
Monday 28th March – Wednesday 30th March, 2005

An International Meeting Held in Conjunction with the
2005 Meeting of the Society of Veterinary Epidemiology and Preventive Medicine (SVEPM)

Edited by
Matthias Greiner, International EpiLab
OIE Collaborating Centre for Research and Training in
Population Animal Health Diagnosis and Surveillance Systems

modified after Doherr, 2005 (presented on the meeting)
Abbreviations

OIE  World Organisation for Animal Health
Se  Diagnostic sensitivity
Sp  Diagnostic specificity
ROC  Receiver operating characteristic
LCM  "Latent class model" for estimation of Se and Sp without “gold standard”
GS  Gold standard, a diagnostic test that is used as index or reference test to describe the relative accuracy of a new test

Contents

Abbreviations
Preface
Objectives of the meeting
Programme details
Discussion points and conclusions
  The role of this meeting in relation to international standardisation of test validation
  Comments on the OIE validation template (see Annex 1)
  Comments on criteria for reviewing validation dossiers
  Some illustrative example for assessing features of validation dossiers
  Recommendations
Disclaimer
Annex 1 The draft OIE assay validation template
Annex 2 Programme of the meeting
Annex 3 Summary of presentations
Annex 4 List of participants
Preface

An inherent property of all diagnostic tests is that they are not free of error. The estimation of diagnostic performance indicators such as sensitivity and specificity is at the core of test evaluation and follows well-known principles that are widely agreed in the veterinary sciences and other disciplines. However, the practical implementation of test evaluation studies often turns out to be a complex task. A sampling design must be chosen that allows estimation of diagnostic performance characteristics for the relevant target population with sufficient statistical certainty. This is further complicated by logistical and financial limitations and by the fact that reference populations and gold-standard methods may be lacking. Therefore, test evaluation requires a truly multi-disciplinary approach.

The First International Meeting on Design and Analyses of Diagnostic Test Evaluation Studies was structured to create a discussion forum that could stimulate the further development and acceptance of international standards in test evaluation. Dr Stephen Walter (McMaster University, Canada) and Dr. Peter Wright (National Centre for Foreign Animal Disease, Canada) were invited as keynote speakers. The undersigned wish to thank all of the speakers for sharing their specific expertise and for contributing to the cross-disciplinary dialogue. Dr Wright represented the OIE at the meeting and was invited to comment on the report. We thank all of the organisations that supported the meeting: the Veterinary and Agrochemical Research Centre (Belgium), the International EpiLab (Denmark), The Royal Veterinary and Agricultural University (Denmark), the Aristotle University of Thessaloniki (Greece), the University of California, Irvine (USA), the SAC Epidemiology Research Unit (UK), Idexx-Laboratories (USA), Cedi-Diagnostics (Netherlands), Institute Pourquier (France) and Guildhay Limited (UK). Last but not least it is our special pleasure to acknowledge the support from the OIE and the Society for Veterinary Epidemiology and Preventive Medicine (SVEPM), who invited us to present the results of the meeting to the delegates of their 2005 annual meeting.
This report summarises the main presentations with emphasis on all aspects related to international standardisation of assay validation. We hope that this meeting and forthcoming events of this nature will constructively support the development of international guidelines for practical and quality assay validation.

July, 2005

The Organising Committee of the Meeting
Marios Georgiadis, Greece
Matthias Greiner, Denmark
George Gunn, UK
Wesley Johnson, USA
Koen Mintiens, Belgium
Nils Toft, Denmark
Objectives of the meeting

The aim of the meeting was to create an environment that fostered the development and dissemination of quantitative methodology relevant to diagnostic test evaluation. The meeting was constructed to serve as an interface between the developers of the statistical methods and the subject matter specialists (test developers, evaluators, auditors, test users). It is intended to organise follow-up meetings if deemed necessary. This first meeting had three objectives, each of which was handled on one day. The first objective was to establish an informal network of researchers interested in quantitative aspects of diagnostic test evaluation. Day 1 of the meeting was essentially a kick-off meeting of this potential network and allowed interested scientists to present and discuss their work on a technical level.

The second objective was to facilitate discussion on current needs for the development of particular methodologies that would address specific practical problems. This was accomplished through a panel discussion in which relevant questions were posed and discussed. In particular, the draft OIE assay validation template (Annex 1), which was drawn up at the 2nd FAO/IAEA-OIE Consultants Meeting on “OIE Guidelines for Validation and Certification of Diagnostic Assays for Animal Infectious Diseases” (9-12 December, Vienna, Austria), was reviewed and comments were drafted. Following the panel discussion, participants formed smaller breakout groups and prepared specific comments on relevant sections of the OIE draft template. The practical outcome of Day 2 was the identification of needs and knowledge gaps.

The third objective was to disseminate the information about the use of existing methods through a session that was open to a general audience of potential users of these methods. Day 3 took the format of a pre-SVEPM workshop (see also: http://www.svepm.org.uk/). The presentations and discussions of the first two days of the meeting were summarised and a keynote presentation was given on guidelines for and limitations in diagnostic test evaluation and the use of no-gold standard methods.
Programme details

The meeting took place from the 28th to the 30th of March, 2005. The full programme is given in Annex 2. The programme contained 28 presentations, including three invited keynotes. A summary of 10 presentations is given in the Annex 3. Round table discussions and working groups were also part of the meeting. Forty-four delegates from twelve countries joined the meeting (see Annex 4). On the third day, the conclusions from the meeting were presented to more than 50 delegates of the annual meeting of the Society for Veterinary Epidemiology and Preventive Medicine (SVEPM).

Discussion points and conclusions

The role of this meeting in relation to international standardisation of test validation

It was found that the participants covered a broad geographical area and institutional range of affiliations, as well as, a broad range of technical expertise (see Fig. 1. and Annex 4: List of participants). Thus, even in the absence of any formal endorsement from the OIE or any other international organisation, it was felt that the group could provide useful comments on the suggested OIE template for test validation and engage in meaningful discussions on criteria to be used for reviewing validation dossiers.

Fig. 1. International Meeting on The Design and Analysis of Diagnostic Test Evaluation Studies. 2005: A place of dialogue among six professional areas of expertise.
Comments on the OIE validation template (see Annex 1)

Ad 1.1. Test method
"Species and specimen" should read "Species and specimens".

Ad 3. Validation -- stage 1
Terminology for analytical characteristics is in conflict with terminology from other ISO standards, e.g. the standard for accreditation of microbiological laboratories of the European Co-operation for Accreditation (publication EA-04/10).

Ad 3.1. Calibration
Guidelines are defined for ELISA tests and not directly applicable for PCR, etc. What if reference preparations are not available to the investigator?

Ad 3.2. Repeatability
Repeatability measures depend on the nature of the data: binary, ordinal, continuous.

Ad 3.3. Analytical specificity
Suggested change: Document cross-reactivity by comparing samples from animals infected with organisms with similar clinical presentations and/or organisms that are genetically closely related.

Ad. 3.4. Analytical sensitivity
Clarify that this is about the lower detection limit. The possibility to work with spiked samples should be mentioned.

Ad 4.1. Reference animals
The "reference" status must relate to purpose of testing. A standard set of relevant information should be collected and summarised for all animals involved at this stage of validation. There is no need to repeat this list of information for negative and positive animals. The standard information set should include: species, age, sex, breed, and information on other factors that are known as influential factors for the Se and Sp of the test, date and place of sampling, immunological status, vaccination history and disease history, pathognomonic and/or surrogate tests used to define status of animals or prevalence within population, description how the reference status was derived.

Ad 4.1.3. Experimental animals
The validation should not solely be based on experimental animals. The time point of sample collection (days post-infection) must be indicated. Sources and history of experimental animals should be described. Data from experimental animals should be presented as time-series. The use of repeated observations from the same animals using statistical models that require independent observations is not acceptable. This includes also the calculation of simple indices such as Se and Sp.
Ad 4.2. Threshold determination

The method and the samples used for selecting a cut-off value must be documented. The selection of the cut-off will typically reflect the intended purpose of the test. Nevertheless, it is recommended to conduct also an ROC analysis to show the potential performance in other epidemiological settings.

