Evaluating Participant Performance in Qualitative EQA Schemes

United Kingdom National External Quality Assessment Service for Microbiology

Dr Vivienne James

UK NEQAS

Qualitative Data

- Categorical measurement expressed by means of a natural language description
  - Nominal e.g. organism name/identity, genotype, presence/absence, positive/ negative
  - Ordinal e.g. 1+, 2+, 3+ (can be ordered) but have no algebraic relationship

- *'There is no such thing as qualitative data. Everything is either 1 or 0'*
  - Fred Kerlinger, Quantitative researcher, Miles and Huberman 1994; Qualitative Data Analysis

Ways of handling qualitative data

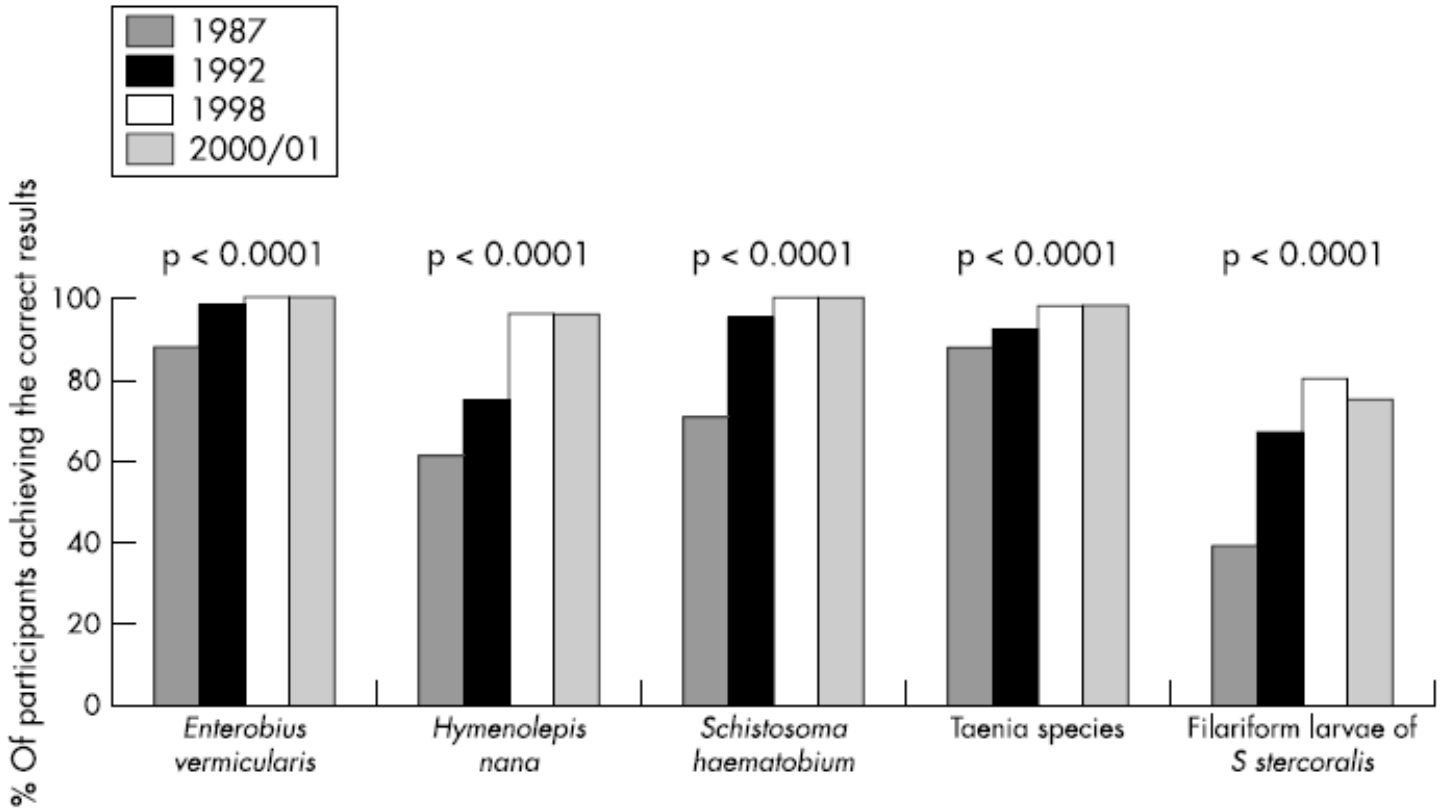
- Use of surrogates
  - Number of participants
  - % laboratories making the correct identity
- Identify significant patterns
  - Changes in practice
- Compare categories
  - Changes in categories
- Apply a numerical score

