Lepr Rev (1985) 56, 177-191

# SPECIAL ARTICLE

# Quality control of skin smear services in leprosy programmes: preliminary experience with inter-observer comparison in routine services

# A J DE RIJK,\* T NILSSON† & M CHONDE‡

\*Royal Tropical Institute, Subdepartment for Health Care and Epidemiology, Health Services Consultant, Amsterdam, the Netherlands

*†All Africa Leprosy and Rehabilitation Training Centre, ALERT, Head Hospital Laboratory, Addis Ababa, Ethiopia ‡National Tuberculosis/Leprosy Programme, Ministry of Health, Senior Laboratory Technician, Dar es Salaam, Tanzania* 

Accepted for publication 11 February 1985

Summary A description of a systematic approach to periodical re-examination of samples of skin smears for leprosy taken in the routine services is given.

Procedures and recordings are described in detail and examples given of test runs in Tanzania and Ethiopia.

Scoring of quality control results is done against three indicators. The effect of the application of various criteria is shown in the two test runs. The exercise was experienced as stimulating and quite revealing.

# Introduction

With the introduction of multiple drug therapy (MDT) regimens, the bacteriological assessment of patients by means of skin smears has become more important. The result of a skin smear examination can be decisive for the choice of the treatment regimen and for the duration of treatment.

In many programmes both the quantity and quality of skin smear examinations need to be considerably improved.

There is need for regular and systematic quality control of the skin smear services.

The quality and validity of bacteriological information in routine services depend on: 1, choice of the sites from which skin scrapings are taken; 2, the way in which the skin is cut/slit and the tissue scraped; 3, spreading of the material on the glass slide; 4, fixation of the specimen on the slide; 5, quality of the reagents used

for the staining; 6, appropriateness of the staining procedures; 7, skill and accuracy of the person performing the staining; 8, quality of the microscope (strength of optics, cleanliness, brightness of the source of light, etc.); 9, skill and accuracy of the person who does the microscopic reading; 10, reporting of skin smear results from the laboratory, and accurate recording on the individual patient record card.

In the leprosy control programmes of Tanzania and Ethiopia some preliminary experience has been gained with a method of quality control based on inter-observer comparison of skin smear slides sampled from routine services. The method assesses eight (Nos. 2–9) of the ten aspects mentioned above which influence the quality of the bacteriological information.

# Method

Laboratories were given instruction to preserve, in an appropriate box, all examined skin smear slides, positives and negatives, for up to 3 months\* after examination.

**1** SAMPLE SELECTION

Once every three months<sup>†</sup> a sample of 6 slides is taken for re-examination by a reference reader. As laboratory technicians seldom travel, selection is usually done by a (regional) leprosy control supervisor (LCS). The slides of the sample are documented on a 'Form for Quality Control of Skin Smear Examinations' (Figure 1) made out in duplicate, and the two sheets are marked A and B.

Selection of slides is done from the laboratory register book. Three positives and three negatives are selected. The method of selection is described on the back of the form, together with all further handling instructions.

In column VI the BI results found in the Service Laboratory are filled in on sheet B but not on sheet A: one BI value per smear.

Sheet B remains with the LCS.

Sheet A, without BI values in column VI, is taken or sent, together with the sample slides, to the Reference Laboratory.

# 2 REFERENCE LABORATORY

The Reference Laboratory Technician (RLT) re-examines the smears and gives judgements on the quality of smearing in column II, and of staining in column III.

\* The staining of smears kept for periods longer than 3 months may fade so that a reference reader would not be able to see as many bacilli as were visible when the specimen was fresh.

† Because it is not feasible to carry out this procedure four times a year, in Tanzania's National Programme the target has now been set at twice a year.

Possible gradings are: good, fair or poor. In column IV the RLT writes a comment for all smears or stainings of poor or fair quality, preferably with suggestions for improvement.

The result of the quantitative assessment of the AFB seen is entered in column V: one BI value for each smear.

