INT J TUBERC LUNG DIS 2(9):756–765

Limitations and requirements for quality control of sputum smear microscopy for acid-fast bacilli

A. Van Deun,* F. Portaels†

*Damien Foundation, Dhaka, Bangladesh, †Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium

SUMMARY

SUMMARY: Sputum microscopy for acid-fast bacilli (AFB) is considered to be the most appropriate method for case-finding in a tuberculosis (TB) control programme. It is usually carried out by general technicians, often after minimal training. Quality control of their results therefore seems indispensable. The methods advocated for quality control are reviewed. Controls by culture leave too much uncertainty because of big differences in technical characteristics of the methods. Sets of smears sent out by a central laboratory can only be used to assess capability. Rechecking routine smears allows daily performance to be appraised and may be a strong motivation, but feasibility may be a problem. Based on our experience, we describe the technical requirements for cross-checking of routine smears. Counter-checking slides with discordant results is crucial for accurate assessments. A sample size should strike a balance between statistical accuracy and the man-power needed. Indicators for evaluation are proposed that allow discrimination of error gradings, to be used in a phased manner with priority at first being given to false negatives and false positives that pass the threshold for clinical decision-making. Estimates of critical values with suggestions about their interpretation are placed in the context of supervising TB laboratories.

KEY WORDS: *Mycobacterium*; quality control; stains and staining

SMEAR-POSITIVE CASES constitute the highest priority for tuberculosis (TB) control. In high-prevalence countries, diagnosis is often made by paramedics on the sole evidence of acid-fast bacilli (AFB) microscopy. However, while all other aspects of TB control are well defined and controlled, the organisation of AFB microscopy is usually left to the general laboratory services.

Motivation among technicians is often low. The high demands on time and effort necessary for reliable AFB microscopy may counterbalance its low cost and simplicity. In particular, little work has been done[1] regarding large-scale quality control; some rules are regularly quoted, but they seem not to have been tested in the field. Moreover, interpretation of the indicators remains vague.

We discuss alternative methods for quality control showing their respective advantages, disadvantages and possible use, in conjunction with experience obtained in the Rwanda National Tuberculosis Programme (NTP) and the Damien Foundation Bangladesh TB Projects over the last 6 years. Technical requirements have been derived which should allow frequently encountered pitfalls to be avoided, and are incorporated in a proposed method for quality control. We also specificy its indications, limitations and feasibility.

REVIEW OF METHODS ADVOCATED FOR EXTERNAL QUALITY CONTROL

Comparison with culture for mycobacteria on the same or another specimen

This method has serious drawbacks because of technical problems and differences in test performance between culture and smear in the detection of TB bacilli. The ratio of smear to culture positivity on the same specimen is often estimated to be around only 40% to 60%.[2,3] Even in highly endemic countries with many smear-positive cases, this ratio is not much different. Maximum values reported from South Africa,[4] Kenya[5] or India[6,7] were 57% to 75%. The remaining minimum of 25% specimens contains too few bacilli to be consistently detectable by microscopy.[4,7]

On the other hand, cultures may be negative while AFB microscopy is positive (M+/C−) in 3% to 7% of specimens.[3,4,8] Various reasons for this have been reported: mainly non-viable bacilli accounting for up to 30% of such results in follow-up examinations.[9,10] In these studies, only about half of these M+/C− were finally considered as false-positives of microscopy.[4,9–11] If culture cannot be done very nearby, transport delays will evidently cause a higher proportion of false-negative cultures.

Correspondence to: Dr Armand Van Deun, Medical Director, Damien Foundation Bangladesh, P O Box 6038, Gulshan, Dhaka 1213, Bangladesh. Tel: (+880) 2 870 903. Fax: (+880) 2 883416. e-mail: dfavd@citechco.net

Article submitted 1 December 1997. Final version accepted 22 April 1998.

[A version in French of this article is available from the IUATLD Secretariat in Paris.]

Thus culture can not be recommended as a method of external quality control to target the quality of individual microscopists. The wide gap between the technical performance characteristics of both methods does not allow poor microscopy to be identified except in extreme cases. Besides, culture facilities are still scarce in endemic countries, and their development is not considered a priority.

However, periodic plotting and comparison of the positivity rates of smears and cultures may be a good method of internal quality control in large laboratories, as described by Petersen,[12] and Allen.[13] The average ratios of M+/C+, as well as discordant specimens, can be compared with those of other laboratories in a similar context, giving a fairly crude indication of the quality of both tests.

Reading, or staining plus reading, of centrally prepared slides with known results

The main advantage of this method is the relative ease of its execution. However, it provides only an evaluation of capability, and not of performance under routine conditions. The set-up is such that these slides will be examined thoroughly, perhaps even by several people pooling their knowledge and spending a lot of time on the exercise. Thus large laboratories, or technicians used to handling large amounts of smears, can be expected to score high in the test.[14] Because of high workloads and fatigue, they often make more mistakes in their routine work compared to laboratories processing moderate numbers with regular positives. Also, the very important step of smearing (and sometimes staining as well) is not included in this control procedure.

