

## Prospective Evaluation of Criteria for Microbiological Diagnosis of Prosthetic-Joint Infection at Revision Arthroplasty

BRIDGET L. ATKINS,<sup>1,2</sup> NICHOLAS ATHANASOU,<sup>3,4</sup> JONATHAN J. DEEKS,<sup>5</sup> DERRICK W. M. CROOK,<sup>2</sup> HAMISH SIMPSON,<sup>4,6</sup> TIMOTHY E. A. PETO,<sup>2</sup> PETER McLARDY-SMITH,<sup>4</sup> ANTHONY R. BERENDT,<sup>2,4\*</sup> AND THE OSIRIS COLLABORATIVE STUDY GROUP†

*Public Health Laboratory,<sup>1</sup> Department of Pathology,<sup>3</sup> Nuffield Department of Orthopaedic Surgery,<sup>6</sup> and Bone Infection Unit,<sup>4</sup> Nuffield Orthopaedic Centre, Academic Unit of Microbiology and Infectious Diseases, John Radcliffe Hospital,<sup>2</sup> and Centre for Statistics in Medicine, Institute of Health Sciences,<sup>5</sup> Oxford, United Kingdom*

Received 6 February 1998/Returned for modification 30 March 1998/Accepted 30 June 1998

**A prospective study was performed to establish criteria for the microbiological diagnosis of prosthetic joint infection at elective revision arthroplasty. Patients were treated in a multidisciplinary unit dedicated to the management and study of musculoskeletal infection. Standard multiple samples of periprosthetic tissue were obtained at surgery, Gram stained, and cultured by direct and enrichment methods. With reference to histology as the criterion standard, sensitivities, specificities, and likelihood ratios (LRs) were calculated by using different cutoffs for the diagnosis of infection. We performed revisions on 334 patients over a 17-month period, of whom 297 were evaluable. The remaining 37 were excluded because histology results were unavailable or could not be interpreted due to underlying inflammatory joint disease. There were 41 infections, with only 65% of all samples sent from infected patients being culture positive, suggesting low numbers of bacteria in the samples taken. The isolation of an indistinguishable microorganism from three or more independent specimens was highly predictive of infection (sensitivity, 65%; specificity, 99.6%; LR, 168.6), while Gram staining was less useful (sensitivity, 12%; specificity, 98%; LR, 10). A simple mathematical model was developed to predict the performance of the diagnostic test. We recommend that five or six specimens be sent, that the cutoff for a definite diagnosis of infection be three or more operative specimens that yield an indistinguishable organism, and that because of its low level of sensitivity, Gram staining should be abandoned as a diagnostic tool at elective revision arthroplasty.**

The bacteriological diagnosis of infection generally depends upon the isolation of a recognized pathogen from a clinical specimen, the nature and quality of which affect the validity and utility of the culture results. Specimens taken from sites with colonizing flora, such as the throat or surgical wounds, are of less diagnostic value than those obtained from normally sterile sites such as joint spaces, the pleural cavity, the cerebrospinal fluid, the blood, or deep tissue. However these “sterile-site” specimens can be obtained only by puncturing the skin, a tissue which is heavily colonized with microorganisms. Samples must then be transported to the diagnostic laboratory for processing and culture.

The pathogens responsible for primary deep-space infections are usually distinct from normal commensal organisms on the skin, but it is precisely these commensal organisms on the skin that most commonly infect implantable biomedical devices (34). In this situation it is usually impossible to decide, simply from the identity of the organism, whether it is clinically significant or a contaminant derived from the skin of the patient, the medical staff obtaining the sample, or the laboratory staff processing it.

These issues are of paramount importance in prosthetic joint infection (PJI). The development of prosthetic joints has been one of the biomedical success stories of the century, with major health-economic and quality-of-life benefits. As a consequence, the number of joint replacements performed has increased each year, with approximately 200,000 primary hip and knee replacements performed in the United States and 75,000 performed in the United Kingdom in 1995. Despite improvements in operative techniques, a proportion of prostheses fail, and infection is one of the most serious causes of this.

PJIs have been classified as “early” and “late,” although to these categories have been added the classifications of “acute hematogenous” and “positive intraoperative cultures,” the latter indicating a group with positive cultures without prior suspicion of infections (36). Early infections present acutely with overt wound infection; late infections present with worsening pain in the joint accompanied by loosening of the prosthesis at the bone-cement interface and sometimes by sinus tract formation with chronic discharge. It is usually necessary to remove the prosthesis and administer antibiotics, either locally or systemically, to eradicate the infection; this is commonly followed by reimplantation of another prosthesis to restore function. There is controversy over the best surgical strategy, with some advocating a one-stage exchange and others recommending a two-stage procedure (for reviews, see references 12 and 34). Other important questions concern the necessary duration and route of administration of antibiotics in the management of acute and chronic infections. Answers to such questions will require multicenter studies, an essential prerequisite for which is an accurate and robust means of diagnosing infection.

This is of particular importance because infection is not the

\* Corresponding author. Mailing address: Bone Infection Unit, Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford OX3 7LD, United Kingdom. Phone: 00 44 1865 227779. Fax: 00 44 1865 227740. E-mail: tony.berendt@noc.anglo.nhs.uk.

† The members of the OSIRIS (Oxford Skeletal Infection Research and Intervention Service) Collaboration Study Group are M. Benson, A. Carr, D. Collopy, P. Cooke, J. Kenwright, P. McClardy-Smith, H. Simpson, R. de Steiger, R. Gundle, K. Willett, B. Atkins, A. Berendt, I. Bowler, C. Conlon, D. Crook, A. Emptage, T. Peto, J. Deeks, and N. Athanasou.

only cause of prosthetic loosening, and it may be difficult to distinguish septic from aseptic loosening either pre- or peroperatively. The preoperative diagnosis of PJI relies on a combination of clinical history, examination, and investigations including erythrocyte sedimentation rate, C-reactive protein analysis, plain radiography, isotope scans, and the microscopy and culture of joint aspirates. Despite numerous studies, none of these diagnostic methods has achieved satisfactory specificity or sensitivity in common practice (11, 12, 14, 24). This may in part be because of a lack of standardization in the criterion standard(s) used to diagnose infection when tests are evaluated in studies or in day-to-day practice. The presence of a sinus tract or of intra-articular pus at operation with organisms visible on Gram staining are specific criteria for defining PJI but are very insensitive (2). Microbiological criteria and culture methods are nonstandardized; histological appearances may be helpful, with a number of studies reporting that the presence of acute inflammation (more than five neutrophils per high-power field) in periprosthetic soft-tissue specimens is specific for infection, provided that the patient does not have inflammatory joint disease (2, 10, 22, 24).

