INTERNATIONAL JOURNAL OF LEPROSY Volume 61, Number 2
Printed in the U.S.A.

Extent, Origin, and Implications of Observer Variation in the Histopathological Diagnosis of Suspected Leprosy[1]

Paul E. M. Fine, Charles K. Job, Sebastian B. Lucas, Wayne M. Meyers, Jörg M. Pönnighaus, and Jonathan A. C. Sterne[2]

Accurate diagnosis is an essential component of research on any disease since errors can lead to serious biases or misinterpretations ([3, 12]). Recognition of these implications has stimulated a series of publications on problems of leprosy diagnosis in recent years ([4, 5, 7, 13]).

Histopathology has been called upon repeatedly and increasingly to assist in leprosy research, and has been incorporated as an important element within the protocols of several recent drug evaluations and vaccine trials ([1, 2, 11]). As part of an evaluation of the contribution of histopathology to leprosy diagnosis in the field, we carried out a study in 1985 comparing independent assessment, by three experienced histopathologists, of 200 biopsies collected from leprosy suspects in Malawi, Central Africa ([5]). The proportion of biopsies considered to show definite evidence of leprosy ranged from 39% to 58% among the three observers. The study also found that an appreciable proportion (which varied from 11.5% to 38.5% among the three histopathologists) of biopsies from individuals suspected of having leprosy were considered to show evidence consistent with, but not confirmatory of, leprosy. This was of interest in that most publications reporting histopathological "confirmation" omit any reference to an uncertain diagnostic category. With regard to classification of leprosy lesions, the study revealed a high level of agreement between the histopathologists, except for a marked difference in their use of the indeterminate category. This finding suggested that the differences reported between studies and populations in proportions of cases classified as indeterminate may be more a reflection of diagnostic criteria than of real differences between the study populations.

An important question raised by this investigation was the extent to which differences observed between histopathologists in their diagnoses of leprosy reflected true between-observer differences, and to what extent they reflect hidden variation within-observers. The former would imply consistent differences of opinion between histopathologists in their interpretations of particular findings. In contrast, within-observer variation would imply general difficulties experienced in the examination and/or the evaluation of some biopsies, which are then revealed as day-to-day variations in conclusions.

We report here a study designed to distinguish between-observer variation from within-observer variation in the histological diagnosis of leprosy. The study entailed double examination of 100 biopsies by each of three histopathologists, under blind conditions.

MATERIALS AND METHODS

One-hundred-twenty biopsies were selected from those routinely collected from individuals ascertained as being suspected of having leprosy within the LEPRA Evaluation Project/Karonga Prevention Trial

[1] Received for publication on 3 November 1992; accepted for publication in revised form on 16 March 1993.

[2] P. E. M. Fine, V.M.D., Ph.D.; S. B. Lucas, F.R.C.Path.; J. A. C. Sterne, Ph.D., London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K. C. K. Job, M.D., F.R.C.Path., Pathology Research Department, GWL Hansen's Disease Center, Carville, Louisiana 70721, U.S.A. W. M. Meyers, M.D., Ph.D., Division of Microbiology, Armed Forces Institute of Pathology, Washington, D.C. 20306-6000, U.S.A. J. M. Pönnighaus, Dr.Med., D.T.P.H., Lepra Evaluation Project, P. O. Box 46, Chilumba, Karonga District, Malawi.

TABLE 1. *Coding protocol used by histopathologists in this study.*

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

Diagnosis
- 1 = Leprosy confirmed beyond reasonable doubt, e.g.:
  - a) Presence of AFB especially in protected sites
  - b) Infiltration by inflammatory cells and/or granuloma and damage or destruction of nerve tissue
  - c) Etc.
- 2 = Suggestive of but not diagnostic of leprosy, e.g.:
  - a) Granulomatous infiltration without definite nerve involvement and absence of features of other granulomatous disease
  - b) Selective inflammation of either perineural tissue and/or sweat glands and/or arrector pili muscle
  - c) Etc.
- 3 = Pathological and possibly due to leprosy (but also possibly due to other diseases), e.g.:
  - a) Granulomatous infiltration not involving nerves at all
  - b) 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

Classification, e.g.:
I, TT, BT, BB, BL, LL

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

(LEP/KPT) in Karonga District, Northern Malawi ([14]). The only selection criterion was that all suspects had been examined by the same medical officer (JMP) who graded his level of confidence in the diagnosis of leprosy, on the basis of clinical evidence alone, as: 1 = "possibly"; 2 = "to be considered seriously"; 3 = "most likely"; 4 = "extremely likely"; or 5 = "certain." Skin biopsies were obtained with a disposable 4-mm punch. Split-nerve biopsies were also included. The specimens were fixed in formol-Zenker and shipped by air to England ([9]).

Fifteen similar skin biopsies were obtained (by Dr. T. J. Ryan) from dark-skinned, long-term residents of the United Kingdom, with various dermatological conditions, but for whom there was no suspicion of leprosy.

Two slides, each with 4–8 serial sections, were made from each biopsy; one was stained with hematoxylin and eosin (H&E), one by the Wade-Fite method. For Round 1, slide pairs from 110 Malawi leprosy "suspects" and from the 15 U.K. residents were coded and sent in batches to the three histopathologists (CKJ, SBL, WMM). The histopathologists examined the slides independently and reported their findings using the same protocol (Table 1). After all 125 slide pairs had been examined by each histopathologist, the slides were returned to London. Ten Malawi slide-pairs were removed, including all damaged ones, and 10 others substituted for them in order to reduce the likelihood that the histopathologists would recognize the specimens. The 125 pairs were then re-ordered, given different code numbers, and again sent to the three histopathologists for independent examination and reporting (Round 2).

The histopathologists were sent the slides with no clinical or other information whatsoever relating to their origin.

