Volume 54, Number 4
Printed in the U.S.A.

Comparability Among Histopathologists in the Diagnosis and Classification of Lesions Suspected of Leprosy in Malawi[1]

Paul E. M. Fine, Charles K. Job, A. Colin McDougall, Wayne M. Meyers, and Jorg M. Ponnighaus[2]

It may be said that difficulties in diagnosing leprosy before the onset of typical peripheral nerve damage are widely recognized but rarely admitted—widely recognized insofar as leprologists have experienced and often discuss difficulties in concluding whether one or another lesion is indeed attributable to *Mycobacterium leprae*; rarely admitted insofar as most publications end up grouping people as either "cases" or "non-cases" with no reference to any uncertainty in deciding upon the diagnosis. Textbooks speak of the cardinal signs of leprosy (anesthesia, thickened nerves, skin lesions, acid-fast bacilli—by convention at least two of the first three, or the fourth, should be present for the diagnosis to be made [1]), but the assessment of the first three of these signs can be difficult, in particular under field conditions. In the face of these difficulties, histopathology has been called upon increasingly in recent years to assist in the diagnosis as well as the classification of leprosy. The reliance upon histology is particularly evident in research. A considerable proportion of current publications claim that cases under study were histologically "confirmed." Reference to the classification of cases according to histopathological criteria defined by Ridley and Jopling ([10]) is now almost universal in leprosy research.

Despite the widespread acceptance of histopathological evidence in leprosy, there have been few critical evaluations of the validity of this method in the hands of different investigators. There are a small number of published studies on the relationship between clinical and histopathological findings, particularly with regard to classification, but none of these was designed as a formal independent comparison ([5, 7, 12, 13, 15]). We are aware of no studies comparing the opinions of different histopathologists in the diagnosis of leprosy.

Given the difficulty and importance of arriving at a diagnosis of leprosy, this absence of critical studies is surprising. It contrasts with a growing literature on the critical evaluation of histopathological assessment of other diseases, in particular cancers ([2, 3, 4, 8, 11, 14]), which has developed methods for measuring and improving validity and comparability between diagnosing histopathologists.

Among the outstanding questions concerning the histopathological diagnosis and classification of leprosy are the following: a) To what extent do histopathologists agree and/or experience difficulties in arriving at a diagnosis of leprosy? b) If they do experience such difficulties, would it be useful to introduce a scale describing the level of certainty of the diagnosis when it is made? c) To what extent are differences in prevalence of leprosy, as reported in different studies, due to differences in diagnostic criteria used by different histopathologists? d) How comparable is the Ridley-Jopling system in the hands of different histopathologists? e) To what extent does the availability or absence of clinical information influence a histopathologist's assessment? f) What is

[1] Received for publication on 27 March 1986; accepted for publication in revised form on 13 June 1986.

[2] P. E. M. Fine, V.M.D., Ph.D., School of Hygiene and Tropical Medicine, Keppel Street, London WC1 7HT, England. C. K. Job, M.D., F.R.C.Path., F.A.M.S., Chief, Pathology Research Department, National Hansen's Disease Center, Carville, Louisiana 70721, U.S.A. A. C. McDougall, M.D., F.R.C.P., Consultant in Clinical Research (Leprosy), The Slade Hospital, Oxford OX3 7JH, England. W. M. Meyers, M.D., Ph.D., Chief, Division of Microbiology, Armed Forces Institute of Pathology, Washington, D.C. 20306-6000, U.S.A. J. M. Ponnighaus, Dr.Med., D.T.P.H., Director, Lepra Evaluation Project, P.O. Box 46, Chilumba, Karonga District, Malawi.

TABLE 1. *LEP histopathology report protocol as used in this investigation.*

LEP biopsy grading	
I. Biopsy quality	
S	= Satisfactory.
U	= Unsatisfactory (cite reason).
II. Diagnosis	
1A	= Leprosy confirmed beyond reasonable doubt.
1B	= Leprosy almost certain but slight element of doubt remains.
(1)	= Unable to select between 1A and 1B.
2A	= Consistent with but not diagnostic of leprosy.
2B	= Pathological and possibly due to leprosy, but lacking specific diagnostic criteria. Implies less evidence of leprosy than in 2A.
(2)	= Unable to select between 2A and 2B.
3A	= Definitely pathological but completely nonspecific.
3B	= Normal or near-normal tissue—minimal changes only.
(3)	= Unable to discriminate between 3A and 3B.
4	= Pathological but indicative of a specific disease other than leprosy—if so, please specify.
5	= "Other"—any unusual or unforeseen circumstance—if so, please specify.
III. Classification	
e.g., Ind, TT, BT, BB, BL, LL.	
IV. Bacillary content	
N	= No bacilli found within usual examination period.
BI (1–6)	= Bacterial index, if bacilli found.
D	= Doubtful findings (e.g., a few granules).

the optimal method to aggregate all available evidence—clinical, historical, bacteriological and histological—in arriving at a final diagnosis of leprosy?

We report here a study designed to investigate the first five of these questions. The sixth and final problem will be addressed in a separate publication (Ponninghaus, Fine and Bliss; unpublished data). The present investigation is based upon the independent examination, by three histopathologists, of 200 biopsies collected in the Lepra Evaluation Project (LEP) in Northern Malawi.