In 1995 we established a multidisciplinary team and a dedicated unit for the management and study of musculoskeletal infection. With the long-term aim of improving the preoperative diagnosis of PJI and of participating in multicenter clinical trials, we have attempted to define reliable criteria for the microbiological diagnosis of infection at revision arthroplasty. We have performed a prospective study to assess the significance of positive cultures and the effect of sample number on our ability to diagnose PJI.

#### MATERIALS AND METHODS

**Patient recruitment.** The study was performed at the Nuffield Orthopaedic Centre, Oxford, United Kingdom, which is a secondary referral hospital for elective orthopedics in southern Oxfordshire and a tertiary referral center for complex procedures including hip and knee revision surgery. All revisions performed between October 1994 and February 1996 (17 months) were studied prospectively. Any patient who had removal of one or both components of the prosthesis was included. To evaluate microbiological results, we used the presence of acute inflammatory cells in histological specimens as the criterion standard (2, 10, 22, 24). Patients were excluded if their histology results could not be interpreted due to underlying rheumatoid arthritis or other inflammatory joint disease.

**Tissue sampling.** Antibiotics were stopped (if they were previously taken) at least 2 weeks preoperatively, and samples were taken before prophylactic antibiotics were given. The participating surgical teams were requested to send a standard set of samples for culture and histology at the time of prosthesis removal. For prosthetic hips these were (i) a swab of joint fluid obtained upon entering the capsule, (ii) capsular tissue, (iii) acetabular membrane, (iv) femoral membrane, or (v) other tissue (e.g., granulation tissue) if the tissue was abnormal. For prosthetic knees these were (i) a swab of joint fluid obtained upon entering the capsule, (ii) capsular tissue, (iii) femoral membrane, (iv) tibial membrane, or (v) other abnormal material. If only one component was revised, multiple samples of relevant periprosthetic tissue were requested. Each sample was taken with a different set of instruments, including a fresh scalpel blade where necessary, to reduce the risk of cross contamination, and each specimen was placed into a separate sterile universal bottle prior to transport to the diagnostic laboratory.

**Microbiological specimens.** Specimens were processed at the Oxford Public Health Laboratory in a class 2 laminar-flow safety cabinet by aseptic technique. Samples were disrupted by vigorous manual agitation with sterile glass beads in sterile diluent. Aliquots of the resulting tissue suspension were inoculated onto chocolate agar plates (incubated in CO<sub>2</sub>) and blood agar plates (incubated in CO<sub>2</sub> and anaerobically). Gram staining was performed with a portion of the sample, and the remainder of the sample was inoculated into Robertson's cooked meat broth. The plates were examined daily for 7 days; broths were subcultured terminally at 5 days or sooner if they were turbid. Organisms were identified by standard methods and antibiotic sensitivities were determined by the method of Stokes et al. (35), which is a comparative disc diffusion method (5, 35). The zone of inhibition of the test organism around an antibiotic disc is compared to the zone of inhibition of a standard, antibiotic-sensitive control strain. If the zone size of the test strain is no less than 3 mm smaller than that of the control strain, it is recorded as susceptible. This method differs from the standard disc diffusion methods of the National Committee for Clinical Laboratory Standards, which

depend on zone size measurements that are interpreted against species- and antibiotic-specific zone sizes. However, there is no major difference in the performance of the method of the National Committee for Clinical Laboratory Standards and the methods of Stokes et al. (4, 35). Organisms of the same species were deemed indistinguishable if they had the same colonial morphology, the same basic biochemical features as determined by the biochemical profiles generated from appropriate API kits (bioMérieux Vitek Inc., Hazelwood, Mo.), and an identical extended antibiogram (13 antibiotics were tested). For subsequent analysis of each patient, the number of samples yielding an indistinguishable organism was recorded, together with the total number of samples sent for that patient.

**Specimens for histological examination.** Multiple specimens were also taken at surgery for histological examination and were fixed in a 10% solution of 40% formaldehyde. They were routinely processed at the Nuffield Orthopaedic Centre Pathology Laboratory, and tissue was embedded in paraffin. Five-micrometer sections were cut and stained with hematoxylin-eosin. Tissue sections were examined by a single, experienced, osteoarticular histopathologist (N.A.), who was unaware of the microbiological findings. On the basis of the histological changes in these samples, an opinion was given on whether the histological appearances were considered to be those associated with infection. This decision was made on the basis of the degree of infiltration by neutrophil polymorphonuclear leukocytes as outlined in previous studies (10, 24, 25), which have shown that the presence of at least five neutrophils per high-power field is very strongly correlated with significant bacteriological growth.

**Statistical methods.** Sensitivities, specificities, and likelihood ratios (LRs) were used to summarize the performance of the alternative diagnostic criteria. The clinical implications of the test results were investigated by a process of Bayesian probability revision (32), with the LRs being used to estimate posttest probabilities of disease over a range of pretest probabilities (19). An LR expresses the ratio of the chance that a given diagnostic test result would be observed for a patient with the target disorder to the chance that it would be observed for a patient without the target disorder (19). It can be calculated from sensitivities and specificities by the relationship  $LR = \text{sensitivity} / (1 - \text{specificity})$ , and it can be used to calculate the posttest odds of a positive result by the relationship  $\text{posttest odds} = LR \times (\text{pretest odds})$ . The posttest odds can then be converted back into a posttest probability by the relationship  $\text{probability} = \text{odds} / (1 + \text{odds})$ .