The protocol used for reporting results (Table 1) was identical to that published in

TABLE 2. *Distribution of diagnostic certainty codes reported by three histopathologists, by round.*

Diagnostic certainty code	A		B		C	
	Round		Round		Round	
	1	2	1	2	1	2
1	55	42	55	52	30	29
2	11	10	12	15	7	8
3	4	15	12	13	11	12
4	19	20	18	16	32	39
5	1	0	0	0	1	1
6	9	12	3	4	19	11
7	0	0	0	0	0	0
Blank[a]	1	1	0	0	0	0
Total	100	100	100	100	100	100

[a] Specimen considered unsatisfactory, and no diagnostic certainty code assigned (see Table 3).

1986 ([5]), having been devised in a workshop following our previous comparability study. It employs a four-part code: first, an assessment of whether the biopsy preparation was considered satisfactory (S) or unsatisfactory (U) for histopathological purposes; second, a diagnostic "certainty" grading descending from a most certain category (1) to suggestive but not diagnostic (2) to pathological and possibly due to leprosy (3) to nonspecific (4), normal (5), indicative of some other disease (6) or unforeseen category (7); third, a classification [tuberculoid (TT), borderline tuberculoid (BT), etc.] to be assigned if leprosy was diagnosed; and lastly, an indication of bacillary load (bacterial index, BI).

The results were linked to the LEP/KPT clinical examination data for analysis. A special effort was made to follow up all individuals whose biopsies were subject to disagreement between the histopathologists during Round 1.

At the conclusion of the study the investigators met for 3 days in London, and reviewed together those biopsies on which there had been appreciable disagreement. An effort was made to identify the reasons for these disagreements.

In addition to cross tabulations, we present Kappa statistics as a measure of agreement between the several assessments ([6]). This statistic is calculated as Kappa (K) = $(P_o - P_e)/(1 - P_e)$, where P_o = the observed proportion which agree, and P_e = the proportion of results which would be expected to agree by chance, contingent upon the marginal totals. The Kappa statistic thus describes the extent of agreement between observers beyond the agreement expected by chance alone. It has been widely recommended and employed in studies of the reproducibility of histopathological diagnosis of cancers ([8, 17]).

The complete protocol was agreed upon in full by all participants prior to the investigation. The three histopathologists are identified only by code (A, B and C) in this report. First and second round results by each of the histopathologists are referred to as A-1, A-2, B-1, B-2, C-1, C-2, respectively.

RESULTS

We concentrate here upon those 100 biopsies (slide pairs) from Malawi leprosy "suspects" which were examined twice by each histopathologist. Ninety-nine biopsies were from skin and one was from nerve. Ninety-eight of the Malawi biopsy slide pairs were considered satisfactory in all six examinations. The other two were considered suitable for evaluation by all histopathologists at least once.

The frequency distributions of diagnostic codes reported for all 100 multiply-examined Malawi biopsies are given in Table 2. Considered overall, histopathologist B had the greatest, and histopathologist C the least, propensity to find evidence of leprosy in the biopsies. Each of the three histopathologists reported fewer definite leprosy diagnoses (1 = "leprosy confirmed beyond reasonable doubt") in the second than in the first round, but this shift away from certainty was appreciable only for histopathologist A, who considered 55 biopsies confirmatory in Round 1, but only 42 in the second round. Histopathologists B and C decreased their numbers of definite diagnoses between Rounds 1 and 2 from 55 to 52 and from 30 to 29, respectively. There is also evidence of increased caution on the part of histopathologist C insofar as he used code 6 (pathological and indicative of a specific disease other than leprosy) less often, and code 4 (pathological but nonspecific) more often, in the second than in the first rounds. Diagnosis code 7 ("unforeseen circumstances") was never used, and code 5 ("normal tissue") was used for only 3 (0.5%) of the

TABLE 3. *Examples of agreement within and between histopathologists in diagnostic certainty scores.*[a]

Within A	A-2				
	1	2	3	4+	Total
1	41	6	4	4	55
2	0	2	3	6	11
A-1 3	0	1	1	2	4
4+	1	1	7	21	30
Total	42	10	15	33	100

Within B	B-2				
	1	2	3	4+	Total
1	50	5	0	0	55
2	1	8	3	0	12
B-1 3	0	2	8	2	12
4+	1	0	2	18	21
Total	52	15	13	20	100

Within C	C-2				
	1	2	3	4+	Total
1	26	3	1	0	30
2	2	4	1	0	7
C-1 3	1	1	4	5	11
4+	0	0	6	46	52
Total	29	8	12	51	100

Between A and B	B-1				
	1	2	3	4+	Total
1	47	6	1	1	55
2	2	2	4	3	11
A-1 3	1	0	1	2	4
4+	5	4	6	15	30
Total	55	12	12	21	100

Between A and C	C-1				
	1	2	3	4+	Total
1	30	6	7	12	55
2	0	0	2	9	11
A-1 3	0	1	0	3	4
4+	0	0	2	28	30
Total	30	7	11	52	100

Between B and C	C-1				
	1	2	3	4+	Total
1	30	7	6	12	55
2	0	0	3	9	12
B-1 3	0	0	1	11	12
4+	0	0	1	20	21
Total	30	7	11	52	100

[a] A-1, A-2, B-1, B-2, C-1, C-2 refer to first and second round results for histopathologists A, B, and C, respectively. Two biopsies not given certainty scores because they were considered unsatisfactory (see Table 2) are included here in the 4+ category (no evidence of leprosy).

600 reports. Since codes 4, 5, 6, 7 (and blank, if specimen unsatisfactory) all imply no evidence of leprosy, they are aggregated together as "4+" in subsequent analyses and tables.

