DIAGNOSIS OF SPONDYLITIS BY EXTERNAL ^{85}Sr COUNTING: A COMPARISON OF ANALYTIC PROCEDURES

Lars Lindberg

Malmö General Hospital, University of Lund, Malmö, Sweden

and Melvin S. Schwartz

The Hospital for Special Surgery, affiliated with the New York Hospital, Cornell University Medical Center, New York, New York

The assessment of activity in tuberculous or septic spondylitis is sometimes difficult even for the experienced clinician. To determine whether or not isotope techniques can be of diagnostic value in these cases, we have performed external counting over the spine of patients injected with ^{85}Sr. The purpose of this investigation was to determine the activity or stage of the spinal infection, and for this reason we only used subjects with verified diagnoses.

The results have been published by Felländer and Lindberg (*1*); the agreement between the clinical assessment and the result of the isotope examination has been very good. The method is also of considerable practical diagnostic value as a complement to the usual clinical and radiological assessment. In general, cases of spondylitis with a high isotope uptake have shown an active infectious process, and all cases of spondylitis with normal isotope uptake have been healed or quiescent. In an intermediate group where the isotope uptake was only a little higher than normal it has often been difficult to assess the activity of the infection even though the usual clinical as well as isotope methods have been used. For this reason we decided to analyze the results of the external counting to attempt to sharpen the differentiation of the two groups.

CLINICAL MATERIAL

Normals. In assessing the counting results, normal values were taken from 20 individuals, 19–62 years old, who did not have any subjective or objective evidence of spinal disease. All were examined radiographically, and a moderate degree of spondylosis deformans, which was noted in some of the older cases, was accepted as being normal (*2*).

Patients. Sixty-five patients with tuberculous and 26 patients with nontuberculous lesions were studied. From this clinical material only the values coming from lesions clinically considered clearly active or quiescent were taken for statistical analysis. Lesions with doubtful infectious activity were omitted.

TECHNIQUE OF EXTERNAL COUNTING OF ^{85}Sr

The patients received intravenously 1 μCi of carrier-free ^{85}Sr per kilogram body weight with a maximum dose of 50 μCi. The counting was performed 2 weeks later using a scintillation detector with a 2-in. crystal, a pulse-height analyzer and a scaler. A 12-deg lead collimator was used. This proved to be the most suitable type in practice (*1*).

Counting was performed in the following way with the patient lying prone: the spinous processes of L-4 and C-7 were identified by palpation. The distance between these points was divided into 10 equal lengths, and the resulting 11 points were marked. Caudally, from the L-4 point, one more point was marked at the same distance as the others. A total of 12 points was therefore used.

ANALYSIS PROCEDURES

Several methods of analysis were investigated.

Peak method. Used by Felländer and Lindberg (*1*). In this method two patterns of the normal curve are used (Figs. 1 and 2).

In constructing these curves the connected counting values are plotted on the ordinate and the spinal distances on the abscissa with the most caudal point at the origin. Along the drawn curve lie the arithmetically averaged counting values. A confidence interval corresponding to 95% (approximately 2σ)*

Received March 5, 1968; revision accepted Sept. 24, 1968.

For reprints contact: L. Lindberg, Dept. of Orthopedic Surgery, Malmö General Hospital, Univ. of Lund, Malmö, Sweden.

* 2 σ has been used to characterize the 95% confidence interval. 2 σ corresponds to the "t" value for 19 deg of freedom at the 0.05 level of probability.