**Mathematical modeling.** For the purposes of the model, culture-positive samples obtained from patients for whom the histology suggested the presence of infection were considered to be true positives. Culture-positive samples obtained from patients for whom the histology suggested aseptic loosening infection were considered to be false positives.

A subject-specific true-positive rate was calculated by averaging the proportion of culture positive samples from patients with positive histology. Similarly, a subject-specific false-positive rate was calculated by averaging the proportion of culture-positive samples from patients with negative histology. This approach was taken to ensure that the estimates were not biased toward the results for patients from whom high numbers of specimens were available, who may represent a more severely diseased group.

A standard binomial expansion was used to compute the probabilities of all possible test results for between one and seven samples, based on the estimates of rates of true-positive and false-positive results. LRs were calculated for each combination of test results and sample size, and Bayesian probability revision was used to investigate their implications for clinical decision making. In addition, receiver operator characteristic curves were plotted for each of the three possible cutoffs of one, two, or three culture-positive specimens by using sensitivities and specificities predicted on the basis of varying sample numbers.

#### RESULTS

**Patient details.** There were 334 hip and knee revisions performed in the 17-month period. Histology results were not received for 25 patients, and for a further 12 patients histology results could not be interpreted for the purposes of this study due to active rheumatoid arthritis or other inflammatory joint disease, leaving 297 patients available for analysis. The patients ages ranged from 29 to 95 years (mean age, 70 years). There were 178 females and 119 males. There were 253 hip and 44 knee revisions. Forty-one patients (13.8%) had an infected prosthesis on the basis of histological findings.

**Three or more culture-positive samples are strongly predictive of infection.** A range of organisms was isolated (Table 1). We used the numbers of culture-positive and culture-negative specimens obtained from patients with positive and negative histology results to calculate the corresponding sensitivities, specificities, and LRs. The proportion of infected patients over-

TABLE 1. Organisms isolated from multiple samples at revision arthroplasty<sup>a</sup>

Organism	No. of organisms					
	One pos		Two pos		Three pos	
	H+	H–	H+	H–	H+	H–
CoNS	1	20	0	2	9	1
CoNS-MIX	2	5	0	2	10	0
<i>Staphylococcus aureus</i>	0	1	0	0	2	0
<i>Streptococcus</i> spp.	0	1	0	0	2	0
<i>Enterococcus</i> spp.	0	1	0	1	0	0
<i>Peptostreptococcus</i> spp.	0	0	1	0	0	0
<i>Corynebacterium</i> spp.	0	12	0	0	2	0
<i>Propionibacterium</i> spp.	1	2	0	1	0	0
<i>Acinetobacter</i> spp.	1	0	0	0	0	0
<i>Pseudomonas</i> spp.	0	0	0	0	1	0
<i>Bacteroides fragilis</i>	0	0	1	0	1	0
Total	5	42	2	6	27	1

<sup>a</sup> Abbreviations: One pos, isolation of an organism from one of multiple samples; Two pos, isolation of an indistinguishable organism from two samples; Three pos, isolation of an indistinguishable organism from three or more samples; H+, histology positive; H–, histology negative; CoNS, coagulase negative staphylococcus; CoNS-MIX, coagulase-negative staphylococcus mixed with other strains or organisms; an indistinguishable coagulase-negative staphylococcus is present in one, two, or three samples.

all (14%) was used as the pretest probability from which a posttest probability of infection could be calculated.

A single culture-positive specimen was found to be of no diagnostic value (LR = 0.7; posttest probability of infection, 10.6%). If the cutoff for a diagnosis of infection is set at one or more culture-positive specimens, the LR is 4.3 (sensitivity, 83%; specificity, 81%; posttest probability of infection, 40.7%). Increasing the cutoff to two or more specimens growing the same organism raises the LR to 25.9, which is highly predictive of infection (sensitivity, 71%; specificity, 97%; posttest probability of infection, 80.6%). However, these favorable results are due largely to the impact of patients with three or more culture-positive specimens. When the limited number of patients with exactly two positive specimens was analyzed ( $n = 8$ ), the LR was found to be low (2.1; posttest probability of infection, 25.2%). By contrast, with a cutoff of three or more specimens growing the same organism, the test becomes extremely powerful; LR = 169; sensitivity, 66%; specificity, 99.6%; posttest probability of infection, 96.4%).

**The number of samples sent by surgeons is weakly predictive of infection.** Operations were performed by 12 orthopedic surgical teams, one of which performed 133 operations (45% of the total). The remaining teams performed a mean of 15 operations each. Surgeons were requested to provide four or five independent specimens, as described in the Materials and Methods section. In total, 1,206 specimens were obtained from the 297 patients, with a mean number of 4.06 specimens per patient. There was variation within and between surgical teams in the consistency of sampling technique. We were concerned that these inconsistencies might reflect bias on the part of the surgeon, with fewer samples being sent from joints believed not to be infected. This could affect the validity of the study, because low sample numbers might reduce the chance of diagnosing infection by either histology or culture, while high sample numbers might increase this chance.

To assess this, we stratified the patients into three bands; those from whom only one or two samples were received, those from whom between three and six samples were received, and those from whom seven or eight samples were received. We

calculated LRs for the presence of a positive histology result on the basis of the number of samples obtained. The results, shown in Table 2, indicate a small but definite systematic trend favoring a final diagnosis of infection if multiple samples had been sent.

**Sampling bias has only minor effects on the overall performance of the test.** Because we were concerned about the potential importance of the findings presented above, we analyzed the diagnostic value of different numbers of positive samples in two subgroups (Table 3). Group B is the subgroup ( $n = 213$ ) comprising only those patients from whom three, four, five, or six specimens were taken. Group C is an alternative subgroup ( $n = 239$ ) which excludes all patients operated on by surgeons with erratic sampling behavior on the basis of an interquartile range for the number of specimens sent of greater than 2. Group C thus contains only patients operated on by surgeons who sent between three and five specimens at least 50% of the time. Group A is the whole group ( $n = 297$ ), as discussed above, and is included as a comparator. The three groups have very similar LRs (and hence posttest probabilities of infection), suggesting that the samples are drawn from a relatively homogeneous group, irrespective of sampling behavior.