# **3** COMPARISON

When the RLT has completed the re-examination, sheet A is handed to the LCS who, preferably in the presence of the RLT, copies the BI-values found in the Service Laboratory (Column VI of sheet B) onto sheet A. The BI values in column VI are then compared with those of column V and differences entered in the appropriate spaces of column VII (lower, same or higher).

All entries on sheet A in columns II, III and IV are also to be copied onto sheet B. The totals of the various findings are calculated and entered in the bottom line of both sheets A and B.

Sheet A then remains at the Reference Laboratory so that the performance of the Service Laboratory Technician concerned can be monitored over a period of time by comparison of the scores of subsequent examinations.

## **4** FEEDBACK TO THE SERVICE LABORATORY

Sheet B, together with the six slides, is taken by the LCS, to the Service Laboratory, where all findings and comments of the Reference Laboratory are then discussed with the Service Laboratory Technician concerned. Where results differ significantly the technician is advised to re-examine the smears concerned. Possible causes of the difference or imperfection may then be identified and corrected. The LCS may have to re-instruct the person who takes the smears how to avoid blood, how to spread the specimen properly on the glass-slide, how to fix the smear properly, etc.

Suggestions for improvement should already have been indicated by the RLT in column IV of the form. For example, the Technician 'simply' may have to increase his/her accuracy, or to correct the staining method (e.g. filter the carbolfuchsine before use), or the microscope may need cleaning, etc.

In cases of major discrepancies or persistent errors the Service Laboratory Technician may need re-training, preferably at the Reference Laboratory.

### 5 MONITORING AT HEADQUARTERS

Finally the LCS sends a summary of the scores of each Service Laboratory Technician-quoting the totals found in the bottom lines of the forms—to the headquarters of the programme (e.g. Ministry of Health), so that the programme coordinator is kept informed of quality control efforts and of results achieved.

#### A J de Rijk et al. 180

FORM FOR QUALITY CONTROL OF SKIN SMEAR EXAMINATIONS Sheet: ×

								code										
Name Service Lab.:							Iotal number of slides of last 8 weeks from which sample was taken (all read by same examiner)											
Name Exar	niner	:			Τ				Number of slides in last 6 months examined								_	
rosition/designation: District:							by this	same e>	aminer									
- Lab. Auxiliary						1	lotal nu	mber of	leprosy s	slides	hand	led i	in thi	s		-1		
- Lab Leobaician								(Period:	ry our	ing the las	5L O II	onths			)	L		
- cm).					_				(							)		
Sample Assessment at the REFERENCE LAR								BORATORY	TORY SERVICE LABORATORY									
1		11 111 1								v	VI	VII						
Slide identif-		Gual of	ity	Q	of	tу	Comment o	or Advic	e	Ref.	Serv.		LOWER		SAME	ŀ	HI GHE	R
lcation		smear	100	SI	aini	nņ				Lab. BI	Lab. BJ	2	2	1	0	1	2	7
	good	Tair	poor	good	TAIT	poor						4.5	- <u>2</u>	1		1	· ·	<i>,</i> ,
** taken by																		
,																		
						-			,									
taken by									· · · · ·									
													1					
													· · ·					
																		1.1
taken by																		
												-						
=												-						
				-		-												
taken by																		
							-											
taken by												-						-
				-														
				-		-										$\vdash$		
taken by																		
				10%										40.2		1.12		
Totals:							Smears read:											
** taken by	: wr	ite H	nere	the	initi	als	of the			Score	in %							
person w	ho t	ook t	he s	pecin	nen.													
Sample take	en fr	or se	ervio	e la	b. or	n dat	e	This	sample of	slides	was taken	by/is	to b	e ret	urned	to:		
Sample arri	ved	at re	efere	ence	lab.	on d	ate	Name:	Name:									
Despatch of	for	m and	sam	ple	from			Posit	ion:									
reference laboratory on date						Address:												

At time of sample selection this form is to be filled in DUPLICATE. Sheets to be marked as A and B. For instructions see the back of this form.

#### INSTRUCTIONS FOR INTER-OBSERVER COMPARISON OF SKIN SMEAR EXAMINATIONS

### I TWO SHEETS

This form is to be made out in duplicate. The 2 sheets are to be marked A and B in the right top corner. In column I the identification numbers of the selected slides and the initials of the makers are entered on both sheets.