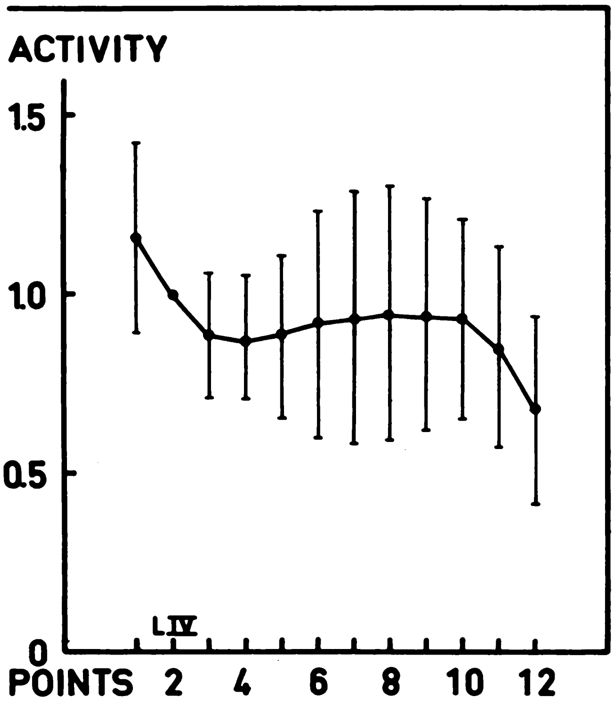

FIG. 1. Normal curve with L-4 as reference point.

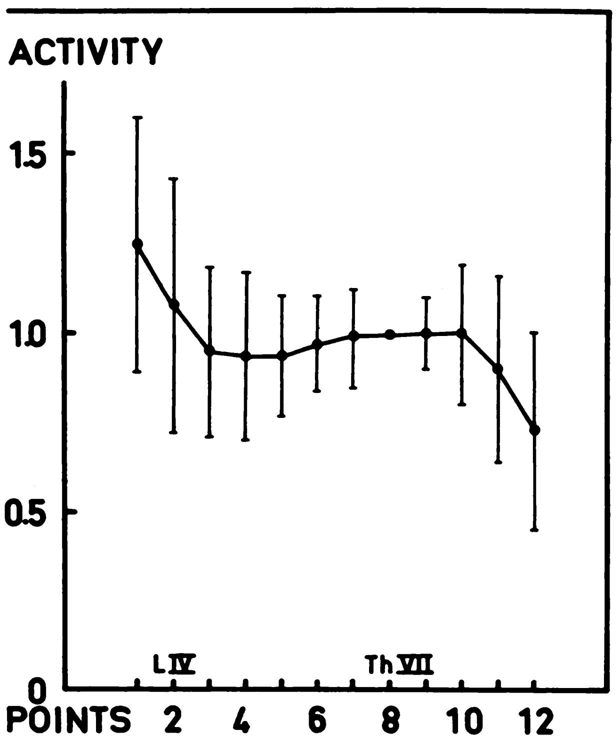

FIG. 2. Normal curve with Th-7 as reference point.

has been drawn in at each point. Curve 1 is used with spondylitis in the thoracic spine and Curve 2 with spondylitis in the lumbar spine.

In constructing a patient's curve, one of the points—2 or 8—is used as reference value. The counting rate at that point is given the numerical