In order to evaluate the nature and extent of the variation in diagnostic certainty interpretations, we examined the variation in certainty scores given by each histopathologist between the first and second round, and between each pair of histopathologists (A and B, A and C, B and C) within the first and within the second round. Table 3 illustrates the observed distributions, which are summarized in Table 4 in terms of "close agreement" (counting 1, 2, 3 and 4+ as separate codes) or as "general agreement" (in which adjacent codes 1 and 2, 2 and 3, 3 and 4+ are also considered as agreements). Not surprisingly, agreement within observers is higher (K = 0.49 to 0.75 for close agreement, K = 0.69 to 0.98 for general agreement) than that between observers (K = 0.31 to 0.52, K = 0.44 to 0.71 for close and general agreement, respectively). Histopathologist A changed his diagnostic score by at least one for 35 biopsies, and by at least two for 16 biopsies. B changed by at least one for 16 biopsies and at least two for only 1 biopsy. C changed by at least one for 20 and by at least two for 2 biopsies, respectively. On the other hand, agreement between histopathologists was higher between A and B, and between A and C, than between B and C. This was because B was most inclined to find definite evidence of leprosy; whereas C was the least so inclined. Thus, the histopathologist who appeared to

TABLE 4. *Summary of agreement in diagnostic certainty score within and between histopathologists.*[a]

Comparison	Close agreement		General agreement	
	Proportion	Kappa	Proportion	Kappa
A-1–A-2	65	0.46	84	0.66
B-1–B-2	84	0.75	99	0.98
C-1–C-2	80	0.68	98	0.96
A-1–B-1	65	0.43	85	0.65
A-1–C-1	58	0.37	72	0.47
B-1–C-1	51	0.30	73	0.48
A-2–B-2	67	0.52	86	0.69
A-2–C-2	63	0.46	85	0.69
B-2–C-2	51	0.32	72	0.46

[a] Close agreement requires precise agreement on certainty score (1, 2, 3 or 4+); general agreement accepts 1 and 2, 2 and 3, 3 and 4+ as agreements. A-1, A-2, B-1, B-2, C-1, C-2 refer to first and second round results for histopathologists A, B, and C, respectively.

be least consistent internally appeared to be the most consistent externally.

The relationship between internal ("within") and external ("between") disagreements for the three observers is shown in another manner in Table 5. This table shows the numbers of biopsies (out of the total 100) for which each histopathologist changed his diagnostic score between the first and second rounds (diagonal in table), or for which each histopathologist differed from either of his colleagues (within Round 1 in upper right of table, within Round 2 in lower left of table). Table 5A enumerates all disagreements; 5B shows disagreements by at least 2 on the certainty code scale. For histopathologist A, there was little difference between the numbers of internal (with himself) and external (with his colleagues) disagreements; histopathologists B and C disagreed with their colleagues two to three times as often as they disagreed with themselves.

Acid-fast bacilli (AFB) were reported on at least 1 examination in 40 of the 100 biopsies. On only six biopsies were AFB reported in all six examinations. The frequency distribution of the number of AFB-positive examinations per biopsy is shown in Figure 1. The numbers of biopsies in which AFB were reported varied by examiner and by round (shown in Table 6), ranging from 9 (histopathologist A, Round 2) to 33 (histopathologist B, Round 1).

TABLE 5. *Distribution of disagreements in diagnostic certainty score within and between histopathologists.*[a]

(a) DISAGREEMENT BY AT LEAST ONE

	A	B	C
A	35	35[b]	42[b]
B	33[c]	16	49[b]
C	37[c]	49[c]	20

(b) DISAGREEMENT BY AT LEAST TWO

	A	B	C
A	16	15[b]	28[b]
B	14[c]	1	27[b]
C	15[c]	28[c]	2

[a] Figures represent numbers of disagreements, out of 100 biopsies, between specified pairs of histopathologists. Shaded diagonal cells show numbers of times each histopathologist changed his diagnostic scale by at least 1 (a) or 2 (b) between the first and second round.
[b] First round.
[c] Second round.

The distribution of classification codes used by the histopathologists for biopsies considered confirmed beyond reasonable doubt is shown in Table 7. Most of the leprosy in this series was judged to be border-

TABLE 6. *Numbers of biopsies (slide pairs) in which acid-fast bacilli (AFB) were reported, by histopathologist and by round.*

Histopathologist	No. biopsies reported to have AFB	
	Round 1	Round 2
A	14	9
B	13	21
C	14	15

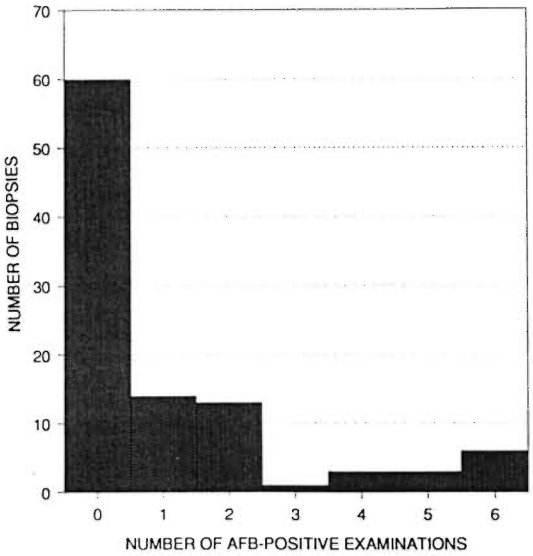

FIG. 1. Frequency distribution of the number of examinations in which AFB were reported, per biopsy. Each biopsy was examined six times (twice by each histopathologist).

line tuberculoid, but there were clear differences in the classifications used by the three observers. A used only four categories and was the most willing to employ the indeterminate classification (for 22 biopsies in the first round, for 6 in the second). B and C used 10 and 7 categories, respectively, but used the indeterminate category infrequently. Histopathologist C was most inclined to use the TT category. Most inconsistencies were minor. A classification divergence from BT to BL arose with only one biopsy, from a patient in reversal reaction with bacilli in a nerve. It is of interest that all three histopathologists indicated difficulty in deciding between BT and BL in this instance. Of the 28 individuals considered to have "clinical leprosy" (deserving antileprosy treatment on clinical grounds alone, i.e., clinical certainty 4 or 5), 21 (75%) were classified as BT and 2 as polar tuberculoid on clinical grounds alone.