#### **2 SAMPLE SELECTION**

The selection of the slides for re-examination is done on the basis of the laboratory register, usually by a leprosy control supervisor (LCS) preferably not by laboratory staff.

#### 2.1 Of one examiner

Slides to be selected in one sample should preferably have been stained and examined all by one laboratory technician. If more persons examine leprosy smears a separate sample may be taken and separate forms filled for each examiner.

#### 2.2 Sampling of negatives and positives

The sample should consist of six slides. Three with negative smears and three with a variety of positives (which will probably also contain some negative smears). Ideally the sample should contain some smears found to be highly positive (BI = 6, 5 or 4), some moderately positive (BI = 2 or 3) and some with BI = I.

#### 2.3 Representative for the period

The 6 slides should be chosen from the total examined during the last 8 weeks, taking a few from the slides examined during the early, middle and late parts of that period.

#### **3 FURTHER HANDLING**

After sample selection at the Service Laboratory and checking of the slide numbers (entered in column I) the BI results are filled in column VI on sheet B but not on sheet A. Sheet A goes with the slides to the Reference Laboratory. Sheet B remains with the LCS.

#### **RECORDING AT THE REFERENCE LAB** 4

4.1 Quality of smearing and of staining The Reference Laboratory Technician (RLT) will re-examine the smears and fill the columns II and III on sheet A with crosses for either good, fair or poor. In column IV the RTH must write a comment for all smears or stainings of poor quality. Preferably an advice should be given on how to improve.

4.2 BI-value

In column V the BI value found for each smear is entered.

### 4.3 Comparison of Service Lab with Reference Lab

When the RTH has completed the re-examination, sheet A with completed columns will be handed to the LCS who will, preferably in the presence of the RTH, copy the BI-values found in the Service Lab (found in VI of sheet B) on sheet A and then compare the BI values in column VI with those of column V.

Differences are entered in the appropriate subcolumns of VII (lower, same or higher). All other entries on sheet A (columns II, III and IV) are also to be copied on sheet B. Sheet A then remains with the RLT.

#### 5 FEEDBACK TO SERVICE LAB

Sheet B together with the slides is to be taken, by the LCS, to the Service Lab, where all findings are then discussed with the Service Lab personnel concerned. In case of considerable difference it is advisable that the technician examines those smears again. Appropriate action for improvement should be taken where and when necessary.

In case the smear taking was not satisfactory this must be taken up with the person who took the smears.

#### 6 REPORTING FROM REGION TO MINISTRY

Of all comparisons done during 6 months the RLT will make a summary report for the LCS, giving for each Service Lab in the region the total 'scores' of the bottom line of sheet A. This should be done in a table, with all districts mentioned. The LCS is requested to check this report, to make further comments where appropriate and to send one copy of the report to the central coordinator of the programme.

### Figure 1.

# 182 *A J de Rijk* et al.

Results and changes in scores are reviewed periodically so that supervisory activity and/or re-training activity can be directed accordingly.

# **Findings and comments**

### 1. QUALITY OF SMEARING AND STAINING

In Tanzania the regional leprosy control supervisors (LCS) and the reference laboratory technicians (RLT) of eight regions were asked to try out this method of quality control.

The preliminary results reported here are based on sheets B which were returned to headquarters by three of the participating regions.

Judgements on quality of smearing and quality of staining have been summarized in Table 1. In Tanzania the present routine is to take 3 smears per patient. A sample of 6 slides thus consists of 18 smears.

