Reproductive performance, pain recurrence and disease relapse after conservative surgical treatment for endometriosis: the predictive value of the current classification system

Paolo Vercellini1,4, Luigi Fedele2, Giorgio Aimi1, Olga De Giorgi1, Dario Consonni3 and Pier Giorgio Crosignani1

1First Department of Obstetrics and Gynaecology, 2Second Department of Obstetrics and Gynaecology, University of Milano and 3Istituti Clinici di Perfezionamento, Milano, Italy
4To whom correspondence should be addressed at: Clinica Ostetrica e Ginecologica, Istituto “Luigi Mangiagalli”, Università di Milano, Via Commenda, 12-20122 Milano, Italy. E-mail: paolo.vercellini@unimi.it

BACKGROUND: To assess the predictive value of the current classification of endometriosis in terms of response to surgical treatment, we studied to what extent disease stage, lesion type and lesion site were associated with post-operative pregnancy rate, symptom recurrence and disease relapse. METHODS: A total of 729 women with endometriosis undergoing first-line conservative laparoscopic surgery were included. Data on age at surgery, disease stage according to the revised American Fertility Society (AFS) classification, anatomical characteristics of endometriotic lesions, fertility status and types and severity of pain symptoms were collected. RESULTS: Minimal endometriosis was present in 222 patients, mild in 106, moderate in 197 and severe in 204. The cumulative probability of pregnancy at 3 years from surgery in 537 infertile women was 47% (51% at stage I, 45% at stage II, 46% at stage III and 44% at stage IV; log-rank test, $\chi^2 = 1.50, \ P = 0.68$). The cumulative probability of moderate or severe dysmenorrhoea recurrence in 425 symptomatic subjects was 24% (32% at stage I, 24% at stage II, 21% at stage III and 19% at stage IV; log-rank test, $\chi^2 = 6.39, \ P = 0.094$). The cumulative probability of disease relapse was 12% (3% at stage I, 11% at stage II, 11% at stage III and 23% at stage IV; log-rank test, $\chi^2 = 24.95, \ P = 0.0001$). Using Cox’s multivariate proportional hazards regression analysis, no association was observed between endometriosis stage or lesion type and lesion site and any of the considered study outcomes. CONCLUSIONS: The current classification of endometriosis has an inadequate predictive value with regard to the major clinical outcomes.

Key words: classification/endometriosis/infertility/pelvic pain/recurrence

Introduction

The management of various disorders, cancer in particular, benefits from the adoption of a classification. This allows immediate comprehension of the severity of the condition, helps to guide therapeutic strategies, permits the formulation of a reliable prognosis and is an indispensable scientific tool in clinical trials (Canis et al., 1993). A staging system should be based on the natural history of the disease, its local invasiveness, the presence of metastases and lymphatic involvement in the case of cancer or functional and organic consequences in the case of endocrine, metabolic or non-cancerous conditions. The sine qua non in designing a staging system is a proven progression through sequential steps of increasing severity that are causally linked with the outcome of interest (Canis et al., 1995).

Whether these criteria are met in the case of endometriosis is unclear. Moreover, the classification has not been demonstrated to be predictive with regard to response to treatment in terms of reproductive results, pain reduction and recurrence rate by stage (Brosens et al., 1993a). This is not only of academic interest but has considerable practical implications, for example, when deciding whether to seek a spontaneous conception or to undergo IVF–embryo transfer in infertile women, suggesting simple observation versus adjuvant treatments after conservative surgery in symptomatic subjects and advising against delays in the search for pregnancy when the risk of disease recurrence is particularly high (Canis et al., 1989, 1992; Palmisano et al., 1993; Schenken, 1998; Roberts and Rock, 2003).

In an attempt to disentangle the issue, we analysed the association between disease stage according to the revised American Fertility Society (AFS) classification (1985) and post-operative outcomes in a large cohort of women undergoing conservative surgery for endometriosis. It was deemed that by performing
survival analysis (which compares reliably the time-to-event at different stages) as well as multivariate analysis (which assesses the effect of the various lesions considered in the scoring system simultaneously) on an adequately sized study population, we would obtain robust evidence on which to base a definite evaluation of the predictive value of the currently adopted classification of endometriosis.