Ad 4.3. Performance estimates

No conclusion was reached whether the wording "one or all of the methods described below may be used to generate performance estimates" is too weak. The use of inappropriate methodology (for example the use of imperfect reference tests without further efforts to control for any biases) should be strongly discouraged.

Ad 4.3.1. Diagnostic Se and Sp -- with defined reference animals

The status of the reference animals must relate back to test purpose. Pure convenience samples should be strongly discouraged. Point estimates and 95% exact binomial confidence intervals or analogue Bayesian credibility intervals should be given for all estimates. The recommended sample size (300 for each population for the example given) may be difficult to realise for stratified analyses or Bayesian analyses. The notion of cumulative evidence (i.e. gathering of data and re-calculation of the indices) could be explored. All methods should be documented with references. The use of descriptive and analytical methods for continuous data should be encouraged. This includes but is not limited to ROC analysis.

Ad 4.3.2. Diagnostic Se and Sp -- without defined reference animals

It was noted that guidelines for the use of latent class models (LCM) or other methods that allow the estimation of Se and Sp in the absence of a gold standard are lacking. As the statistical models become complex, an Expert Review Panel should be established that provides assistance for reviewing the dossiers. The OIE glossary should be updated with the terminology related to such newer statistical methods. Some ad hoc guidelines would be useful to streamline submissions that use LCMs. The guidelines should give recommendations on proper ways to conduct and describe the sampling from the population(s), the purpose of testing, the characteristics of other tests included in the analysis, the appropriate choice of model and the estimation methods based on peer-reviewed literature. It is suggested to use the term 'reference (or index) test' when referring to the GS test.

Ad 4.3.3. Agreement between tests

This option cannot generally be recommended because the agreement is a function of many parameters. No proper inference can be made about Se and Sp of the test in question.

Comments on criteria for reviewing validation dossiers

It is expected that the scenarios encountered in submitted validation dossiers will be too diverse and too complex to allow the formulation of explicit guidelines. This means that the appropriateness of any particular study design and analysis method will have to be assessed on a case-by-case basis.
However, some basic requirements or elements will be common to all validation dossiers (e.g. sample size plan, calculation of confidence intervals). For those requirements, some guidance can be given. Other elements will be unique to a single dossier. In both cases, we may want to assess the approaches used to determine or define these elements. Some approaches or methods should be encouraged as good practice, while others should be discouraged as not acceptable. In some cases, a critical element may be missing altogether which would of course be unsatisfactory. Lastly, some approaches may be novel enough that only a consensus by the Expert Review Panel will determine acceptability. This results in a set of 8 categories of features (Tab. 1).

Tab. 1. Eight categories for features of test validation dossiers and action taken.

<table>
<thead>
<tr>
<th>A priori assessment</th>
<th>Common element</th>
<th>Unique element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfactory</td>
<td>Accept as good practice</td>
<td>Accept and add to &quot;positive list&quot;</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>Reject as unacceptable practice</td>
<td>Reject and add to &quot;negative list&quot;</td>
</tr>
<tr>
<td>Unaccounted</td>
<td>Reject as missing</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Uncertain</td>
<td>Resolve by Expert Review Panel</td>
<td>Resolve by Expert Review Panel</td>
</tr>
</tbody>
</table>

It should be possible to classify common elements into good and unacceptable practices based on existing knowledge. Those common acceptable and unacceptable elements could be documented on a "positive" and a "negative" list, respectively. The unique elements may be treated in a similar way. The positive and negative list with annotated comments would serve as a guide to the applicant with respect work to be completed before re-submission. Important elements unaccounted for or missing in the dossier, would lead to rejection. The “negative list” could contain examples of such unacceptable omissions, whereas the “positive list” could give some guidance of how do deal with the issue appropriately.

To resolve the last category of uncertain or novel approaches will require more in-depth assessment and a consensus from the Expert Review Panel.

Acceptable methods for estimation of the Se and Sp could be listed in the “positive list”.

Clear rules are needed to define whether an element is acceptable or not. Table 2 contains some suggested criteria and decision options for the suggested categories.
Tab. 2. Suggested categories and criteria for assessing the validity of conclusions from validation dossiers.

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
<th>Decision</th>
</tr>
</thead>
</table>
| 1: Good practice or standard approach | Coherence: agreement with published, peer-reviewed principles or recommendations  
Consistency: identical or comparable methods & principles used to address problems of the same nature  
Clear communication: results presented clearly, including critical appreciation of any assumptions or limitations of the methods | To be accepted                |
| 2: Good practice, extended beyond standard approach | As 1 except that no peer-reviewed published material exists to support the approach  
Requires additional documentation of the method to the level of peer-reviewed paper | To be accepted on a case-by-case basis |
| 3: Not acceptable | Lack of coherence: not supported by published, peer-reviewed principles or recommendations or other documentation  
Lack of consistency: elements of the design or analysis lack internal consistency or are contradicting  
Lack of clear communication: results not presented clearly, information missing, critical assumptions or limitations not addressed | Revision and re-submission recommended |

Some illustrative example for assessing features of validation dossiers

For the common features of validation studies, some recommendations could be given. Tab. 3. contains an illustrative example for this.
### 4.1. Populations from which samples are drawn

<table>
<thead>
<tr>
<th>Feature</th>
<th>Good practice or standard approach</th>
<th>Good practice, extended beyond standard approach</th>
<th>Not acceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representativeness</td>
<td>Study population is part of the target population for application of the test</td>
<td>Multiple populations, including experimental animals</td>
<td>Study population not related to target population; issue of external validity not addressed</td>
</tr>
<tr>
<td>Case definition</td>
<td>There exists a de-facto (gold) standard diagnosis established using an acceptable protocol or A plausible concept for a latent case definition exists that accounts for the communality of multiple diagnostic tests in LCM analysis</td>
<td>Multiple case definitions (e.g., stages, strains) addressed in the analysis</td>
<td>No case definition given or Case definition does not correspond with purpose of testing or Multiple case definitions ignored in the analysis</td>
</tr>
</tbody>
</table>

**Recommendations**

The following priorities were suggested to further support the international standardisation of test validation.

- Guidelines should be written where those are lacking, particularly on the use of LCM.
- Strategic plans for knowledge transfer, particularly on the issue of LCM, should be developed.
- The OIE validation template and related documentation should be further disseminated.
- The dialogue between and within the different expert groups should be continued.
- The OIE validation template should be put into practical application and improved where necessary.
- Procedures for re-submission (re-analysis) should be considered.
- The feasibility, practicability and efficiency of all procedures should be considered as critical constraints.
- The submission of a validation dossier, following the OIE guidelines should financially be feasible for the test developer.
- The OIE guidelines need to be compared with the ISO 17025 guidelines to identify discrepancies.
Disclaimer

All care was taken to report the discussion accurately. However, the Organising Committee explicitly reserves the right to provide in this report its own opinion and understanding on all discussion items.

Annex 1
The draft OIE assay validation template

Annex 2
Programme of the meeting

Annex 3
Summary of presentations

Annex 4
List of participants
Annex 1

OIE Validation Template

A draft version of December 2003 of the “OIE Validation template” was used as reference for the presentation and discussion. It is noted that the development of the template is still in progress. More information and the current version of the template should be obtained from http://www.oie.int/vcda/eng/en_background_VCDA.htm.