Handling qualitative data: Use of surrogates

Review of the Parasitology schemes: 15 years

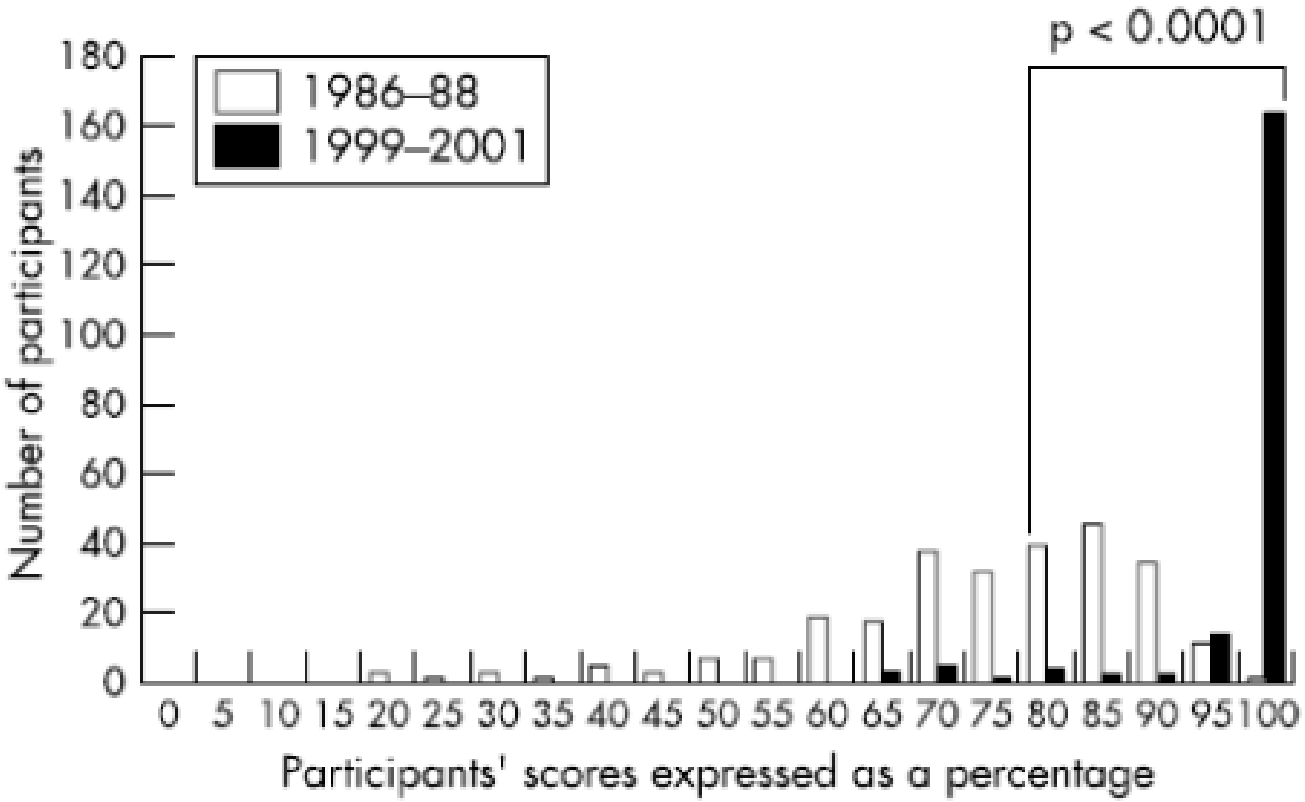
- Faecal and blood parasitology schemes introduced in 1986
- Identification of parasites and stage as ova, cysts, larvae
- Comparison of reported result with the assigned value/identity
- % of participants reporting the correct result