This method might best be reserved for tests at the end of a training session, or for a survey aimed at identifying individuals with insufficient technical knowledge in need of retraining. However, for slides sent out to the field, where the method measures suitability of equipment and materials together with technical competence, retraining may not always be the solution to poor results.

If carefully homogenised, diluted and smeared, as described by Smithwick,[15] such smears can be used to assess correctness of quantification more accurately than is possible by controlling routine slides.

It might also be used for quality control in large or national laboratories, where known samples can be inserted unobtrusively into a routine series, and then give a fair idea of daily performance.[16] Otherwise, quality control at this level may be difficult to achieve.

Control reading of a random selection of routine slides

This is the method of choice for continuous monitoring and evaluation of a control programme. In principle, it allows an appraisal of day-to-day performance and the identification of some causes of error. If performed regularly and with feed-back, it may have the effect of strongly motivating the peripheral technicians.

The following rules are commonly cited for this method:

- all positive and 10% of negative slides should be checked;
- slides should be randomly selected by a supervisor, and sent on to a laboratory at a higher level of the service, which is in turn controlled in the same way by the next level;
- controls are performed blind, the controller being unaware of the results at the lower level;
- slides must be checked without prior processing using the same type of microscopy as at the peripheral level;
- a supervisor finally compares both results and determines the rate of false positives and false negatives, considering the higher level result automatically as correct.

Today, this method has been widely adopted. However, on applying the above rules, numerous problems have been encountered, and not all of them are common knowledge.

PROBLEMS ENCOUNTERED WITH THE CONTROL OF ROUTINE SMEARS

In our experience, the regular control of routine smears is difficult to organise, and entails a heavy workload. The first difficulty is to make sure that all slides are being kept, that they are properly identified, and that controls are really blind. The sample size must be chosen with care to make this type of quality control feasible as well as reliable. The main technical problem is that controls in their turn bring about a number of false results, sometimes to the extent of making interpretation impossible.

Gold standard?

Using methods with much higher sensitivity or different specificity as gold standard, as in the case of culture, leaves too much uncertainty as to the causes and normally to be expected rates of discrepancies identified. A method with essentially the same sensitivity and specificity—but much less dependent on the human factor—would be ideal, but none has yet been identified. Re-examination using a fluorescence microscope after overstaining of Ziehl-Neelsen (ZN) smears with auramin would come very near, but it produces inconsistent results and cannot be recommended. Therefore the same technique must be used, but by another person who has good equipment available. However, even when the controller is a technician with higher qualifications, it should not be automatically assumed that the performance is better or even that no mistakes are made at all; a person cannot be promoted to the level of gold standard.

Our experiences in different settings and with controllers of different levels of experience and motivation have shown that even the best technicians do make mistakes, especially when overloaded. Table 1 shows error rates (defined below under 'Errors and indicators') for our Bangladesh projects. From this Table, it can be observed that errors were made at both levels, with some rates higher for the controllers than for the peripheral centres.

Hence smears with results that are discordant between the periphery and the first controlling level must be counterchecked at a second controlling level. Without this, one must allow for an unavoidable margin of discordance between the results of two microscopists.[7] It will then be impossible to know who made the error or to determine their rates. Omission of counterchecks also runs the risk of rendering individual feed-back counterproductive: falsely accusing peripheral microscopists of mistakes made by the controllers should be avoided as much as possible, in order to obtain the full benefit of quality control.

Number of slides to be selected for control

The rule of checking all positives originates from the wish to confirm all diagnoses. However, strictly speaking, this can not be done: microscopy can only demonstrate AFB (not necessarily TB bacilli), and misidentification of a specimen or slide cannot be detected by reading. The second positive specimen required before starting treatment constitutes a far better confirmation in this respect, besides adding to the diagnostic yield.[2,8] Also, a systematic sample of 10% of negatives is not justified.

Points to consider with regard to sample size are the following:

- Workload for the controllers constitutes the main limiting factor of the method: overload will render this quality control unfeasible as well as unreliable. The statistical accuracy desired must be balanced against the capacity of the service and requirements for technical accuracy. Sample size will usually have to be limited to the absolute minimum required.
- Since thresholds for considering a result as truly positive have been set high for most NTPs (10 AFB/100 fields in the IUATLD scale,[17] or 4 AFB/100 in the WHO scale[18]), clinically significant false positives are rare and often clustered. They are seen mainly with a few inexperienced microscopists or unusable microscopes. In principle, no such error should occur, so every one of them must be considered significant, and statistical considerations do not come into play.
- By contrast, some false negatives are to be expected. The rate of false negatives will vary not only with the quality of the microscopy, but also with the positivity rate among suspects. A proportional number of smears with low bacillary content has often been missed at first screening,[7] but by chance a few of these will be found to be positive at repeat examination. Table 1 shows rates of false negatives for the best of our Bangladeshi microscopists of under 1%, with 15–20% positivity among sus-