<table>
<thead>
<tr>
<th>1. Background Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Test Method</td>
</tr>
<tr>
<td>• Disease</td>
</tr>
<tr>
<td>• Type of Method</td>
</tr>
<tr>
<td>• Target analyte</td>
</tr>
<tr>
<td>• Species and specimen</td>
</tr>
<tr>
<td>• Name of kit (if applicable)</td>
</tr>
<tr>
<td>1.2 Intended purpose(s) of test</td>
</tr>
<tr>
<td>• Population freedom (declaration)</td>
</tr>
<tr>
<td>• Animal freedom (trade)</td>
</tr>
<tr>
<td>• Eradication/control</td>
</tr>
<tr>
<td>• Investigation of clinical signs</td>
</tr>
<tr>
<td>• Prevalence estimate (risk analysis)</td>
</tr>
<tr>
<td>• Immune status</td>
</tr>
<tr>
<td>1.3 Applicant</td>
</tr>
<tr>
<td>• Name and complete contact information</td>
</tr>
<tr>
<td>• Job title within organization</td>
</tr>
<tr>
<td>• Type of organization (i.e. commercial, institutional or governmental)</td>
</tr>
<tr>
<td>1.4 Scientific contact</td>
</tr>
<tr>
<td>• Name and complete contact information</td>
</tr>
<tr>
<td>• Job title within organization</td>
</tr>
<tr>
<td>1.5 Accreditation or certification status of laboratory</td>
</tr>
<tr>
<td>• OIE Quality Standard, ISO/IEC 17025, ISO/IEC 9000 series, GLP/GMP, etc.</td>
</tr>
<tr>
<td>1.6 Intellectual property</td>
</tr>
</tbody>
</table>
2. Test Method
   2.1 Protocol
      • Test method protocol must include:
        * Introduction
          - Test Method
          - Fitness for Intended Purpose(s)
          - Definitions
        * Equipment and Instrumentation
        * Reagents
          - Chemicals
          - Biologicals
        * Preparation for Test
          - Preparation of sample
          - Preparation of reagents
          - Preparation of equipment and instrumentation
          - Preparation of laboratory personnel
        * Performance of Test
          - Test procedure
        * Interpretation of Results
          - Test controls
          - Test results
        * References
        * Appendices
   2.2 Kit configuration (if Commercial)
      • Samples per kit
      • Production capacity (theoretical and actual)
3. Validation – Stage I

3.1 Calibration

- Dose-Response Curve
  - Specify linear operating range
- Calibration against reference reagents
  - International, i.e. OIE, WHO, FAO, etc.
  - or
  - In-house, i.e. selection of strong positive, weak positive and negative reference reagents from dose-response curve

3.2 Repeatability

- Repeatability data
  - Minimum of 3 in-house samples representing activity within linear range of assay.
    i.e. from strong positive to negative (as per Section 3.1 above)
  - Within run – test each sample in quadruplicate
  - Between run – minimum 20 runs (total), 2 or more operators, preferably on separate days (note – all runs must be independent of each other)
  - Between serials – repeat above for each of 3 production batches (serials or lots) of kit, where applicable
  - Data should include means, SD’s, UCL/LCL’s on both raw and normalized test values

3.3 Analytical specificity

- Cross-reactivity, near neighbour data
  - Document cross-reactivity by comparing samples from animals infected with organisms with similar clinical presentations and organisms that are genetically closely related
- Type/group specificity data
  - Documentation affirming serotype or group specificity

3.4 Analytical sensitivity

- Specify standard of comparison (i.e. currently accepted test method)
- Comparison may include:
  - End point titrations
  - Earliest time of detection post-exposure
  - Duration of detection post-exposure (if applicable)
4. Validation – Stage II

4.1 Reference Animals

4.1.1 Negative reference animals (Note: negative refers to lack of exposure to or infection with the agent in question)

- Complete description
  - Age, sex, breed, etc.
  - Immunological status
  - Relatedness to intended target population
  - Selection criteria including historical, epidemiological and/or clinical data
  - Pathognomonic and/or surrogate tests used to define status of animals or prevalence within population
  - Sampling plan and procedures

4.1.2 Positive reference animals (Note: negative refers to known exposure to or infection with the agent in question)

- Complete description
  - Age, sex, breed, etc.
  - Immunological status
  - Relatedness to intended target population
  - Selection criteria including historical, epidemiological and/or clinical data
  - Pathognomonic and/or surrogate tests used to define status of animals or prevalence within population
  - Sampling plan and procedures

4.1.3 Experimental animals

- Complete description
  - Age, sex, breed, etc.
  - Immunological status
  - Relatedness to intended target population

- Exposure
  - Inoculum – source, dose, etc.
  - Type of exposure – inoculation, aerosol, contact, etc.
  - Sampling plan and procedures

4.2 Threshold determination
• Complete description of method used
  - empirical, ROC, mean ± SD, etc
  - descriptive statistics, frequency distribution diagrams, etc

4.3 Performance Estimates
• Depending on available resources, one or all of the methods described below may be used to generate performance estimates
• Irrespective of the method chosen, the standard method(s) of comparison should be run in parallel on all samples, i.e. the test methods in current use

4.3.1 Diagnostic sensitivity and specificity estimates – with defined reference animals
• Conventional method using reference animals (see 4.1.1 and 4.1.2)
  - Assuming a minimum sensitivity and specificity of 75% with an allowable error of ± 5% in the estimate at a level of confidence of 95%, number of reference animals required is 300 for each population
  - Individual animals must be selected from negative and positive reference populations
  - Include 2x2 table, calculations for diagnostic sensitivity and specificity including error and confidence
  - Include same calculations for other tests if being compared to the test in question

4.3.2 Diagnostic sensitivity and specificity estimates – without defined reference animals
• Complete description of model used
  - Bayesian inference, latent class analysis, etc.
  - Describe rationale, priors, supporting data
  - Population selection criteria, including prevalence estimates
• Other test methods in evaluated should also include the standard method of comparison
• Using best available priors, choose test populations with appropriate prevalences and select animals in sufficient numbers to generate estimates of sensitivity and specificity with an allowable error ± 5% at a level of confidence of 95%

4.3.3 Agreement between tests
• Complete description of test methods in comparison
  - Presumptive vs confirmatory tests
  - Relatedness of analytes
  - Potential biases
• Complete description of samples tested
  - Source of samples may include experimental animals sequentially sampled over time
- May also include animals or herds defined by reactivity in confirmatory tests or multiple presumptive tests and sampled over a period of time
- Describe measures of agreement and explanations for results not in agreement

5. Validation – Stage III

5.1 Laboratory Selection
- Selection criteria for candidate laboratories
  - Location, i.e. country
  - Status, i.e. regional, national, provincial/state
  - Level of expertise, familiarity with technology
  - Accreditation status
- Number of laboratories included
  - Minimum of 3 laboratories, should also include OIE Reference Laboratory, if possible

5.2 Evaluation panel
- Description of test panel
  - Selection criteria, number of samples (minimum of 20)
  - Sample volume, allowable number of repeats
  - Panel composition, i.e. number of replicated, range of analyte concentrations/reactivities
  - Sample processing requirements, i.e. extractions, spiking, serial dilutions, preservatives, sterilization
  - Coding of unknown (blind) samples
  - Frequency of testing

5.3 Reproducibility
- Description of type of data/interpretation
  - Qualitative (categorical)
  - Quantitative or semi-quantitative data
  - Single dilution vs titration
- Description of type of analysis
  - Pre-determined limits, consensus, Youden plots
- Descriptive statistics
  - Include mean, sd, range of results
  - Should include controls, as well as blind samples
  - Number and proportion of accepted/rejected runs should be included
6. Validation – Stage IV
6.1 Laboratories
   • List laboratories where this test method is in current use
     - Location, i.e. Country
     - Status, i.e. Regional, national, provincial/state
     - Accreditation status

6.2 Test applications
   • For each laboratory
     - Indicate purpose of test, see Section 1.2
     - Integration with other tests
     - Status test, i.e. official test, supplementary, etc
     - Throughput, i.e. daily, monthly, annual
     - Turn-around-times

6.3 International reference reagents
   • List type and availability of international reference reagents
     - Source
     - Negative, weak/strong positive reference reagents
     - Other key biologicals, e.g. antigens, antibodies, etc

