



NATIONAL SEROLOGY REFERENCE LABORATORY, AUSTRALIA

**EVALUATION OF SEROLOGICAL ASSAYS:
AN APPROACH SUGGESTED BY NRL**

Contact: Mr Wayne Dimech
Project Manager
National Serology Reference Laboratory, Australia
4th Floor, Healy Building
41 Victoria Parade
FITZROY VIC 3065

Tel: +613 9418 1111
Fax: +613 9418 1155
Email: wayne@nrl.gov.au

Index

Page	Section	Topic
1	1.	Background
	1.1	History
2	1.2	Value of Centrally Coordinated Evaluations of Serological Assays
3	1.3	NATA View on Centralising Kit Evaluations
4	2.	AIM
	2.1	Laboratory Participation
5		Laboratory Participation continued
6	3.	Methods
	3.1	Samples
7	3.2	Consumables
	3.3	Follow-up Testing
	3.4	Evaluation Parameters
8		Evaluation Parameters continued
9	3.5	Procedure
10	3.6	Data Collection Methods
11	3.7	Analysis
12	3.8	Reporting
13	4	References



1. BACKGROUND

1.1 History

The National Serology Reference Laboratory, Australia (NRL) has been evaluating assays for diagnostic serology for nearly two decades. It has contributed significantly to the understanding of the suitability and use of HIV and HCV assays for both blood screening and diagnostic laboratories. During this time, the NRL has acquired expertise in accomplishing serological assay evaluations.

At an Annual General Meeting of the Australian Society for Microbiology (ASM), Serology Special Interest Group (SSIG) the NRL and the SSIG agreed to collaborate to deliver information about the analytical performances of assays to diagnostic laboratories. One of the actions arising from that meeting was for NRL to produce guidelines for the proposed kit evaluations. The evaluation criteria proposed by NRL are set out in this document. Information was drawn from a range of published documents ^{1,2,3,4,5,6,7,8}. The NRL will collaborate with diagnostic laboratories, in particular with members of the ASM SSIG, to facilitate multicentre evaluations of assays. It is intended that groups of interested laboratories will use these guidelines as the basic method for carrying out evaluations of serological and similar assays. The NRL will provide technical and practical support to the laboratories. It is keen to establish working parties to investigate the most appropriate target assays for the early evaluations. The NRL also proposes that laboratories performing independent, in-house kit evaluations or validations could use these guidelines.

1.2 Value of Centrally Coordinated Evaluations of Serological Assays

Meaningful evaluations of serological assays involve commitment of significant time and resources. Each evaluation requires the development of a sound, scientific protocol that includes implementation, methods and analysis. The collection and storage of significant numbers of samples as well as the production of results and reports should be managed through planning within a quality management system. Results should be made available for



peer review. Most busy general laboratories have little time for evaluating assays using a model such as that suggested here.

Reports on completed assay evaluations would be useful to laboratories in a number of circumstances including:

- reagents unexpectedly and permanently becoming unavailable through their withdrawal by the reagent's manufacturer;
- reagents becoming temporarily unavailable;
- multiple reagents for the same market needing to be assessed;
- assessment of new reagents on those suspected to be performing poorly;
- changing the supplier of reagents e.g. after the purchase of, or a change to, new instrumentation;
- local outbreak of an infection or disease that requires a laboratory to implement a new testing method rapidly.

There are many kit evaluations performed in both public and private laboratories. Often the results are not scientifically valid because of poor experimental design e.g. they are conducted in limited numbers and types of sample, or in samples of unknown status. The cost of these studies is absorbed by the testing laboratory or recouped by manufacturers through higher kit prices. The results of many studies are never reported. Larger evaluations may be presented at state or national scientific conferences the results being traceable only via the abstracts that often do not contain complete details of the methods or the findings. These abstracts are often quickly lost to follow up.

1.3 NATA's View on Centralizing Kit Evaluations

The Australian Standard ISO/IEC 17025: 1999 states that "The range and accuracy of the values obtainable from validated methods (e.g. the uncertainty of the results, detection limits, sensitivity of the method, linearity, limit of repeatability and/or reproducibility, robustness



against external influences and/or cross-sensitivity against interference from the matrix of the sample/test objective) as assessed for the intended use, shall be relevant to the clients' needs"⁹. NATA regulations oblige all laboratories to show scientific evidence to support their selections of assays. This evidence must be based on appropriate scientific validation with the results documented in a manner that can be traced and retrieved. Therefore, a well-structured, central evaluation would fulfil the NATA requirements. However, two further important issues exist.