Table 1 Damien Foundation, Bangladesh: quality control of sputum smears, 1996

A. Unequivocal results: high false positives (HFP) and high false negatives (HFN)*

	Controlled level: periphery				Controlled level: first controller			
	Totals checked		HFP†	HFN†	Totals checked		HFP†	HFN†
Project	Pos.	Neg.	No. (%)	No. (%)	Pos.	Neg.	No. (%)	No. (%)
A	1325	894	6 (0.5)	17 (1.9)	1354	854	4 (0.3)	10 (1.2)
B	1076	1426	8 (0.7)	26 (2.0)	1064	1397	3 (0.3)	5 (0.4)
C	1401	1830	3 (0.2)	13 (0.7)	1382	1911	0	36 (1.9)
D	479	974	10 (2.1)	23 (2.4)	497	954	0	10 (1.1)
Total	4281	5124	27 (0.6)	79 (1.5)	4297	5116	7 (0.2)	61 (1.2)

B. Results under the cut-off: scanty false positives (SFP) and scanty false negatives (SFN)

	Controlled level: periphery			Controlled level: first controller		
Project	Number of reportedly scanty checked	SFP† No. (%)	SFN† No. (%)	Number of reportedly scanty checked	SFP† No. (%)	SFN† No. (%)
A	83	10 (0.7)	13 (1.5)	96	6 (0.4)	33 (3.9)
B	45	4 (0.4)	28 (2.0)	88	4 (0.4)	14 (1.0)
C	130	11 (0.7)	19 (1.0)	65	2 (0.1)	69 (3.6)
D	32	4 (0.8)	32 (3.3)	31	3 (0.6)	14 (1.5)
Total	290	29 (0.6)	92 (1.8)	280	15 (0.3)	130 (2.5)

*Cut-off used to declare a smear positive: 4 AFB per 100 high power fields. Below this the result is called scanty.
†For definitions of HFP, HFN, SFP, SFN, see Table 3.

pects, and a similar proportion of follow-up smears. But in programmes with less positivity among suspects, acceptable error rates would be proportionally lower. For false negatives, quality control should aim at discriminating between these unavoidable errors inherent to the technique, and unsatisfactory performance. This is done by choosing a critical value above which action is required. In practice, normal and critical values for false negatives need to be derived from the performance of the best centres, as well as from mean performance and its standard deviation. Furthermore, they may need to be revised after a period of regular quality control, with improving microscopy.

For any applicable statistical method, sample size will then depend mainly on the level of these two values and the margin between them, besides the statistical power and confidence level desired. The size of the population, in this case the total number of negative slides from which the sample is drawn, is of relatively little importance. The sample size needed barely increases once this population exceeds 1000. For this reason it is not appropriate to define required sample size as a percentage.

Is blind checking necessary for quality control?

Blind checking is a must for any objective control, and it will also prevent cheating. It allows an evaluation of the reliability of the first controlling level, by comparison of their false-negative rates for scanty or low positive smears with those determined for the controlled centres. If the first controllers are also responsible for routine AFB microscopy at their own centre, these blind control-readings can be considered as part of their routine work. Comparison with concordant results from the peripheral centres and with the results of the counterchecks for discordants may then be sufficient as quality control for the centres where first-level controls are being done—always on condition that a countercheck of slides with discordant results is included.

The technician performing the counterchecks has to search long enough in order to find the AFB, or to be able to exclude the presence of AFB with high probability. It is helpful when both results are known (quantified), so as to motivate the technician to search long enough in case of a report of rare bacilli or for accurate quantification of unevenly distributed AFB. If it is not apparent to which level the respective contradictory results belong, this check is in effect still blind.

Restaining of slides to be controlled

The classical recommendation is to control smears in the condition in which they are, so that the staining quality can also be evaluated. In fact, assessment of the quality of the original staining may only be needed when controls have indicated poor performance, and will often need a supervisory visit to the centre for full investigation of the causes. On the other hand, a failure of quality control was reported by Allen in the detection of false-negative results caused by an inadequate cold staining method,[13] since the bacilli were evidently not visible to the controllers. In Bangladesh, rapid fading of the red fuchsin-stain after adequate ZN has frequently been encountered, and was found to be due to a combination of high humidity and heat.[19] Restaining prior to quality control readings thus seems indispensable to avoid gross errors on the part of the controllers, but it also considerably increases the workload.

PROPOSED SYSTEM FOR QUALITY CONTROL UNDER PROGRAMME CONDITIONS

Proposals regarding sample size and sampling

Sample size is important for reportedly negative slides to ascertain that the false negative rate stays in the normal range, and does not surpass a critical value. The 'Lot Quality Assurance Sampling' method (LQAS[20]) is proposed as adequate for this aim. If controls of the sample show not more than the corresponding maximum allowed number of errors, we can be statistically sure that the critical value has not been surpassed, but without knowing the exact error rate. This allows small samples to be used, only obtaining the most essential information about individual centres. For the totals of the service, global sample size will probably be high enough to obtain more accurate overall error rates.