Faecal parasitology: examining for helminths



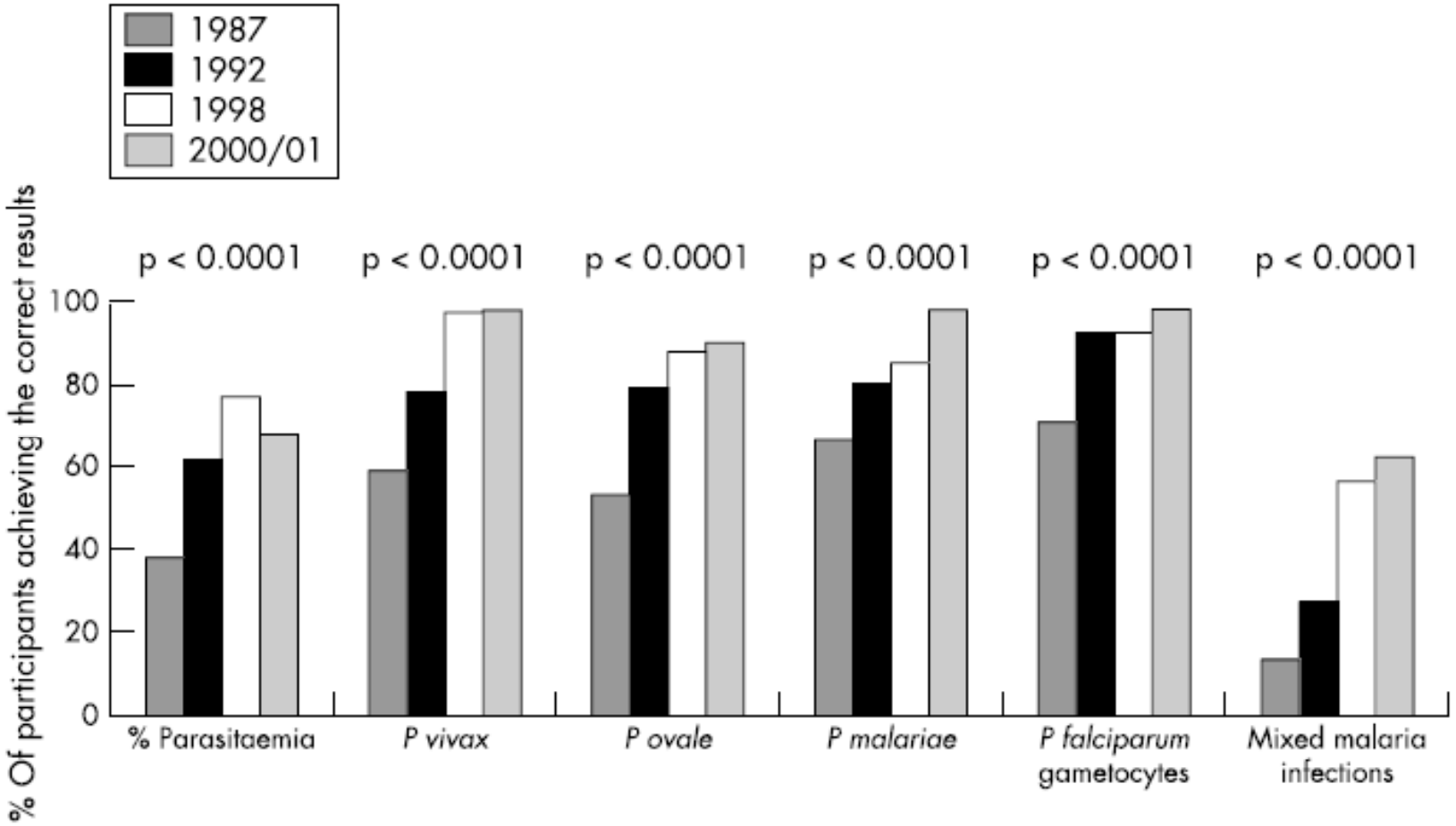
Kettelhut *et al.* Journal of Clinical Pathology 2003

Faecal parasitology: overall performance UK participants subscribing since start of scheme