MATERIALS AND METHODS

All the biospies in this study were collected (by JMP) during the course of a total population survey for leprosy in Karonga District, Northern Malawi ([6] and Ponnighaus, Fine, Bliss, Sliney, Bradley and Rees; unpublished data). They were obtained under local anesthesia using a 4-mm punch (Steiffel Laboratories) from the most active area of lesions, nearly all of which were considered as at least possibly attributable to leprosy. In each case, the clinician graded his level of confidence in the diagnosis of leprosy based on a complete physical examination, but in the absence of any biopsy information, as: 1 = "certain"; 2 = "extremely likely"; 3 = "most likely"; 4 = "to be considered seriously"; or 5 = "possibly" ([6] and Ponnighaus, Fine and Bliss; unpublished data). The biopsies were fixed in formol-Zenker and shipped by air to Oxford, England, where they were embedded and cut into sections. At least one slide with an average of 10 sections was stained with hematoxylin and eosin (H&E) and one by the Fite-Faraco (FF) method from each biopsy. The study of these biopsies, as reported here, consisted of two phases.

Phase 1. The same 100 pairs of slides from 100 biopsies were examined independently by each of three histopathologists (CKJ, ACM, WMM), each in his own laboratory. In this part of the investigation, each biopsy was accompanied by a brief clinical note (by JMP), generally describing the age and sex of the biopsied individual, the site and size of the lesion, and whether anesthesia was present. Examples of these notes are as follows:

(51)—"The 4-mm punch biopsy is from the raised edge of a hypopigmented lesion on the back of this woman from Tanzania. There seemed to be anesthesia along the edges of the 2 × 1 cm large oval lesion."

(113)—"This young man was found by

TABLE 2. *Diagnostic certainty categories used by histopathologists X, Y, and Z in phase 1 and phase 2 of the study.*[a]

Diagnostic category	Phase 1			Phase 2		
	X	Y	Z	X	Y	Z
1A	44	23	46	55	25	56
1B	10	5	—	6	12	—
(1)	—	2	—	1	11	—
2A	12	7	—	14	8	—
2B	2	24	8	12	27	15
(2)	—	7	—	—	4	—
3A	17	22	36	9	12	22
3B	1	7	8	—	—	6
(3)	—	2	—	—	—	—
4	14	1	2	3	1	1
5	—	—	—	—	—	—
Totals	100	100	100	100	100	100

[a] Numbers in table refer to the number of times each category was used. See Table 1 for definitions of each category.

my staff with a swelling covering the forehead, the nose and most of the right cheek. Both zygomatic nerves seemed enlarged and the right one was tender. There was perhaps just slight mouth drop on the right side. Duration 3 weeks. (Dr. _____) and myself thought the most likely diagnosis was probably BT leprosy in (type I) reaction. The two specimens are from the forehead."

(137)—"The two specimens are from a faintly hypopigmented lesion on the right upper arm with raised edges. No anesthesia."

(284)—"He received treatment at Kochirira in 1965 and thereafter for several years at Karonga hospital. He was now found with several new macules. Slit skin smears from the macules were positive (BI 1+) several solids and fragments."

Each histopathologist reported his findings using the same protocol (see Table 1). The protocol specifies a four-part code: first, an assessment of whether the biopsy was considered satisfactory (S) or unsatisfactory (U); second, a "certainty" grading descending in steps from a most-certain category (1A = "leprosy confirmed beyond reasonable doubt") to an assessment that the lesion was not leprosy but something else (4 = "pathological but indicative of a disease other than leprosy"); third, a classification (TT, BT, etc.) to be assigned if leprosy was diagnosed; and, lastly, an indication of bacillary load (BI). The wording of the protocol was not discussed among the histopathologists before the study.

Phase 2. This was identical to phase 1, except that only one histopathologist had access to clinical notes and the other two read the slides totally "blind" of any information.

All results were sent directly to the London School of Hygiene and Tropical Medicine for compilation and analysis. The results of phase 1 were made available to the collaborating histopathologists before the beginning of phase 2.

RESULTS

The results of the two phases will be described separately. The histopathologists are identified only as X, Y, and Z. Histopathologists X and Y lacked clinical notes in phase 2.

Phase 1. All of the biopsies were considered sufficiently satisfactory to be reported upon by each of the histopathologists.

The leprosy diagnostic certainty results for all three histopathologists are summarized separately in Table 2 and together in Table 3. A clear difference is evident in the preference of each histopathologist for different categories, Y using the most and Z the least (Table 2). If we, for the sake of simplicity, consider categories 1A, 1B or 1 as indicative of "definite leprosy," categories 2A, 2B and 2 as "possibly leprosy," and categories 3A, 3B, 3 or 4 as "no evidence of leprosy," we find estimates from 30% (by Y) to 54% (by X) for the proportion with definite leprosy in this series (Table 3). The percent agreement on each of these categories between each pair and among all three histopathologists is set out in Table 4. It is seen that whereas pairs of histopathologists agreed by this overall criterion on between 57% and 72% of biopsies, only 49% of the biopsies were assigned the same level of certainty by all three histopathologists.