Table 2 compiles selected data from LQAS tables, and illustrates what has been said concerning the main influencing parameters. The sample size required increases rapidly with decreasing critical value, and also to a fair extent with increasing numbers of errors allowed. This means that it will rarely be feasible to use very low critical values, and we suggest starting with a 5% false negative rate. When the cut-off of the IUATLD or even the WHO scale is used to define false negative, this is in fact already a high rate, but it may be suitable for a start-up period in programmes where positivity rates among suspects are not too low. At this cut-off, errors by the controllers in declaring false negatives should be extremely rare, so the number of errors that are allowed to be found in the sample can be put at only 1 or 2. At zero error allowed, the sample size can be kept even smaller, but this puts higher demands on specificity. No single 'false false negative' due to controllers' mistakes should occur then.

On the contrary, opting for a 90% or 95% confidence level does not make a big difference for sample size in Table 2, and letting the population vary from 1000 to 5000 (or more, not shown) even less. Evidently, therefore, the 95% confidence level should be

Table 2 Sample size required for quality control: Lot Quality Assurance Sampling[20]

A. Critical value not surpassed with 95% confidence, total no. of registered negative results 5000 (in parentheses: same for 1000 total registered negative results).

No. of false negatives allowed	Critical value		
	1.25%	2.50%	5%
None	234 (212)	118 (112)	58 (57)
Maximum 1	367 (324)	186 (174)	93 (90)
Maximum 2	486 (417)	246 (227)	123 (118)
Maximum 3	596 (501)	303 (275)	152 (145)
Maximum 4	701 (578)	357 (321)	179 (170)

B. Critical value not surpassed with 90% confidence, total no. of registered negative results 5000 (in parentheses: same for 1000 total registered negative results).

No. of false negatives allowed	Critical value		
	1.25%	2.50%	5%
None	181 (168)	91 (87)	45 (44)
Maximum 1	303 (274)	153 (145)	76 (74)
Maximum 2	414 (366)	209 (195)	105 (101)
Maximum 3	518 (449)	262 (241)	131 (126)
Maximum 4	619 (527)	314 (285)	157 (150)

preferred, as the small gain in lower sample size at 90% does not justify the increased probability of failure to recognise poor performers. Finally, for the sake of operational simplicity, a fixed sample size independent of the turn-over of the centres can be adopted, for instance one appropriate to a turn-over of 5000 negative smears per year, or a figure that comes closest to the average for the centres.

Based on these and earlier considerations, we propose the following system:

- Total sample should be matched against the turn-over of a period of, for example, one year. Using Table 2, this means about 120 slides with negative results, which will suffice if most centres record fewer than 5000 negatives yearly. Sample size for positives is not defined statistically (and would have to be very high indeed, because of the very low critical value). As explained earlier, a modest sample will already allow those who make these errors frequently to be identified. We propose, as a rule of thumb, to sample the same number as that for negatives. Scanty (doubtfully) positive smears, with numbers of AFB under the threshold for positivity, are at the limit sensitivity of the method, and the repeatibility and hence reliability of their control is lower. Nevertheless it is important to include them in the sample, since comparison of error rates in this group allows the attention given to the controls, and hence the reliability of the quality control itself, to be evaluated. Scanties are best sampled as part of the positives, proportionally to their occurrence in the laboratory register. This will produce a well-balanced sample for the ever-needed assessment of quality control reliability, while automatically a bigger number will be checked when there is a special problem, such as confusion with artefacts or contamination.
- If most of the centres of the service are small, e.g. an annual turn-over of less than 1000 slides on average, a smaller fixed number must be chosen to be checked. However, it should be realized that the total workload for quality control then increases, expressed as proportion of routine smears controlled—one of the disadvantages of over-decentralization.
- At a low prevalence of positive specimens, the proposed critical value of 5% is too high. A lower one may be chosen, but this means a much bigger sample size. Alternative solutions are to include in the sample a sub-group with a necessarily higher positivity rate, such as follow-up smears at 2 months or X-ray diagnosed cases (if these diagnoses are not too bad). However, the simplest solution may be to choose a lower cut-off for positivity and thus false negative, only for the controls.
- Selection of the sample should be done by supervisors (not necessarily laboratory people) visiting the centre. If peripheral technicians are asked to select the slides themselves and send them to the higher level, the sample will often not be random at all. Small samples should be taken at each visit, adding up to the total required over one year.

It is best to start by making a list of slide numbers and results from the laboratory register, picking the slides randomly according to their results. All slides put on the list are then taken from the slide boxes. In this way, it will be obvious if all slides have really been kept. However, this method will be too time-consuming if numerical order has not been respected in slide storage.

Organisation of the controls at peripheral and intermediate level (Figure)

First level control will usually be done at a district or regional laboratory. However, it is a misconception that this should be assigned to a more qualified technician, e.g. the Chief Laboratory Technician. As screening is a tough, boring job, a person with high qualifications is neither necessary nor desirable. Perfectly good microscopes are a must for the controllers. They should screen the same number of fields as required from the routine centres, not more.

