

The revised American Fertility Society classification of endometriosis: reproducibility of scoring*

John A. Rock, M.D.†‡
ZOLADEX Endometriosis Study Group§||

Department of Gynecology and Obstetrics, Division of Reproductive Endocrinology and Infertility, Emory University
School of Medicine, Atlanta, Georgia

Objective: To assess the reproducibility in staging endometriosis using the revised American Fertility Society (AFS) classification of endometriosis.

Design: Visual documentation of laparoscopies of 315 women with endometriosis was scored by the investigators and a blinded reviewer.

Setting: Patients from private practice institutional setting.

Participants: Patients who participated in a multicenter trial to study the efficacy and safety of a GnRH agonist (GnRH-a).

Interventions: Laparoscopic visual documentation of the extent of endometriosis before and after 6 months of GnRH-a therapy.

Main Outcome Measure: The reproducibility of the AFS classification system comparing scoring during laparoscopy and by a blinded reviewer.

Results: Good to fair agreement scoring endometriosis between the investigator and the blinded reviewer was noted.

Conclusions: Visual documentation may be used to determine the stage of endometriosis using the revised AFS classification guidelines. Fertil Steril 1995;63:1108-10

Key Words: Revised AFS classification of endometriosis, endometriosis, reproducibility

The modern era of classification of endometriosis can be dated to the 1970s when the advent of diagnostic laparoscopy, together with other technologi-

cally sophisticated diagnostic and research tools, provided the means to detect and document mild degrees of endometriosis with subtle appearances and unusual locations. These advances fostered expectations for the development of more detailed and precise definitions of the disease, well-documented

Received May 16, 1994; revised and accepted December 16, 1994.

* Supported by a grant from ZENECA Pharmaceuticals, Wilmington, Delaware.

† This manuscript was prepared by John A. Rock, M.D.

‡ Reprint requests and present address: John A. Rock, M.D., Department of Gynecology and Obstetrics, P.O. Box 21246, Emory University School of Medicine, Atlanta, Georgia 30322 (FAX: 404-727-8609).

§ ZOLADEX is a trademark, the property of ZENECA Limited, Wilmington, Delaware.

|| The ZOLADEX Endometriosis Study Group included the following investigators: E. James Aiman, M.D., Medical College of Wisconsin, Milwaukee, Wisconsin; David H. Barad, M.D., and Michael A. Feinman, M.D., Albert Einstein College of Medicine, Dobbs Ferry, New York; George Betz, M.D., Kaiser Medical Center, Denver, Colorado; W. Paul Dmowski, M.D., Elmhurst, Illinois; Phillip C. Galle, M.D., Southern Illinois University School of Medicine, Springfield, Illinois; Alvin F. Goldfarb, M.D., Philadelphia, Pennsylvania; Robert Hemmings, M.D., Royal Victoria

Hospital, Montreal, Quebec, Canada; Ekkehard Kemmann, M.D., University of Medicine and Dentistry of New Jersey, New Brunswick, New Jersey; Andre Lemay, M.D., Ph.D., and Rodolphe Maheux, M.D., Hospital St. Francois d'Assise, Quebec City, Quebec, Canada; L. Russell Malinak, M.D., Houston, Texas; George Maroulis, M.D., Tampa, Florida; Kamran Moghissi, M.D., and Kenneth Ginsburg, M.D., Hutzel Hospital, Detroit, Michigan; Eldon Schriock, M.D., and Sandra Carson, M.D., University Physicians Foundation, Memphis, Tennessee; Robert W. Shaw, M.D., and K. Geraldine McSweeney, M.D., The Royal Free Hospital, London, England; Melvin Taymore, M.D., Brookline, Massachusetts; Dan Tulchinsky, M.D., The Malden Hospital, Malden, Massachusetts; A. Albert Yuzpe, M.D., University Hospital, London, Ontario, Canada; and John A. Rock, M.D., Emory University School of Medicine, Atlanta, Georgia.

Table 1 Type and Quality of Laparoscopic Documentation*

Category	Prestudy (n = 203)	Week 24 (n = 204)	Total (n = 407)
Type of documentation			
Videotape	128 (63)	133 (65)	261 (64)
Photograph or slide	75 (37)	71 (35)	146 (36)
	n = 204	n = 204	n = 208
Quality of documentation			
Readable	88 (43)	83 (41)	171 (42)
Readable with difficulty	116 (57)	121 (59)	237 (58)

* Values are number of patients (n) with percentages in parentheses.

guidelines for treatment, and good predictive value for treatment outcome.

The revised American Fertility Society (AFS) classification of endometriosis (1), published in 1985, represented an important advance for endometriosis staging. Since its publication, the revised AFS classification of endometriosis has become widely accepted as the standard classification scheme, and it has been used in most studies reporting results of endometriosis staging (2). Despite the popularity of the revised AFS classification scheme, only one published study has examined its reproducibility (2).

The purpose of the present study was to determine the reproducibility of the revised AFS classification by investigators who performed scoring during laparoscopy and for an observer who performed scoring during a blinded review of visual documentation.

MATERIALS AND METHODS

The patient population consisted of 315 women with endometriosis who participated in a multicenter study comparing the efficacy and safety of the GnRH agonist goserelin acetate implant (ZOLADEX; ZENECA Pharmaceuticals Group, Wilmington, DE) and danazol (Danocrine; Sanofi Winthrop Pharmaceuticals, New York, NY).