**Lack of utility of Gram staining in diagnosing infection.** For many samples, positive cultures were obtained only from the enrichment broths, suggesting that the number of organisms present in the sample was low. This impression was confirmed by the very poor sensitivity of Gram staining, the diagnostic impact of which we evaluated. Its sensitivity, measured against a positive culture result for the same sample, was 6%, with a specificity of 99.7% (LR = 20). Measured against a positive histology result for a patient, Gram staining of a single sample has a sensitivity of 12% and a specificity of 98.8% (LR = 10). Thus, although a positive Gram staining result does predict infection, the sensitivity is so poor that multiple samples from a large number of infected patients must be examined for each positive result. Furthermore, management was not altered by the Gram staining result, since patients were treated empirically, once samples were obtained, if the clinical history or examination of frozen sections of specimens obtained during surgery were suggestive of infection.

**Age distributions of prostheses with positive culture or histology results.** Infection is known to be associated with early failure of a prosthesis, and hence, the majority of patients with positive microbiology results should have had revisions within a few years of implantation. By contrast, the probability of contamination should be constant, so that patients with false-positive microbiology results should constitute a roughly constant proportion of the total number of revisions performed, irrespective of the age of the prosthesis. The date that the prosthesis was inserted was available for 281 patients (95% of patients), and we plotted the number of patients against the

TABLE 2. Surgical sampling behavior correlates with presence of infection

No. of specimens taken	No. of patients with the following histology result:		LR for positive result	Posttest probability of infection (%)
	Positive	Negative		
1 or 2	2	59	0.21	3.3
3, 4, 5, or 6	29	184	0.98	13.6
7 or 8	10	13	4.80	43.4
Total	41	256		



TABLE 3. Diagnostic value of different numbers of positive specimens

Microbiology result	Group (no. of patients) <sup>a</sup>	No. of patients with the following histology result:		LR for positive result	Posttest probability of infection (%)
		Positive	Negative		
Three or more specimens positive (same organism)	A (297)	27	1	169	96.4
	B (213)	18	1	114	94.8
	C (239)	23	1	144	95.8
Two or more specimens positive (same organism)	A	2	6	2.1	25.2
	B	1	4	1.6	20.4
	C	1	5	1.3	17.2
One positive specimen	A	5	42	0.74	10.6
	B	5	33	0.96	13.3
	C	4	34	0.73	10.5
No growth from any specimen	A	7	207	0.21	3.3
	B	5	146	0.22	3.4
	C	5	166	0.19	3.0

<sup>a</sup> Group A, whole group; group B, patients for whom three to six samples were sent; group C, patients operated on by most consistent surgeons (interquartile range for number of samples sent per patient, <2).

age of the prosthesis for histology-positive and -negative patients and separately for patients with none, one or two, and three or more specimens that were culture positive (Fig. 1). The similarities between the curve for patients with histology-positive specimens and that for patients with three or more culture-positive specimens is to be expected, but it is also apparent that the curve for patients with one or two culture-positive specimens is as predicted if the majority of these are positive because of contamination.

**Mathematical modeling of the number of samples that should be taken.** Even though our study was based on data for nearly 300 patients, the estimates of test performance for different diagnostic strategies could not be investigated reliably due to insufficient data. The simple mathematical model offered an alternative approach to scrutinizing the predictive behavior of alternative diagnostic rules.

The results, shown in Table 4 and Fig. 2, demonstrate that the model predicts an excellent combination of sensitivity and specificity if five or six samples are taken and a cutoff of two or more culture-positive samples is used to diagnose infection. Use of a cutoff of three or more positive samples gave extremely high specificity, but in order to achieve a satisfactory sensitivity, the number of specimens that the laboratory would need to process would be impractical. Conversely, by using a cutoff of one culture-positive sample, for a given number of samples the test was always more sensitive but lacked specificity.

## DISCUSSION

We have prospectively evaluated the criteria for the microbiological diagnosis of infection at elective revision arthroplasty. We have demonstrated that the isolation of indistinguishable microorganisms from three or more independent operative specimens is highly predictive of the presence of acute inflammatory cells in specimens examined histologically, our definition of infection. We have further used simple mathematical modeling to demonstrate that the culture of five or six specimens obtained during surgery is necessary to produce a diagnostic test that is both sensitive and specific. This appears to be because this condition is associated with an extremely low burden of microorganisms, as indicated by the lack of utility of

Gram staining, the need for the use of enrichment broths to obtain positive cultures for many samples, and the frequent occurrence of culture-negative samples from infected patients.

The key assumption underlying our approach is that histology is a reliable means of diagnosing infection. We believe this is correct, based on a combination of previously published data from studies with animals and clinical studies. In a canine model of prosthetic hip joint infection, there was a strong correlation between infection and an acute inflammatory response (26, 27). This correlation was less strong if only limited portions of randomly selected specimens of tissue were sent, suggesting that infection and the inflammatory response were multifocal. A similar relationship between the presence of acute inflammatory cells and infection has also been found in a number of studies with human tissue obtained during surgery (6, 9, 22–25, 28, 29). Our decision to use histology is also based on the known unreliability of clinical features in predicting infection (3). Although some cases of prosthetic joint infection are clinically obvious, many patients who come for elective revision arthroplasty do not have a history of acute infection, and for many patients the clinical presentation is indistinguishable from early aseptic loosening (16, 36). For these reasons we believe that the presence of acute inflammatory cells in histologically examined specimens is a good criterion standard for the diagnosis of infection at elective revision arthroplasty, provided that individuals with known inflammatory joint disease are excluded and sufficient numbers of samples are examined by an experienced pathologist. Previous studies have used various criteria for the microbiological diagnosis of infection, ranging from the isolation of organisms from one direct culture or two or more broth enrichment cultures (2, 8, 12, 21, 34) to requiring four of five, five of six, or even five of five specimens (20) to yield the same organism. Although several investigators have mentioned the need to culture multiple samples in order to increase sensitivity and to overcome the problem of contamination (1, 3, 8, 16, 20, 21, 25, 30, 31, 34, 36), a standardized method has been neither defined nor validated. Our prospective study indicates that the isolation of an organism from at least three of several independent specimens is diagnostic of infection.