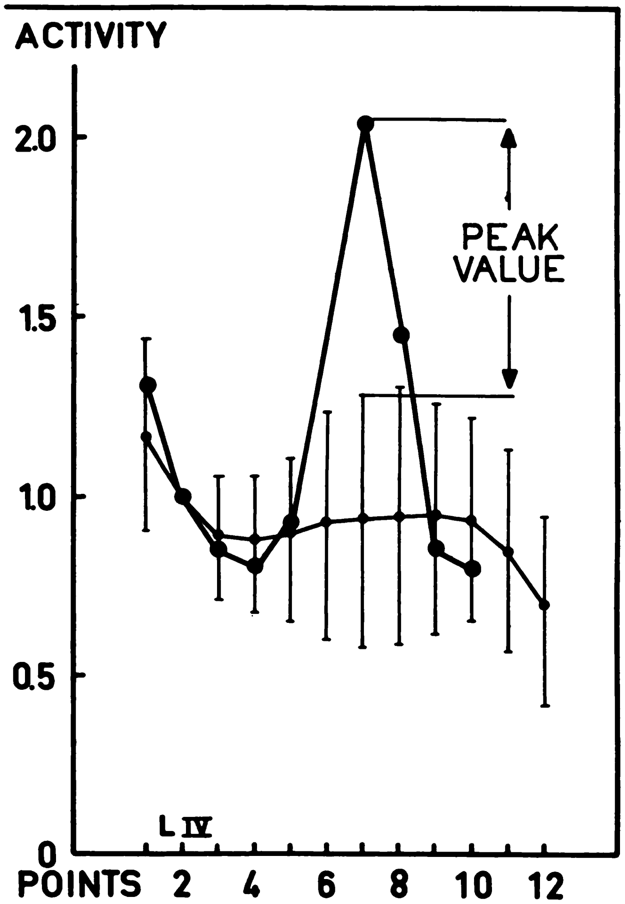

FIG. 3. Peak value of curve for patient with tuberculous lesion in Th-7.

value 1, and the other counting values are expressed as ratios of that value. In normal Curve 1, Point 2, corresponding to L-4, and in normal Curve 2, Point 8 corresponding approximately to Th-7, were chosen as the reference points.

In this method the increased strontium uptake in the lesion was measured in terms of the increase of the ratio at the "peak value" above the upper limit of the normal value (Fig. 3).

With this method a good correlation was found in tuberculous spondylitis between clinical assessment and the result of isotope studies with peak values between 2.87 and 0.28. Good correlation between the skeletal lesions considered healed and the results of the isotope counting was also found. In the group of patients in which peak values were between 0.28 and 0, the clinical assessment was more difficult. In septic spondylitis the clinical assessment cannot be made in the same way as in the tuberculous cases, but in these cases it was found that patients with "peak values" over 0.55 were under medical care and that patients with peak values under 0.55 had been released from medical care.

Method of point comparisons. The observed points at which counts were taken were standardized by several different procedures. For each of the 20 normal individuals, two curves were constructed in the manner described under "peak method," one with Point 2 and one with Point 8 chosen as reference points. When the values for the points had been standardized, each separate point was examined in connection with the measurements on the 20 normal individuals by a probit method of analysis. The probit estimates of mean and standard deviation were quite close and in good agreement with the calculated values for these parameters for each of the 12 points. Hence it can be reasonably assumed that the 20 normal individuals formed a group consistent with a normal distribution, and this would be true for each of the 12 points.

From these baseline values determined on the normals, it is possible to investigate the patient groups with regard to the counting values at each point by setting limits about the normal point values that are equivalent to the average ±1 s.d., average ±1.5 s.d. and average ±2 s.d., respectively. These ranges were applied to the patient group and evaluated for diagnostic accuracy in an effort to find the optimal range to use as a diagnostic criterion.

When the calibration procedures described above were established, both the patient group and the normal group were examined by the same diagnostic criteria, person by person, to estimate diagnostic accuracy. It was then possible to construct a table in which the agreement or disagreement with clinical diagnosis on the basis of point values ±1, 1.5 or 2 s.d. could be summarized for both normal and patient groups. It should be noted that an important criterion for the diagnosis of pathology is the presence of consecutive abnormal points. In general this implies more definitely the presence of pathology than does the appearance of an isolated abnormal point.

Summaries are included in Table 1 in which the number of agreements and the number of disagreements with clinical judgment are given for normal and patient groups. It can be seen by inspection that the criterion of either average ±1.5 s.d. or average ±2.0 s.d. is satisfactory. Chi-square analysis indicates that the average ±1 s.d. is preferable for the Point 2 curve.*

Method of interval comparisons. Another approach that was investigated was the measurement of slope or change in uptake values from point to point. This can be estimated in at least two ways. Simple differences can be calculated; in this way for 12 points 11 interval differences are available for each of the 20 normal individuals. These differences are evaluated by the probit method used in the 11 point comparison method. The analyses show once again good consistency with the bell-shaped distribution, and the standards of average difference between two adjacent points (or average interval difference) ±1 s.d., ±1.5 s.d. and ±2 s.d. were investigated.