For 46 biopsies there were appreciable differences of diagnostic opinion, with histopathological diagnostic certainty codes ranging from 1 to 3 or 1 to 4+ or 2 to 4+. Sixteen of these individuals were started on antileprosy treatment at the time of the initial biopsy. Twenty-seven of the remaining individuals were traced 2 to 3 years after the biopsy was taken. In 18 (including 11 whose initial biopsy had been considered to show definite evidence of leprosy by at least one histopathologist) the lesions had totally disappeared. Of 9 individuals with persistent lesions, 3 were later registered as leprosy patients.

TABLE 7. *Distribution of classification codes used by three histopathologists in this study for biopsies in which leprosy was considered confirmed beyond reasonable doubt.*[a]

Classification	A-1	A-2	B-1	B-2	C-1	C-2
IND	22	6	4	2	3	3
TT	5	4	1	2	8	10
TT/BT	—	—	3	1	—	1
BT	26	31	39	41	15	12
IND/BT	—	—	3	—	—	—
BT/BB	—	—	—	3	1	—
BT/BL	—	—	—	—	1	1
BB	—	—	1	—	—	—
BB/BL	—	—	2	2	—	—
BL	2	1	1	1	2	2
BL/LL	—	—	1	—	—	—
LL	—	—	—	—	—	—
	55	42	55	52	30	29

[a] A-1, A-2, B-1, B-2, C-1, C-2 represent first and second results of histopathologists A, B and C, respectively.

Comparisons of diagnostic certainty between the clinician and the histopathologists were based on a schema (Table 8) which defines various combinations of clinical and

TABLE 8. *Schema for comparing clinical and histopathological diagnostic certainty.*

Clinical certainty	Biopsy certainty				Totals
	1	2	3	4	
Certain	I				7
Extremely likely				VI	21
Most likely		II			19
Consider seriously					29
Possibly	V		III		19
Other (not leprosy)				IV	5

I = Agreement on definite or very strong evidence of leprosy; **II** = Agreement on moderate evidence of leprosy; **III** = Agreement on slight evidence of leprosy; **IV** = Agreement on no evidence of leprosy; **V** = Relative clinical underdiagnosis (or histopathological overdiagnosis); **VI** = Relative clinical overdiagnosis (or histopathological underdiagnosis); (**I** + **II** + **III** + **IV**) = "Overall agreement"; (**V** + **VI**) = "Overall disagreement."

TABLE 9. *Comparison of clinical and histopathological diagnosis according to the schema of Table 8.*

	Clinical certainty compared with					
	A-1	A-2	B-1	B-2	C-1	C-2
I (Agreement on definitely leprosy)	25	21	24	23	19	18
II (Agreement on moderate evidence of leprosy)	2	4	4	5	6	6
III (Agreement on slight evidence of leprosy)	2	11	8	9	7	7
IV (Agreement on no evidence of leprosy)	14	12	8	8	19	19
"Overall agreement"	43	48	44	45	51	50
V (Relative clinical underdiagnosis or histopathological overdiagnosis)	39	27	40	40	12	14
VI (Relative clinical overdiagnosis or histopathological underdiagnosis)	18	25	16	15	37	36
"Overall disagreement"	57	52	56	55	49	50

histopathological results as representing either agreement, or else disagreement in the direction either of clinical underdiagnosis (or histopathological overdiagnosis), or of clinical overdiagnosis (or histopathological underdiagnosis). Observed combinations are shown in Table 9, revealing that each of the histopathologists agreed with the clinical diagnosis on approximately 50% of the biopsies, and that the differences lay primarily in the direction of histopathological overdiagnosis (or clinical underdiagnosis) with histopathologist B and histopathological underdiagnosis (or clinical overdiagnosis) with C. Considering only the 28 biopsies from individuals with strong clinical evidence of leprosy (clinical certainty 4 or 5), the range of agreement in definite leprosy (histopathological certainty S-1) ranged from 18 (64%, histopathologist C in Round 2) to 25 (89%, histopathologist A in Round 1).

Of the U.K. biopsies, one was judged to show some evidence of leprosy (S-2) on one examination by one histopathologist. A second, which was considered to show evidence of leprosy by one pathologist, was considered unsatisfactory for diagnosis by the other two.

DISCUSSION

The results reported here document further the difficulty of arriving at an accurate diagnosis of lesions suspected to be due to leprosy, and provide an insight into the problems experienced by histopathologists involved in the diagnostic process. That such problems exist has been demonstrated in previous studies ([5]). Thus, the essential task is to understand how they arise and what might be done to minimize them.

We recognize at the outset that the precise numbers which arose in this investigation are a function of the specific clinician, biopsies and histopathologists involved. Different combinations of observers and material would have produced different results. We thus concentrate upon patterns and implications rather than on specific numerical results.

In the present series of 100 multiply and repeatedly examined biopsies, the proportion considered to show definite (certainty code = 1 = "leprosy confirmed beyond reasonable doubt") evidence of leprosy ranged from 29 to 55 among the three histopathologists (Table 2). The 100 biopsies included 28 from individuals with "clinical leprosy" (in the sense of deserving antileprosy treatment on clinical grounds alone, i.e., clinical certainty categories 4 or 5), 67 from individuals with suspected or possible leprosy (clinical certainty categories 1, 2 or 3), and 5 from individuals with no clinical suspicion of leprosy (clinical certainty = 0). Most of these were actively detected in the course of population surveys, and thus might be assumed to have posed a more difficult challenge to the histopathologists than biopsies taken from individuals with more advanced disease ([16]). On the other hand, it is

worth nothing that the almost twofold range of definite diagnoses, from 29% to 55%, is based on only three histopathologists, each with many years of experience. It is inevitable that a broader sample of histopathologists would further increase this range of results.