6.4 Inter-laboratory testing programs
   • Describe programs involving inter-laboratory comparisons utilizing this test method
     - National, International
   • Describe eligibility and number of laboratories participating

6.5 International recognition
   • List internationally-recognized reference laboratory responsible for this test method and/or biologicals
   • List international standards containing this test method
   • List international programs employing this test method
# PROGRAMME

**MONDAY 28 MARCH 2005**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Chair/Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>08.30 – 08.40</td>
<td>Introductory session</td>
<td>Mintiens Welcome</td>
</tr>
<tr>
<td>08.40 – 09.55</td>
<td>Georgiadis Diagnostic test validation: Why do we need the math?</td>
<td></td>
</tr>
<tr>
<td>09.55 – 11.10</td>
<td>Johnson Introduction to Bayesian models and methods for diagnostic test data.</td>
<td></td>
</tr>
<tr>
<td>11.10 – 11.30</td>
<td>Coffee break</td>
<td></td>
</tr>
</tbody>
</table>

**Open Session 1** Chair: M. Greiner

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Chair/Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.30 – 11.55</td>
<td>Johnson Introduction to Bayesian models and methods for diagnostic test data.</td>
<td></td>
</tr>
<tr>
<td>11.55 – 12.20</td>
<td>Toft Continuous-data diagnostic testing of multi-stage diseases.</td>
<td></td>
</tr>
<tr>
<td>12.20 – 12.45</td>
<td>Toft Diagnostic properties of clinical observers.</td>
<td></td>
</tr>
<tr>
<td>12.45 – 13.10</td>
<td>Baadsgaard A Bayesian approach to the precision of clinical observers.</td>
<td></td>
</tr>
<tr>
<td>13.10 – 14.40</td>
<td>Lunch</td>
<td></td>
</tr>
</tbody>
</table>

**Open Session 2** Chair: M. Georgiadis

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Chair/Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.40 – 15.05</td>
<td>Engel Evaluation of diagnostic tests in the absence of a gold standard – models and applications.</td>
<td></td>
</tr>
<tr>
<td>15.05 – 15.30</td>
<td>Doherr Three field data-based examples of using various test evaluation techniques with and without a gold standard.</td>
<td></td>
</tr>
<tr>
<td>15.30 – 15.55</td>
<td>Johnson Bayesian estimation of herd level test characteristics under different sampling designs.</td>
<td></td>
</tr>
<tr>
<td>15.55 – 16.15</td>
<td>Coffee break</td>
<td></td>
</tr>
</tbody>
</table>

**Open Session 3** Chair: N. Toft

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Chair/Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.40 – 17.05</td>
<td>Verloo Validating and interpreting tests when the aim is to decrease the uncertainty on the serological status rather than the infection status.</td>
<td></td>
</tr>
<tr>
<td>17.05 – 17.30</td>
<td>Johnson Bayesian hierarchical estimation of between and within herd prevalences accounting for staging and without cutoffs or a gold standard.</td>
<td></td>
</tr>
<tr>
<td>17.30 – 18.30</td>
<td>Round table discussion 20.00 Dinner</td>
<td></td>
</tr>
</tbody>
</table>

---

**Organising Committee:**

- M. Georgiadis Aristotle University of Thessaloniki, Greece
- M. Greiner International Epilab, Copenhagen, Denmark
- G. Gunn Scottish Agricultural College, Inverness, Scotland
- W. Johnson University of California, Davis, California
- K. Mintiens Veterinary and Agrochemical Research Centre, Brussels, Belgium
- N. Toft Royal Veterinary and Agriculture University, Copenhagen, Denmark

**Address:**

- c/o K. Mintiens Veterinary and Agrochemical Research Centre B-1180 Brussels, Belgium
- Tel: +32 2 379 06 26
- Fax: +32 2 379 06 70
- Email: komin@var.fgov.be

---

[http://www.dina.kvl.dk/~nt/dte](http://www.dina.kvl.dk/~nt/dte)
### PROGRAMME (2)

**Tuesday 29 March 2005:**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session/Panel</th>
<th>Chair</th>
</tr>
</thead>
<tbody>
<tr>
<td>08.30 – 09.40</td>
<td><strong>Keynote presentation</strong></td>
<td>M. Greiner</td>
</tr>
<tr>
<td>09.40 – 10.30</td>
<td><strong>Keynote presentation</strong></td>
<td>K. Mintiens</td>
</tr>
<tr>
<td>10.30 – 11.20</td>
<td><strong>Round table discussion</strong></td>
<td></td>
</tr>
<tr>
<td>11.20 – 11.40</td>
<td><strong>Coffee break</strong></td>
<td></td>
</tr>
<tr>
<td>11.40 – 12.05</td>
<td><strong>Open Session 4</strong></td>
<td>N. Toft</td>
</tr>
<tr>
<td>12.05 – 12.30</td>
<td><strong>Open Session 5</strong></td>
<td>W. Johnson</td>
</tr>
<tr>
<td>12.30 – 12.55</td>
<td><strong>Open Session 6</strong></td>
<td>K. Mintiens</td>
</tr>
<tr>
<td>16.05 – 16.15</td>
<td><strong>Coffee break</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Keynote presentation Chair: M. Greiner**

08.30 – 09.40  Stephen Walter  Use of latent class methods with epidemiologic data: a review.

**Keynote presentation Chair: K. Mintiens**

09.40 – 10.30  Peter Wright  The OIE template for assay evaluation dossiers: a new approach towards international and methodological harmonization.

10.30 – 11.20  **Round table discussion about the OIE template for assay evaluation dossier**

11.20 – 11.40  **Coffee break**

**Open Session 4 Chair: N. Toft**

11.40 – 12.05  Elfinger  Virus isolation, PCR and Directigen testing in the 2002 LowPath avian influenza outbreak in Virginia USA.

12.05 – 12.30  Wapanaar  Use of a bovine Neospora caninum ELISA in hemolyzed and non-hemolyzed fox and coyote samples.

12.30 – 12.55  Bak  A panel of meat juice samples for use in quality control of Salmonella.

12.55 – 14.25  **Lunch**

**Open Session 5 Chair: W. Johnson**


14.50–15.15  Ebel  Time value of serologic surveillance evidence.

15.15–15.40  Mintiens  The evaluation of available diagnostic kits to be used in official animal control programs in Belgium: a case study.

15.40–16.05  Tavornpanich  Determination of optimal strategies for detection of paratuberculosis infected dairy herds using stochastic simulation models.

16.05 – 16.15  **Coffee break**

**Open Session 6 Chair: K. Mintiens**

16.15 – 16.40  Gelbmann  Role of the European Food Safety Authority in evaluation of diagnostic tests.

16.40 – 17.05  Gunn & Humphry  Consequences of improved test validation at individual animal-level for test validity at herd-level and subsequent surveys of herd prevalence: using paratuberculosis as an example.

17.05 – 18.05  **Round table discussion**  K. Mintiens

20.00  **Dinner**
**International Meeting On The Design and Analysis Of Diagnostic Test Evaluation Studies**
The Golf View Hotel, Nairn, Inverness, Scotland
Monday 28th March – Wednesday 30th March, 2005

**PROGRAMME (3)**

**WEDNESDAY 30 MARCH 2005**

<table>
<thead>
<tr>
<th>Time</th>
<th>Speaker</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.30 - 10.00</td>
<td>Wright</td>
<td>Fit for purpose: a laboratory perspective</td>
</tr>
<tr>
<td>10.00-10.30</td>
<td>Greiner</td>
<td>Ad hoc recommendation of criteria to assess the validity of test evaluation studies. Report from the workshop.</td>
</tr>
<tr>
<td>10.30 - 10.45</td>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>10.45 – 11.00</td>
<td></td>
<td>Coffee break</td>
</tr>
<tr>
<td>11.00 – 11.20</td>
<td>M. Georgiadis</td>
<td>Methodological developments in the area of no-gold standard methods for the evaluation of diagnostic tests. Report from workshop.</td>
</tr>
<tr>
<td>11.40 - 12.00</td>
<td></td>
<td>Discussion</td>
</tr>
</tbody>
</table>
Annex 3

Summary of presentations (alphabetical first author)

**A panel of meat juice samples for use in quality control of Salmonella**
Hanne Bak and Vibeke Sørensen

**Description of the problem that was addressed**
The level of Salmonella infection in the Danish slaughter pigs is monitored with an indirect ELISA, which detects antibodies against Salmonella in meat juice samples and express results as OD%. Two factors made an improvement of the quality control of this ELISA necessary: The control samples on the individual microtiter plates were positive and negative pig sera even though the test samples were meat juice, and the analyses were recently changed from double to single determination, which gave an increased risk of overlooking systematic errors. Therefore the objective for the present study was to establish a control panel of meat juice samples to be used as an expansion of the existing quality control system.