Visual documentation was recorded on VHS videotapes, 35 mm photographs, or 35 mm slides at 17 sites. Videotape (64%) was the most popular method of recording (Table 1). During visual recording, all pelvic structures were visualized and all lesions were measured with a graduated probe to allow precise documentation of the implant size or the degree of ovarian enlargement. In addition, mobilization of the pelvis was required to determine the extent of disease.

Visual documentation was defined as unreadable or readable with difficulty when insufficient views of the pelvis were provided, measurements were not obtained with a graduated probe, or technical diffi-

culties provided videotapes, photographs, or slides of poor quality. Visual documentation that did not allow classification of extent of disease, as recommended in the revised AFS classification, was deemed unreadable. The remaining readable documentation was included in the analysis.

Laparoscopy was performed on 315 patients before therapy and after 6 months of therapy. Six patients with stage IV, revised AFS scored for endometriosis who were phase III study protocol violators were excluded from the analysis (3). Visual documentation was not provided for 105 patients. All documentation obtained at the pretherapy and post-therapy evaluations for the remaining 204 patients was submitted to a blinded reviewer; the reviewer classified the quality of the visual documentation as readable, readable with difficulty, and unreadable. Documentation considered by the blinded reviewer to be unreadable or readable with difficulty (58%) was excluded from analysis, because a classification of the extent of disease could not be accurately performed. Documentation classified as readable for 88 patients (43%) at the pretherapy evaluation and for 83 patients (41%) at the post-therapy evaluation was scored by the blinded reviewer and included in the analysis (Table 1).

The kappa statistic (4), which measures the association between two raters when data are on a categorical scale, was used to compare the assignment of the revised AFS classification performed by 22 participating investigators during laparoscopy and by the blinded reviewer. Characterization of the different range of values for kappa with respect to the degree of agreement they suggest were based on the estimates of Landis and Koch (5). Values of kappa > 0.75 represented excellent agreement beyond chance, values between 0.40 and 0.75 represented fair to good agreement beyond chance, values < 0.40 represented poor agreement beyond chance. An overall kappa was calculated to test the hypothesis that reproducibility is above chance expectation; a one-sided test was used to test the hypothesis that kappa = 0.

RESULTS

Table 2 presents the distribution of scores by disease stage according to the revised AFS classification for the investigators and the blinded reviewer. The kappa statistic was 0.44, indicating fair to good agreement. The overall kappa of 0.898 was significant ($P < 0.0001$).

DISCUSSION

The revised AFS classification is the most widely used staging system for endometriosis, yet the

Table 2 Joint Ratings of Endometriosis Scoring by Investigators and Blinded Reviewer According to the Revised AFS Classification*

Investigator Classification: Stage	Blinded Reviewer Classification: Stage			
	I	II	III	IV
I	27	11	2	0
II	7	25	12	1
III	2	9	41	18
IV	0	0	1	3

* A total of 159 pairs of ratings are included.

reproducibility of this system has not been established. A study by Hornstein and colleagues (2) represents the only other investigation to examine the reliability of the revised AFS classification. In that study, the authors determined the intraobserver and interobserver variability of the scoring system using videotapes of laparoscopies of patients with endometriosis viewed under standardized conditions by five observers. They reported that there was good reproducibility in three of five subcomponents of the system, but that intraobserver and interobserver variability were high for ovarian endometriosis and cul-de-sac subscores.

The present investigation was conducted to determine the reproducibility of the revised AFS classification when the extent of endometriosis was scored by investigators during laparoscopy and later by a blinded reviewer. To compensate for agreement by chance, the kappa statistic was applied to measure the association between the two raters. Our results suggest that there was reproducibility of the revised AFS classification when assignments to disease stage were made by the investigators and the blinded reviewer. The kappa statistic was 0.44, which indicated fair to good reproducibility beyond chance. In addition, the overall kappa was statistically significant ($P < 0.01$); its magnitude (0.898) indicated good agreement beyond chance.

Several limitations are evident with this type of

reproducibility study. Because this study was performed at multiple study sites, it was impossible to standardize the conditions for recording the visual documentation. As a result, the laparoscopies were recorded on three types of documentation (videotapes, photographs, or slides) and the documentation produced was of variable color and contrast. A second factor was that the blinded reviewer was limited by the quality of the visual documentation and the details recorded by each of the participating investigators. To address this problem, all documentation deemed inadequate to allow for classification was omitted from analysis. Finally, the kappa statistic is generally used to compare the association between two raters, but in this analysis it was used to compare the association between a host of raters (the investigators) and one rater (the blinded reviewer). Thus, the analysis did not provide a measure of association between individual investigators and the blinded reviewer. However, it was useful in providing an overall measure of the reproducibility between the investigators and the blinded reviewer.

Acknowledgments. Acknowledgment is given to David Guzick, M.D., and Sharon Petty, M.D., for review of the manuscript, Roger D. Aitchison, Ph.D., for statistical analysis, and Gary Dorrell, M.S., for editorial assistance.

REFERENCES

1. The American Fertility Society. Revised American Fertility Society classification of endometriosis. *Fertil Steril* 1985; 43:351-2.
2. Hornstein MD, Gleason RE, Orav J, Haas ST, Friedman AJ, Rein MS, et al. The reproducibility of the revised American Fertility Society classification of endometriosis. *Fertil Steril* 1993;59:1015-21.
3. Rock JA, Truglia JA, Caplan RJ, and the Zoladex Endometriosis Study Group. Zoladex (goserelin acetate implant) in the treatment of endometriosis: a randomized comparison with danazol. *Obstet Gynecol* 1993;82:198-205.
4. Fleiss JL. The measurement of interrater agreement. *Statistical methods for rates and proportions*. New York: John Wiley & Sons, 1981:212-36.
5. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.