Criteria for good, fair and poor had not been given, but as the RLTs are themselves practical teachers of AFB microscopy, it was assumed that they were prepared and capable of making such judgements. Scores per sheet showed that

		Sr (+	nearin fixatio	g on)	Staining					
Region	District	Good	Fair	Poor	Good	Fair	Poor			
I	А	3		15	3	3	12			
	В	7	4	7	7	1	10			
	С	0	4	14		18				
	D	8	9	1		9	9			
	Ε	11	5	2	9	6	3			
II	F	. 9	9		6	12				
	G	6	11	1	4	12	2			
	Н	10	6	2	6	11	1			
III	J	14	3	1	14	3	1			
	Κ	14	4		17	1				
	L	14	4		15		3			
	Μ	12	4	2	12	4	2			
Total sc	ores	108	63 209/	45	93 429/	80	43 20%/			
		30%	29%	21%	43%	31%	20%			

**Table 1.** Quality of smearing and quality of staining. Scores on samples of 18 smears per district, for 12 districts (A–M) in 3 regions of Tanzania.

judgements on both smearing and on staining had been made separately for each smear.

The reference readers did not show any tendency to overscore the vague category of 'fair'. Although the judgements are arbitrary and cannot be standardized, the authors believe that the scores given provide useful feedback for both the persons concerned and for programme management. Only after repeated use of the same method will it be clear as to what extent these exercises lead to improvements.

The method is vulnerable to personal bias.

The RLT may be inclined to give many scores 'good' because he/she has trained and supervised the technicians.

The samples are not presented anonymously. Even if the names of the workers concerned were not filled in on sheet A, it is likely that they could be identified by means of the slide identification numbers. Therefore, if an RLT were to be personally biased towards a certain worker (in favour or against), this could influence the judgements given in columns II, III and IV. Because criteria cannot be standardized, and because of the possibility of personal bias, it is not considered valid to use the scores for comparing the performance of various districts or regions.

One RLT had not given any comments in column IV, but the two others had given quite a variety of comments or advice for improvements: 'don't decolourize too much, smear too thick, smear too thin, blood!' etc.

The totals of scores per sample at the bottom of each sheet (Figure 1) offer an easy means of monitoring trends in the change of quality per technician.

In Ethiopia, results were similar.

# 2. DIFFERENCES IN BI-VALUES

Comparison of results in the three regions in Tanzania, showed considerable differences in the BI-values of Service Laboratories and Reference Laboratories. On several occasions smears read in the Service Laboratory as BI = 1 were given BI = 5 by the Reference Laboratory Technician (RLT). A summary of the differences found is given in Table 2. How to judge these results? What degree of difference should be considered acceptable?

It is not difficult to decide that in samples D, E and J (in Table 2) the two readings showed far too much difference.

It is probably also not difficult to agree that samples F, H and K show differences so minute that the correlation of these readings should be considered quite satisfactory.

But how is one to decide, in such a complex series of scores, where to draw the line between acceptable and unacceptable?

Various possibilities were explored. Here, a construction is presented in which the correlation is judged by three indicators.

District	No. of smears	L -≥3	ower - 2	- 1	Sat 0	me + 1	Higher $+2 + \ge 3$		
		/All		1					
A	18			1	10	3	1	3	
В	18	2	3	3	7	2	1		
С	14*	1	1	3	7	2			
D	18	5	1	4	6	2			
E	18	4	1	2	7	3	1		
F	18			4	14				
G	18			4	9	5			
Н	18			1	14	3			
J	17*	8	4	2	3				
K	18			1	14	2	1		
L	18		3	3	12				
М	18	3	1	4	8	2			
Totals Proportion	211	23	14	32 15%	111	24	4	3	
. reportion	100%	18%	0	/0	79%	••/0		3%	

**Table 2.** Differences of BI-values. BIs in twelve Service Laboratories, compared to those of Reference Laboratories were: lower, same or higher.

\* Four smears of sample C and one of sample J were not examined as these specimens had either been washed off or were considered of too poor quality to be examined.

(a) Proportion of full correlation. In one project (where BI readings were made by very experienced examiners) the proportion of full correlation (i.e. readings 'same' or difference 0) was 75%.

For the Tanzania test run a criterion of  $\ge 50\%$  full correlation was applied.

(b) If a difference of 1 mark BI to either side is considered to be of little significance and therefore acceptable, one might also consider the three central columns together (-1, 0, +1) to represent 'a measure of correlation' between the two readers. This proportion should be much larger than 50%, e.g. 80%.