For practical purposes we called two organisms indistin-

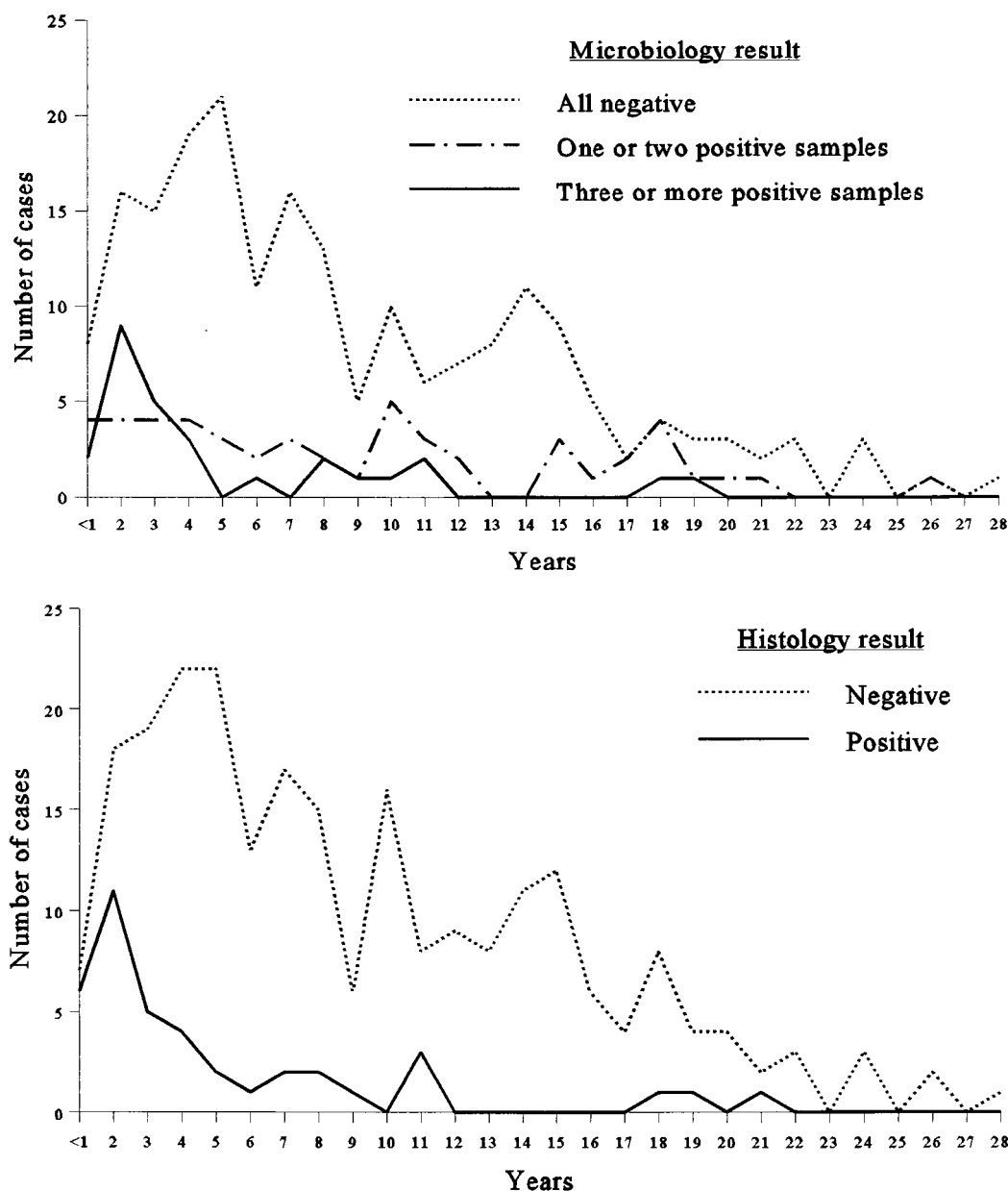


FIG. 1. Microbiology and histology results according to the duration that the prosthesis had been in situ.

guishable on the basis of simple laboratory tests and the extended antibiotic sensitivity pattern. Care is needed with this approach, because the antibiotic sensitivity pattern is not an infallible way of distinguishing strains, particularly among the coagulase-negative staphylococci, the dominant group of both pathogens and contaminants in this series. Molecular methods have shown that the same antibiogram may be shared by two different strains, and conversely, organisms expressing different antibiograms may appear to be indistinguishable by genotyping. Furthermore, multiple strains can be present in the same infected joint prosthesis (7, 17, 21, 25). Nonetheless, our results do show that in a routine laboratory, simple methods of identification appear to be adequate in most cases.

We found that a substantial proportion of samples (35%) obtained from patients who were infected on the basis of histological criteria failed to yield organisms, despite an ex-

tended culture regimen that included an enrichment broth suitable for the recovery of many fastidious organisms. Even for those patients for whom three or more specimens were culture positive, it was usual to have at least one other specimen with no growth. These data reinforce the multifocal and low-grade nature of this infection, which means that sampling error becomes an important reason for false-negative results. Furthermore, the possibility of sampling bias needs to be considered.