A good microscopist may be expected to process an average of 25 slides per day, restaining included. Screening 100 fields before declaring negative seems little for an 8-hour working day. However, experience has shown that good quality control is not possible at a consistently high rate, and that the technician most probably also has routine smears from the capture area to examine. So if possible, quality control should be further decentralized, so that each controller has only 10 centres (quality control sample of less than 2500 slides per year). At an incidence of smear positive pulmonary tuberculosis of around 1 per thousand, these 10 centres could, and ideally would, cover a population of 2 million. This means that one half-time controlling microscopist can handle a population of 2 million. These assumptions correspond to our own experience in Bangladesh, but would also be appropriate in other situations, such as Tanzania.[21]

Counterchecks can be centralised in one or very few centres. These may be the highest levels of the service, but not necessarily. Only slides with results that are discordant between the periphery and first controlling level should be included. The countercheck results are considered as final; they determine at which level (periphery or first controller) the mistake was made. With reasonably good peripheral and first-level control technicians, 5%–10% of those can be expected. This means that one technician at this level can check 10 first-level controllers (each processing a maximum of 2500 quality control slides annually), corresponding to 20 million population, based on the above assumptions.

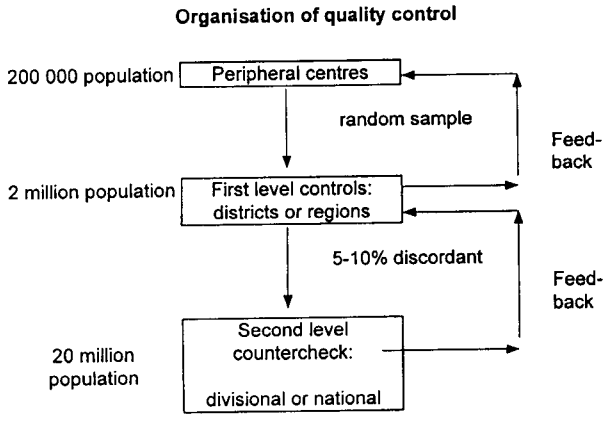

Figure Organisation of the controls at peripheral and intermediate level.

Errors and indicators

Quality control should only consider the presence or absence of AFB. Trying to recognise atypical mycobacteria is nonsense in countries with high TB prevalence where disease due to nontuberculous mycobacteria is almost inexistent,[22] and would only lead to confusion for people working at field level. So, even if the controller feels sure that what is observed is not TB, this should not be considered as an error as long as the bacilli correspond to the description of AFB. Of course, remarks as to the presence of possible contaminants are necessary.

Logically, a quality control system should start by focusing on the gross errors above the cut-off or threshold for positivity. Scanty results below this threshold are of minor importance to the clinician, and their countercheck is more difficult. However, error rates in this zone give an indication of the reliability of the controls, so these discordants should always be counterchecked, too. At a later stage, once the gross deficiencies in the centres have been resolved, quality control can be refined by including false negatives below the clinical threshold for positivity to see if the critical value has been surpassed, e.g., to keep the critical value high enough for feasible sample sizes. Quantification errors will be the last priority.

Errors should thus be classified according to their importance. We have used a classification that is explained in Table 3. Indicators are high false negative (HFN), high false positive (HFP), quantification false positives (QFP) and quantification false negatives (QFN), all of which affect patient management. Less important for clinical management are scanty false positive (SFP), scanty false negative (SFN) and gross quantification errors (QE).

With the LQAS method, the number of HFN (or HFN plus SFN at a later stage) will indicate whether performance is satisfactory. Likewise, for HFP the actual number of errors is sufficient.

Furthermore, error rates are calculated mainly to get an idea of the global performance of the service. The denominator in the calculation of rates is the total of smears in the quality control series reported as positive (for HFP) or negative (for HFN), by the controlled level. For SFP, the sum of scanty and high positives should be used, and for SFN the total of the reportedly negative samples. The rates should be calculated for the peripheral centres as well as for the first controlling level in one and the same series. Totals to be used as denominators will be slightly different, except where there is 100% concordance between the results of these two levels.

Table 3 Definition of the indicators used

Registered result	Final result at control	Indicator (rate)
Positive	Negative	HFP
Positive	Scanty positive	QFP
Negative	Positive	HFN
Negative	Scanty positive	SFN
Scanty positive	Negative	SFP
Scanty positive	Positive	QFN
High positive +3	Low positive +1 or 4–9/100	QE
Moderately positive +2	Low positive 4–9/100	QE
Low positive +1 or 4–9/100	High positive +3	QE
Low positive 4–9/100	Moderately positive +2	QE

HFP = high false positive; QFP = quantification false positive; HFN = high false negative; SFN = scanty false negative; SFP = scanty false positive; QFN = quantification false negative; QE = quantification error.
Note: Cut-off for positivity used is at 4/100 high-power fields (HPF) in all the definitions.
Scanty positive indicates 1–3 AFB per 100 HPF (doubtful result).

Interpretation of quality control results

An isolated HFP is usually due to clerical error[16,23] at the controlled centre or made during sampling or controls for quality control. If more than a single one is found, identification and registration are probably being done carelessly.

Higher rates of HFP (usually together with SFP) are typical of inexperienced microscopists who have not yet seen TB bacilli regularly. This is observed more often in centres detecting a positive case only infrequently (less endemic countries, an over-decentralised laboratory network, NTPs using a low threshold for positivity). A combination of all these factors was found to be responsible for the often poor performance in the Rwanda NTP (Table 4), where excluding centres with excessive HFP still left an unacceptable proportion of low false positives (+1) that could have been avoided by using the IUATLD scale rather than the American Thoracic Society (ATS) scale (+1 in the ATS scale is equivalent to +2 for the IUATLD).