(c) If the differences in a sample are acceptable according to criteria 1 and 2, confirming that at least 80% of all differences remain within the range of the three central columns, it is still important whether the remaining scores are in the next nearest columns of 2 marks difference, or represent readings of 3 or more marks difference. Those further deviations are given a heavier weight in the assessment by the third indicator for correlation: variance.

Variance is the sum ( $\Sigma$ ) of the square values of the differences ( $D^2$ ) divided by the number of observations (N):

Variance = 
$$\frac{\Sigma D^2}{N}$$

Of each sample of smears the variance of the observations of the one reader compared to the readings of the reference reader can thus be calculated:

*Example*. In sample A (see Table 2) the variance of 18 observations was:

Variance = 
$$\frac{1^2 + (10 \times 0^2) + (3 \times 1^2) + (1 \times 2^2) + (3 \times 3^2)}{18}$$
$$= \frac{1 + 0 + 3 + 4 + 27}{18} = \frac{35}{18} = 1.944.$$

The above-mentioned three indicators of correlation have been applied on the findings given in Table 2 and results are shown in Table 3. For variance, a value of less than 1 was adopted as a criterion.

In Ethiopia the test run was conducted within the laboratory of ALERT, where large numbers of skin smear specimens are examined both from hospital patients and field patients. In the first 9 months of 1984 as many as 12,000 slides were examined. The laboratory has 12 technicians who deal with skin smears. A periodical re-examination of slides, picked randomly from the boxes of the

Table 3. Judgements on correlation according to various criteria

		I	IJ		П	I			
District	Same	≥ 50%	CC*≥	<b>80%</b>	Varia	nce < 1	Conclusion		
А	10/18	+	14/18	INS	1.944	_			
В	7/18	INS†	12/18		2.166	_			
С	7/14	+	12/14	+	1.286	INS			
D	6/18	INS	12/18	_	3.056				
E	7/18	INS	12/18		2.722				
F	14/18	+	18/18	+	0.222	+	+		
G	9/18	+	18/18	+	0.500	+	+		
Н	14/18	+	18/18	+	0.222	+	+		
J	3/17	INS	5/17	_	5.294	_			
K	14/18	+	17/18	+	0.389	+	+		
L	12/18	+	15/18	+	0.833	+	+		
Μ	8/18	INS	14/18		2.056	-			
12	7/	7/12		2	1.700	5/12	5/12		
					"		42%		

\* CC = the number of BI-readings falling in the three Central Columns (between the vertical lines in Table 2).

 $\dagger$  INS = insufficient correlation when found for the first time. In subsequent columns indicated with a minus (-).

various technicians, was carried out for 11 readers. This paper gives results of seven readers.

The reference reader was the Head of the Laboratory (TN).

Technicians were coded, A, B, C to G. Samples usually consisted of 6 slides with 4 smears each. For each technician four, five or six samples were examined. In the case of technician F, only two samples were taken. Findings on 30 samples with a total of 681 smears examined are given in Table 4. Table 4 shows deviations that are much less than those in Table 2. Of the 681 examinations in Ethiopia, the reference reader fully agreed with the BI values of as many as 455 smears (67%).

The bottom line of totals indicates that 93% of the readings remained within the three central columns, that is, the range of not more than 1 BI mark difference to either side  $(0\pm 1)$ .

For the 211 smears of Tanzania districts (Table 2) these proportions were 53% and 79% respectively. However, it was to be expected that seven laboratory technicians working as colleagues together in one laboratory of a leading referral and teaching hospital should produce better standardized readings than others working on their own in 12 different district hospitals.

In the further columns of Table 4, the correlation was assessed by the same three criteria as those applied in the test run in Tanzania (Table 3). Samples E1 and E2 are the only ones failing on all three criteria. Sample G2 is an interesting one. Its deviation is acceptable according to the rather tough third criterion, but fails on the seemingly mild first criterion of getting the same BI-value in at least half of the smears (same  $\geq 50\%$ ). The third criterion for correlation, the variance, particularly 'punishes' deviations beyond 1 BI mark and this works out progressively for every further mark because of the squaring of the differences. By this criterion, samples A1, B1 and F2 have insufficient correlation or, in other words, too much deviation.

