INTER- AND INTRATESTER RELIABILITY OF RADIOGRAPHIC MEASUREMENTS OF SPONDYLOLISTHESIS

G. CAPASSO[1], N. MAFFULLI[1,2], V. TESTA[1]

Thirteen radiographic methods to measure spondylolisthesis were assessed to determine :
1. whether or not the different methods were correlated with each other ;
2. their precision in detecting vertebral displacements ;
3. inter- and intrameasurer reliability among six orthopedic surgeons performing radiographic measurements.
Several methods were found to be correlated with each other in detecting spondylolisthesis on radiographs. However, on simulated displacements only three of them showed good correlation ($r \geq 0.95$, $p < 0.01$). Intra- and intergroup correlations among the orthopedic surgeons were equally good ($r > 0.90$; $p < 0.01$).

Keywords : radiography ; vertebral deformity ; olisthesis ; lumbosacral spine ; spondylolysis.
Mots-clés : radiographie ; déformation vertébrale ; olisthésis ; colonne dorso-lombaire ; spondylolyse.

INTRODUCTION

Spondylolisthesis is a form of vertebral instability (11), with most of the slipping occurring during the early stages of pubertal growth (1, 12, 23). Radiographic assessment is of primary importance in studying and monitoring its evolution (11, 15). Evaluation of a vertebral olisthesis is a relatively simple procedure. Much care must be put into it, especially when one is evaluating a patient in the initial stages of the disease. Only a few critical papers have been published on the terminology of spondylolisthesis and on the measurements to be applied in its monitoring (6, 26).

Radiographic measurements of olisthesis can be divided into three groups :

a. evaluation of the presence and entity of spondylolisthesis ;

b. evaluation of the general attitude of the lumbosacral spine ;

c. evaluation of the associated anatomopathological vertebral deformity.

Recently, interest has arisen in the accuracy and reproducibility of several methods for the radiographic measurement of vertebral deformities such as scoliosis (5, 10, 18) and spondylolisthesis (7, 8, 21).

Taking into account only the measurements cited in group A, we have found 13 measurement methods : Burckhardt (2), Capasso and Zanchini (3), Capener (4), Finneson (9), Laurent and Einola (12), Marique (14), Meschan (16), Meyerding (17), Newman (19), Pipino (20), Sim (22), Taillard (23), Ulmann (25). Each of them is based on different geometrical concepts, and the entity of the spondylolisthesis measured is thus different.

The purpose of this study was to report our experience with :

[1] University of Naples, First Medical School, First Institute of Orthopaedics and Traumatology, Via S. Andrea delle Dame, 4. 80138 Naples, Italy.

[2] The Hospital for Sick Children, Department of Orthopaedics, Great Ormond Street, London WC1N 3JH, United Kingdom.

Correspondence and reprints : N. Maffulli, Department of Orthopaedics, Newham General Hospital, Glen Road, Plaistow, London E13 8RU, United Kingdom.

a. the measurement methods' correlation among each other ;

b. their precision in detecting vertebral displacements ;

c. inter- and intratester reliability of the 6 orthopedic surgeons performing the measurements.

MATERIAL AND METHODS

Copies of 30 lateral standing radiographs of differing quality taken of patients who had been admitted for assessment and/or treatment of one-segment lumbar or lumbosacral spondylolisthesis were made from original films taken from the material of the clinical images archive of the First Institute of Orthopaedics, University of Naples. Six doctors divided in three groups of two, according to their stage and choice of career, took part to the study. They were :

a. fully trained clinical staff orthopedic surgeons ;
b. fully trained academic staff orthopedic surgeons ;
c. third-year residents.

The measurers were blinded as to the dates of the radiographs and the identities of the patients. They were instructed to measure each radiograph using in a random order all the 13 different methods outlined in the Introduction section. A small modification (the calculation of a percentage index) was applied to the Ulmann method (25). The participating doctors were asked to measure the radiographs as they would in a clinic, using their own markers and protractors. After each measurement, the marker lines were to be wiped, and another measurement to be carried out on the same radiograph. The result of each measurement was recorded by the measurer on a prestamped sheet which contained the identity of the measurer, the date of the measuring session, and an alphabetical list of all the methods. A second measuring session, using the same methods and radiographs, was performed within one week from the first.