Some HFN are unavoidable. With high positivity rates among suspects, less than 1% of HFN can be considered an excellent result, and even up to 3% may still be quite acceptable.[7,24] In highly endemic countries, 5% or more of false negatives should be considered excessive.

Higher rates will be seen mainly with microscopists who are overloaded. In this case, it may not be possible to improve performance, since a balance has to be struck between the numbers of slides to be checked and accuracy. The solution in this case would be to increase man-power or to improve equipment (i.e., install fluorescence microscopy). False negatives may also point to a technical problem (a bad microscope or stain, poor smearing, poor eyesight, or a complete lack of practical training). If all these problems have been excluded, many false negatives just indicate lack of motivation and a poorly done job.

Heavily positive slides repeatedly being reported as negative might be due to bad registration. Otherwise, this can indicate deliberate cheating, grossly inadequate technique or total neglect. Nonsense results, i.e., almost all results are HFP and HFN, occur when

Table 4 Rwanda NTP: results of quality control per case-detection group, 1992–1993

Centres grouped according to no. of smear positive cases detected	Total smear positive detected 1992–1993	No. of centres	False results (%)*					
			False positives† +2/+3/+4	+1	<3/300	False negatives +2/+3/+4	+1	<3/300
A. All centres included								
Less than 1 case per quarter	216	74	5	11	4	1	1	0
Less than 1 case per month	510	40	2	6	1	1	1	0
Less than 1 case per week	1 253	28	3	11	3	2	1	0
At least 1 case per week	2 120	8	1	2	0	6	1	3
Sum of all centres	4 099	150	3	8	2	2	1	1
B. Centres with excessive false positive excluded								
Less than 1 case per quarter	159	60	1	3	1	1	1	0
Less than 1 case per month	472	37	1	3	0	1	1	0
Less than 1 case per week	869	21	2	6	0	3	1	0
At least 1 case per week	2 120	8	1	2	0	6	1	3
Sum of all centres	3 620	126	1	4	0	2	1	1

*Denominators: sum of all positive + scanty checked (false positive), respectively all negative checked (false negative).
†Scale of the American Thoracic Society (ATS) used in the NTP.

there is no notion whatsoever of AFB, or where a microscope is unusable.

Extremely high rates of false negatives (mainly SFN) may also be found because of contamination of stains with saprophytic AFB, when restaining is done prior to control (own data, not shown). It indicates that the carbolfuchsin solution at the controlling laboratory, or the counterstain or rinsing water used by the peripheral centre, contains AFB.

In principle the rates of SFP and SFN cannot be interpreted fully without a countercheck. Almost always a scanty positive will then be confirmed as scanty or clearly positive. AFB may have been missed by either of the two technicians, at the peripheral centre or during control. Acceptable rates are difficult to define since they will depend on the positivity rate among suspects (and during follow-up). In our situation, low rates, of 1–2%, have no special meaning and should simply be ignored.

Reliability of this quality control method

The number of scanty positives identified at any level and confirmed during quality control is in itself already an indicator of the quality at that level. If they are almost totally absent, microscopy has been done very superficially (or quantification is not at all respected). The controls will usually then show a high rate of SFN. If not, it may be that controls have been done just as poorly. In that case, very few scanty positives will have been identified, some at the peripheral centre and some at the first controlling level, but hardly any by both. The SFN rates of the centres and of the first controller, and their ratio, enable verification as to whether quality control results are reliable.

Finally, if no discordants (HFP, HFN, SFP and SFN) have been identified at all in checking at least 100 slides, one should conclude that the first controls have not been done blind, or that the sample has not been random.

UTILITY AND LIMITATIONS OF QUALITY CONTROL

Quality control must be an integral part of the supervision of AFB microscopy laboratories. Supervisors should choose the sample and take care of the feed-back, including investigation into the causes of errors.

To make optimal use of this quality control exercise, feed-back to the centres is indispensable. It includes sending back slides for which errors were found, along with the complete list. Microscopists must be given the chance to show what they have interpreted as AFB, or be shown the AFB they have missed. Simply telling them that they were wrong will just add to their confusion, and in fact will often be met with disbelief. Moreover, the cause of the errors should be discovered while giving feed-back. Some sources of errors may have been identified already by the controllers: artefacts and reused scratched slides, poor smearing technique, insufficiently thorough microscopy.

Further checks during visits may yield additional information. A low proportion of scanty positive smears should be found in cases under treatment, at 2 months or later. In our setting, this is true of around 10% of results in this group. Quite high rates were reported by Rieder for a population treated with a powerful regimen under strictly observed conditions.[25]

Plotting the proportion of positive and scanty positive results each on a graph allows deteriorating performance to be recognised.[12] This kind of control might be used as surveillance once a satisfactory level of performance has been reached by rechecking routine smears. It should then be complemented by occasional random sampling and checking of some of the scanty positive slides, or of a large number of reportedly negative slides in case the graph indicates deteriorating performance.