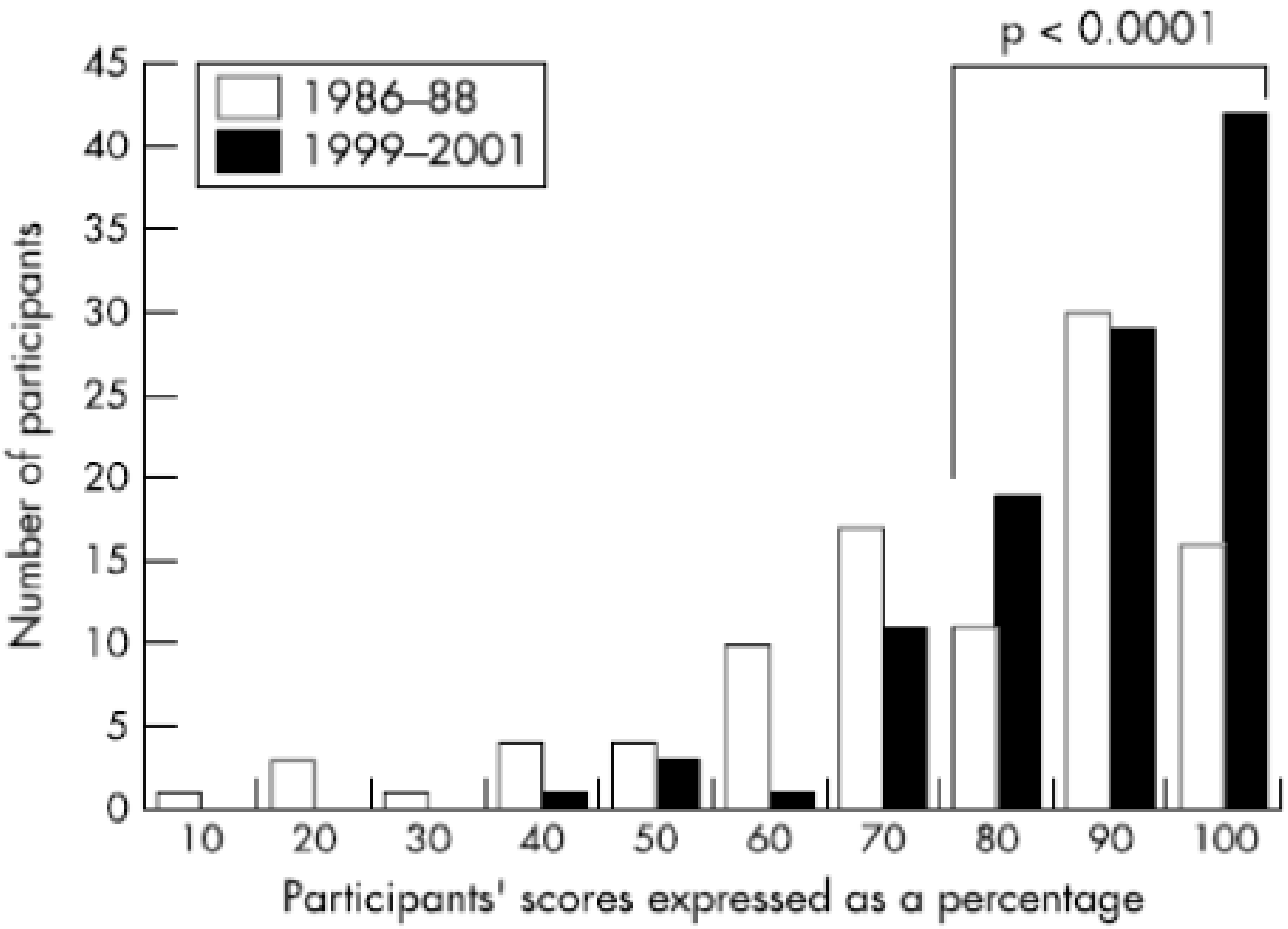
UKNEQAS

Blood parasitology: comparison of participant performance



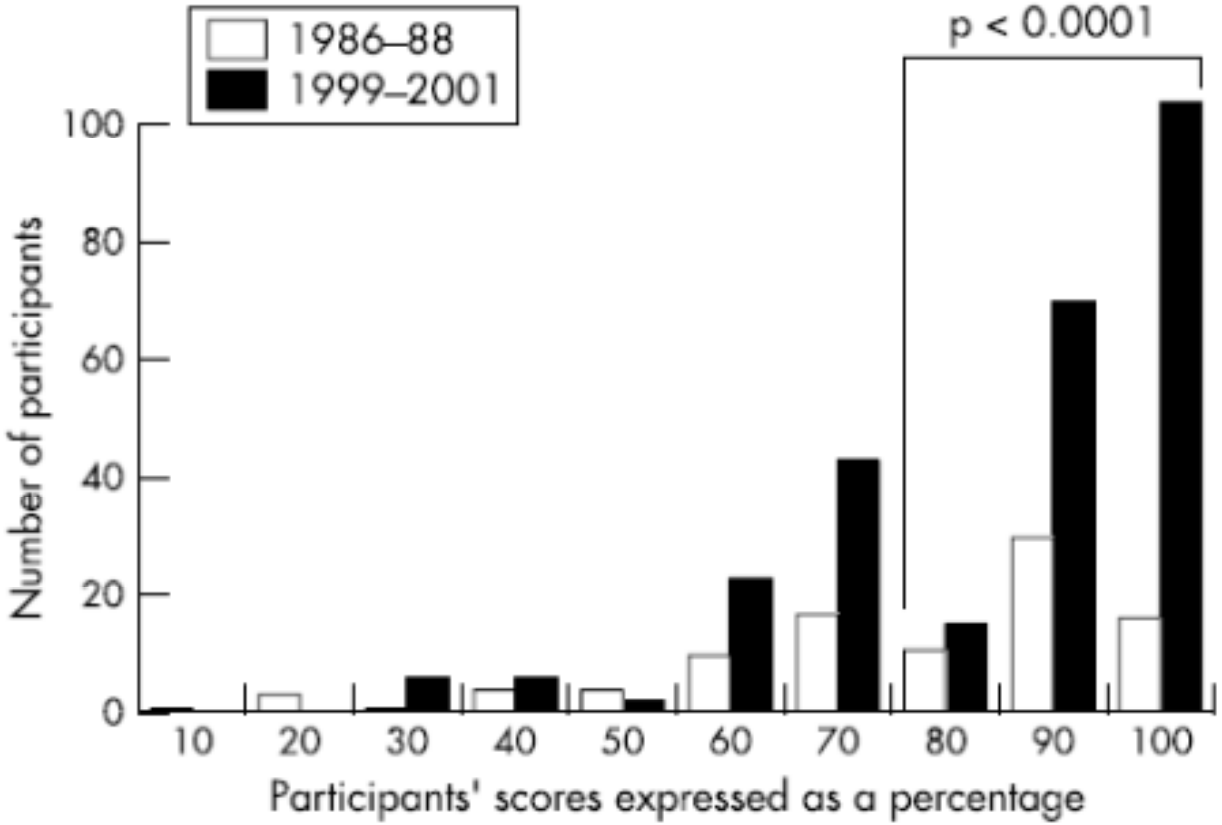
UKNEQAS

Blood parasitology: overall performance of UK participants subscribing since start of scheme