- A laboratory must show that it can reproduce the evaluation results using the method selected. Mechanisms for demonstrating staff competence in this area could be achieved by the use of a validation panel and/or evidence of training by the manufacturer's customer support staff.
- Further validation with selected specimens may be required if the laboratory serves a population that differs significantly from that tested in the central evaluation.

2. AIM

This document aims to propose a protocol for characterizing the performances of diagnostic assays used in infectious diseases serology.

The NRL will seek groups of interested staff in various laboratories to establish relevant working parties. These working parties will design specific evaluation protocols for a target analyte, help to collect the appropriate samples for the evaluation and participate in testing. The NRL will provide the resources for data analysis and report writing. NRL will maintain a quality plan incorporating all aspects of the assay evaluation including authorship of publications. The responsibilities of the group's members for each separate evaluation will be determined during the development of the specific protocol.



3. METHODS

3.1 Coordination

There are three situations in which this protocol could be utilised. They are when:

1. The NRL performs the evaluations in-house without collaboration from other laboratories.
2. Diagnostic laboratories collaborate with the NRL as working parties focused on evaluations for specific analytes. The form of collaboration would be negotiated and would depend upon the number and needs of laboratories, the resources available and the amount of autonomy required by the laboratories involved. Each collaboration would be considered and documented separately.
3. An individual laboratory or group of laboratories would evaluate assays utilising this protocol without involvement of the NRL.

Specific needs of laboratory staff will influence the range of, and priorities for, the evaluations. It is acknowledged that diagnostic laboratory staff have considerable expertise in specific areas of infectious diseases serology and are keen to contribute to scientific endeavour. However their labour and consumable resources are often very limited. The infrastructure in routine laboratories is more suited to large volume testing and the range of samples available is large. There are several laboratories, primarily within the Public Health Laboratory Network, that maintain reference tests that are not generally available. These tests are vital to the success of the evaluations as in many cases, it is the results of these tests that will provide the status of the samples tested.

Laboratories that choose to perform evaluations independently of the working parties may also request assistance from the NRL. This assistance may be in the form of sample



collection and storage, customisation of the protocol to suit the analyte under investigation or the analysis of data. The NRL would also be prepared to develop or help develop and monitor a quality plan for each evaluation.

3.2 Samples

All evaluations require well-characterised samples of known status. Access to samples is often a limiting factor for laboratories performing evaluations, because they may be rare or the status difficult to determine. Some useful samples may be commercially available. However they are expensive and may be beyond the resources allocated to the evaluation. The NRL has a sample bank containing significant numbers of samples that may be used for evaluations. To perform the proposed broad range of evaluations, the sample bank will require additional samples to be collected for each, specific evaluation. In forming collaborating working groups, preference will be given to those laboratories that can aid in the collection of appropriate sera that will help expand the evaluation sample bank over time. The NRL has approval from several ethics committees throughout Australia to request additional blood samples from individuals who test reactive to analytes of interest during routine pathology testing. Further ethics committees' approvals will be sought from specific institutions where necessary.

All samples will be stored and aliquotted at the NRL and transported appropriately. Panels for evaluations will be constructed from stored aliquots and distributed under suitable transport conditions to testing laboratories participating in any evaluation. The remaining aliquots of samples will be retained and made available for future evaluations or validations as required. The NRL now has the resources that enable it to maintain substantial banks of samples in good condition and make them available to collaborating laboratories.

A range of samples will be required for the evaluations for example known positive and negative samples, samples taken over the entire course of the infection and samples known to be falsely reactive. All samples will be characterised to obtain a true status. This will



usually be achieved by testing the samples using a “gold standard” assay, many of which are only performed by Public Health Reference Laboratories. “Gold standard” has been described by the NCCLS as “a non-specific term that indicates that a process or material(s) is the best available approximation of the truth”². It is acknowledged that many “gold standard” results may not be definitive. If necessary, follow up testing such as Western blots or other appropriate methods may be used to define the status of samples whose test results are difficult to interpret. The choice of reference testing will be determined by the working parties and defined in the quality procedures.

Several of the evaluation parameters (e.g. linearity for quantitative assays, and reproducibility) require testing known dilutions of a, highly reactive sample. A single, highly reactive sample will be diluted in a matrix consisting of human serum or plasma devoid of the analyte being evaluated. For quantitative assays, multiple dilutions of the sample should be made to reflect the full range of clinically possible levels from high to undetectable. The sample dilutions will be made in sufficient quantity to allow testing by all assays within a single evaluation series to enable comparisons between assays.