Quality control by cross-checking of routine smears has its limitations as well. For a single slide it is not possible to be absolutely certain of a negative control result if the original result was low or scanty positive. Even if counterchecking includes 1000 fields, 5 to 10 times more have not been examined which may contain just one clump of bacilli. This must be seen as a limitation of the quality control method. Only when checks on a series of scanty slides are consistently negative for the same technician can it be concluded that the results are false.

Absolute accuracy is impossible due to the absence of a suitable gold standard. Not all points determining the final result are controlled, for instance false negatives due to poor quality of sputum may not be recognised. The controls cannot detect misidentifications, which might even be deliberate in circumstances where a thorough system of control co-exists with a strong stimulus to increase case-finding: one positive sample is sufficient to produce hundreds of positive slides. Finally, contamination of smears with mycobacteria from any source cannot always be accurately identified by the controllers.

The feasibility, and especially the cost-effectiveness, of quality control will depend mainly on its careful organisation and the extent to which it is recognised as a priority. Necessary input will vary considerably depending on the epidemiological situation, the degree of decentralization of the microscopy network, and socio-economic conditions. The main costs are the salaries, slide-boxes, some stains and small administrative and transport costs. As an indication, the 13 000 controls performed in our Bangladesh projects over 1997, employing the equivalent of three full-time workers, are estimated to have cost less than $6 000 US, or well under 1% of total project costs.

CONCLUSION

Quality control of sputum AFB microscopy in peripheral laboratories is feasible under programme condi-

tions, but it requires careful organisation, good discipline and considerable input of man-power. The system used will depend on the stage of implementation.

By way of survey, centrally prepared slides may be sent out for staining and examination in the peripheral centres. This will mainly allow those in need of more training, and centres with unusable equipment, to be identified.

This should be followed by the regular control of a sample of routine slides with feed-back on individual results and identification of the cause of the errors during supervisory visits. If done correctly, quality control will be the first step towards solving any problems. Together with the motivation of the technicians resulting from regular controls, this will allow for a gradual improvement in the quality of the microscopy network.[14,26] Once a satisfactory level has been reached, it may be possible for bigger centres to replace this quality control by a system of internal control, with monthly plotting of the proportion of positive and scanty positive results.

Acknowledgements

This study was supported by the Damien Foundation (Brussels). We thank Dr E Declercq, Dr H L Rieder and Prof S R Pattyn for critical comments. We also thank Dr Rosemary Croft for assistance in the preparation of the manuscript.

References

1 Arnadottir Th. Assessment of tuberculosis: is 'rapid assessment' needed? Tubercle Lung Dis 1995; 76: 375–376.
2 Urbanczik R. Present position of microscopy and of culture in diagnostic mycobacteriology. Zbl Bakt Hyg 1985; A 260: 81–87.
3 Kubica G P. Correlation of acid-fast staining methods with culture results for mycobacteria. Bull Int Union Tuberc 1980; 55: 117–124.
4 Levy H, Feldman C, Sacho H, van der Meulen H, Kallenbach J, Koornhof H. A reevaluation of sputum microscopy and culture in the diagnosis of pulmonary tuberculosis. Chest 1989, 95: 1193–1197.
5 Kinyanjui M G, Githui R G, Kamunyi M J, Omwega P G, Waiyaki P G, Muthami L N. Quality control in the laboratory diagnosis of tuberculosis. E Afr Med J 1991; 68: 3–9.
6 Raj Narain, Subba Rao M S, Chandrasekhar P, Pyarelal. Microscopy positive and microscopy negative cases of pulmonary tuberculosis. Am Rev Respir Dis 1971; 103: 761–773.
7 Toman K. Tuberculosis, case-finding and chemotherapy. Geneva; WHO 1979: pp 14–18.
8 Blair E B, Brown G L, Tull A H. Computer files and analysis of laboratory data from Tuberculosis patients: II. Analyses of 6 years' data on sputum specimens. Am Rev Respir Dis 1976; 113: 427–432.
9 Pollock H M, Wieman E J. Smear results in the diagnosis of mycobacterioses using blue light fluorescence microscopy. J Clin Microbiol 1977; 5: 329–331.
10 Rickman T W, Moyer N P. Increased sensitivity of acid-fast smears. J Clin Microbiol 1980; 11: 619–620.
11 Maso Dominguez J, Saida Vivas E. Smear-positive and culture-negative results of routine sputum investigations for the detection and therapy control of pulmonary tuberculosis. Tubercle 1977; 58: 217–220.
12 Petersen K F. Methods for internal quality control in the mycobacteriology laboratory. Zbl Bakt Hyg I Abt Orig 1983; A255: 503–510.
13 Allen J L. A modified Ziehl-Neelsen stain for mycobacteria. Med Lab Sc 1992; 49: 99–102.
14 Küchler R. Die Zuverlässigkeit der bakteriologischen Tuberkulosediagnostik. Ergebnisse der externen Qualitätsprüfungen von 1991 und 1992. Pneumologie 1993; 47: 670–677.
15 Smithwick R W. Preparation of acid-fast microscopy smears for proficiency testing and quality control. J Clin Microbiol 1978; 8: 110–111.
16 Aber V R, Allen B W, Mitchison D A. Laboratory studies on isolated positive cultures and the efficiency of direct smear examination. Tubercle 1980; 61: 123–133.
17 International Union Against Tuberculosis. Technical guide for sputum examination for tuberculosis by direct smear microscopy. 3rd ed. Paris; IUAT 1978: p 13.
18 World Health Organization Tuberculosis Programme. Managing tuberculosis at district level, a training course. Supporting laboratory services. Geneva; WHO. pp 18–20.
19 Van Deun A. Rapid fading of carbolfuchsin stained AFB under extreme conditions of temperature and humidity. Int J Tuberc Lung Dis 1997; 1: 384–385.
20 Lemeshow S, Hosmer D W, Klar J, Lwanga S K. Lot quality assurance sampling. In: Lemeshow S, Hosmer D W, Klar J, Lwanga S K. Adequacy of sample size in health studies. Chicester; John Wiley & Sons (on behalf of WHO) 1990: pp 24–28.
21 Ipuge Y A I, Rieder H L, Enarson D A. The yield of acid-fast bacilli from serial smears in routine microscopy laboratories in rural Tanzania. Trans Roy Soc Trop Med Hyg 1996; 90: 258–261.
22 Portaels F. Epidemiology of mycobacterial diseases. In: M. Schuster, ed. Clinics in dermatology. Vol 13. New York; Elsevier Science Inc., 1995: 207–222.
23 Gordin F, Slutkin G. The validity of acid-fast smears in the diagnosis of pulmonary tuberculosis. Arch Pathol Lab Med 1990; 114: 1025–1027.
24 Takahashi M. Tuberculosis in Nepal: case-finding and quality control of sputum-smear examination. Kekkaku 1994; 69: 475–482.
25 Rieder H L. Sputum smear conversion during directly observed treatment for tuberculosis. Tubercle Lung Dis 1996; 77: 124–129.
26 Murray P R, Elmore C, Krogstad D J. The acid-fast stain: a specific and predictive test for mycobacterial disease. Ann Intern Med 1980; 92: 512–513.

