Quality in systematic literature review

Parry J.E.[1]*, Clifton-Hadley R.[2], Broughan J.[1], Goodchild T.[1], Upton P.[1], Collins J.[3], Pritchard E.[2], and Downs S.H.[1] for the Meta-analysis Working Group[2]

[1] Centre for Epidemiology and Risk Analysis, [2] Statutory and Exotic Bacteria and [3] VLA Library, The Veterinary Laboratories Agency, New Haw, Addlestone, Surrey, KT153NB, UK
*J.E.Parry@vla.defra.gsi.gov.uk

Veterinary Laboratories Agency®

INTRODUCTION

We seek to answer the following question "What are the best estimates of the classical test characteristics of sensitivity and specificity for diagnostic tests for bovine tuberculosis in cattle in GB?" by conducting a meta-analysis on data extracted from the systematic review of studies that have measured test performance of diagnostic tests for *M. bovis* in cattle. To identify test performance data (sensitivity and specificity) we needed to conduct a systematic literature review which would identify relevant references whilst minimising selection and publication biases.

METHODS

Systematic literature comprising 2 parts: **STAGE 1** – a search for references and review of abstracts or titles to identify references that could contain estimates of test performance and **STAGE 2 –** detailed review of eligible references to reject or collect data from accepted references using a standardised format.

1. DEFINING A COMPREHENSIVE LITERATURE SEARCH FRAME

1. Electronic databases including: Web of Knowledge (which simultaneously searches Web of Science 1995- Current contents 1998-,CAB abstracts 1910-, Medline 1950-); Dialogue (which simultaneously searches Embase 1974-, Agricola 1970-, Agris 1975-); Procite database of bTB references maintained by the Statutory and Exotic Bacterial Diseases Programme at the VLA
2. Bibliographies of references reviewed
3. References known to members of the Working Group
4. Reports held by laboratories and research institutions not in the peer reviewed scientific press
5. Research reported at conferences in the recent past and, as yet, unpublished

FINAL SEARCH STRING FOR SEARCH OF ELECTRONIC DATABASES

(bovine tuberc* or mycobacterium bovis*) or ((mycobact* not (paratub* or johne*))
AND
(bovin* or cattle or cow or cows or calf or calves or buffa)
AND
(test* or screen* or diagn* or eia or elisa or pcr or polym* chain react* or lympho* or Interferon or skin or rapid or detect* or peptid* or cervical or caudal or sicct or antibody* or necroscopy or necropsy or survei* or sensitivi* or specifici* or perform* or eval* or valid* or accura* or confirmatory)

The sensitivity of different search strings was tested against a panel of 65 references reporting diagnostic test performance between 1937 and 2007

2. DEFINING INCLUSION CRITERIA AND TESTING REVIEWER AGREEMENT AT STAGE 1

INCLUSION CRITERIA

Primary research reporting sensitivity and/or specificity of a diagnostic test for bovine tuberculosis on cattle or buffalo or providing data enabling the statistics to be calculated

AGREEMENT STUDY

To determine whether good agreement in references that should pass through stage 1 could be achieved five agreement studies were conducted with four reviewers. In each study a random sample of the references identified by test search strings were selected for review by abstract or title alone (where an abstract was not available).

A	Study	No. of Refs*	References that passed stage 1				IC†	Agreement‡
			Rev A	Rev B	Rev C	Rev D		
	1	98	5	6	7	-	0.59	Moderate
	2	99	3	2	3	3	0.29	Poor
	3	100	9	7	6	10	0.84	Almost perfect
	4	100	5	3	4	2	0.46	Moderate
	5	500	34	-	15	-	0.55	Moderate

*randomly selected from results of electronic searches 1 and 2. †Measure of the reliability of a single rater (from 2-way anova model measuring variability between references and between reviewers) ‡ (Landis and Koch 1977)

B	Study	References that pass stage 1			
		Rev A	Rev A+B	Rev A+B+C	Rev A+B+C+D
	1	5	8	10	-
	2	3	5	7	7
	3	9	9	10	10
	4	5	7	8	8
	5	34	-	35	-

It was demonstrated that the inclusion of greater than 3 reviewers did not increase the number of references which passed to stage 2 (B). The use of three reviewers at stage 1 would have been preferable but only two reviewers were available for this task. To ensure the sensitivity of stage 1 it was decided that any reference considered 'borderline' by either reviewer would be discussed and re-reviewed with a joint decision on inclusion or exclusion.