As well as the assay standards and controls, all test runs will contain a QC sample that will be supplied by the NRL. QC samples will be calibrated against International Standards where available. As a guide to expected reactivity, reference ranges for the QC sample will be calculated prior to the evaluation testing.

3.5 Evaluation Parameters

When an evaluation is performed under the auspices of the NRL, each evaluation will require the development of a specific protocol by the working group. This protocol will be reviewed by NRL prior to the commencement of testing. The protocol will comply fully with the requirements of the NRL’s Quality Management System and other policies. The selection of



appropriate parameters for each evaluation depends upon the existence of a suitable “gold standard”, reference method or international standard, the availability of appropriate samples and the clinical implications of testing. The evaluation will comprise a range of investigations.

3.5.1 Sensitivity: The sensitivity is the proportion of true positive samples that are correctly identified by the assay as reactive. The number of samples used will be determined by the working party, during the development of specific protocols. However the number used should establish the sensitivity within reasonable confidence intervals (Table 1). The results of the sensitivity testing will indicate the value of a reactive result. Where appropriate, the sensitivity panel will include a population of characterised true positive samples with low reactivity to determine the ability of the assay under evaluation to detect low concentrations of the target analyte in undiluted clinical samples.

3.5.2 Specificity: The specificity is the proportion of true negative samples that are correctly identified by the assay as non-reactive. The number of samples used will be determined by the working party and will depend on the confidence intervals required of the results.

3.5.3 Within-run precision: The precision of an assay is a measure of the variation in results of the same sample when tested multiple times. The within-run precision measures the variation in results when a single sample is tested repeatedly in a single, assay run thereby involving the least opportunity for introducing variation. Results from a minimum of 20 sample replicates is required to analyse statistically the results from this testing. The results of the within-run precision will indicate the degree of random error inherent within an analytical run. Within-run precision may be estimated at different concentrations of the target analyte across the dynamic range of the assay.

3.5.4 Between run precision: To establish the between-run variation, the same samples as those used to determine within run precision will be tested in a number of separate



runs. The protocol would specify the number and frequency of replicate tests. The same batch of reagent must be used. The between-run precision results indicate the degree of random error inherent in the method from run-to-run.

3.5.5 Linearity: The linearity experiment confirms that a quantitative method is linear over the clinical range of results expected in a population. Multiple replicates of a series of dilutions of a highly reactive sample are tested in a single run and the test results are compared to the expected results. The sample should be calibrated against an international or national standard where one is available.

3.5.6 Lower detection limits: The detection limit is the lowest concentration of the target analyte that is detectable consistently in the assay under evaluation. The analysed data from the linearity studies will be used to estimate the lower limit of detection. The lower detection limit is an indication of the analytical sensitivity of an assay.

3.5.7 Seroconversion sensitivity: A seroconversion panel is a series of samples obtained from an individual over the course of acute infection. Usually there are no more than 10 samples per panel. Testing a seroconversion panel allows assessment of the ability of an assay to detect the target analyte early in the disease process .

3.5.8 Cross-Reactivity: Some samples can produce falsely reactive results in assays due to non-specific binding. Where available a panel of samples, known to cause non-specific reactivity in immunoassays, will be tested to estimate the proportion of false reactivity produced by the assay under investigation.

The number of tests, including controls will be approximately 800 for an assay to be used for diagnostic (not blood screening) purposes. Numbers will vary and be determined for each assay depending on the intended use and the population in which it will be tested.



3.6 Procedure

All samples will be tested and all test runs will be validated according to the assay manufacturer's package insert. Results from test runs that are invalid will not be included in the analysis. If any deviation from the package insert is directed by the manufacturer, this will be noted in the final report. The evaluation will replicate the practices that will be used in the field. All manufacturers will have an opportunity to identify a laboratory to evaluate their assays. The NRL will work with manufacturers and the designated laboratories to ensure the staff are adequately trained and the instrumentation conforms to the NATA/ISO 15189 maintenance guidelines. Where possible, microtitre plate enzyme immunoassays will be performed in a single laboratory to reduce variation between washers and readers.

If the standards or QC results are "out of control" as specified in the specific protocol developed by the working party, the testing will be halted and the NRL and the manufacturer will be informed. The NRL will liaise with the manufacturer to determine the cause and implement corrective action where necessary. Testing will re-commence only when all parties are satisfied the assay is 'in control'.

A manufacturer's representative will be invited to visit the testing laboratories to ensure that the test is performed according to manufacturer's instructions and to review results prior to their being submitted to NRL. However all data must be passed to NRL, without any alteration, whether valid or not, by the testing laboratories for analysis and reporting.