UK NEQAS

Blood parasitology: overall performance all UK participants



UK NEQAS

Handling qualitative data: Significant patterns

Review of mycobacterium culture scheme

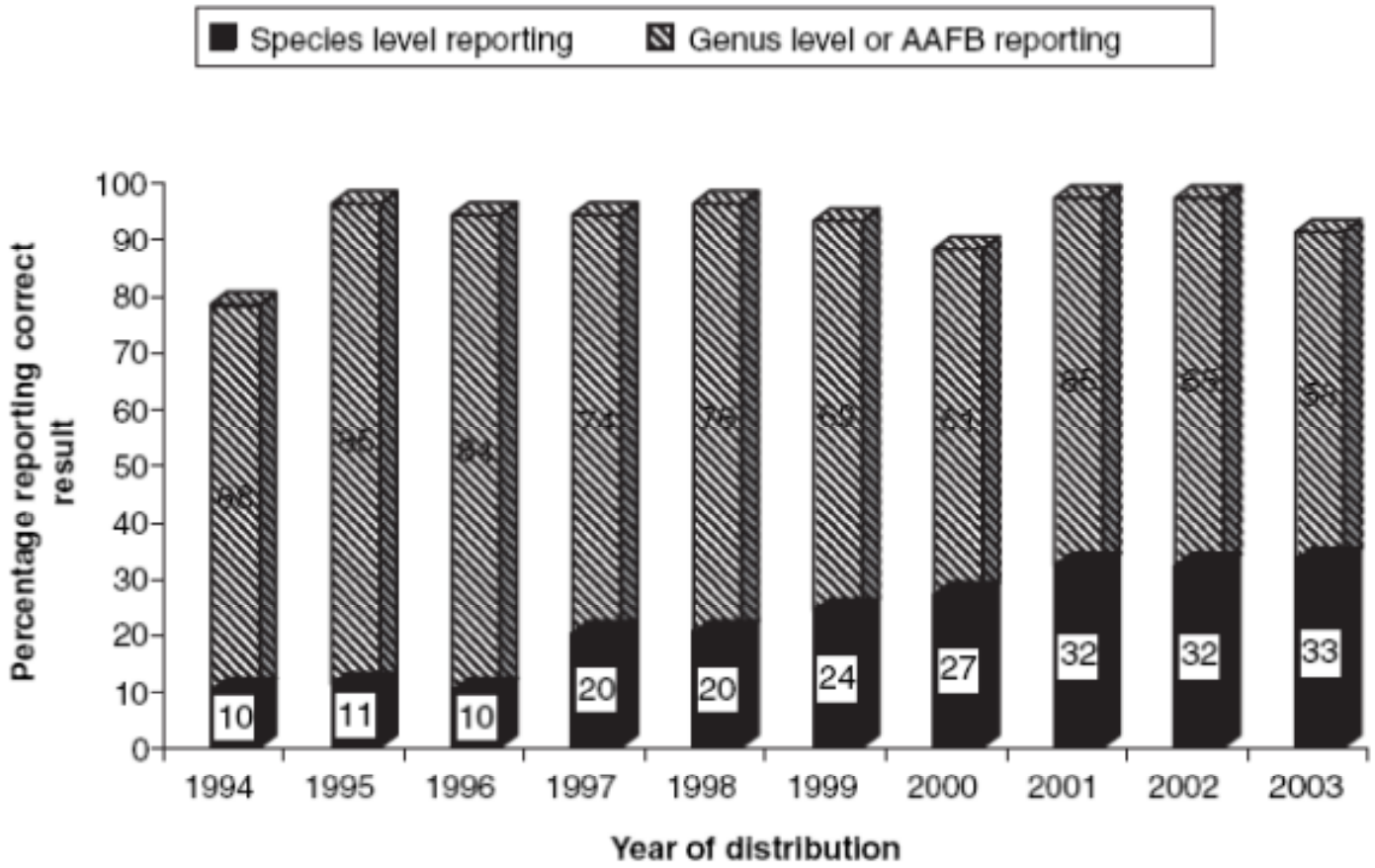
- Introduced in 1993
- Participants report on the culture results and identify to genus or species level
- Range of different culture media used
- UK standard method recommends culture for 10 to 12 weeks to have confidence in correct report of a negative result
- Time to identification of culture positive dependant on
  - Species
  - Strain
  - Bacterial load
  - Method

10 year review

- % participants reporting correct results
- Centre for Disease Control recommendation
  - Time to reporting