For the sake of sample A1, the block for BI-values in Table 4 has been given an extra column on each side. This was to show the one smear that had been read with a difference of 4 BI marks. For the calculation of the third criterion it makes a lot of difference whether the deviation is 3 or 4 or 5 BI marks, namely  $D^2$ becoming either 9 or 16 or 25!

In Table 4 the second criterion  $(cc \ge 80\%)$  does not disqualify any of the 30 samples that had not already been declared insufficient by the first criterion. With all three criteria applied, 6 out of the 30 samples had too much deviation. The final correlation is then 24/30 or 80%.

# Discussion

# 1. FEASIBILITY OF THE METHOD

Requirements: boxes to store slides; proper instructions and tools for durable

Table 4. Quality control of skin smear reading in	30 samples of 7 technicians of ALERT, Ethiopia.
Differences of BI-values compared to Reference	Reader. Correlation of the two readings assessed
by 3 criteria.	

Service	<b>BI-Values</b>									Correlation							
reader and sample nr	-4 -:	3 -2	-1	Centra column 0 same	1  s  +1	+ 2	+ 3	+ 4	Nr of smears	I Same≩	≥ 50%	II CC≥8	0%	I Varia	II nce < 1	Conclusion	
A 2 3 4		1	4 4 1 2	11 18 16 19	2 4 2	1	1	I	19 24 23 23	11/19 18/24 16/23 19/23	+ + + +	17/19 22/24 21/23 23/23	+ + +	1.368 0.500 0.783 0.174	INS† + + +	+++++++	
B 2 3 * 4		4	6 7 1 2	12 13 19 19	2 2	13			24 20 24 24	12/24 13/20 19/24 19/24	+ + + +	20/24 20/20 22/24 21/24	+ + +	1.000 0.350 0.458 0.583	INS + + +	++++++	
1 2 C 3 4 5		2 2 1	7 4 8 5	14 17 12 14 24	1 1 2 4				24 24 23 23 24	+++++++++++++++++++++++++++++++++++++++		+++++++++++++++++++++++++++++++++++++++		0.667 0.542 0.609 0.391 0.000	+ + + +	+ + + +	
D 2 3 4	1	2 1	2 3 5	16 20 13 19	3 1 5				23 24 20 24	+++++++++++++++++++++++++++++++++++++++		+++++++++++++++++++++++++++++++++++++++		0.565 0.167 0.900 0.208	+ + + +	+ + + +	
E 3 4 5 6	1	6 4 1	6 8 5 2 1 1	9 9 15 19 13 19	2 2 3 3 6 1	1			24 24 24 24 20 22	9/24 9/24 + + +	INS INS	- + + +		1.708 1.250 0.500 0.208 0.350 0.273	- + + +	+++++++++++++++++++++++++++++++++++++++	
F 1 2		2	6 3	12 13	5 4	1	1		24 24	+		++++++		0.625 1.000	+ INS	+	
1 2 G 3 4 5	1	1	4 10 9 5	19 7 14 14 16	1 1 1 2				24 19 24 22 16	+ 7/24 + +	INS	+++++++++++++++++++++++++++++++++++++++		0.208 0.789 0.417 0.727 0.000	+ + + +	+ + + + + +	
Totals	3	30	121	455	60	9	2	1	681	27/30		28/30		0.584	25/30	24/30	
Propor- tion	5%	,	18%	67% 93%	9%		2%		100%							80%	

\* The calculations of the correlation criteria I and II have only been shown for the first 8 samples, but these were thereafter, in case of positive results, left out to keep the table more easily readable.

† INS = insufficient correlation when found for the first time, in subsequent columns indicated with a minus (-).

slide identification (diamond pencil); suitable forms (e.g. of the format shown in Figure 1); and appropriate explanation of purpose and method.

If these requirements are met, the method of quality control described above is feasible. It is crucial that personnel concerned should experience quality control as a necessary routine procedure of support and assistance, rather than as a threat.