The results of each measurement were entered in an IBM compatible PC and analyzed using a commercially available statistical program [Systat (13)]. Descriptive statistics were calculated. Results were analyzed using linear regression and one- and two-way analysis of variance (ANOVA). The Wilcoxon matched-pairs signed-ranks test was used to test the results of pairs of observers. Kappa statistics were used as a further index of the quality of measurement (24). Given the perceived need of great accuracy, significance level was set at $p < 0.01$.

Correlation among the different methods was calculated by entering the mathematical average of the results of each method for all the radiographs. Intergroup correlation was calculated using, for each group, the mathematical average of the measurements performed with the same method.

The measuring capability of all the methods used in this study, independently from the correlation among them, was tested as follows : A designer produced three anterior spondylolistheses of different entity, based on normal radiographs, moving L5 by a number of millimeters known only to him. All the 13 methods were used, for a total of 234 measurements.

RESULTS

A total of 2340 radiographical measurements were performed. All were considered for statistical analysis.

Intragroup variability of the orthopedic surgeons

The intragroup variability was low. Linear regression analysis gave a correlation of 0.92 (group A), 0.93 (group B) and 0.89 (group C) ($p = 0.005$). Kappa coefficients were within 0.70 and 0.75.

Correlation among the 13 methods

The Meschan (16) method, involving angular measurements, was excluded from statistical analysis as it is not homogeneous with the others. The Meyerding (117) and Newman (19) methods were excluded as they give only a grossly approximate value. The correlation among the remaining 10 methods is shown in table I.

Intergroup variability

All methods were used. Intergroup variability was acceptable (kappa coefficient ≥ 0.60). The results of linear regression analysis are shown in table II.

Accuracy of the methods

Intragroup variation was low ($0.93 < r < 0.95$), and so the results were averaged and tested as outlined in the intragroup variability section.

Table I. — Linear regression correlation among the different methods in measuring vertebral displacement in 30 radiographs

	1	2	3	4	5	6	7	8	9	10
1	1									
2	- 0.018	1								
3	0.92	0.056	1							
4	0.93	0.32	- 0.91	1						
5	0.23	0.80	0.48	- 0.08	1					
6	0.85	0.49	0.88	- 0.07	0.70	1				
7	0.79	0.22	0.96	- 0.08	0.68	0.88	1			
8	0.99	0.07	0.95	- 0.09	0.35	0.90	0.85	1		
9	- 0.16	- 0.46	- 0.52	0.17	- 0.88	- 0.52	- 0.73	- 0.64	1	
10	- 0.21	0.77	0.10	0.32	0.89	0.31	0.36	- 0.09	- 0.85	1

1 = Ulmann ; 2 = Pipino ; 3 = Marique ; 4 = Burckhardt ; 5 = Taillard ; 6 = Laurent and Einola ; 7 = Capener ; 8 = Finneson ; 9 = Capasso and Zanchini ; 10 = Sim.

Table II. — Linear regression correlation of intergroup measurements to assess a known olisthesis

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 vs. 2	0.55	- 0.90	0.95	- 0.57	- 0.54	0.86	1	0.87	0.95	0.64	0.77	0.11	1
1 vs. 3	0.97	0.87	0.92	- 0.94	0.26	0.93	1	- 0.80	0.82	0.56	0.77	- 0.59	1
2 vs. 3	0.57	- 0.87	0.99	- 0.71	0.74	0.98	1	0.92	0.96	0.98	0.90	- 0.65	1

1 = Ulmann ; 2 = Pipino ; 3 = Meschan ; 4 = Marique ; 5 = Burckhardt ; 6 = Taillard ; 7 = Newman ; 8 = Sim ; 9 = Capener ; 10 = Finneson ; 11 = Laurent and Einola ; 12 = Capasso and Zanchini.

The Finneson (9), Laurent and Einola (12), and Marique (14) methods are the ones better correlated ($r \geq 0.95$, $p = 0.003$) with the displacement known to the designer only. The Capener (4), Pipino (20) and Ulmann (25) methods are significantly correlated as well ($a \geq 0.90$, $p = 0.005$, kappa coefficient $0.75 \leq k \leq 0.82$). The other methods showed a lower correlation.