Walton *et al.* Clinical Microbiology and Infection 2005

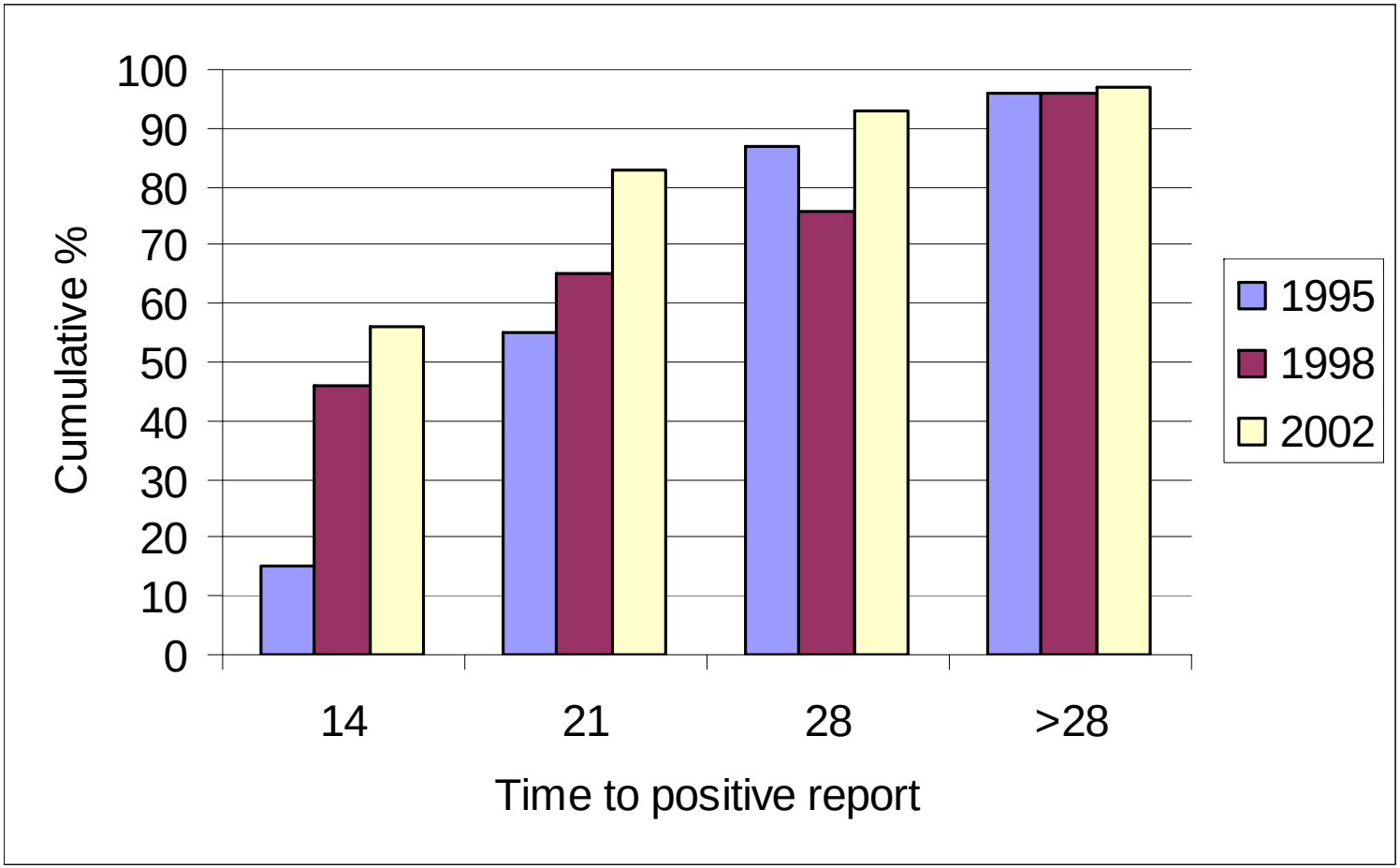
Mean percentage of laboratories correctly reporting *Mycobacterium tuberculosis*



Walton *et al.* Clinical Microbiology and Infection 2005

UK NEQAS

Time to positive reporting



UK NEQAS

Mycobacterium culture scheme Summary

- % participants reporting positive result by 21 days rose from 55% in 1995 to 83% in 2002 and 88% in 2011
- Proportion of non-UK laboratories has increased from 20% in 1995 to 44% in 2002 and 57% in 2011
- Increasingly liquid culture systems have been used; 78% in 2011