An alternative approach is to divide the interval difference by the linear distance separating the two measured points. This can in fact be a more precise tool than interval difference alone. However, certain technical variations which existed in the measurement of the current patient groups make it difficult to assess the distances accurately for each case. Therefore the simple interval method was adopted as the working approach. It may be possible at some other time to refine the interval by the interval-to-distance approach (which would be a true slope measure). The diagnostic approach used below will make use of simple interval measurements. Once again the number of consecutive abnormalities is taken as a more reliable criterion of pathology than the presence of an isolated interval change.

Table 2 shows the number of agreements with clinical judgment for each of the normal and patient groups investigated for each of the relevant ranges. It is apparent that the cleanest separation between normals and patients is effected by the criterion of either average interval difference ±1.5 s.d. or aver-

TABLE 1. DIAGNOSIS BY POINT

		Point 2 curve				Point 8 curve			
		N	S	TB	S+TB	N	S	TB	S+TB
±1 s.d.	Agreement	12	9	10	19	7	12	24	36
	Disagreement	8	2	4	6	13	3	9	12
±1.5 s.d.	Agreement	15	9	7	16	15	13	29	42
	Disagreement	5	2	7	9	5	2	4	6
±2 s.d.	Agreement	19	9	6	15	18	13	29	42
	Disagreement	1	2	8	10	2	2	4	6

N = normals; S = nonspecific spondylitis; TB = tuberculous spondylitis.

* When the data in the table are analyzed by chi-square procedure, the finding of nonsignificance is a meaningful index; a nonsignificant result indicates that there are no serious differences in the percentage of correct diagnoses between the normal and patient groups.

TABLE 2. DIAGNOSIS BY INTERVAL

		N	S	TB	S+TB
±1 s.d.	Agreement	8	23	35	58
	Disagreement	12	3	12	15
±1.5 s.d.	Agreement	18	22	32	54
	Disagreement	2	4	15	19
±2 s.d.	Agreement	19	21	29	50
	Disagreement	1	5	18	23

N = normals; S = nonspecific spondylitis; TB = tuberculous spondylitis.

TABLE 3. DIAGNOSIS BY AREA

		Area I			Area II			Area I + II		
		N	S	TB	N	S	TB	N	S	TB
±1 s.d.	Agreement	18	9	12	13	15	29	31	24	41
	Disagreement	2	2	1	7	0	5	9	2	6
±1.5 s.d.	Agreement	18	9	10	18	15	28	36	21	38
	Disagreement	2	2	3	2	0	6	4	2	9
±2 s.d.	Agreement	19	9	11	18	14	25	37	23	36
	Disagreement	1	2	2	2	1	9	3	3	11

N = normals; S = nonspecific spondylitis; TB = tuberculous spondylitis.

age interval difference ±2.0 s.d. Chi-square analysis indicates that average ±1.5 s.d. is the method of choice.

Method of area comparison. Another diagnostic approach investigated is the computation of areas under the uptake curve of each patient and each normal individual. For this measurement a compensating polar planimeter was used (K&E 4242). On this instrument 1 cm^2 is equivalent to 9.1 units of the planimeter scale. The area measured was taken between the uptake curve for each individual (normal or pathological) and a baseline established by the over-all averages for each point for the 20 normal individuals. Examples of such areas, together with the numerical values obtained, are shown in Fig. 4. The values in each of the two different cases arranged themselves into a pattern that did not deviate significantly from the normal distribution. Again standards were set by taking the average normal area and establishing ranges equal to ±1 s.d., ±1.5 s.d. and ±2.0 s.d. Table 3 summarizes the results for each of the ranges; agreement and disagreement with clinical judgment are listed. Using chi-square analysis, one finds that the best cut-off points are ±1.5 s.d. or ±2.0 s.d., depending on which of the two areas is being used. This is also obvious from gross inspection of Table 3.