A similar comparison may be made between each of the histopathologists and the certainty of the clinical assessment. The categories chosen for this comparison are illustrated in Table 5. Agreement on "definite leprosy" occurred when the clinician considered the case either "certain" or "extremely likely" and the histopathologist as-

TABLE 3. *Three-way tables showing the agreement among histopathologists X, Y, and Z in diagnostic certainty.*[a]

Phase 1

Y	1					2					3					4					5					
X	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
Z 1	26[b]					14	3				2		1													46
2	2					2	1		1		1		1													8
3	2					4	6	4	3		1	3	11	9			1									44
4													1	1												2
5																										0
	30	0	0	0	0	20	10	4	4	0	4	3	14	10	0	0	1	0	0	0	0	0	0	0	0	100
	30					38					31					1					0					

Phase 2

Y	1					2					3					4					5					
X	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
Z 1	39	2				10	5																			56
2	2	1		1		4	6				1															15
3	1	1	1			5	4	4	1			6	4	1												28
4																	1									1
5																										0
	42	4	1	1	0	19	15	4	1	0	1	6	4	1	0	0	1	0	0	0	0	0	0	0	0	100
	48					39					12					1					0					

[a] In this table, categories 1A, 1B and (1) are grouped together as "1"; whereas 2A, 2B and (2) are grouped together as "2," and 3A, 3B and (3) are grouped together as "3." Diagnostic certainties for histopathologists X and Y are in columns, and those for Z are in rows. As an illustration, in phase 1, 26 biopsies were given a certainty grade of 1 by all three histopathologists; whereas 14 biopsies were considered grade 1 by X and Z, but grade 2 by Y.

[b] The numbers representing total agreement by all three histopathologists are underlined.

signed the biopsy to category 1A, 1B or 1. Agreement on "some evidence of leprosy" occurred when the clinician considered the case to be either "most likely" or "to be considered seriously," and the histopathologist assigned the biopsy to category 2A, 2B or 2. Finally, agreement of "no real evidence of leprosy" occurred when the clinician con-

TABLE 4. *Percent agreement among histopathologists X, Y, and Z broken down by certainty category.*[a]

Agree on	Between X and Y	Between X and Z	Between Y and Z	Among X, Y, and Z
	Phase 1			
"Definite leprosy"	30%	42%	26%	26%
"Possibly leprosy"	10%	1%	4%	1%
"No specific evidence of leprosy"	24%	29%	27%	22%
Total	64%	72%	57%	49%
	Phase 2			
"Definite leprosy"	42%	49%	41%	39%
"Possibly leprosy"	15%	7%	10%	6%
"No specific evidence of leprosy"	5%	11%	12%	5%
Total	62%	67%	63%	50%

[a] In this context, codes 1A, 1B, 1 are taken as indicating "definite leprosy"; codes 2A, 2B, 2 as "possibly leprosy;" and codes 3A, 3B, 3 or 4 as "no specific evidence of leprosy."

TABLE 5. *Criteria used in comparison between clinical and histopathological certainty (see Table 6).*

Clinical certainty	Biopsy certainty				
	1	2	3	4	Totals
"Certain"	A		I	J	P
"Extremely likely"	B		K	L	Q
"Most likely"		C	M	N	R
"To be considered seriously"	G	D			S
"Possibly"	H		E	F	T
"Other/not leprosy"					

Cells A + B indicate agreement on "definitely leprosy."
Cells C + D indicate agreement on "some evidence of leprosy."
Cells E + F indicate agreement on "no real evidence of leprosy."
(A + B + C + D + E + F)/(P + Q + R + S + T) = overall agreement between clinician and histopathologist.
Cells G + H indicate relative clinical underdiagnosis (or histopathological overdiagnosis).
Cells I + J + K + L + M + N indicate relative histopathological underdiagnosis (or clinical overdiagnosis).
Cells (G + H + I + J + K + L + M + N)/(P + Q + R + S + T) = overall considerable disagreement between clinician and histopathologist.

sidered the case only "possible" and the histopathologist assigned the biopsy to category 3 or 4. Those cases classified as only "to be considered seriously" or as "possible" by the clinician, but whose biopsies were considered 1A, 1B or 1 by the histopathologist, were considered to indicate considerable disagreement in the direction of relative clinical underdiagnosis (or histopathological overdiagnosis). On the other hand, those cases classified clinically as either "certain," "extremely likely," or "most likely," but which were assigned to categories 3 and 4 by the histopathologist, were considered to indicate considerable disagreement in the direction of relative histopathological underdiagnosis (or clinical overdiagnosis). Both of these groups together constitute the total with considerable disagreement. These comparisons between the clinician and each histopathologist are shown in Table 6. The percentages are calculated excluding five biopsies which were taken for reasons other than leprosy (interestingly enough none of the histopathologists classed these as 1A, 1B or 1, but his-

TABLE 6. *Percent agreement and disagreement among histopathologists X, Y, and Z and clinical assessment of "cases" biopsied because leprosy was suspected on clinical grounds (95 in phase 1 and 99 in phase 2).*[a]