What is the origin and reason for the observed differences? Comparison of the agreement within and between observers provides evidence both of difficulties in arriving at a decision (as reflected in changes made by individual histopathologists between the first and second blind reading of the slides) and of consistent differences of opinion (as reflected in differences between observers above and beyond that attributable to internal inconsistency). Our effort to explore the relative contributions of internal and external inconsistency revealed several things. First, the number of changes of opinion was significantly ($p < 0.05$) greater for one histopathologist in our investigation (A) than for the other two (Tables 2, 3, 4, 5). This appeared to reflect not a random process but a systematic shift of diagnostic opinion between the first and second rounds in the direction of greater caution on the part of histopathologist A. Thus, the number of biopsies he reported as showing definite evidence of leprosy fell from 55 in Round 1 to 42 in Round 2 (Table 2), and the number of biopsies classified as indeterminate fell from 22 to 7 (Table 7). As a consequence of this shift, the internal ("within observer") consistency of histopathologist A was of a similar magnitude to his external consistency with the other two participants in the study. In contrast, the other two histopathologists (B, C) were far more consistent internally than was A, and more consistent internally than they were with each other. Only three times did their second evaluation of a pair of slides differ from their first by more than 1 level on the scale from 1 to 4+ (Tables 3, 4, 5). On the other hand, although both B and C showed little tendency for internal "within observer" disagreement, B consistently reported more definite leprosy than did the other two histopathologists, and C reported the least, with the effect that agreement was worst between these two workers. We thus have an example of two observers working very consistently but within different paradigms.

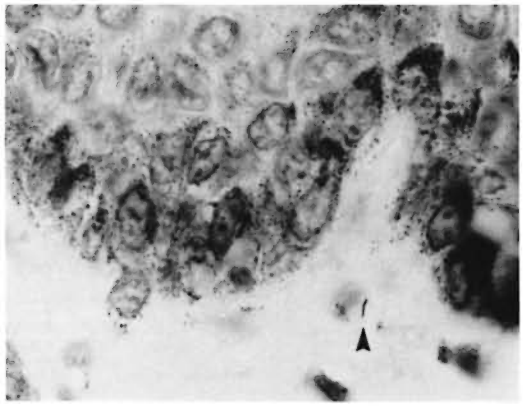

FIG. 2. A single AFB in the subepidermal zone (arrowhead) found by two pathologists, once each (Wade-Fite ×1000).

By chance our investigation thus included three histopathologists with rather different responses to the diagnostic challenge: one whose opinions shifted during the period of the study, the two whose opinions remained consistent but who differed in their readiness to consider a biopsy as showing evidence of leprosy. Inevitably these differences reflect personalities, training, experience and the context of this particular investigation. It was also of interest to note evidence for increased caution (fewer certain S-1 diagnoses by all three histopathologists, greater use of the nonspecific category by C) in the second round diagnoses, which may reflect a natural tendency when a decision is to be compared with one's own earlier conclusion as well as with that of others.

The reasons for these differences were explored in a meeting of the participants at the close of the investigation. None of the histopathologists considered that the differences observed in this investigation were attributable to the reporting protocol (Table 1) *per se.* Although more than 2 years were required for the complete repeat examination of all biopsies, there was no evidence that the staining had faded enough to influence the histopathologists. The long interval further insured against the possibility of their remembering particular sections from one round to the next. The differences are thus attributable to real difficulties in the examination and interpretation of leprosy biopsies.

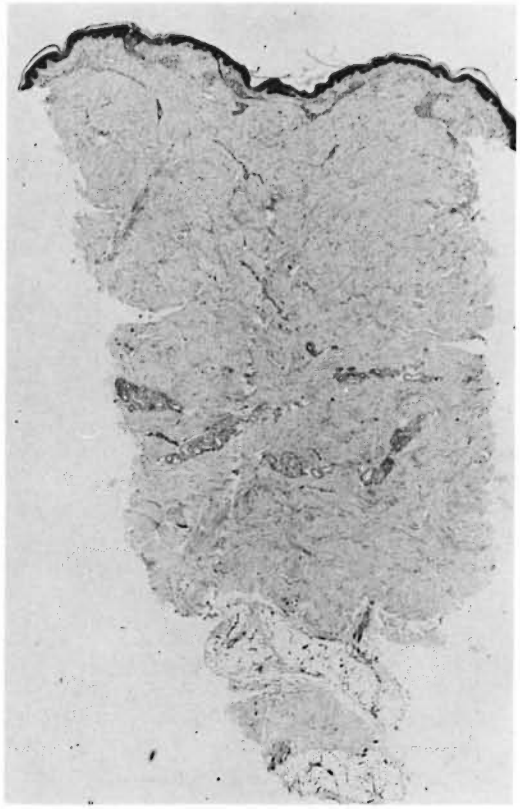

FIG. 3. Punch biopsy of histopathologically confirmed leprosy (same case as Fig. 2) indicating very small amount of upper dermal inflammation; deep nerves were normal (H&E ×20).

The histological evaluation of such biopsies depends upon the detection of AFB and the assessment of patterns of inflammation. The difficulty in finding AFB is evident in Table 6 and Figure 1. Out of 40 biopsies in which AFB were reported at least once, in 15 they were seen in only one of the six examinations, and in 11 biopsies the AFB were seen in only two examinations. This indicates that the amount of time which a histopathologist can afford to spend in examining each biopsy will have a profound influence upon the proportion in which leprosy is diagnosed. It was noted that in several of the cases included in this study, the diagnosis of histologically confirmed leprosy depended entirely upon the finding of one or more AFB.