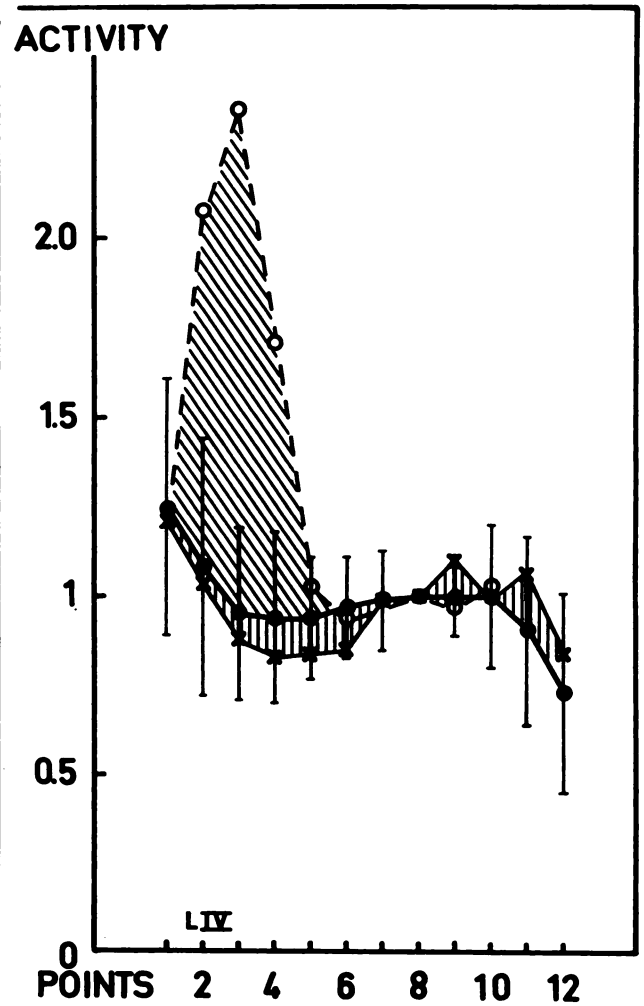

FIG. 4. Example of areas. x——x is curve of individual normal person. Numerical value of area is 60.5 o——o is curve of patient with nonspecific spondylitis of L-3. Numerical value of area is 286.

DISCUSSION

Peak method from Felländer and Lindberg (*1*). At peak values between 2.89 and approximately 0.28 there was clinical evidence of active disease in all but two patients. The high value in one of these could be explained by a focus which was not operated and which appeared to be active on reevaluation of the radiographs. The high value in the second case could not be explained. All spondylitis foci with normal counting values were judged clinically inactive or healed.

The group with counting values between 0.28 and normal was heterogeneous. Here there were foci with both obvious and doubtful clinical activity, foci which were clinically inactive and foci with continuing postoperative bone healing. It was this intermediate group in which judgment was difficult, and there were the problems of correlating the clinical findings in septic spondylitis with the results of isotope counting which led us to make use of methods of assessment other than the peak value. This intermediate group of patients is not included in the clinical material used in the statistical analysis of the methods of points, intervals or areas described above. Only patients with diagnoses considered certain were used. But because it is interesting to see the result when these three methods are used on all the clinical material, the intermediate group is included in the tabulations in the following section.

Methods of points, intervals and areas. The choice of cut-off points or ranges for each of the three diagnostic approaches will quite evidently be between 1.5 and 2.0 s.d. for each of the three types of measurement investigated (point, interval or area). For any of these criteria, the data in Tables 1–3 will confirm the good separation of normal and patient groups as judged by agreement or disagreement with clinical criteria. However, agreement and disagreement with clinical standards is not an absolute measurement. In several instances when disagreement with clinical judgment occurred consistently in all three methods of comparison, reinvestigation of the patient's status is clearly necessary.

This observation indicates a need for a second, more intense scrutiny of the data under analysis here. Each individual can be followed carefully through the three diagnostic methods and the agreements or disagreements tabulated; therefore it is useful to know whether agreement with clinical diagnosis can be established in most cases in two out of the three methods; if so, this will tend to include all three approaches as further diagnostic criteria.

A tabulation of agreements and disagreements with clinical evaluation is presented in Table 4. Point 2 normalization is used for foci in the thoracic spine and Point 8 normalization for foci in the lumbar spine.