RÉSUMÉ

L'examen microscopique des crachats pour recherche de bacilles acido-résistants (AFB) est considéré comme la méthode la plus appropriée pour le dépistage dans le cadre d'un programme de lutte antituberculeuse. Elle est habituellement conduite par des techniciens généraux, souvent après un entraînement minimal. Dès lors, le contrôle de qualité de leurs résultats semble indispensable. L'on revoit les méthodes conseillées pour le contrôle de qualité. Les contrôles par culture laissent trop d'incertitudes en raison des différences importantes dans les caractéristiques techniques des méthodes. L'envoi de lames par un laboratoire central peut également être utilisé pour apprécier la capacité. Le contrôle des lames de routine permet d'apprécier la performance quotidienne et

peut entraîner une excellente motivation mais sa faisabilité peut poser problème. Sur la base de nos expériences, nous décrivons les exigences techniques pour le contrôle croisé des lames de routine. Le contre-contrôle des lames dont les résultats sont discordants est essentiel pour des appréciations précises. La taille de l'échantillon doit chercher un équilibre entre la précision statistique et les besoins en personnel. Des indicateurs d'évaluation sont proposés, qui permettent de discriminer des degrés d'erreur et doivent être utilisés de manière progressive, la priorité étant donnée d'abord aux faux négatifs et aux faux positifs qui dépassent le seuil imposant une décision clinique. Des estimations des valeurs critiques et des suggestions quant à leur interprétation sont proposées dans le contexte de la supervision des laboratoires en matière de tuberculose.

RESUMEN

Se considera que la búsqueda de bacilos ácido-resistentes (BAR) en el esputo es el método más apropiado para búsqueda de casos en un programa de control de tuberculosis (TB). Casi siempre lo efectúan técnicos generales, a menudo con poco entrenamiento. Resulta, por lo tanto, indispensable un control de calidad. Se revisan los métodos aconsejados para el control de calidad. Los controles a través de los cultivos dejan mucha incertidumbre debido a las grandes diferencias en las características técnicas de ambos métodos. Los frotis de esputos enviados a un laboratorio central pueden ser usadas sólo para evaluar la capacidad. El repetir rutinariamente los frotis permite una valoración del trabajo diario y puede ser una buena motivación, pero no siempre es posible. En base a nuestra experiencia describimos los requerimientos técnicos para un control cruzado rutinario de frotis. La confrontación de láminas con resultados discordantes es crucial para evaluaciones seguras. Un tamaño de muestra establecería un balance entre la seguridad estadística y las necesidades prácticas. Se proponen indicadores de evaluación que permiten la discriminación de los grados de error, usados de manera tal que den prioridad a los falsos negativos y falsos positivos que pasan el umbral de las decisiones clínicas. Estimaciones de los valores críticos con sugerencias hacia su interpretación se ubican en el contexto de la supervisión de los laboratorios de TB.