**Materials and methods (including assumptions and computational implementation of the method, if applicable)**
Meat juice was collected according to experience gained through a pilot experiment. Mm. sternomastoideae and Mm. psoas minori were collected from 222 pigs from 5 farms infected with Salmonella. The muscles were cut in pieces and put in containers for extraction of meat juice. 39 pools of meat juice with OD% ranging from 0 to 100 were prepared. Each pool contained 30 ml meat juice from 3-10 pigs. The 39 pools were analysed 20 times during 6 months with at least 2 weeks interval between analyses. Each pool was analysed twice in a single determination on a total of 10 microtiter plates. From the 20 analyses of each pool, CV% was determined, and pools with a coefficient of variation (CV) above 20% would have been excluded from the panel. Mean OD% and standard deviation (SD) were determined for the 39 pools. The panel was implemented as an expansion of the existing quality control system for the ELISA. Upper and lower control limit for approval of test results were set as mean OD% ± 2SD. Furthermore, the CV% of the two single determinations on a microtiter plate should be below the normal value for the test as a measure of systematic errors within the microtiter plates. Monthly analyses of the panel should give test results within these predetermined limits, otherwise actions should be taken to identify and correct the cause of the deviation. In the period where the meat juice panel was under preparation, problems occurred with the automatic ELISA washer. Logistic analysis with the probability of a test sample to have test result positive as the outcome revealed a probability of 2.6% in the lower left corner of the microtiter plates compared to a probability of 8.2% in the lower right corner. These errors would have been detected by regular analyses of a meat juice panel.

**Conclusions**
A panel of meat juice pools was prepared, and regular analyses of such a panel will detect systematic errors giving trends across the microtiter plates.
Evaluation of diagnostic tests in the absence of a gold standard – models and applications

Bas Engel, Annemarie Bouma, Herman van Roermund, Bas Swildens & Willem Buist

Description of the problem that was addressed

The Animal Sciences Group (ASG), in cooperation with the University of Utrecht, has been involved in the evaluation of accuracy of diagnostic tests for various diseases in farm animals in The Netherlands, such as classical swine fever (CSF), EF-positive *Streptococcus suis* serotype 2 strains (*S. suis* for short) in sows, paratuberculosis, avian tuberculosis and foot and mouth disease. In all of these studies, for some or all of the animals the true disease status was unknown, i.e. in all problems there is lack of a gold standard. Interest in these studies was in evaluation of sensitivity and specificity or in the (often intricate) prevalence structure. Here, we will focus on two studies: accuracy of clinical diagnosis of CSF for breeding sows at herd level (Engel et al., 2004a) and accuracy of diagnostic tests for *S. suis* in sows (Engel et al., 2004b). We will give an overview of the statistical analyses performed and the kind of problems that we encountered.

Materials and methods (including assumptions and computational implementation of the method, if applicable)

In the CSF study data from the 1997-1998 CSF outbreak in the Netherlands were analysed. Herds were visited by veterinarians and each visit was coded as a binary variable (negative or positive clinical diagnosis). A feature of special interest in this study was the dependence of herd sensitivity of clinical diagnosis of CSF on the number of days elapsed since virus introduction in the herd. The true status of the herds (positive or negative) was (eventually) known. However, when a herd was visited by a veterinarian, neither the status of the herd at that moment, nor the number of days since virus introduction was known. There was no gold standard, but a probability distribution for the moment of virus introduction could be derived from serum samples collected at the moment of depopulation of the herds. This distribution was incorporated in a logistic regression model for the binary data derived from the veterinarians' reports.

The *S. suis* study involved data from a field study with 3 diagnostic tests that are (possibly) conditionally dependent. No gold standard was available. Because one of the tests, involving bacterial examination, was quite expensive, it was only performed for a subset of the samples. The model was a mixture of distributions for the latent classes of true positives and true negatives. The herd prevalences were the mixture probabilities. We prefer to model joint test results for true positives and true negatives by multinomial distributions, with a parameterization similar to log linear models for contingency tables. Because the data set was incomplete, a particular parameterization was chosen involving a sequence of logistic regression models and binomial distributions rather than multinomial distributions. The model comprised parameters for conditional dependence.

We prefer posterior Bayesian inference for these problems, because it offers an elegant framework to include prior information, and because powerful Markov Chain Monte Carlo (MCMC) algorithms are available that can handle complicated models, sometimes with intricate prevalence structures (e.g. within and between herd prevalences in paratuberculosis) in a fairly easy way. We use the Gibbs sampler, as implemented in the WinBUGS software.

Conclusions

A Bayesian analysis with the latent class model seems to be the best tool available when there is no gold standard available. However, as in many Bayesian analyses, choice of priors can be tricky and unexpected and even counter intuitive patterns may occur in posterior distributions. Although dependence between tests is important when test are combined, for most problems we find it hard to obtain an accurate impression of dependence between tests. Since there is no gold standard, the mixture that is actually fitted is a probabilistic communality of the tests involved. Therefore, in principle, results depend on the suite of tests that are used. Although this seems unavoidable, it does show that there are limitations for analysis of test data in the absence of a gold standard.
Surveillance of animal diseases in small herds
Matthias Greiner, Prof PD Dr med vet MSc

Description of the problem that was addressed
Small herds may present a problem in surveillance for infectious animal diseases because typical levels of a within-herd design prevalence are not directly applicable. Practically important is also the fact that the sample size of animals per herd is constrained by the herd size. Therefore, the chance of detecting an infected herd cannot be improved by increasing the herd sample size. The issue of sampling from small herds is well addressed in the literature. However, the testing of the complete herd rather than sampling a fraction of animals may be the preferred option of surveillance in small herds. The issue of surveillance for small herds and the appropriate statistical approach has not yet been addressed in the literature. Therefore, it was the objective of this study

- to define "small" herds in the context of animal disease surveillance
- to describe freedom-from-disease calculations for small herds
- to review and refine the statement about problems presented by small herds.

Materials and methods
Standard methods of probability theory were used to develop formulas for (a) the probability of detecting individual small herds as infected, given it is infected at a specified level of within-herd prevalence (herd sensitivity), (b) the probability of detecting infection in a sub-population of small herds, given that infection occurs at a specified level of among-herd prevalence (confidence or power) and (c) sample sizes for planning of surveillance that meets a specified level of confidence.

Both a binomial model and a Poisson model was be used for these purposes. The results were validated using a synthetic data set generated by stochastic computer simulation of the surveillance process for a sub-population of small herds.

Conclusions
Small herds can be defined as those smaller than \(2 / \text{(within-herd design prevalence)}\) on the basis that such herds can be expected to have less than 2 (i.e. only one) infected animal. Formulas for the confidence can be given and sample size can be

the Poisson model provides more conservative (lower) estimates of the confidence for a given sample size and should therefore be preferred.
Consequences of improved test validation at individual animal-level for test validity at herd-level and subsequent surveys of herd prevalence: using paratuberculosis as an example

George Gunn & Roger Humphry

Description of the problem that was addressed

Better validation of a test at the level of the individual animal allows us to infer test qualities (primarily sensitivity and specificity) at the herd level. When designing a herd-level prevalence study that will utilise an imperfect diagnostic test, it is necessary to consider the test sensitivity and specificity. We present a practical two-stage approach (Humphry, et al., 2004) that we developed to take into account the imperfections of the test estimated from test validation.