Programme managers must ensure that personnel concerned have been properly trained and instructed in all procedures of skin smear examination, before any attempt is made to introduce this kind of quality control. (See the 10 steps mentioned in the Introduction.)

A refresher course for laboratory technicians is a very appropriate occasion for the introduction of the method. It may be helpful to find out whether quality control by means of inter-observer comparison is done with regard to any other tests in the medical laboratories concerned. If this is not the case the leprosy programme may have to pioneer.

# 2. USEFULNESS

This method of quality control will prove to be useful if, with subsequent testing, the results improve and then can be maintained at a certain level. The preliminary tests reported here cannot yet provide a measure of usefulness. Although the results of consecutive tests for some of the technicians at ALERT showed improvement (e.g.  $C1 \rightarrow C5$  and  $E1 \rightarrow E6$ ), this experience is considered to be too limited to draw conclusions regarding its impact on the quality of work.

# 3. WHO PROVIDES A STANDARD BI READING?

A further limitation of the method is that the reference reader who sets a standard may very well have some 'deviation' in his/her readings. Therefore, it is to be recommended that the various reference readers of a programme from time to time have their readings checked against a common standard.

There are various ways of doing this, for example arrange another interobserver comparison again between the RLTs, either during a national seminar or refresher course, or by mail. It might also be possible\* to provide sets of standard smears with known BI values that can be distributed for a systematic testing of the reference readers.

# 4. WHICH INDICATORS FOR CORRELATION?

In the assessment of these two test runs in Tanzania and Ethiopia the choice of criteria applied was arbitrary. The authors certainly do not assume that these are the only possible, or the best, criteria.

Generally, two steps are needed to set criteria: (a) choose indicators; and (b) choose criteria for each indicator.

\* The Royal Tropical Institute, Amsterdam, recently made one batch of 125 slides (625 smears) of known BI values. These smears were made from homogenous suspensions of various skin biopsies. Regular production of 'standard smears' may be considered.

In these test runs, the following three indicators were used: 1 proportion of BI values 'same' or 'difference zero'; 2 proportion of BI values in three central columns 'CC' or 'difference 0 and +/-1'; and 3 the variance† of the readings = the sum of squares of differences divided by number of observations.

Variance 
$$=\frac{\Sigma D^2}{N}$$
.

One might wonder why *three* indicators are proposed. Would variance alone not have been an acceptable indicator?

The indicators 'proportion same' and 'proportion in central columns' are presented first because these values are immediately *visible* on the form and are therefore expected to be immediately understood and accepted by all concerned.

It is not known to what extent 'variance' will be acceptable in field programmes. If, as in Tables 2 and 3, a large proportion of the samples already show insufficient correlation by mild criteria of the first two indicators, it may not always be necessary to apply the third indicator.

# 5. CRUCIAL VALUE BI = 2, FALSE POSITIVES, FALSE NEGATIVES

We considered for some time a possible fourth indicator, which would take into account the occurrence of false negatives and false positives round the crucial values BI = 1 and  $BI \ge 2$ , because the difference between BI = 1 and BI = 2 can be decisive for the important distinction paucibacillary or multibacillary, with all its consequences for cost and duration of treatment. However, introducing a distinction of false positives and false negatives appeared to be too complicated\* at this stage of the project. It may become relevant only after considerable improvement, measured by the present three indicators, has been achieved.

Because of the importance of BIs (0,) 1 and 2, Laboratory Technicians should be instructed that, when specimens contain only a few bacilli, at least *one hundred* microscopic fields should always be examined and all bacilli found should be reported: e.g. BI = 1 (4 bacilli/100 fields) and BI = 2 (27 bac/100 flds).

In situations where finding a few more or less bacilli would categorize a patient as either paucibacillary or multibacillary, e.g. BI = 1 (9 bac./100 flds) as compared to a BI = 2 (11 bac./100 flds), it should be a rule that the slide is re-examined and preferably new smears are taken aiming at the sites most likely to be positive.

# 6. WHICH CRITERIA

The criteria for each of the three indicators were also chosen arbitrarily. We used:

<sup>†</sup> Some readers might prefer, instead of the variance, the in statistics more often used 'standard deviation'.