If AFB are difficult to find (Fig. 2) but easy to interpret, the converse is true of patterns of inflammation. The great majority of the biopsies included in this study series

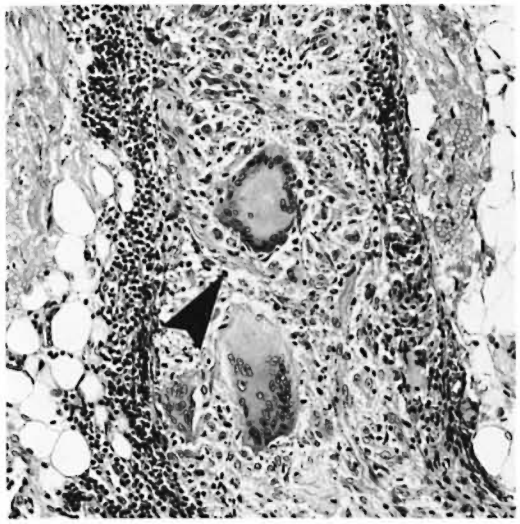

FIG. 4. Unambiguous disruption of deep dermal nerve (arrowhead) by giant cell granulomas (H&E ×200).

had a granuloma fraction of less than 0.1 (i.e., less than 10% of the dermis involved with inflammatory cells) ([15]). This is usual in cases of leprosy sought by active surveillance in Africa, where most cases are paucibacillary ([10]). A typical example of such a biopsy is shown in Figure 3. Discussion of biopsies over which there was disagreement, in the absence of AFB, centered upon the following issues: a) granulomatous infiltration of nerve (whether or not a nerve was disrupted or invaded, Fig. 4); b) intraneural lymphocytic infiltration (evaluation of the extent and significance of lymphocytes within the endoneurium of dermal nerves, Figs. 5, 6); c) thickening of the perineurium (whether or not there was a genuine increase in perineurial cells, Fig. 7); d) Schwann cell hypercellularity (whether or not there was a genuine increase in the number and size of Schwann cell nuclei); e) perineural inflammation (evaluation of the extent and significance of mononuclear cells around a nerve); f) dermal fibroplasia (evaluation of the extent and significance of diffuse fibroblast proliferation in the dermis); and g) epidermal "changes" (for example, the compatibility of epidermal acanthosis with leprosy, Fig. 8). Although the establishment of objective criteria for such patterns remains an elusive goal, all of the histopathologists in this investigation agreed that it would be helpful if more were known

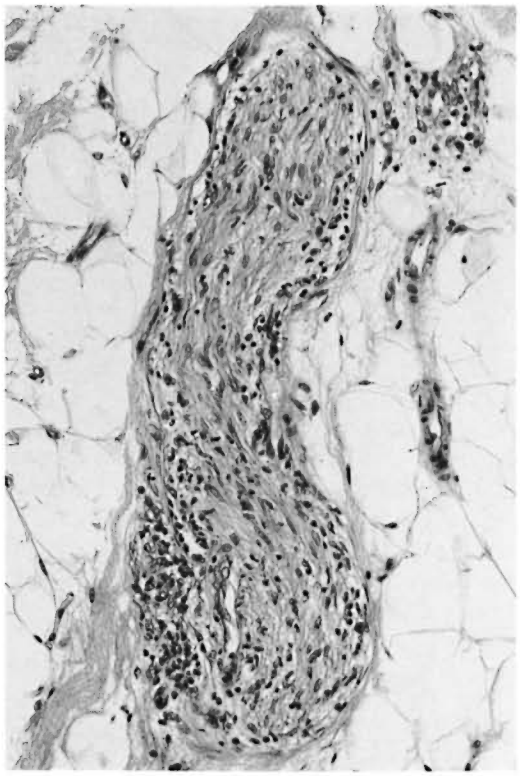

FIG. 5. Deep dermal nerve with intraneural infiltration by macrophages and lymphocytes; all three pathologists agreed this indicated leprosy (H&E ×200).

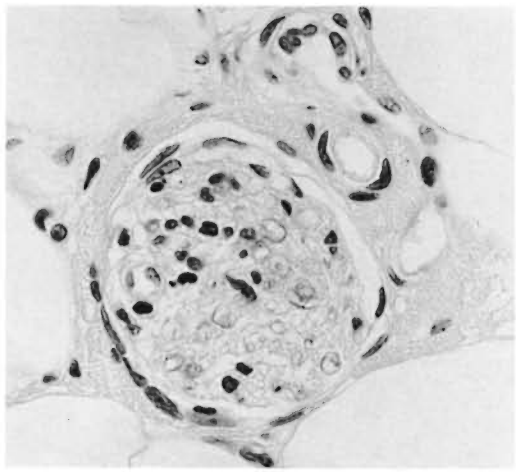

FIG. 6. Deep dermal nerve with intraneural lymphocytes; while histopathologically suggestive only, clinically this case was considered to be leprosy (H&E ×400).