DISCUSSION

Thirteen radiographic methods, based on various geometrical principles, have been used to measure vertebral olisthesis. To the best of our knowledge, no study has been performed to compare them, although some studies have been carried out comparing intra- and inter-observer variation using the Boxall method (7) and 7 methods (21). The results of this study show that both intra- and intergroup variability among the measurers was low. Probably, as all the orthopedic surgeons came from the same training source, albeit at different stages of their careers, the quality of training itself was quite homogeneous, the likely variable being experience.

In the study of the correlation among the various methods, the somewhat surprising result is that the Capasso and Zanchini (3) and the Pipino (20) methods are not significantly correlated with any of the others. The first one is based on completely different principles, as it does not take into account the displacement of the margin of the vertebral body, but the misalignment of their centers. It is not possible to explain why the Pipino method shows a poor correlation with the others, as it is based on similar geometrical assumptions.

Third-year resdents show a slightly lower correlation with the other two groups. This could be

because of the length of their training. The trainees involved in this study were third-year postgraduate students, dedicating mainly to clinical duties. This may have practical implications in training. It may be advantageous to program some more time to be spent in developing the radiographic measurement ability of residents, as most of the measurements are performed by orthopedic surgeons. The environment where the measurements were performed is a privileged one, in that, being a university department, all orthopedic surgeons can dedicate more time to nonclinical and research duties, and much stress is put on the theoretical bases of orthopedics. Probably, in both a research and a clinical environment, special training should be given for radiographic measurements. In order to minimize interobserver variability, radiographical measurements should be performed by one or a few people only, possibly trained together and actually working together.

From the results of this study, the most precise and reliable methods are the Finneson (9), Laurent and Einola (12), and Marique (14) methods. At present, they are the ones routinely used in our department.

Acknowledgements

Many thanks are given to Mr. P. J. Webb, FRCS, Consultant Orthopaedic Surgeon, Royal National Orthopaedic Hospital, Stanmore, Middx, U.K., for helpful discussion, suggestions and critical reading of the manuscript. We are deeply indebted to Mrs. J. Baines-Preece for reviewing the manuscript.