CLARIFYING CRITERIA

After each study agreement between reviewers was measured and the guidance document for review was discussed and criteria clarified. Agreement between reviewers was not high and the principal reason identified through discussion was the variation in the extent and range of information provided in titles and abstracts.

CONDUCT OF STAGE 1 REVIEW

Results from the agreement studies were presented to the project Working Group and it was agreed that since agreement was not high (A) each reference should be reviewed by two reviewers at stage 1 using criteria developed and all references passed by at least one reviewer would be considered eligible for stage 2.

3. STAGE 1
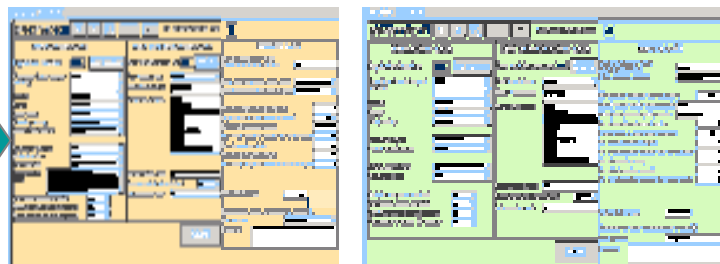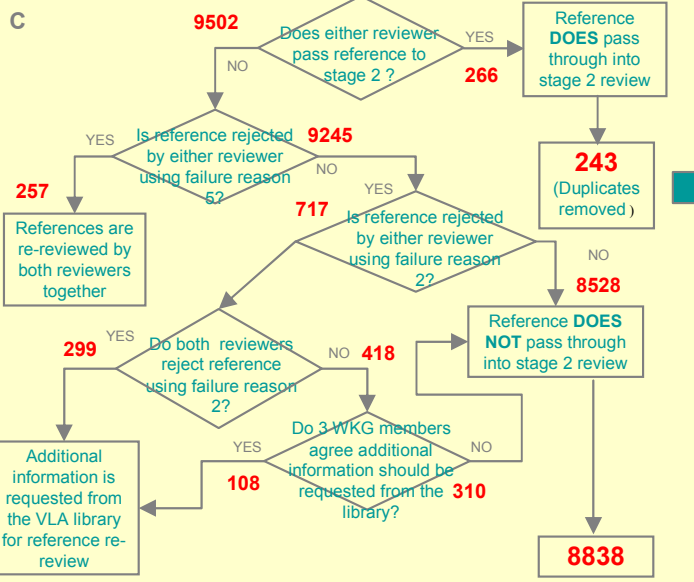#### Selection of references to pass though to full review

9768 abstracts and titles identified in the electronic searches were reviewed by two reviewers. Exclusion criteria for not passing through to stage 2 review were as follows: 1. Wrong subject material; 2. Insufficient information in record (Title only available); 3.Literature review; 4: Surveillance/prevalence report; 5. Not possible to calculate either sensitivity or specificity (borderline reference)

RESULTS: A total of 243 references were eligible for Stage 2 full reference review. Borderline references (failure reason 5) will be re-reviewed simultaneously by the two reviewers. Further reference information is being requested for those containing insufficient to information to review (failure reason 2) are being retrieved.

4. STAGE 2
Full review of references and data collection

References selected as eligible for full review were each randomly allocated to two reviewers from an working group of 17 experts in the field of bovine tuberculosis diagnostics. A database was designed specifically for the standardised data capture (fields agreed by Working Group) and entry for test performance parameters: sensitivity and specificity. Additional fields relating to the animal population, test modification and reference standard were included to capture factors which can influence test performance.

References:
Landis, J. R. and G. G. Koch (1977). "The measurement of observer agreement for categorical data." Biometrics 33: 671-679.