Handling qualitative data: Compare categories

Susceptibility to Rubella

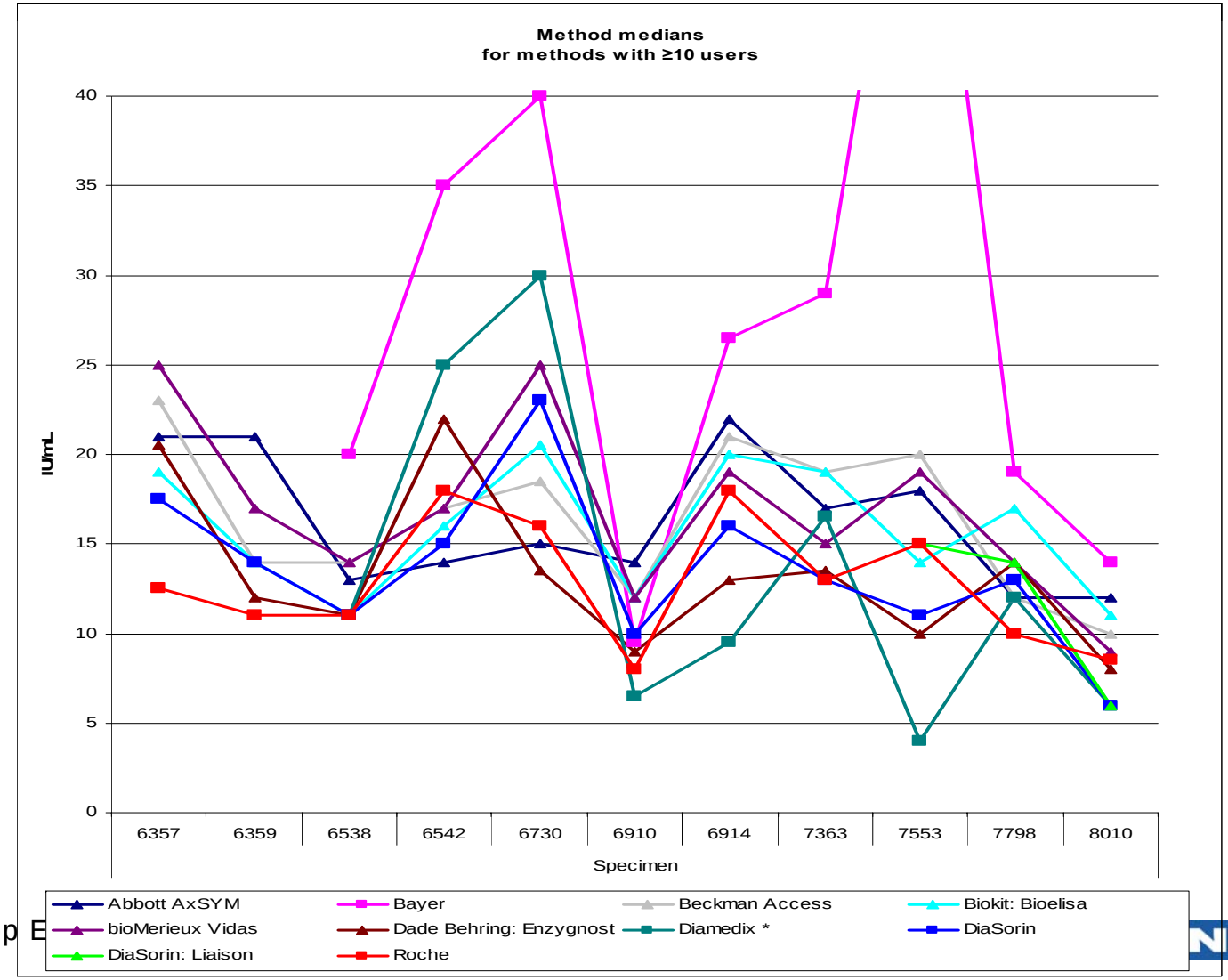
- Historically immunity to rubella was set at the limit of detection of the diagnostic assays
- Changes in practice from Radial Haemolysis through to Reverse Passive Haemagglutination to ELISA resulted in the introduction of a low level positive category where initially clarity about protection from infection was not clear
- In 2001 10 IU/mL cut off set
- Comparison of kits made
  - Implications to management of rash in pregnancy

Low level positive rubella reporting

Spec no.	No. >10 IU/mL	% pos	No. <10 IU/mL	No. numerical data sets	Range	Median for all kits	5% CI	95%CI
6357	363	97.6	9	329	0-70	21	12	29
6359	353	94.6	20	332	0-147	17	10	26
6538	333	91.0	33	342	2-118	13	9	19
6542	364	98.6	5	339	7-71	16	11	25
6730	365	98.7	5	344	0-55	18	12	33
6910	291	82.7	61	348	2-33	12	7	18
6914	360	97.8	8	341	0-150	20	12	29
7363	370	98.4	6	375	2-55	16	11	27
7553	349	93.6	24	373	0-162	17	9	27
7798	337	90.8	34	395	4-38	13	9	21
8010	209	56.2	163	402	0-500	10	5	16

Fagan *et al* 2006

Rubella IgG serology



Rubella IgG serology

- 56.2% to 98.7% of participants reported a positive (>10 IU/mL) result
- Linear regression, taking DiaSorin as the baseline (due to its fairly low mean and large-enough frequency of usage), showed that Bayer produced the highest results (2.1 fold > DiaSorin, 95% CI (2.0-2.3)).
- Overall Roche followed by Diamedix and DiaSorin produced the lowest results.
- However a more recent analysis (3 low level samples) has shown that Roche now gives high results, Bayer (now Siemens) now gives lower results, DiaSorin remains consistently low.