From the summaries in Table 4 it can be seen that for tuberculous spondylitis 56 out of 65 patients are diagnosed correctly by two out of the three methods; for nonspecific spondylitis 24 out of 26 patients are diagnosed correctly by two out of the three methods.

Each of these methods of analysis—point, interval or area—can be automated and the data analyzed for comparison with the established norms by an entirely computerized process. Therefore the use of the three approaches would not be time-consuming.

From Tables 1–3, it is apparent that the method of point estimation and the method of area estimation provide the cleanest separation of patients from normals and therefore are methods of choice. As

TABLE 4. COMPARISON BETWEEN CLINICAL EVALUATION AND POINT INTERVAL AND AREA METHODS

	No. of countings	
Tuberculous spondylitis		
All methods agree with clinical diagnosis:	27	
Two out of three methods agree with clinical diagnosis:	29	Of those in *disagreement:* 16 instances for method of interval estimation; eight instances for method of point estimation; and five instances for method of area estimation.
One of three methods agree with clinical diagnosis:	6	Of those in agreement: one instance for method of area estimation; three instances for method of point estimation; and two instances for method of interval estimation.
None of three methods agree with clinical diagnosis:	3	Nor does method of peaks. (In one case clinical diagnosis now is shown to be wrong.)
Total	65	
Nonspecific spondylitis		
All methods agree with clinical diagnosis:	17	
Two out of three methods agree with clinical diagnosis:	7	Of those in *disagreement:* three instances for method of interval estimation; three instances for method of point estimation; and one instance for method of area estimation.
One out of three methods agree with clinical diagnosis:	1	Of those in agreement: one instance for method of area estimation.
None of three methods agree with clinical diagnosis:	1	Nor does method of peaks.
Total	26	

±1.0 s.d. is used for the Point 2 curve in the method of points.

±2.0 s.d. is used for the Point 8 curve in the method of points.

±1.5 s.d. is used for the method of intervals.

±2.0 s.d. is used for area 1.

±1.5 s.d. is used for area 2.

pointed out before, where each individual patient lies with respect to each of the two methods (and in fact, with respect to the method of interval estimation as well) is now of interest. It may be true that if the interval estimate is converted to a true slope measure, the separation of normal and patient groups with this approach is improved.

The somewhat less satisfactory separation with the method of interval estimation is not entirely a disadvantage, however. An examination of Table 2 shows that this method fails only in the cases of tuberculous spondylitis. Therefore a means of differential diagnosis between tuberculous and nontuberculous spondylitis can be achieved by comparing the results on a given individual when he is evaluated by the method of point estimation and the method of area estimation on the one hand, and the method of interval estimation on the other hand.

SUMMARY

In a previous article we have described a method to assess with the ^{85}Sr external counting whether an infectious spondylitis is active or healed.

In this paper we give a statistical analysis of four different possibilities to assess the results of the countings.

ACKNOWLEDGMENT

This work was supported in part by the U.S. Public Health Service through Research Grant FR-5495 from the General Support Branch, Division of Research Facilities and Resources, National Institutes of Health, and Research Grant AM-9982 from the National Institute of Arthritis and Metabolic Diseases.

REFERENCES

1. Felländer, M. and Lindberg, L.: Clinical use of radiostrontium in evaluation of spondylitis. *J. Bone Joint Surg.* **48A**:1,585, 1966.

2. Hult, L.: The Munkfors investigation. A study of the frequency and causes of the stiff neck-brachialgia and lumbago-sciatica syndromes, as well as observations on certain signs and symptoms from the dorsal spine and the joints of the extremities in industrial and forest workers. *Acta Orthopaed. Scand. Suppl.* **16**, 1954.

3. Hult, L.: Cervical, dorsal and lumbar spinal syndromes. A field investigation of a non-selected material of 1,200 workers in different occupations with special reference to disc degeneration and so-called muscular rheumatism. *Acta Orthopaed. Scand. Suppl.* **17**, 1954.