Materials and methods (including assumptions and computational implementation of the method, if applicable)

The test sensitivity and specificity are assumed to be "known" (although computer simulation could allow uncertainties to be accounted for). Different values for herd sensitivity and specificity are chosen in combination with the assumed a priori herd-level prevalence, desired confidence and tolerance, to calculate the number of herds to be tested. This assumes that there are enough herds in the population to allow a normal approximation to a formula from (Rogan and Gladen, 1978)

Within the herd, Freecalc V2 (Cameron and Baldock, 1998) is used to calculate the number of animals to be tested to achieve the different herd-level sensitivity and specificity values used in the previous calculation. Assumed are herd size, minimum within herd prevalence, test sensitivity and specificity. The two calculations then can be matched to give the number of herds and the number animals per herd to be tested under various different scenarios. The trade-off between the benefit (low number of herds) and the disadvantage (large number of animals per herd and exclusion of small herds) that are associated with achieving high herd-level sensitivity and specificity can then be assessed by the decision maker.

Conclusions

Straightforward theory demonstrates the need for test imperfections to be accounted for by policy makers and this includes survey design. Decision makers are often wary of complex solutions that they do not understand. The method described provides a necessary, relatively easily understood and practical means for calculating the number of animals and herds to sample and still account for test imperfections when creating surveys to estimate herd level prevalence.


The Evaluation of Available Diagnostic Kits to Be Used in Official Animal Control Programs in Belgium: a Case Study


Description of the problem that was addressed

All diagnostic activities related to the official animal disease control programmes of the Belgian Food Agency are performed by recognized diagnostic laboratories.

The diagnostic procedures related to the official animal disease control programmes are standardised for all laboratories, in compliance with the ISO 17025 recommendations.

One of the procedures in the standardisation is the selection of the optimal diagnostic reagents after going through a standard evaluation procedure.

In 2004, the ELISA test kits for serological diagnosis of leucosis virus in bovine serum and tank milk, Brucella abortus in bovine serum, pseudorabies virus in serum of pigs, and Salmonella species in serum of pigs, which were available on the market, were evaluated.

Materials and methods (including assumptions and computational implementation of the method, if applicable)

The standard evaluation procedure is based on the OIE guidelines (template) and differentiates between analytical and diagnostic quality. The evaluation of the analytical quality was performed for all application, whereas the diagnostic quality was only evaluated for the Salmonella species in serum of pigs application.

Test kit producers were invited to provide all available information on the topics (technical manuals), which are listed in the OIE template. The provided information was compared between producers for the same application.

The National Reference laboratory evaluated the analytical characteristics of the kits: calibration against reference serum, analytical specificity, analytical sensitivity, repeatability/reproducibility.

For the evaluation of the diagnostic sensitivity and specificity, field samples were collected and tested by all candidate tests. No-gold standard methods were used, accounting for conditional dependence between tests.

Conclusions

The huge heterogeneity in the design of studies carried out before and after the launch of the kits on the market and the absence of important information in technical manuals do not necessarily allow a strict comparison of available diagnostic kits.

A large amount of the kits failed to meet the legal requirements, which are set by European Legislation.

This study reflects the absence of clear guidelines for the design of diagnostic test evaluation studies.

These guidelines would allow test producers to provide the required information to test users to allow them to evaluate the quality of a test. A mandatory full registration process of diagnostic kits, as for drugs, may be required.

The diagnostic sensitivity and specificity was simultaneously evaluated for four Salmonella test kits, based on 180 field samples.
Determination of optimal testing strategies for detection of paratuberculosis infected dairy herd by simulation modeling

S. Tavornpanich¹, C. Muñoz-Zanzi²³, I. Gardner¹, E. Raizman², S. Wells²
¹ Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, CA.
² Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN.
³ Minnesota Veterinary Diagnostic Laboratory, College of Veterinary Medicine, University of Minnesota, St. Paul, MN.

Description of the problem that was addressed
Correct classification of herd-infection status is a critical component of a successful control program for paratuberculosis, caused by Mycobacterium avium subsp. paratuberculosis (Map). However, a scientifically-sound and cost-effective method that provides high confidence of classification of herds as infected or non-infected for program participants has yet to be determined. The objectives of this study were to determine optimal testing strategies for detection of paratuberculosis at the initial step of herd classification, and to provide decision-maker with insightful information on cost-effectiveness of the available testing methods.

Materials and methods (including assumptions and computational implementation of the method, if applicable)
A simulation model was developed to compare performances of several testing procedures, including cultures of fecal pools and environmental samples in different simulated herd situations. Specific factors that were considered in the evaluation included herd size, within-herd prevalence, expected distribution of shedding in infected cattle, number of randomly selected animals, number of environmental samples, detection limit of testing methods, and costs of testing. Detection probability defined as the probability that an infected herd was correctly classified as infected was derived from the simulation model based on each testing method. Cost-effectiveness analysis was performed to assist decision-maker for determination of optimal testing strategies under different circumstances.

Conclusions
Pooling of fecal cultures resulted in a higher detection probability compared with individual-sample testing, especially for low prevalence. Environmental sampling showed promising results for detection of paratuberculosis in high prevalence herds. Findings from this study can be a useful tool to determine appropriate testing strategies in various situations with differences in herd characteristics and levels of infection.
Validating and interpreting tests when the aim is to decrease the uncertainty on the serological status rather than the infection status

Didier Verloo, Carine Letellier, Gilles Maquet, Luc Demeulemeester, Koen Mintiens & Pierre Kerkhofs

Description of the problem that was addressed
EU directives for import and export of cattle are in part based on the fact that animals should be seronegative against Bovine herpes virus 1.
Out of the available tests two candidate blocking ELISAs and two indirect ELISAs were selected by the Belgian reference lab and, in order to avoid dependency conditional on the serological status, it was decided a priori that a multiple test setup should be based on one blocking ELISA and one indirect ELISA. The primary target of this study was to guide the decision on which blocking and which indirect ELISAs to choose and how to interpret the test results.
The challenge was to validate the four candidate antibody ELISAs in absence of a reference test which could determine the “true serological” status and the interpretation of the parameters in the model.

Materials and methods (including assumptions and computational implementation of the method, if applicable)
Different latent class models were compared using the deviance information criterion (DIC) starting with the simplest model under conditional independence and gradually increased the complexity by including conditional covariances between test pairs. All was done in WinBugs. All priors were noninformative and all models were identifiable.

Conclusions
According to the DIC, the model allowing a dependency between the se and sp of both indirect and both direct ELISAs was the better model. This is also the biologically most plausible model. Tests and test combinations were chosen based on the output of this model.
The latent “diseased” variable is dependent on the tests in the setup and the used model. For this study the latent variable will rather define a sort of seropositive status of the animal instead of the disease status therefore influencing the conditional probabilities se and sp and the post test probabilities. However, as the regulations/decisions are based on the seronegative status of the animal, the estimates from this model are valid within this framework.
We conclude with the remark that, as the latent variable is per definition not observable interpretation of “no-gold standard” models should always be done with care and with complete understanding of the model.
Use of latent class methods with epidemiologic data: a review

SD Walter

Description of the problem that was addressed
Evaluation of diagnostic and screening tests when there is no gold standard, or if there is restricted access to an appropriate gold standard validation.
Identification of design principles and assumptions required to use latent class models for this type of problem.