REFERENCES

1. Boxall D., Bradford D. S., Winter R. B., Moe J. H. Management of severe spondylolisthesis in children and adolescents. J. Bone Joint Surg., 1979, 61-A, 479-485.
2. Burckhardt E. Spondylolisthesis. Schweiz. Med. Wochenschir., 1940, 70, 1093-1098.
3. Capasso G., Zanchini M. A new method for the radiographic evaluation of spondylolisthesis. Ital. J. Orthop. Traum., 1987, 13, 227-233.
4. Capener N. Spondylolisthesis. Brit. J. Surg., 1932, 19, 374-386.
5. Caran D. L., Browne R. H., Birch J. G. Measurement of scoliosis and kyphosis radiographs. Intraobserver and interobserver variation. J. Bone Joint Surg., 1990, 72-A, 328-333.
6. Dandy D. J., Shannon M. J. Lumbo-sacral subluxation (Group I spondylolisthesis). J. Bone Joint Surg., 1971, 53-B, 578-595.
7. Danielson B., Frennered K., Irstam L. Roentgenologic assessment of spondylolisthesis. I. A study of measurement variations. Acta Radiol., 1988, 29, 345-351.
8. Danielson B., Frennered K., Selvik G., Irstam L. Roentgenologic assessment of spondylolisthesis. I. An evaluation of progression. Acta Radiol., 1989, 30, 65-68.
9. Finneson B. E. Low back pain. Lippincott Company, Philadelphia, 1980.
10. Goldberg M. S., Poitras B., Mayo N. E., Labelle H., Bourassa R., Cloutier R. Observer variation in assessing spinal curvature and skeletal development in adolescent idiopathic scoliosis. Spine, 1988, 13, 1371-1377.
11. Kirkaldy-Willis W. H., Farfan H. F. Instability of the lumbar spine. Clin. Orthop., 1982, 165, 110-123.
12. Laurent L. E., Einola S. Spondylolisthesis in children and adolescents. Acta Orthop. Scand., 1961, 31, 45-64.
13. Leland W. SYSTAT : the system for statistics. Systat Inc., Evanston, IL, 1988, pp. 1-959.
14. Marique P. Le spondilolisthesis. Acta Chir. Belg., 1951, Suppl. 3, 3-89.
15. McPhee I. B., O'Brien J. P., McCall I. W., Park W. M. Progression of lumbosacral spondylolisthesis. Australas. Radiol., 1981, 25, 91, 95.
16. Meschan I. Radiographic study of spondylolisthesis with special reference to stability determination. Radiology, 1947, 47, 249-262.
17. Meyerding A. W. Spondylolisthesis. J. Bone Joint Surg., 1931, 13, 39-48.
18. Morrissy R. T., Goldsmith G. S., Hall E. C., Kehl D., Cowie G. H. Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. J. Bone Joint Surg., 1990, 72-A, 320-327.
19. Newman P. H. A clinical syndrome associated with severe lumbosacral subluxation. J. Bone Joint Surg., 1965, 47-B, 472-481.
20. Pipino F. La spondilolistesi. Classificazione, quadro clinico e radiografico, terapia incruenta e valutazione medico-legale. Bari, LV Congresso S.I.O.T., 1970.
21. Shaffer W. O., Spratt K. F., Weinstein J., Lehmann T. R., Goel V. The consistency and accuracy of roentgenograms for measuring sagittal translation in the lumbar vertebral motion segment. An experimental model. Spine, 1990, 15, 741-750.
22. Sim G. P. G. Vertebral contour in spondylolisthesis. Brit. J. Radiol., 1973, 46, 250-254.
23. Taillard W. Etiology of spondylolisthesis. Clin. Orthop., 1976, 117, 30-39.
24. Thompson W. D., Walter S. D. A reappraisal of the kappa coefficient. J. Clin. Epidemiol., 1988, 41, 949-958.
25. Ulmann W. Diagnostic line for determining subluxation of the fifth lumbar vertebra. Radiology, 1924, 2, 305-312.
26. Wiltse L., Winter R. B. Terminology and measurement of spondylolisthesis. J. Bone Joint Surg., 1983, 65-A, 768-772.

SAMENVATTING

G. CAPASSO, N. MAFFULLI en V. TESTA. Betrouwbaarheid van röntgentechnieken bij evaluatie van de spondylolisthesis.

Dertien röntgentechnieken, gebruikt voor de meting van spondylolisthesis, werden geevalueerd ter bepaling van :
— het al dan niet bestaan van een correlatie tussen de verschillende technieken
— de nauwkeurigheid bij het ontdekken van een verplaatsing
— de betrouwbaarheid van de technieken onderling en van de verschillende metingen, volgens een zelfde techniek, bij 6 orthopedische chirurgen die radiologische metingen van de spondylolisthesis uitvoeren.
Verschillende methoden bleken onderling goed gecorreleerd bij het radiologisch opsporen van spondylolisthesis. Bij gesimuleerde verplaatsingen echter bestond er een goede correlatie, alleen bij 3 technieken ($r \geq 0,95$, $p < 0,01$). De verschillende correlaties van de technieken onderling en van de verschillende metingen, volgens dezelfde techniek, bij de orthopedische chirurgen, was correct ($r > 0,90$; $p < 0,01$).

RÉSUMÉ

G. CAPASSO, N. MAFFULLI et V. TESTA. Fiabilité des méthodes radiographiques de mesure du spondylolisthésis.

Treize techniques radiographiques de mesure du spondylolisthésis furent évaluées pour déterminer :
— s'il existe, ou non, une corrélation entre les différentes méthodes
— leur précision dans la détection des déplacements vertébraux
— la fiabilité de différentes mesures dan le temps et des différentes techniques de mesure, employées par 6 chirurgiens orthopédistes qui évaluent radiographiquement le spondylolisthésis.
Les différentes méthodes de détection du spondylolisthésis présentent une bonne corrélation entre elles. Cependant, dans la détection des déplacements simulés, 3 seulement présentaient une corrélation valable ($r \geq 0,95$, $p < 0,01$). La corrélation des différentes mesures dans le temps et entre les différentes techniques, employées par différents chirurgiens orthopédistes, se révéla acceptable ($r > 0,90$; $p < 0,01$).