Handling qualitative data: Apply a numerical score

Benefits of a scoring scheme

- Simplify data – assist participants to assess their performance relative to other labs
- Allow assessment of cumulative performance over a number of rounds
- Comparisons between groups of laboratories
  - Method comparisons
  - Country comparisons

Types of scoring schemes

- Single response (diagnosis)
  - Immune/non-immune; normal/abnormal
  - Weighted degrees of how right, partial identification
- Multiple response (differential diagnosis)
  - Likelihood of each diagnosis
  - Risk e.g Down's syndrome in foetus

Scoring schemes - weighted

Top marks – the better you are the higher your score

UK NEQAS for microbiology

4 point scoring:

- 2 fully correct
- 1 partially correct
- 0 wrong
- -1 grossly misleading

Performance based on the average score

Penalty points – the better you are the lower your score

QCMD

4 point scoring:

- 0 fully correct
- 1 partially correct
- 2 incorrect
- 3 grossly misleading

Performance is based on % correct

Performance monitoring

- Applying a numeric score provides a mechanism for monitoring performance over several rounds
- The score can be subjected to basic statistical analyses
  - Standard errors
  - Ranking

General Bacteriology report: Page 1

Intended Result		Your Report	Your Score
Specimen 0444	*Peptostreptococcus anaerobius*	*Peptostreptococcus* and an unexpected pathogen	-1
Specimen 0445	*Staphylococcus lugdunensis*	Coagulase negative staphylococcus	Not scored
Specimen 0446	*Salmonella* Enteritidis	*Salmonella* Enteritidis	2

Cumulative score information

Total number of specimens sent to you for **UK NEQAS for General bacteriology** over the last 6 distributions is 18
Specimen numbers 0182 0183 0184 0276 0277 0278 0314 0316 0361 0362 0363 0401 0403 0444 0446 have been analysed and scored.

Number of reports returned and scored 15
Number of specimens reported as not examined (not scored) 0
Number of specimens received too late for analysis (not scored) 0
Number of specimens for which no report was received (scored as 0) 0
Your cumulative score for these specimens was 22 out of a possible total of 30
The mean score calculated from the reports returned by **UK** laboratories was 27.46 with a standard error of 2.64.

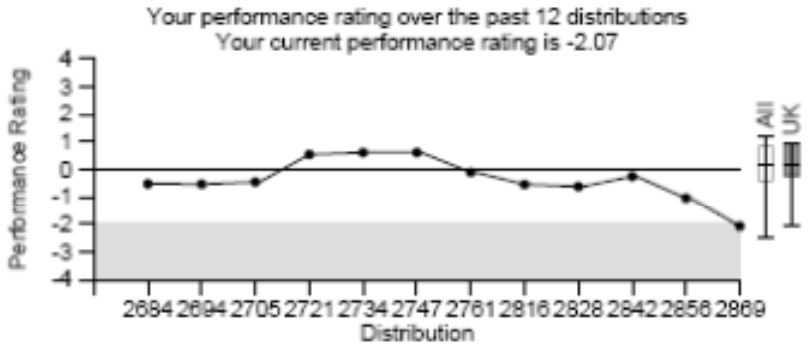
Cumulative score is less than mean score

PR – a form of ranking: Compares other labs examining the same specimens
Country specific if over 10 labs

Performance rating

Your performance rating for **UK NEQAS for General bacteriology** (i.e. the number of standard errors by which your cumulative score lies above or below the mean for **UK** laboratories) is -2.07.
Your performance rating indicates possible poor performance
A performance rating of more than 1.96 standard errors below the mean indicates possible poor performance.

Performance ratings may change if other participants' results are amended.



UK NEQAS

Performance Governance

A brief summary of the relevant analysis is given below and a print-out of the details of your results for the relevant specimens is attached.

	Your total score	Total possible score	Average Score	Your performance rating
Antimicrobial susceptibility	221	230	227.64	-2.50

I realise that Quality Assessment results may not reflect the total performance of a laboratory but they are designed to help the head of the laboratory to assess the accuracy of the procedures carried out by his or her staff……..

Notes of caution when reviewing performance over time

- Categorisation of a fully correct result can alter due to changes in practices/changes in the state of the art
- Scoring is tailored to practice

Is there a need for a universal harmonised scoring scheme?

- Benefits
  - Potential for comparisons internationally
- Concerns
  - Comparing apples with pears
    - Differences in the specimens sent
      - Need for defined standards for each property to be evaluated
      - Need for common sample specifications
- Is the cumulative performance the place to start?
  - Sector specific?
  - Discipline specific?

Summary

- Raw descriptive data can be categorised and comparison made between the categories
- Comparisons can be interpreted
- Changes to the categories can be monitored over time
- Applying a numeric score allows ‘hard’ statistical analysis

Thanks

Colleagues in UK NEQAS

UK National Quality Assurance Advisory Panel

Scheme participants