On examining our data, we found definite evidence of sampling bias, with an increasing proportion of infected patients as more samples were sent. To examine the impact of this on our results, we compared the whole data set with two subsets of cases; one in which the "ideal" number of specimens ( $n = 3$  to 6) was sent and one that included only samples from surgeons with consistent sampling behavior. Results for all groups were

TABLE 4. Probability of histologically positive or negative specimens in relation to microbiological results for one to seven specimens from each patient

Microbiological findings	Probability for histologically positive specimens when the following no. of specimens were taken:							Probability for histologically negative specimens when the following no. of specimens were taken:							LR when the following no. of specimens were taken:						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
All negative specimens	0.41	0.16	0.07	0.03	0.01	0.004	0.002	0.94	0.88	0.83	0.78	0.73	0.68	0.64	0.43	0.19	0.08	0.03	0.02	0.006	0.003
One positive specimen	0.60	0.48	0.29	0.16	0.08	0.04	0.02	0.06	0.12	0.16	0.20	0.24	0.27	0.29	9.7	4.2	1.8	0.78	0.34	0.15	0.06
Two positive specimens		0.35	0.43	0.35	0.24	0.14	0.08		0.004	0.01	0.02	0.03	0.04	0.06		93	41	18	7.6	3.3	1.4
Three or more positive specimens			0.21	0.47	0.67	0.81	0.90			0.0002	0.0009	0.002	0.004	0.007			1,051	512	321	203	134

very similar, suggesting that the samples were all drawn from essentially homogeneous groups of infected and uninfected patients. The removal of data for patients for whom there were too few or too many samples still leaves data for more than 70% of the patients, and the performance of the test is unchanged.

Our impression that this is a condition associated with very low numbers of bacteria is strengthened by the poor performance of Gram staining. Although the rapidity of Gram staining makes it potentially useful as a peroperative test, in our center considerable additional resources would have to be expended to make this possible. It is labor-intensive and has an inherently low level of sensitivity, and on the basis of the results of this study, we recommend that it be abandoned for specimens from elective revision arthroplasties.

The apparently low numbers of organisms and the problem of sampling error mean that a number of culture-negative specimens is probably inevitable. Our modeling suggests that even for an infected patient, when four samples are taken there is still a 3.0% chance that all specimens will be culture negative and a 16% chance that only a single specimen will be culture positive (Table 3). If five specimens are taken, these values fall to 1.0 and 8.0%, respectively. Use of the binomial expansion assumes that all samples are truly independent. When the calculated (expected) probabilities were converted into expected absolute numbers, there was very close agreement with the observed values for the samples from the histology-negative patients (chi-squared = 1.56; degrees of freedom = 3;  $P = 0.66$ ). An exception was that the observed number of patients whose specimens were all culture negative, despite positive histology, was greater than expected (chi-squared = 17.9; degrees of freedom = 4;  $P = 0.001$ ). This may be due to unrecognized inflammatory joint conditions, accidental preoperative antibiotic use, or infection with organisms that cannot be cultured by the methods used. Our technique would fail to isolate *Mycoplasma* spp., L-form bacteria, and *Campylobacter* spp., to name a few, all of which have been described as pathogens in prosthetic joint infection (15, 33, 37). Design of a culture protocol able to accommodate a very wide range of fastidious organisms would be impractical for routine diagnostic laboratories.

These considerations may well explain the discrepancy between the conclusions from our modeling (five or six specimens sent; an infection is diagnosed if two or more specimens are culture positive) and the actual data themselves (infection was unequivocally diagnosed if three or more specimens were culture positive). If specimens from a proportion of patients have nonculturable organisms, the overall performance of the diagnostic rules is reduced by the effect of specimens that can never yield diagnostic information, however many specimens are sent. We are testing this hypothesis in a further prospective study.

Definitions of infection have been offered by using scoring systems that take into account clinical, histological, and microbiological features (13, 18). Such systems have not been validated. We have instead used an approach based on the calculation of LRs, which allows us to take into account clinical epidemiological considerations in interpreting the culture results. We diagnosed infection in 30% of prostheses revised within the first 4 years after implantation but in only 5% of prostheses revised between 4 and 8 years postimplantation. The likelihood ratio can be applied to these different pretest probabilities of infection (Table 5). This is particularly useful because it demonstrates that when the prior probability of infection is high (as in early failure), the isolation of bacteria from only two specimens, while inconclusive, nonetheless

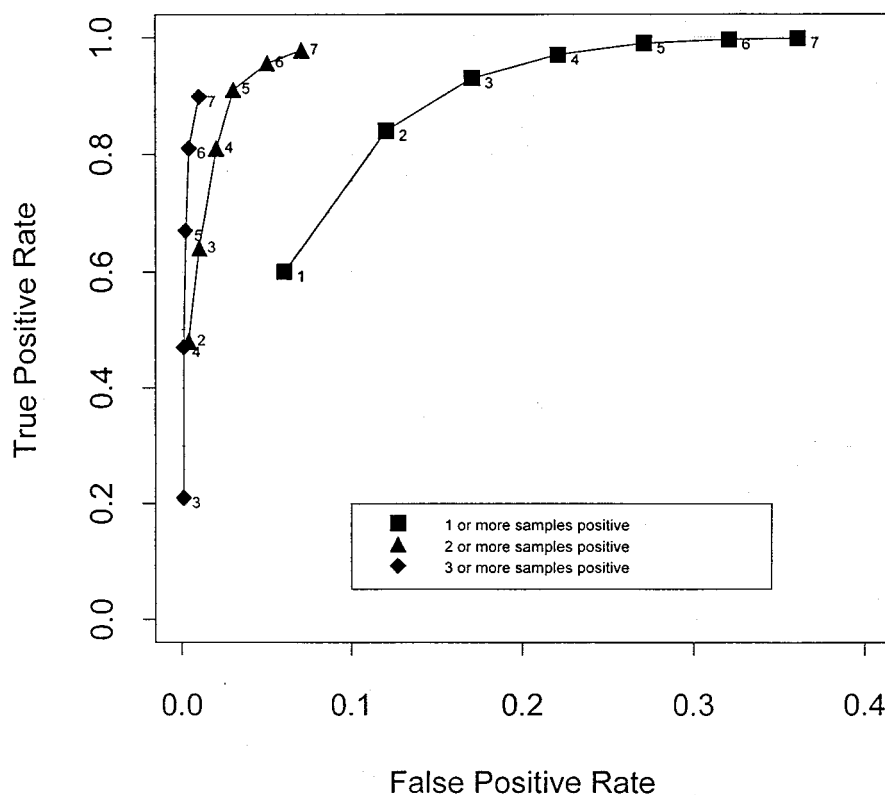


FIG. 2. Receiver operator curves showing the predicted sensitivities and specificities of the test with various numbers of specimens, using cutoffs of one, two, or three specimens positive for the same organism. The numbers of specimens taken are indicated alongside each datum point.

makes the diagnosis of infection more likely. Clinical information is also amenable to this approach, so the pre- and postoperative probabilities of infection could be modified further on the basis of symptoms, macroscopic appearances, and blood tests. We are undertaking further studies to test this hypothesis.