The manufacturer will be invited to contact NRL to discuss any issues arising from the testing procedure prior to the results being submitted. However, once submitted, the analysis will be performed and reported as detailed below.



3.7 Data Collection

All data will be forwarded to the NRL or other organisation analysing the results as hard-copy printouts with or without an electronic copy. Also, testing laboratories will be required to provide copies of their worksheets containing sample identifiers, run data test kit batch number and expiry date, operator and any calculations that are performed on the results. A copy of the printout will be made available to each manufacturer for their own assays only. The data will be stored as defined by NRL procedures. All data from evaluations performed in collaboration with NRL will be owned by the NRL.

3.8 Analysis

The analyses will be performed according to NRL's procedures and work instructions.

3.8.1 Sensitivity and specificity: Estimates of sensitivity and specificity with 95% confidence intervals will be calculated. If possible, when more than one assay for a particular analyte is being assessed, the sensitivity and specificity found for individual assays will be compared by appropriate statistical analysis to determine whether the results differ significantly between assays.

3.8.2 The delta value: The delta values for the results in the positive and negative populations of samples in each assay will be calculated where appropriate. The delta value is a measure of the removal of the populations' results from the cut-off and takes into account the distance of the mean from the cut-off and the distribution around the mean. Once the removal of a population is 3SD or more from the cut-off, the perceived improvement in the significance of the value should be interpreted with caution.

3.8.3 Within-run precision: The mean, median, coefficient of variation (CV) and standard deviation (SD) will be calculated. Any differences between assays will be determined using statistically valid analyses.



3.8.4 Between-run precision: As for the within-run precision.

3.8.5 Linearity: The results of the linearity study will be plotted against the expected results. A line of best fit will be drawn and the slope and intercept calculated. Limits will be demonstrated and the correlation coefficient reported.

3.8.6 Seroconversion sensitivity: The results will be presented in graphical or tabular form relating results obtained to the post exposure period.

3.9 Reporting

The reports will follow formats accepted for scientific writing with Aim, Methods, Results and Conclusion sections. The report will include a brief description of the assay based on information provided by the manufacturer. The conclusions will reflect the differences in the results obtained between assays. As the aim of the project is to detect assays that are significantly different from other kits on the market, there is no intention to otherwise rate the kits or to provide comment on which kits should be selected or used in a clinical setting. The NRL may give further interpretive advice to laboratories if requested to do so.

The results of “gold standard” and supplemental tests will be made available to the manufacturers, along with the raw data pertaining to their assay(s), once testing is complete.

The appropriate forum for publishing the findings of each evaluation will be determined by individual working parties. Publication of the findings in peer-reviewed journals would be encouraged. If peer-review publication is not possible, the summaries of the findings will be presented in News and Reviews On-Line and detailed results made available via the NRL web site. All collaborating laboratories will be recognized appropriately and their authorship in any publication will be included based on their contribution in line with NRL’s publication



procedures. The NRL will encourage participating laboratories to present the results at appropriate scientific meetings.



4. REFERENCES

1. **Heron L.** et al. 1994. Proposed general protocol for the study of serologic diagnostic reagents. *Aust. Micro.* **15**:37-45.
2. **NCCLS.** User protocol for the evaluation of qualitative test performance; Proposed guidelines. NCCLS document EP12-P. Wayne, PA: National Committee for Clinical Laboratory Standards; 2000
3. **NCCLS.** A framework for NCCLS evaluation protocols; Proposed guideline. NCCLS document EP19-P. Wayne, PA: National Committee for Clinical Laboratory Standards; 2000
4. **NCCLS.** Laboratory statistics- Standard deviation; a report. NCCLS document EP13-R. Wayne, PA: National Committee for Clinical Laboratory Standards; 1995
5. **NCCLS.** User demonstration of performance for precision and accuracy; proposed guidelines. NCCLS document EP15-P. Wayne, PA: National Committee for Clinical Laboratory Standards; 1998
6. **NCCLS.** Preliminary evaluation of quantitative clinical laboratory methods; approved guidelines. NCCLS document EP10-A. Wayne, PA: National Committee for Clinical Laboratory Standards; 1998
7. **Empson M.B.** 2002. Statistics in the pathology laboratory: Characteristics of diagnostic tests. *Pathol.* **33**:93-95.
8. **McAdam AJ.** 2000. Discrepant analysis: How can we test a test? *J. Clin. Micro.* **38**:2027-2029.
9. **Haekkel R.** (ed) Evaluation methods in laboratory medicine. VCH, Weinheim, Germany.
10. **Australian Standard.** General requirement for the competence of testing and calibration laboratories. AS ISO/IEC 17025:1999. Standards Australia International.