Materials and methods (including assumptions and computational implementation of the method, if applicable)
Construction of appropriate models and the corresponding likelihoods for typical situations in epidemiology.
Identification of the required number of observations (or constraints) required for estimation of model parameters.
Applications to multi-test diagnostic data, case-control studies, multiple samples, screening with gold standard validation on test positives, and diagnostic risk marker studies.
Other applications include prevalence surveys, sequential diagnostic testing, natural history studies, and meta analysis of diagnostic and screening test data.
Consideration of correlated error models.

Conclusions
Latent class models have enjoyed wide applicability in the evaluation of diagnostic and screening tests using clinical and epidemiologic data. Further research is needed on assessing correlated error structures, and on model behaviour in small sample situations.
Use of a bovine *Neospora caninum* ELISA in hemolyzed and non-hemolyzed fox samples

W. Wapenaar, H. W. Barkema, R. M. O'Handley

**Description of the problem that was addressed**

There is an abundance of foxes and coyotes in Prince Edward Island (Canada), and their role in the epidemiology of *Neospora caninum* is unknown. Additionally, PEI has the highest density of dairy farms in Canada and contact between wild canids and cattle occurs frequently. To estimate the prevalence of *N. caninum* in these wild canids, accurate sero-diagnostic tests are vital. However, with the collection of field samples from these species after they are hunted or trapped, hemolysis of blood samples often occurs. Goal of this study was to investigate the influence of hemolysis in sero-diagnostic tests. Several ELISA’s are registered for use in cattle and / or dogs, we decided to work with a competitive ELISA because of several presumed advantages: it is not species specific, more objective than an IFAT because you measure optical density in stead of assessing fluorescence, and it is easy to perform on a high number of samples.

**Materials and methods (including assumptions and computational implementation of the method, if applicable)**

Sera were collected from 10 ranched foxes that had not been exposed to *N. caninum* infected material for three generations, and were therefore assumed to be negative controls. After collection of blood samples, the foxes were vaccinated with a *N. caninum* vaccine, registered for cattle. Blood samples of the foxes was collected twice after vaccination. Of every sample clean serum was collected before hemolyzing. Thereafter, the sample was mixed and left at room temperature for 4-24 hours. All samples were analyzed in duplicate.

**Conclusions**

The high proportion of false-positives in hemolyzed samples may be a fox-specific problem. Further investigation is being done to find this out. Diluting the sample (it is used undiluted when following manufacturers recommendations) helps to increase specificity but decreases sensitivity. Cleaning up the hemolysis by extracting hemoglobin from the sample is expensive and with that procedure you will be adding other components that might interfere with the cELISA. Although a cELISA seems ideal to use across species, we must remain cautious when using blood samples that were collected in more difficult circumstances, where hemolysis can occur. When conducting research with non domestic species, this often is an issue that should be addressed carefully.
International Validation: Moving towards a graduated process

Peter Wright

Description of the problem that was addressed

In the OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals (the Terrestrial Manual), diagnostic techniques with associated tests are described for each disease. Some tests are not classified as prescribed or alternative tests. But, unless indicated otherwise, all are presumed to be validated, i.e. capable of producing results that properly infer the infection/exposure status of animals. The purpose of implementation of the test, however, is not only to predict infection status of the animal for trade purposes. There are many other reasons for testing, including: seromonitoring, demonstration of freedom from infection, estimation of prevalence of infection for risk assessment, etc. Therefore, test validation should be a process that will demonstrate fitness of that test for a particular use. The current procedure in use by the OIE for the test qualification does not take into consideration all the different purposes for laboratory testing for animal diseases. Even for trade purposes, there is no guideline available that specifies what is required in a test dossier that is submitted to the OIE for evaluation. OIE has received requests from many Member Countries and also from commercial test manufacturers to provide clear guidelines and much broader recognition of diagnostic tests as fit for specific purposes, not only for trade.


Conclusions

The above meeting followed the adoption of Resolution XXIX at the OIE General Session, Paris, May, 2003. This Resolution pertains to the need for development of new OIE procedures for the validation and certification of diagnostic assays for infectious animal diseases. At the Vienna meeting, participating experts were from international organisations, public institutions, and importantly, also private companies. The conclusions and recommendations of the meeting addressed three main issues:

The standard template to be used when validating and submitting a test to the OIE for evaluation,
The procedure used to evaluate the data submitted to the OIE, and
The establishment of reference material collections.

The template discussed at this meeting was developed from input at the Vienna meeting and subsequently adopted by the OIE.
Fit for Purpose: A Laboratory Perspective

Peter Wright

Description of the problem that was addressed

It has become obvious that it is necessary to improve the current system for the qualification and certification of diagnostic assays for infectious animal diseases. This is why the OIE Collaborating Centre for ELISA and Molecular Techniques in Animal Disease Diagnosis, have convened a consultants meeting on the validation and certification of diagnostic assays for infectious animal diseases. This meeting was meant to elicit discussions on two main areas:

Validation with respect to ‘fitness for purpose’;
The process by which the assay (kits/reagents) can be certified by the OIE for a purpose.


Conclusions

Animal disease management is carried out for economic, public health, and environmental reasons. Risk assessment is the key component in disease management. An important factor in risk assessment is evaluation of animals and their products. Diagnostic testing is an important activity in this process is useful only if it is applied within specific contexts. Therefore, testing can be classified as to its fitness for purpose. The purposes can be classified into six broad categories.

1) Demonstrate population ‘freedom’ from infection (prevalence apparently zero)
2) Demonstrate freedom from infection or agent in individual animals or products for trade purposes
3) Eradication of infection from defined populations
4) Confirmatory diagnosis of clinical cases
5) Estimate prevalence of infection to facilitate risk analysis (surveys, classification of herd health status, implementation of disease control measures)
6) Determine immune status in individual animals or populations (post-vaccination)

Fitness for purpose discussed at this workshop was developed from input at the Vienna meeting and subsequently adopted by the OIE.
Annex 4
List of Participants

## Belgium

<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Address</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerts, Prof Marc</td>
<td>Limburg Universitair Centrum, Center for Statistics</td>
<td>Universitaire Campus Diepenbeek 3590, Belgium</td>
<td><a href="mailto:marc.aerts@luc.ac.be">marc.aerts@luc.ac.be</a></td>
</tr>
<tr>
<td>Deluyker, Dr Hubert</td>
<td>EFSA, AHAW</td>
<td>10, Rue de genève Evere 1140, Belgium</td>
<td><a href="mailto:hubert.deluyker@efsa.eu.int">hubert.deluyker@efsa.eu.int</a></td>
</tr>
<tr>
<td>Faes, Dr Christel</td>
<td>Limburg Universitair Centrum, Center for Statistics</td>
<td>Universitaire Campus Diepenbeek 3590, Belgium</td>
<td><a href="mailto:Christel.faes@luc.ac.be">Christel.faes@luc.ac.be</a></td>
</tr>
<tr>
<td>Hens, Dr Niel</td>
<td>Center for Statistics, Limburg Universitair Centrum</td>
<td>Universitaire Campus, Building D Diepenbeek 3590, Belgium</td>
<td><a href="mailto:NIEL.HENS@LUC.AC.BE">NIEL.HENS@LUC.AC.BE</a></td>
</tr>
<tr>
<td>Mintiens, Dr Koen</td>
<td>Veterinary and Agrochemical Research Centre, CCVD</td>
<td>Groeselenberg, 99 Brussels 1180, Belgium</td>
<td><a href="mailto:komin@var.fgov.be">komin@var.fgov.be</a></td>
</tr>
<tr>
<td>Verloo, Dr Didier</td>
<td>European Food Safety Authority (EFSA), BIOHAZ</td>
<td>Rue de Geneve, 10 Brussels 1140 Brussels, Belgium</td>
<td><a href="mailto:didier.verloo@telenet.be">didier.verloo@telenet.be</a></td>
</tr>
</tbody>
</table>