It is possible that our diagnostic rules could be modified to take account of organisms that are not part of the normal skin flora and that have a known pathogenic potential. Such cases of infection might prove to be easier to diagnose on the basis of the need to use a smaller number of samples. Unfortunately, the numbers of individual species of organism isolated were too small to calculate organism-specific LRs. In any event, because the identity of the organism is rarely known in advance, it is unlikely that the number of samples taken during surgery can be modified on a patient-by-patient basis without affecting the performance of the diagnostic test.

We conclude that the isolation of indistinguishable microorganisms from three or more independent specimens, to be

taken as part of a standard set of five or six specimens, is an accurate and practical microbiological definition of infection at revision arthroplasty. We stress that the use of meticulous sampling technique for the prevention of the cross contamination of samples is central to this approach and that these observations do not apply to acute infections of prostheses, in which the bacterial load is higher and for which the diagnostic issues are more straightforward. Culture techniques must make use of enrichment media because of a very low burden of microorganisms in most samples, and Gram staining should be abandoned for specimens from elective revisions. In centers where histology is unavailable (and perhaps when histology cannot be used), our microbiological criteria can be used alongside clinical data to make the diagnosis of infection more reliable. We have used these findings to modify our practice. Patients with definite infections by these criteria are treated by a two-stage revision protocol that includes appropriate antibiotics for 6 weeks before a new prosthesis is reinserted. Patients with fewer than two microbiologically positive samples and

TABLE 5. Probability of infection as a function of the time that the prosthesis was in situ and the number of specimens positive

Time (yr) that prosthesis was in situ (no. of patients) <sup>a</sup>	Pretest probability of infection (%)	Posttest probability of infection (%)			
		All specimens negative (LR = 0.21)	One specimen positive (LR = 0.74)	Two specimens positive (LR = 2.1)	Three or more specimens positive (LR = 169)
<2 (43)	39.5	12	33	58	99
2-4 (50)	18	4	14	32	97
4-10 (97)	8.2	2	6	16	94
>10 (91)	6.6	1.5	5	13	92

<sup>a</sup> Specimens from 281 patients were tested.

negative histology have early reinsertion of a prosthesis and receive only prophylactic antibiotics.

Although our findings are specific to infected prosthetic hips and knees, we speculate that they may also hold true for other low-grade chronic infections wherever pathogens and commensal organisms overlap. This includes infections of other prosthetic joints and implantable devices, fracture fixations and nonunions, and other forms of chronic osteomyelitis including vertebral and contiguous osteomyelitis. Finally, our observations highlight the importance of considering sampling error and bias whenever the result of a diagnostic test is a qualitative one that depends on the analysis of multiple samples. This is the case for some other microbiological tests (such as blood cultures) and for the histological diagnosis of multifocal inflammation or neoplasia, particularly lymph node metastasis in patients with cancer and gastric, colonic, or cervical metaplasia.

#### ACKNOWLEDGMENTS

This project was supported by funds donated by the Microbial Diseases Trust Fund, Oxford Radcliffe Hospital, Headington, Oxford.

This study was done while A. Berendt was a Lister Institute Research Fellow.

We thank E. Coles for assistance in preparing the manuscript.