	Phase 1 Between clinician and			Phase 2 Between clinician and		
	X	Y	Z	X	Y	Z
			Agree on			
"Definitely leprosy"	(31)[b] 32.6%	(20) 21.1%	(26) 27.4%	(37) 37.4%	(32) 32.3%	(37) 37.4%
"Some evidence of leprosy"	(6) 6.3%	(17) 17.9%	(5) 5.3%	(13) 13.1%	(19) 19.2%	(7) 7.1%
"No real evidence of leprosy"	(17) 17.9%	(18) 18.9%	(20) 21.1%	(4) 4.0%	(5) 5.1%	(12) 12.1%
Overall agreement	(54) 56.8%	(55) 57.9%	(51) 53.7%	(54) 54.5%	(56) 56.6%	(56) 56.6%
			Disagree on			
Relative clinical underdiagnosis	(12) 12.6%	(6) 6.3%	(11) 11.6%	(9) 9.1%	(4) 4.0%	(4) 4.0%
Relative histopathological underdiagnosis	(4) 4.2%	(5) 5.3%	(12) 12.6%	(5) 5.1%	(6) 6.1%	(11) 11.1%
Total considerable disagreement	(16) 16.8%	(11) 11.6%	(23) 24.2%	(14) 14.1%	(10) 10.1%	(15) 15.2%

[a] Categories are described in text and in Table 5.
[b] In each cell, the number of biopsies is given in parentheses.

TABLE 7. *Classifications provided by histopathologists X, Y, and Z.*[a]

	Phase 1						Phase 2					
	X Certainty		Y Certainty		Z Certainty		X Certainty		Y Certainty		Z Certainty	
	1	2	1	2	1	2	1	2	1	2	1	2
I	4	10	0	1	2	0	21	8	0	2	2	0
I/TT	0	0	0	2	0	0	0	0	0	0	0	0
TT	13	0	3	6	11	0	15	1	1	4	6	0
TT/BT	0	0	11	7	8	0	0	0	12	2	14	0
I/BT	0	2	0	0	0	0	0	0	1	2	0	0
BT	35	2	12	5	22	1	25	2	32	27	30	0
BT/BB	0	0	0	0	0	0	0	0	1	0	1	0
BB	0	0	0	0	1	0	0	0	0	0	1	0
BB/BL	0	0	1	0	0	0	0	0	1	0	0	0
BL	2	0	0	0	1	0	1	0	0	0	0	0
BL/LL	0	0	2	0	0	0	0	0	0	0	0	0
LL	0	0	0	0	1	0	0	0	0	0	2	0
Totals	54	12	29	21	46	1	62	11	48	37	56	0

[a] Numbers are the total numbers of classifications provided. The number of classifications are given separately for biopsies given certainty codes 1A, 1B, 1 (="1") vs certainty codes 2A, 2B, 2 (="2").

topathologist X classed one of them as 2A). Using these criteria, we find that the overall agreements between the clinician and each histopathologist were quite similar although there were differences in the distribution of their agreements among the "definite," "some," and "no real" evidence categories. There was greater variation between the levels of considerable disagreement between the clinician and different histopathologists, 12% with histopathologist Y and 24% with histopathologist Z.

Classifications were provided for different numbers of biopsies by each of the histopathologists as shown in Table 7. With one exception (a 1B biopsy for histopathologist Y), each histopathologist provided a classification for all biopsies given certainty 1A, 1B or 1. In addition, histopathologist X provided a classification for all 2A biopsies but no 2B biopsies; whereas histopathologists Y and Z provided classifications for only some of the biopsies coded as 2A, 2B or 2 on the certainty scale. Different preferences are evident, with the TT/BT category being used frequently by Y and Z but never by X. Histopathologist X used the "indeterminate" classification far more frequently than either Y or Z, but most (10 out of 14) of the biopsies so assigned were in certainty group 2.

Tables 8, 9, and 10 show the correlations between classifications by each of the histopathologists. The overall agreement is good, with the exception of one biopsy (113) classed as BT ("in downgrading reaction") by X, but as BL/LL by Y and as LL by Z.

Phase 2. Again, all biopsies were considered by all three participants to be satisfactory for reporting purposes. The diagnostic certainty results and agreement between the histopathologists are presented in Tables 2, 3, and 4. There were similar preferences for diagnostic categories in phase 2 as in phase 1, with histopathologist Z preferring the fewest categories and Y the most (Table 2). All three found the proportion of biopsies indicative of definite leprosy (codes 1A, 1B or 1) higher in phase 2 than in phase 1. Again, histopathologist Y reported the lowest number of definite biopsies (48, compared to 30 in phase 1) and X found the largest number (62, compared to 54 in phase 1). Histopathologist Z considered 56 biopsies as showing definite evidence of leprosy in phase 2, compared to 46 in phase 1.

Given the differences between phases 1 and 2 in the proportion of biopsies consid-

TABLE 8. *Comparison between classifications assigned by histopathologists X and Y.*[a]

		Classification by histopathologist Y													
		I	I/TT	TT	TT/BT	I/BT	BT	BT/BB	BB	BB/BL	BL	BL/LL	LL	?	Total
Classification by histopathologist X	I	2	2			2	23	1						12	43
	I/TT														0
	TT			8	11		9							1	29
	TT/BT														0
	I/BT														0
	BT			4	18		32					1		9	64
	BT/BB														0
	BB														0
	BB/BL														0
	BL									2		1			3
	BL/LL														0
	LL														0
	?	1		2	2	1	12								18
	Total	3	2	14	32	3	76	1	0	2	0	2	0	22	157

[a] Includes all biopsies in phase 1 and phase 2 which were classified by either histopathologist X or Y.

ered to show definite evidence of leprosy, it is interesting to note that the overall agreement on diagnostic certainty was very similar in phase 2 to that in phase 1 (Table 4). Once again, the certainty agreement was best between histopathologists X and Z. All three histopathologists agreed on the diagnostic certainty of 50% of the biopsies in phase 2.