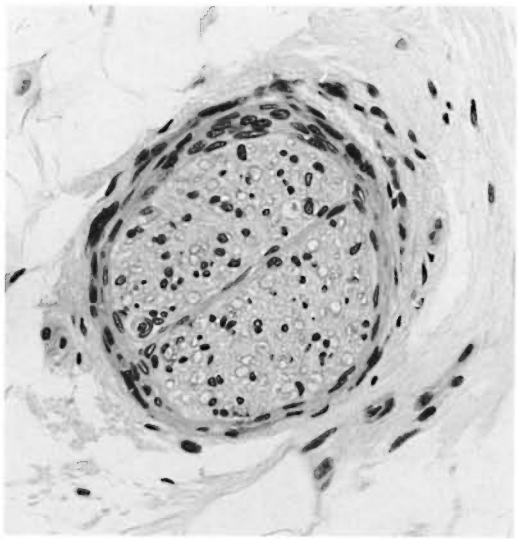

FIG. 7. Deep dermal nerve with normal endoneurium but a prominent perineurium; the three pathologists were divided as to its significance for the diagnosis of leprosy (H&E ×300).

of the nerve-related histopathology of nonleprosy conditions.

It was evident that there were considerable differences of opinion among the participants in this study over the use and implications of the term indeterminate. For some, the term was restricted to biopsies in which a small number of AFB was found in association with nongranulomatous inflammation; whereas another was willing to make a diagnosis of indeterminate leprosy on the basis of pathological changes, such as those discussed above, without detecting AFB. It is clear that this term still lacks a generally agreed upon definition.

In conclusion, we would note the severity of the challenge put to the histopathologists in this investigation. The differences and difficulties revealed must be interpreted in the context of the investigation as a whole. Never have leprosy clinicians been put to so stern a test! Histopathologists involved in the diagnosis of leprosy may wish to consider the implications of these findings for the emphasis placed upon finding of AFB, and for future studies to define better the nerve-related pathology in nonleprosy conditions. Difficulties notwithstanding, the contribution of histopathology to the diagnosis of leprosy remains unchallenged.

SUMMARY

Identical slides from 100 biopsies obtained from individuals suspected of having

FIG. 8. Acanthotic epidermis with interface dermatitis. The lesion was considered to be leprosy by two of the three pathologists; the third considered that this epidermal lesion argued strongly against leprosy. Clinically, the lesion was not thought to be leprosy, and disappeared spontaneously (H&E ×100).

leprosy, ascertained in a total population survey in Malawi, were examined twice, independently, by three histopathologists. Results were reported in a standard protocol, and were compared among themselves and with a standardized clinical assessment of each "suspect." The proportion of biopsies considered to show definite evidence of leprosy ranged from 29 to 55 among the six evaluations (twice by each of three histopathologists). Comparisons of variations within and between histopathologists revealed three different patterns. Two of the pathologists were very consistent as individuals, but differed markedly between themselves in that one was the least inclined and the other the most inclined to report definite evidence of leprosy. The third pathologist was less consistent, reporting appreciably more definite leprosy on the first than on the second examination of the same biopsies. Although acid-fast bacilli (AFB) were reported on at least 1 examination in 40 of the biopsies, they were observed in all six examinations of only six of the biopsies. There was greater agreement regarding classification than regarding diagnosis, except with reference to the indeterminate category which was employed more frequently by one histopathologist than by the other two. A workshop of participants at the end of the investigation highlighted several reasons for the variations observed. The fact that AFB were reported in only nine biopsies by one histopathologist but in 33 by another reveals the importance of the examination method and time in arriving at a diagnosis of leprosy. The differences in the interpretation of cellular evidence of inflammation revealed the need for further studies of nerve-related pathology in nonleprosy conditions to serve as a reference against which to judge possible evidence of leprosy *per se*.

RESUMEN

Tres histopatólogos examinaron, en 2 ocasiones diferentes, laminillas idénticas preparadas a partir de 100 biopsias tomadas de individuos bajo sospecha de tener lepra. El estudio se hizo en Malawi. Los resultados se registraron en un protocolo estándar y se compararon entre sí y con los datos clínicos correspondientes a cada individuo bajo sospecha. La proporción de biopsias consideradas para decidir la existencia de evidencias definitivas de la lepra varió de 29 a 55 entre las 6 evaluaciones (dos por cada uno de los 3 histopatólogos). Los histopatólogos fueron individualmente muy consistentes pero difirieron marcadamente entre ellos. Mientras que uno de ellos fue el más inclinado a reportar evidencias definitivas de la lepra, otro fue el menos inclinado, en tanto que el tercero fue menos consistente, reportando apreciablemente más lepra definitiva en el primer examen que en el segundo examen de la misma biopsia. Aunque los bacilos ácido resistentes (AFB) fueron reportados en cuando menos 1 examen, en 40 de las biopsias, éstos fueron observados en los 6 examenes de sólo 6 de las biopsias. Hubo más concordancia en relación a la clasificación que al diagnóstico, excepto que la categoría indeterminada fue empleada más frecuentemente por un histopatólogo que por los otros dos. La discusión conjunta de los participantes al final de la investigación apuntó varias razones de las discrepancias observadas. El hecho que los AFB fueran reportados en solo 9 biopsias por un histopatólogo y en 33 por otro, revela la importancia del método de examen y el tiempo requerido para llegar al diagnóstico de lepra. Las diferencias en la interpre-

tación de las evidencias celulares de la inflamación, revelaron la necesidad de ampliar los estudios sobre la patología relacionada a nervios en condiciones no leprosas, con el fin de tomar los hallazgos como referencia contra la cual se pueda hacer el estudio diferencial de la lepra.