\* In the LEPRA Programme of Malaŵi this criterion is already in use!

# 190 A J de Rijk et al.

for (1) 50%, for (2) 80%, and for (3) the variance to be below 1. However, it is quite possible that a programme manager chooses other criteria, e.g. for (1) 60%, for (2) 90% and for (3) variance <0.7. With the latter, 8 more of the 30 samples of ALERT (see Table 5) would have been picked out as insufficient, so that the overall correlation would have been only (16/30 =) 53%. Comparison of Tables 4

Service				В	I-Valu	es					Correlation						
and sample nr.	-4	-3	-2	-1	Central olumn 0 same	s + 1	+ 2	+3	+4	Nr. of smears	I Same ;	≥ 60%	II CC≥	90%	II Varian	I ce < 0·7	Conclusion
$A \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}$			1	4 4 1 2	11 18 16 19	2 4 2	1	1	1	19 24 23 23	11/19 18/24 16/23 19/23	INS† + + +	17/19 22/24 21/23 23/23	- + +	1 · 368 0 · 500 0 · 783 0 · 174	- + INS +	+ +
1 B 2 3 * 4		1	4 1	6 7 1 2	12 13 19 19	2 2	1 3			24 20 24 24	12/24 13/20 19/24 19/24	INS + + +	20/24 20/20 22/24 21/24	- + + INS	1.000 0.350 0.458 0.583	- + + +	+ +
1 2 C 3 4 5			2 2 1	7 4 8 5	14 17 12 14 24	1 1 2 4				24 24 23 23 24	INS + INS + +		+ + + +		0.667 0.542 0.609 0.391 0.000	+ + + +	+ + +
$D \frac{1}{3} \frac{1}{4}$		1	2 1	2 3 5	16 20 13 19	3 1 5				23 24 20 24	+ + + +		+ + + +		0·565 0·167 0·900 0·208	+ + INS +	+ + +
1 2 E 3 4 5 6		1	6 4 1	6 8 3 2 1 1	9 9 15 19 13 19	2 2 3 3 6 1	1			24 24 24 24 20 22	INS INS + + + +		+ + + +		1.708 1.250 0.500 0.208 0.350 0.273	- + + +	+ • + + +
F 1 2			2	6 3	12 13	5 4	1 1	1		24 24	INS INS		+ _		0.625 1.000	+ -	
1 G 3 4 5		1	1	4 10 9 5	19 7 14 14 16	1 1 1 2				24 19 24 22 16	+ INS INS + +		+ + + +		0·208 0·789 0·417 0·727 0·000	+ + INS +	++++++
Totals Pro- por- tion		3	30	121	455 67% 93%	60 9%	9	2	1	681 100%	20/30		24/30		0.584	21/30	16/30 53%

**Table 5.** Quality control of skin smear reading. Assessment of correlation in the same 30 samples as in Table 4. By same indicators, but with more strict criteria.

\* The calculations of the correlation indicators I and II have only been shown for the first 8 samples, but were thereafter left out, for the sake of readability.

+ INS = insufficient correlation when found for the first time, in subsequent columns indicated with a minus (-).

and 5 gives an impression of the effects of the various criteria. When for the category 'same', a proportion of 60% had been required, Table 5 shows that of ALERT's 30 samples, four (C1, C3, F1 and G3) would have been disqualified by this first criterion, while the differences remained within acceptable range according to the two other indicators/criteria.

Applied to the findings of 12 district laboratories in Tanzania, the stricter criteria used in Table 5 would have disqualified two more samples, so that only the three samples F, H and K would have 'passed' as sufficient.

# Conclusion

Although indicators and criteria are thus still a matter for discussion, it can be concluded that in our preliminary experience this inter-observer comparison of skin smears was found to be a stimulating exercise, revealing differences in BI reading of which the personnel concerned were not aware. More experience with the method will be needed to produce evidence of its value in routine services.

# Acknowledgment

Full acknowledgment is given to ALERT, Addis Ababa, Ethiopia and the Ministry of Health of Tanzania for permission to reproduce forms and data in this paper.