The comparability between clinical and histopathological diagnostic certainty in phase 2 is presented in Table 6. Here the percentages are calculated excluding an attempted nerve biopsy which failed to include a nerve. The percent overall agreement between the clinician and each histopathologist was quite similar to that observed in phase 1. The same was true for percentage of disagreement, with the exception of the relationship between the clinician and histopathologist Z. Although Z disagreed with the clinician more often than did either X or Y in phase 2, as well as in phase 1, there was a fall in this disagreement (from 24.2% to 15.2%) between phase 1 and phase 2, a fall explained entirely by fewer disagreements in the direction of relative clinical underdiagnosis (or histopathological overdiagnosis) in phase 2 (4.0%) than in phase 1 (11.6%).

The classifications provided for the phase 2 biopsies are shown in Table 7. Histopathologist Z provided a classification only for biopsies classified as certain leprosy (1A); whereas X and Y also provided classifica-

TABLE 9. *Comparison between classifications assigned by histopathologists X and Z.*[a]

		Classification by histopathologist Z												
		I	I/TT	TT	TT/BT	BT	BT/BB	BB	BB/BL	BL	BL/LL	LL	?	Total
Classification by histopathologist X	I	3			5	11	1	1				1	21	43
	I/TT													0
	TT			10	4	11							4	29
	TT/BT													0
	BT	1		7	11	29						1	15	64
	BT/BB													0
	BB													0
	BB/BL													0
	BL							1		1		1		3
	BL/LL													0
	LL													0
	?				2	2								4
	Total	4	0	17	22	53	1	2	0	1	0	3	39	142

[a] Includes all biopsies in phase 1 and phase 2 which were classified by either histopathologist X or Z.

TABLE 10. *Comparison between classifications assigned by histopathologists Y and Z.*[a]

		Classification by histopathologist Z													
		I	I/TT	TT	TT/BT	I/BT	BT	BT/BB	BB	BB/BL	BL	BL/LL	LL	?	Total
Classification by histopathologist Y	I						1							2	3
	I/TT						2								2
	TT			4	2		3							5	14
	TT/BT			7	7		14							4	32
	I/BT													3	3
	BT	2		6	12		28	1	1					26	76
	BT/BB												1		1
	BB														0
	BB/BL								1				1		2
	BL														0
	BL/LL										1		1		2
	LL														0
	?	2			1		5								8
	Total	4	0	17	22	0	53	1	2	0	1	0	3	40	143

[a] Includes all biopsies in phase 1 and phase 2 which were classified by either histopathologist Y or Z.

tions for several biopsies considered to show some evidence of leprosy (2A, 2B, 2). Once again we note the preference of the TT/BT category by histopathologists Y and Z, and its total avoidance by X. Histopathologist X used the indeterminate classification far more often than Y or Z, for 29% of all biopsies and for more than a third of all the biopsies he considered to have definite evidence of leprosy (1A, 1B or 1).

A major difference in classification arose in only a single phase 2 biopsy (284), which was classified as indeterminate by histopathologist X, BT/BB by Y, and LL by Z (Tables 8, 9, and 10).

DISCUSSION

People differ. This is true of diseased individuals in terms of their lesions, both macroscopic and microscopic. It is also true of clinicians and histopathologists in terms of their observations and interpretations of these lesions. The study reported here provides an attempt to assess the magnitude of such differences as they affect the diagnosis of leprosy by three histopathologists, each of whom has spent many years involved in the histopathological diagnosis of leprosy. Without a doubt, different results would have arisen if other individuals had been involved or if a different series of biopsies had been used. What is important is thus to consider the nature and trends of these disagreements in order to be able to apply these results to the improvement of leprosy diagnosis in the future.

One of the fundamental findings of this study is the frequent use of the 2A, 2B or 2 diagnosis category by all three histopathologists (20% of all biopsies by X, 38.5% by Y, 11.5% by Z). This category, implying a histopathological picture which was consistent with but not pathognomonic of leprosy, implies a degree of uncertainty on the part of the histopathologists analogous to that experienced by clinicians when assessing leprosy in the field. Given the frequency of use of the category in this study, it is of interest that the unclear diagnosis category appears so infrequently in the literature.

The proportions of biopsies classed as showing strong or definite evidence of leprosy (diagnostic categories 1A, 1B, 1; see Table 1) ranged from 30% to 54% in phase 1 and 48% to 62% in phase 2 (39% to 58% overall). In this context, it should be recalled that the biopsies were obtained from lesions ascertained in a total population survey, and included a range from minimal to well-advanced disease. Of the 200 individuals included in this investigation, leprosy was considered certain or extremely likely on clinical grounds in 82 (35 in phase 1 and 47 in phase 2). Considering these 82 patients alone, the proportions of biopsies classed as showing strong or definite evidence of leprosy ranged from 63% to 83% (Tables 5 and 6) among the three histopathologists.