RÉSUMÉ

Des préparations identiques en provenance de 100 biopsies obtuenues chez des personnes suspectes de lèpre, découvertes lors d'une enquête dans une population entière au Malawi ont été examinées deux fois de manière indépendante par trois pathologistes. Les résultats ont été rapportés selon un protocole standard et ont été comparés entre eux et avec une évaluation clinique standardisée de chaque suspect. La proportion des biopsies considérées comme montrant des signes clairs de lèpre s'étalait de 29 à 55 pour les six évaluations (deux fois par chacun des trois histopathologistes). La comparaison des variations intra- et inter-histopathologistes a révélé trois tendances différentes. Deux des pathologistes étaient très cohérents avec eux-mêmes, mais différaient de manière marquée entre eux en ce sens que l'un était le moins enclin et l'autre le plus enclin à rapporter des signes clairs de lèpre. Le troisième pathologiste était moins cohérent, rapportant nettement plus de cas de lèpre lors du premier que lors du second examen des mêmes biopsies. Bien que des bacilles acido-résistants (BAR) aient été rapportés pour au moins un examen dans quarante des biopsies, ceux-ci ont été observés pour seulement six biopsies lors des six examens. Il y avait un meilleur accord concernant la classification que concernant le diagnostic, sauf pour la catégorie indéterminée, qui était utilisée plus fréquemment par un pathologiste que par les deux autres. Une réunion de travail des participants, tenue à la fin de la recherche, a mis en humière différentes raisons pour les variations observées. Le fait que des BAR aient été rapportés pour seulement neuf biopsies par un histopathologiste mais pour trent-trois par un autre révèle l'importance de la méthode d'examen et du facteur temps avant d'arriver à un diagnostic de lèpre. Les différences dans l'interprétation des signes cellulaires d'inflammation ont révélé le besoin pour des études supplémentaires des pathologies non-lépreuses en rapport avec les nerfs pour servir de pointe de comparaison afin de juger des signes possibles de lèpre.

Acknowledgment. The LEPRA Evaluation Project/Karonga Prevention Trial (LEP/KPT) has been supported primarily by the British Leprosy Relief Association (LEPRA) with assistance from the International Federation of Anti-Leprosy Organizations (ILEP) and the WHO/UNDP/World Bank Special Programme for Research and Training in Tropical Diseases. We thank The Heiser Program for a special grant to support the workshop meeting for participants in this investigation, Dr. T. J. Ryan for providing the nonleprosy control biopsies, Drs. Ryan and P. McKee for participating in the workshop and providing helpful comments on the project, and Helen Berrett for assistance in organizing the workshop and typing the manuscript.

REFERENCES

1. Bagshawe, A., Scott, G. C., Russell, D. A., Wigley, S. C., Merianos, A. and Berry, G. BCG vaccination in leprosy: final results of the trial in Karimu, Papua New Guinea, 1963–1979. Bull. WHO **67** (1989) 389–399.
2. Boerrigter, G., Ponninghaus, J. M., Fine, P. E. M. and Wilson, R. J. Four year follow-up results of a WHO recommended multiple drug regimen in paucibacillary leprosy patients in Malawi. Int. J. Lepr. **59** (1991) 255–261.
3. Copeland, K. T., Checkoway, H., McMichael, A. J. and Holbrook, R. H. Bias due to misclassification in the estimation of relative risk. Am. J. Epidemiol. **105** (1977) 488–495.
4. Cree, I. A., Srinivasan, T., Krishvian, S. A. R., Gardiner, C. A., Melita, J., Fisher, C. A. H. and Beck, J. S. Reproducibility of histology in leprosy lesions. Int. J. Lepr. **56** (1988) 296–301.
5. Fine, P. E. M., Job, C. K., McDougall, A. C., Meyers, W. M. and Ponnighaus, J. M. Comparability among histopathologists in the diagnosis and classification of lesions suspected of leprosy in Malawi. Int. J. Lepr. **54** (1986) 614–625.
6. Fleiss, J. L. *Statistical Methods for Rates and Proportions.* London: John Wiley & Sons, 1973.
7. Gupte, M. D., Villishayee, R. X., Nagaraju, B., Ramalingam, A., Lourdusamy, G. and Kannan, S. Inter-observer agreement and clinical diagnosis of leprosy for prophylaxis studies. Indian J. Lepr. **62** (1990) 281–295.
8. Heenan, P. J., Matz, L. R., Blackwell, J. B., Kelsall, G. R. H., Siagh, A., Ten Seldam, R. E. J. and Holman, C. D. J. Inter-observer variation between pathologists in the classification of cutaneous malignant melanoma in Western Australia. Histopathology **8** (1984) 717–729.
9. Jones, R. L. and Ponnighaus, J. M. The transport of histopathological specimens by airmail. Lepr. Rev. **53** (1982) 67–68.
10. 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. **55** (1987) 88–98.
11. Pattyn, S. R., Groenen, G., Janssens, L., Kugkens, L., Mputu, L. B. and The Collaborative Study Group for the Treatment of Leprosy in Zaire. A controlled therapeutic trial in paucibacillary leprosy comparing a single dose of rifampicin with a single dose of rifampicin followed by one year of daily dapsone. Lepr. Rev. **62** (1991) 179–185.
12. Ponnighaus, J. M. and Fine, P. E. M. Sensitivity and specificity of the diagnosis and the search for risk factors for leprosy. Trans. R. Soc. Med. Hyg. **82** (1988) 803–809.

13. PONNIGHAUS, J. M., FINE, P. E. M. and BLISS, L. Certainty levels in the diagnosis of leprosy. Int. J. Lepr. **55** (1987) 454–462.
14. PONNIGHAUS, J. M., FINE, P. E. M., BLISS, L., SLINEY, I. J., BRADLEY, D. J. and REES, R. J. W. The Leprosy Evaluation Project, an epidemiological study of leprosy in Northern Malawi. I. Methods. Lepr. Rev. **58** (1987) 359–375.
15. RIDLEY, D. S. *Skin Biopsy in Leprosy.* 2nd edn. Documenta Geigy. Basel: CIBA-GEIGY, 1985.
16. RIDLEY, D. S. and LUCAS, S. B. The use of histopathology in leprosy diagnosis and research. Lepr. Rev. **60** (1989) 257–262.
17. SILKOCKS, P. B. S. Measuring repeatability and validity of histological diagnosis—a brief review with some practical example. J. Clin. Pathol. **36** (1983) 1269–1275.