#### REFERENCES

- Ahnfelt, L., P. Herberts, H. Malchau, and G. B. J. Andersson. 1990. Prognosis of total hip replacement: a Swedish multicenter study of 4,664 revisions. *Acta Orthop. Scand.* **61**(Suppl. 238):1–26.
- Athanasou, N. A., R. Pandey, R. DeSteiger, D. Crook, and P. McLardy-Smith. 1995. Diagnosis of infection by frozen section during revision arthroplasty. *J. Bone Jt. Surg. Br. Vol.* **77B**:28–33.
- Brause, B. 1989. Infected orthopedic prostheses, p. 111–127. In A. L. Bisno and F. A. Waldvogel (ed.), *Infections associated with indwelling medical devices*. American Society for Microbiology, Washington, D.C.
- Brown, D. F. J. 1990. The comparative methods of antimicrobial susceptibility testing—time for a change? *J. Antimicrob. Chemother.* **25**:307–312.
- BSAC Working Party. 1991. A guide to sensitivity testing. *J. Antimicrob. Chemother.* **27**(Suppl. D).
- Charosky, C. B., P. G. Bullough, and P. D. Wilson. 1973. Total hip replacement failures: a histological evaluation. *J. Bone Jt. Surg. Br. Vol.* **55A**:49–58.
- Crichton, P. B., L. A. Anderson, G. Phillips, P. G. Davey, and D. I. Rowley. 1995. Subspecies discrimination of staphylococci from revision arthroplasties by ribotyping. *J. Hosp. Infect.* **30**:139–147.
- Cuckler, J. M., A. M. Star, A. Alavi, and R. B. Noto. 1991. Diagnosis and management of the infected total joint arthroplasty. *Orthop. Clin. N. Am.* **22**:523–530.
- Fehring, T. K., and J. A. McAlister, Jr. 1994. Frozen histologic section as a guide to sepsis in revision joint arthroplasty. *Clin. Orthop.* **304**:229–237.
- Feldman, D. S., J. H. Lonner, P. Deai, and J. D. Zuckerman. 1995. The role of intraoperative frozen section in revision total joint arthroplasty. *J. Bone Jt. Surg. Am. Vol.* **77A**:1807–1813.
- Fitzgerald, R. H., Jr. 1992. Total joint arthroplasty sepsis. Prevention and diagnosis. *Orthop. Clin. N. Am.* **23**:259–264.
- Garvin, K. L., and A. D. Hanssen. 1995. Current concepts review: infection after total hip arthroplasty. *J. Bone Jt. Surg. Br. Vol.* **77A**:1576–1588.
- Garvin, K. L., E. A. Salvati, and B. D. Brause. 1988. Role of gentamicin-impregnated cement in total joint arthroplasty. *Orthop. Clin. N. Am.* **19**:605–610.
- Gristina, A. G., and J. Kolkin. 1983. Currents concepts review. Total Joint replacement and sepsis. *J. Bone Jt. Surg. Am. Vol.* **65A**:128–134.
- Haley, S., J. Paul, D. W. Crook, and S. H. White. 1990. *Acinetobacter* spp. L-form infection of a cemented Charnley total hip replacement. *J. Clin. Pathol.* **43**:781. (Letter.)
- Hamblin, D. L. 1993. Diagnosis of infection and the role of permanent excision arthroplasty. *Orthop. Clin. N. Am.* **24**:743–749.
- Hope, P. G., K. G. Kristensson, P. Norman, and R. A. Elson. 1989. Deep infection of cemented total hip arthroplasties caused by coagulase negative staphylococci. *J. Bone Jt. Surg. Br. Vol.* **71B**:851–855.
- Hughes, P. W., A. E. Salvati, P. D. Wilson, and E. L. Blumenfeld. 1978. Treatment of subacute sepsis of the hip by antibiotics and joint replacement. Criteria for diagnosis with evaluation of twenty-six cases. *Clin. Orthop. Relat. Res.* **141**:143–157.
- Jaeschke, R., G. Guyatt, and D. L. Sackett. 1994. Users guide to the medical literature: how to use an article about a diagnostic test. What are the results and how will they help me in caring for my patients? *JAMA* **271**:703–707.
- Kamme, C., and L. Lindberg. 1981. Aerobic and anaerobic bacteria in deep infections after total hip arthroplasty. *Clin. Orthop.* **154**:201–207.
- Kristinsson, K. G., P. G. Hope, P. Norman, and R. Elson. 1989. Deep infections associated with total hip arthroplasties caused by coagulase negative staphylococci—pathogenesis and microbial diagnosis. *J. Bone Jt. Surg. Br. Vol.* **71**:329–337.
- Lonner, J. H., P. Desai, P. E. Dicesare, G. Steiner, and J. D. Zuckerman. 1996. The reliability of analysis of intra-operative frozen sections for identifying active infection during revision hip or knee arthroplasty. *J. Bone Jt. Surg. Am. Vol.* **78A**:1553–1558.
- Mirra, J. M., R. A. Marder, and H. C. Amstutz. 1982. The pathology of failed total joint arthroplasty. *Clin. Orthop.* **170**:175–183.
- Mirra, J. M., H. C. Amstutz, S. M. Matros, and R. Gold. 1976. The pathology of the joint tissues and its clinical relevance in prosthesis failure. *Clin. Orthop.* **117**:221–240.
- Perdreau-Remington, F., D. Stefanik, G. Peters, C. Ludwig, J. Rutt, R. Wenzel, and G. Pulverer. 1996. A four year prospective study on microbial ecology of explanted prosthetic hips in 52 patients with “aseptic” prosthetic joint loosening. *Eur. J. Clin. Microbiol. Dis.* **15**:160–165.
- Petty, W., S. Spanier, J. J. Shuster, and C. Silverthorne. 1984. The influence of skeletal implants on incidence of infection. *J. Bone Jt. Surg. Br. Vol.* **67A**:1236–1244.
- Petty, W., S. Spanier, and J. J. Shuster. 1988. Prevention of infection after total joint replacement. *J. Bone Jt. Surg. Br. Vol.* **70A**:536–539.
- Pipino, F., L. Molfetta, L. Resta, A. Facilone, O. Brandonisio, and M. Altamura. 1989. Bacteriological and histological study of 40 loose cemented hip prostheses. *Ital. J. Orthop. Trauma* **15**:481–490.
- Pizzoferrato, A., F. Fiori, and L. Savarino. 1981. Microbiological investigation on 161 cases of hip endo-arthroprosthesis failure. *Chir. Organi. Mov.* **66**:297–307.
- Rand, J. A., B. F. Morrey, and R. S. Bryan. 1984. Management of the infected total joint arthroplasty. *Orthop. Clin. N. Am.* **15**:491–504.
- Rand, J. A., R. S. Bryan, B. F. Morrey, and F. Westholm. 1986. Management of infected total knee arthroplasty. *Clin. Orthop.* **205**:75–85.
- Sackett, D. L., R. B. Haynes, G. H. Guyatt, and P. Tugwell. 1991. The interpretation of diagnostic data, p. 69–152. In *Clinical epidemiology. A basic science for clinical medicine*, 2nd ed. Little, Brown & Co., Boston, Mass.
- Sneller, M., F. Wellborne, M. F. Barile, and P. Plotz. 1986. Prosthetic joint infection with *Mycoplasma hominis*. *J. Infect. Dis.* **153**:174–175.
- Steckelberg, J. M., and D. R. Osmon. 1994. Prosthetic joint infections, p. 259–290. In A. L. Bisno and F. A. Waldvogel (ed.), *Infections associated with indwelling medical devices*, 2nd ed. American Society for Microbiology, Washington, D.C.
- Stokes, E. J., G. L. Ridgway, and M. W. Wren. 1993. *Clinical microbiology*, 7th ed. Edward Arnold, London, United Kingdom.
- Tsukayama, D. T., R. Estrada, and R. B. Gustilo. 1996. Infection after total hip arthroplasty. *J. Bone Jt. Surg. Br. Vol.* **78A**:512–523.
- Yao, J. D. C., M. C. N. Herman, and I. Campbell. 1993. Prosthetic hip joint infection due to *Campylobacter fetus*. *J. Clin. Microbiol.* **31**:3323–3324.