With reference to classification, we found a high degree of correlation among the three histopathologists (Tables 8, 9, and 10). Of

the 200 biopsies included in this study, only two led to major disagreements. One of these was from a patient considered to be in (type 1) reaction clinically, which may explain why it was classified as BT ("in downgrading reaction"), BL/LL and LL by the three histopathologists (see clinical note 113 in Materials and Methods, Phase 1). An additional factor might be that the Fite-Faraco stain of this slide had faded considerably by the time it was examined by histopathologist Y, and even more so by the time the slide was examined by histopathologist X. The second was from a patient who had received antileprosy treatment for many years, and in whose biopsy the inflammatory infiltrate was minimal, which may explain why it was classified independently as I, BT/BB and LL (see clinical note 284 in Materials and Methods, Phase 1). In addition, there were obvious preferences for and against certain classifications by the participants. The TT/BT category was used frequently by histopathologists Y and Z, but never by X. In contrast, the "indeterminate" classification was used frequently by histopathologist X but rarely by Y and Z. Histopathologist Y used the indeterminate classification three times, but never in a biopsy classed as definite leprosy (1A, 1B or 1). Histopathologist Z used the indeterminate classification four times, but only for biopsies considered to show definite evidence of leprosy (1A). There was no overlap between the three indeterminates of histopathologist Y and the four of Z. In contrast, histopathologist X used the indeterminate classification 43 times (21.5% of all biopsies in the study), including for 25 out of 116 biopsies (21.6%) which he considered to have definite evidence of leprosy. This finding is particularly interesting insofar as it suggests that the major differences in the proportion of leprosy cases classed as indeterminate, as reported in different studies and in different parts of the world, may be due entirely to differences in diagnostic terminology.

One of the important questions addressed by this investigation is the extent to which clinical information may influence an histopathologist's interpretation of biopsy material. In phase 1 of this study each histopathologist had access to a brief clinical note; whereas in phase 2 this information was available only to histopathologist Z. If the clinical information had influenced the histopathologists, we would expect that the agreement between X and Z and between Y and Z should have been higher in phase 1 (when all had access to the information) than in phase 2 (when only Z had such access). Scrutiny of Table 4 shows no evidence of such a trend. Another way to test the hypothesis is by examining the correspondence between the clinician and the histopathologist. If the clinical note had made a difference, one would expect that the agreement between X or Y and the clinician would be higher in phase 1 than in phase 2, but the agreement between Z and the clinician should have been the same in phase 1 and phase 2. We find no strong evidence for such an effect. On the other hand, we see in Table 6 that histopathologist Z disagreed with the clinician less often in phase 2 than phase 1. This may suggest that the phase 2 biopsies were slightly "easier" for the histopathologists. If so, this could have counteracted any disadvantage suffered by histopathologists X and Y because of their not having the clinical notes. In summary, then, the results are not inconsistent with the hypothesis that clinical information had some influence on these histopathologists, but they suggest that any such influence was small.

All of these results must be seen in the context of the design of this study. All of the histopathologists worked independently of one another and used the same protocol for reporting results. Although these features of the study were ideal, the design had certain failings. Some of the differences which arose in establishing and comparing the certainty of the diagnosis of leprosy are probably attributable to the wording of the coding protocol (Table 1). Differences in interpretation would be expected to arise, in particular since the histopathologists did not discuss the wording among themselves prior to the study. In one sense this makes our findings "realistic," insofar as the majority of histopathologists in the world work independently and rely upon their interpretation of certain published descriptions and terminology to achieve a modicum of consistency. On the other hand, the wording used in this study was probably not optimal, and greater comparability among the his-

topathologists could have been achieved if the wording had been "better." This issue is addressed in the Appendix to this paper.

Another weakness of the design of this study was the absence of any "blind" duplicate or repeat biopsies. Their inclusion in the study would have permitted an assessment of the degree of intra-observer variation, and would have facilitated the assessment of the influence of clinical information on biopsy interpretation. It is hoped to carry out a study including such "blind" duplicates in the future.

We have presented the results of our study at their face value. We have not attempted to measure the sensitivity or specificity of diagnoses made by each histopathologist since this is a difficult problem, requiring reference to an absolute standard, and will be discussed in a separate publication (Ponnighaus, Fine and Bliss; unpublished data). Derived statistics such as the comparability *Kappa* have not been used since they are not widely understood ([14]). More importantly, it should be recognized that the numerical values reported here are not the most important result of this study. They are a function of the histopathologists, the biopsies and the coding conventions. Other combinations would have yielded different numerical values. What is important is the clear demonstration that the use of histopathology in the diagnosis of "early" leprosy (i.e., early with respect to evolution) is not a simple matter, even when the biopsies are read by people with many years of experience. There is room for improvement in the standardization of criteria and of terminology.

SUMMARY

Identical slides from 200 biopsies obtained from individuals suspected of having leprosy during the course of an epidemiological survey in Northern Malawi were examined sequentially and independently by three histopathologists, using a standard protocol to report their findings. Their results are compared among themselves and with a standardized clinical assessment of each subject. There was more agreement among the histopathologists as to classification of leprosy cases than there was on the diagnosis of leprosy. The proportion of biopsies considered to show definite evidence of leprosy varied from 39% to 58% among the histopathologists. An appreciable additional proportion of biopsies (11.5% to 38.5% for the three histopathologists) was considered to show evidence suggestive but not pathognomonic of leprosy. Although there was, in general, good agreement on classification, the proportion of biopsies considered to show evidence of indeterminate leprosy varied from 1.5% to 21.5% among the three histopathologists. This suggests that some of the reported differences in the prevalence and proportion of indeterminate leprosy in different populations is due to terminology alone. A follow-up meeting of the study participants revealed that many of the differences in diagnosis certainty were due to difficulties in interpreting evidence of nerve involvement. It is recommended that greater attention be paid to the difficulties of diagnosing leprosy on histopathological as well as clinical grounds. A revised standard protocol for reporting histopathological evidence of leprosy is presented.