## Canada

<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Address</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berezowski, Dr John</td>
<td>Alberta Agriculture Food and Rural Development, Food Safety Division</td>
<td>6909 - 116 St Edmonton T6H 4P2 Alberta, Canada</td>
<td><a href="mailto:John.Berezowski@gov.ab.ca">John.Berezowski@gov.ab.ca</a></td>
</tr>
<tr>
<td>Walter, Dr Stephen</td>
<td>McMaster University, Canada, Clinical Epidemiology and Biostatistics</td>
<td>1200 Main St West Hamilton L8N 3Z5 Ontario, Canada</td>
<td><a href="mailto:Walter@mcmaster.ca">Walter@mcmaster.ca</a></td>
</tr>
<tr>
<td>Wapenaar, Dr Wendel</td>
<td>University of Prince Edward Island, Atlantic Veterinary College, Department of Health Management</td>
<td>550, University Avenue Charlottetown C1A4P3 Prince Edward Island, Canada</td>
<td><a href="mailto:wwapenaar@upei.ca">wwapenaar@upei.ca</a></td>
</tr>
<tr>
<td>Wright, Dr Peter</td>
<td>National Centre for Foreign Animal Disease, Canadian Food Inspection Agency</td>
<td>1015 Arlington Street T2200 Winnipeg B3E 3m4 Manitoba, Canada</td>
<td><a href="mailto:pwright@inspection.gc.ca">pwright@inspection.gc.ca</a></td>
</tr>
</tbody>
</table>

## Denmark

<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Address</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baadsgaard, DVM Niels</td>
<td>Danish Institute of Agricultural Sciences, 8830, Denmark</td>
<td><a href="mailto:NielsP.Baadsgaard@agrsci.dk">NielsP.Baadsgaard@agrsci.dk</a></td>
<td></td>
</tr>
<tr>
<td>Bak, Mrs Hanne</td>
<td>Danish Institute for Food and Veterinary Research, Serology Bulowsvej, 27 Copenhagen 1790, Denmark</td>
<td><a href="mailto:hba@dfvf.dk">hba@dfvf.dk</a></td>
<td></td>
</tr>
<tr>
<td>Ekeroth, Dr Lars</td>
<td>Danish Institute for Food and Veterinary Research, Dep. of Veterinary Diagnostics and Research Bulowsvej, 27 Copenhagen 1790, Denmark</td>
<td><a href="mailto:Lek@dfvf.dk">Lek@dfvf.dk</a></td>
<td></td>
</tr>
<tr>
<td>Greiner, Dr Matthias</td>
<td>Danish Institute for Food and Veterinary Research, International Epilab Moerkhoj Bygade 19 Soeborg 2860, Denmark</td>
<td><a href="mailto:mgr@dfvf.dk">mgr@dfvf.dk</a></td>
<td></td>
</tr>
<tr>
<td>Toft, Dr Nils</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**International Meeting On The Design and Analysis Of Diagnostic Test Evaluation Studies**

**The Golf View Hotel, Nairn, Inverness, Scotland**

**Monday 28th March – Wednesday 30th March, 2005**

<table>
<thead>
<tr>
<th>Country</th>
<th>Name</th>
<th>Email Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td><a href="mailto:nt@dina.kvl.dk">nt@dina.kvl.dk</a></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>Virtala, Dr Anne Maija</td>
<td><a href="mailto:anna-maija.virtala@helsinki.fi">anna-maija.virtala@helsinki.fi</a></td>
</tr>
<tr>
<td>France</td>
<td>Biteau-Coroller, Ms Fabienne</td>
<td><a href="mailto:biteau@cirad.fr">biteau@cirad.fr</a></td>
</tr>
<tr>
<td></td>
<td>Guillossou, DVM Stephane</td>
<td><a href="mailto:s.guillossou@synbiotics.fr">s.guillossou@synbiotics.fr</a></td>
</tr>
<tr>
<td></td>
<td>Klewer, Mrs Kristine</td>
<td><a href="mailto:K.Klewer@institut-pourquier.fr">K.Klewer@institut-pourquier.fr</a></td>
</tr>
<tr>
<td>Germany</td>
<td>Schroeder, Dr Heinz</td>
<td><a href="mailto:heinz.schoeder@bioscreen-ms.de">heinz.schoeder@bioscreen-ms.de</a></td>
</tr>
<tr>
<td>Greece</td>
<td>Georgiadis, Dr Marios</td>
<td><a href="mailto:mariosg@vet.auth.gr">mariosg@vet.auth.gr</a></td>
</tr>
<tr>
<td>New Zealand</td>
<td>Glossop, Dr Jamie</td>
<td><a href="mailto:J.c.glossop@massey.ac.nz">J.c.glossop@massey.ac.nz</a></td>
</tr>
<tr>
<td></td>
<td>Heuer, Dr Cord</td>
<td><a href="mailto:c.heuer@massey.ac.nz">c.heuer@massey.ac.nz</a></td>
</tr>
<tr>
<td>Switzerland</td>
<td>Doherr, Dr Marcus</td>
<td><a href="mailto:marcus.doherr@itn.unibe.ch">marcus.doherr@itn.unibe.ch</a></td>
</tr>
<tr>
<td></td>
<td>Ochs, Dr Hansueli</td>
<td><a href="mailto:Hansueli.Ochs@bvet.admin.ch">Hansueli.Ochs@bvet.admin.ch</a></td>
</tr>
<tr>
<td>The Netherlands</td>
<td>Dopfer, Dr Dorte</td>
<td><a href="mailto:Dorte.dopfer@wur.nl">Dorte.dopfer@wur.nl</a></td>
</tr>
<tr>
<td></td>
<td>Engel, Dr Bas</td>
<td><a href="mailto:Bas.Engel@wur.nl">Bas.Engel@wur.nl</a></td>
</tr>
<tr>
<td></td>
<td>Jacobs, Dr Liesbeth</td>
<td><a href="mailto:Liesbeth.Jacobs@wur.nl">Liesbeth.Jacobs@wur.nl</a></td>
</tr>
<tr>
<td></td>
<td>Orsel, DVM Karin</td>
<td><a href="mailto:Orsel.DVM@wur.nl">Orsel.DVM@wur.nl</a></td>
</tr>
</tbody>
</table>
United Kingdom

Elliott, Dr Heather
Institute for Animal Health,
Ash Road Pirbright
GU24 ONF Surrey, UK
heather.elliott@bbsrc.ac.uk

Gibbens, Mrs Jane
DEFRA, Vet Surveillance Division, ROOM
603A
1A Page Street London
SW1P 4PQ , UK
Jane.gibbens@defra.gsi.gov.uk

Gunn, Mr George
SAC Epidemiology Research Unit, Animal
Health Group, R&D
Drummondhill Stratherrick Road Inverness
IV2 4JZ , UK
svepm05@sac.ac.uk

Humphry, Dr Roger
SAC Epidemiology Research Unit, Animal
Health Group, R&D
Drummondhill Stratherrick Road Inverness
IV2 4JZ , UK
svepm05@sac.ac.uk

Iversen, DVM Johanne
Veterinary Laboratories Agency (VLA),
Centre of Epidemiology and Risk Analysis
Woodham lane New Haw
KT15 3NB Addlestone, UK
j.iversen@vla.defra.gsi.gov.uk

Mpoko, Dr David
Guildhay Limited, R&D
6 Riverside Business Centre, Walnut Tree
Close Guildford

United States of America

Ebel, Dr Eric
USDA - Veterinary Services,
2150 Centre Ave MS 3 E 20 Fort Collins
1 CO, USA
eric.d.ebel@usda.gov

Elvinger, Dr François
Virginia - Maryland Regions College of
Veterinary Medicine, Large Animal Clinical
Sciences
Virginia Tech Blacksburg
Virginia, USA
Elvinger@vt.edu

Johnson, Dr Wesley
University of California, Irvine, USA
Wjohnson@uci.edu

Tavornpanich, Dr Saraya
University of California, Davis, Epidemiology
and Medicine
School of Veterinary Medicine Tupper Hall,
2108 Davis
95616-8737 CA, USA
stavornpanich@ucdavis.edu