RESUMEN

Durante un estudio epidemiológico en Malawi del Norte, tres histopatólogos, usando un protocolo de reporte estandar, examinaron independientemente, laminillas idénticas de 200 biopsias obtenidas de individuos con sospecha de lepra. Sus resultados fueron comparados entre sí y con la evaluación clínica estandarizada de cada sujeto. Hubo más concordancia entre los histopatólogos en cuanto a la clasificación de los casos que en cuanto al diagnóstico de la lepra. La proporción de biopsias con evidencias definitivas de lepra varió del 39% al 58% entre los histopatólogos. Una apreciable proporción adicional de biopsias (del 11.5% al 38.5% para los 3 histopatólogos) se consideraron con evidencias sugestivas pero no patognomónicas de la enfermedad. Aunque en general hubo buena concordancia en cuanto a clasificación, la proporción de biopsias con evidencias de lepra indeterminada varió del 1.5% al 21.5% entre los 3 histopatólogos. Esto sugiere que algunas de las diferencias reportadas en la prevalencia y proporción de la lepra indeterminada en diferentes poblaciones solo se deben a la terminología. En una reunión de los participantes del estudio se estableció que muchas de las diferencias en el diagnóstico se debieron a dificultades en la interpretación de las evidencias de afección nerviosa. Se recomienda que se ponga más atención a las dificultades de diagnóstico de la lepra tanto a nivel histopatológico como clínico.

Se presenta un protocolo estandar para reportar las evidencias histopatológicas de la lepra.

RÉSUMÉ

Trois histopathologistes, utilisant un protocole standard pour noter leurs observations, ont étudié de manière séquentielle et de façon indépendante, des lames identiques provenant de 200 biopsies recueillies chez des personnes soupçonnées d'être atteintes de lèpre. Cette étude a été menée au cours d'une enquête épidémiologique au Malawi du Nord. Les résultats ont été comparés entre eux, de même qu'avec une évaluation clinique standardisée de chaque individu. On a constaté une concordance plus forte entre les histopathologistes pour ce qui regarde la classification des cas de lèpre; cette concordance est moins prononcée pour le diagnostic de lèpre. La proportion de biopsies considérées comme présentant des signes indubitables de lèpre a varié de 39% à 58% selon les histopathologistes. Une proportion supplémentaire notable de biopsies (11,5% à 38,5% pour les trois histopathologistes) ont été considérées comme présentant des signes qui suggèraient la lèpre, sans en être pathognomoniques. Malgré la concordance généralement bonne pour ce qui concerne la classification, la proportion de biopsies considérées comme présentant des images de lèpre indéterminée a varié de 1,5% à 21,5% chez ces trois histopathologistes. Cette observation donne à penser que certaines des différences qui sont rapportées quant à la prévalence et à la proportion de la lèpre indéterminée dans diverses populations, pourraient être dues simplement à la terminologie. Une réunion subséquente des participants à l'étude a révélé qu'une grande part des incertitudes diagnostiques était due aux difficultés rencontrées pour interpréter les manifestations d'atteinte nerveuse. On recommande dès lors d'accorder une plus grande attention aux difficultés du diagnostic de la lèpre, tant au point de vue histopathologique qu'au point de vue clinique. On propose un nouveau protocole standard pour consigner les observations histopathologiques de la lèpre.

Acknowledgments. This investigation arose as part of the Lepra Evaluation Project (LEP) and was funded primarily by LEPRA, the British Leprosy Relief Association. Additional special funding for the follow-up meeting was provided by the Heiser Foundation and is gratefully acknowledged. The authors wish to thank the people of Karonga District, Northern Malawi, and the entire LEP staff without which this study would not have taken place. Particular thanks are due to Mr. N. Maine for assisting in the analysis of the results and to Ms. B. Slessor for preparation of the manuscript.

REFERENCES

1. Bryceson, A. and Pfaltzgraff, R. E. *Leprosy.* 2nd ed. London: Churchill Livingstone, 1979.
2. Correa, P., O'Connor, G. T., Bernard, C. W., Axtell, L. M. and Meyers, M. H. International comparability and reproducibility in histologic subclassifications in Hodgkin's disease. J. Natl. Cancer Inst. **50** (1973) 1429–1435.
3. Evans, D. M. D., Shelley, G., Cleary, B. and Baldwin, Y. Observer variation and quality control of cytodiagnosis. J. Clin. Pathol. **27** (1974) 945–950.
4. Heenan, P. J., Matz, L. R., Blackwell, J. B., Kelsall, G. R. H., Singh, A., Ten Seldam, R. E. J. and Holmann, C. D. J. Inter-observer variations between pathologists in the classification of cutaneous malignant melanoma in western Australia. Histopathology **8** (1984) 717–729.
5. Jerath, V. P. and Desai, S. R. Diversities in clinical and histopathological classification of leprosy. Lepr. India **54** (1982) 130–134.
6. McDougall, A. C., Ponnighaus, J. M. and Fine, P. E. M. The histopathological examination of skin biopsies from an epidemiological study of leprosy in northern Malawi. Int. J. Lepr. (1987) (in press).
7. Meyers, W. M., Heggie, C. D., Kay, T. L., Staple, E. M. and Kvernes, S. The Ridley-Jopling five-group classification of leprosy—correlations of parameters of the classification in 1429 leprosy patients. Int. J. Lepr. **47** (1979) 683–684.
8. Non-Hodgkin's Lymphoma Pathologic Classification Project. Summary and description of a working formulation for clinical use. Cancer **49** (1982) 2112–2135.
9. Ridley, D. S. *Skin Biopsy in Leprosy.* 2nd ed. Basle: Documentia Geigy, 1985.
10. Ridley, D. S. and Jopling, W. H. Classification of leprosy according to immunity; a five-group system. Int. J. Lepr. **34** (1966) 255–273.
11. Schmoeckel, C. How consistent are dermatopathologists in reading early malignant melanomas and lesions "precursor" to them? An internation survey. Am. J. Dermatopathol. **6** Suppl. (1984) 13–24.
12. Sehgal, V. N., Koranne, R. V., Nayyar, M. and Saxena, H. M. K. Application of clinical and histopathological classification of leprosy. Dermatologica **161** (1980) 93–96.
13. Sehgal, V. N., Rege, V. L. and Rey, M. Correlation between clinical and histopathological classifications in leprosy. Int. J. Lepr. **45** (1977) 278–280.
14. Silcocks, P. B. S. Measuring repeatability and validity of histological diagnosis—a brief review with some practical examples. J. Clin. Pathol. **36** (1983) 1269–1275.
15. Singh, K., Iyengar, B. and Singh, R. Variations in clinical and histopathological classification of leprosy—a report and a possible explanation. Lepr. India **55** (1983) 472–479.

APPENDIX

At the conclusion of this study, the participants met for two days in order to discuss the results. Twenty-six pairs of slides over which there had been substantial

disagreement concerning the certainty of the diagnosis of leprosy were reviewed. In each case, the histopathologist who had given the highest certainty grading reexamined the slide and tried to present the evidence on which his certainty grading was based.

In only one case could the crucial evidence no longer be found, within a reasonable period of time. With reference to the remaining 25 biopsies, the three histopathologists reached complete agreement in ten. Some disagreement remained for 15 biopsies. The main reasons for the remaining disagreements were: a) whether or not a particular infiltrate contained the remains of a nerve; b) whether or not, if there was undoubtedly a nerve, the infiltrate was substantial enough to permit a definite diagnosis of leprosy; and c) whether or not particular minimal findings should be considered as possibly due to leprosy. It was felt that it would not be possible to resolve all of the remaining differences although, in some cases, special staining procedures might have provided sufficient additional information to bring complete agreement within reach.

After this workshop, the participants were joined on the third day by Professor K. P. W. J. McAdam (Department of Clinical Tropical Medicine, London School of Hygiene and Tropical Medicine); Dr. D. S. Ridley (formerly Hospital for Tropical Diseases, London); and Dr. T. J. Ryan (Department of Dermatology, John Radcliffe Hospital, Oxford) in order to discuss the broader implications of this study. It was agreed that the results accumulated thus far raised important issues relevant to leprosy research. Insofar as some of the discrepancies may have been attributable to the wording of the original coding protocol (Table 1), the group discussed and agreed upon an improved version of the certainty scale. The revised protocol, analogous to one recently published by Ridley ([12]), is reproduced here as Table A-1, along with examples of criteria which may be used to assign the categories implying different levels of evidence of leprosy. These are only examples, and are by no means intended as an exhaustive set of criteria. A follow-up study is now planned in order to assess the usefulness of this revised protocol. The participants encourage other workers in the leprosy field to consider using this protocol when reporting and analyzing histopathological evidence of leprosy and/or to comment critically upon it.

TABLE A-1. *Revised biopsy grading protocol.*

I. Biopsy quality
- S = Satisfactory.
- U = Unsatisfactory—if so, please explain.

II. Diagnosis
- 1 = Leprosy confirmed beyond reasonable doubt:
  - e.g., i) Presence of AFB especially in protected sites.
  - ii) Infiltration by inflammatory cells and/or granuloma and damage or destruction of nerve tissue.
  - iii) Etc.
- 2 = Suggestive of but not diagnostic of leprosy:
  - e.g., i) Granulomatous infiltration without definite nerve involvement and absence of features of other granulomatous disease.
  - ii) Selective inflammation of either perineural tissue and/or sweat glands and/or arrector pili muscle.
  - iii) Etc.
- 3 = Pathological and possibly due to leprosy (but also possibly due to other diseases):
  - e.g., i) Granulomatous infiltration not involving nerves at all.
  - ii) Etc.
- 4 = Pathological but (completely) nonspecific.
- 5 = Normal or near-normal tissue (skin with no significant lesion).
- 6 = Pathological and indicative only of a specific disease other than leprosy—if so, please specify.
- 7 = "Other"—any unusual or unforeseen circumstance—if so, please specify.

III. Classification
- e.g., I, TT, BT, BB, BL, LL.

IV. Bacillary content
- N = No bacilli found.
- BI (1–6) = Bacterial index, if bacilli found.
- D = Doubtful findings (e.g., a few granules).

Examples of codes as reported:	S,1,TT,N	S,2,–,N	S,6,–,N (onchocerciasis)
	S,1,BL,3	S,3,–,N	