Materials and methods
We included in the analysis 729 consecutive women with any type of endometriotic lesion undergoing first-line conservative surgery via laparoscopy at the First Department of Obstetrics and Gynecology of the University of Milan and for whom complete follow-up information was available (Table I). Data were collected on age at surgery, BMI, parity, infertility and pain symptoms. Patients with a previous clinical or endoscopic diagnosis of endometriosis, or those with other diseases that might cause pelvic pain (chronic pelvic inflammatory disease, pelvic varices and genital malformations) or who participated in controlled clinical trials were excluded from this study. Other exclusion criteria were treatment for endometriosis other than non-steroidal anti-inflammatory drugs in the 6 months before study entry, previous abdominal surgery except appendectomy, a diagnosis of gastrointestinal, urological or orthopaedic diseases with potential pain irradiation to the pelvic area and known psychiatric disturbances. The overall median follow-up time was 32 months.

To evaluate the effect of surgery on reproduction, we considered 537 patients who had tried to conceive without success for more than 12 months and had no major causes of infertility other than endometriosis. Before the operation, these women had undergone a routine infertility workup, including plasma progesterone determinations in the luteal phase, hysterosalpingography and semen analysis of the partner. Women with persistent anovulation, bilateral tubal occlusion or severe semen abnormalities of the partner (in this case, \( <10 \times 10^6 \) sperm/ml, \(<30\%\) progressive motility at the third hour and \(<30\%\) typical forms) were excluded from the fertility evaluation, as were those with other diseases that might affect reproduction, such as leiomyomas or uterine malformations, and those who underwent IVF–embryo transfer post-operatively.

To evaluate variations in pelvic pain, we considered subjects with moderate or severe symptoms of over 6 months' duration before surgery. Each patient was asked to complete a questionnaire on the presence and severity of dysmenorrhoea, deep dyspareunia and menstrual pelvic pain graded according to a 0- to 3-point multidimensional categorical rating scale, modified from the one devised by Biberoglu and Behrman (1981), which defines dysmenorrhoea according to loss of work efficiency and need for bed rest (absence of pain, 0; some loss of work efficiency, mild, 1; in bed part of 1 day, occasional loss of work, moderate, 2 and in bed for 1 or more days, incapacitation, severe, 3), non-menstrual pain according to various degrees of discomfort and use of analgesics (absence of pain, 0; occasional pelvic discomfort, mild, 1; noticeable discomfort for most of the cycle, moderate, 2 and pain persisting during the cycle or requiring strong analgesics, severe, 3) and deep dyspareunia according to limitation of sexual activity (no discomfort, 0; tolerated discomfort, mild, 1; intercourse painful to the point of interruption, moderate, 2 and intercourse avoided because of pain, severe, 3).

The women were also requested to grade the severity of dysmenorrhoea, non-menstrual pelvic pain and deep dyspareunia using a 100-mm visual analogue scale (VAS), the left extreme of which indicates the absence of pain and the right one indicates the presence of pain as bad as it could be; a score of 1–50 was considered mild pain, 51–80 moderate pain and 81–100 severe pain. Threshold points defining different categories of pain were chosen based on a previous correlation analysis (Vercellini et al., 1996) with the Biberoglu and Behrman multidimensional categorical rating scale (1981). The severity of symptoms was dichotomised between pain absent (no or mild symptom) and pain present (moderate or severe symptom) on both scales.

Conservative surgery at laparoscopy was performed with a three- or four-puncture technique using mechanical instruments and electrosurgery only. Adhesions were sectioned with scissors, the ovaries were completely mobilized and endometriomas were evacuated, rinsed with normal saline and excised by means of counter-traction applied on the pseudo-capsule and normal gonadal cortex withatraumatic forceps. Peritoneal implants were excised or coagulated with low-power unipolar current. Haemostasis was finally achieved with limited application of bipolar current. Endometriosis was staged according to the revised AFS classification (1985), and the anatomical characteristics of lesions were described in detail. In the cases not personally operated by the authors, the original staging was verified by G.A. who systematically reviewed the surgical descriptions and diagrams made at the time of laparoscopy. Pictures and video recordings were not always available. Inclusion in stage I was not limited to women with subtle lesions only. The distribution of the study population according to disease stage is summarized in Table I.

Post-operatively, all patients were monitored regularly every 6 months. At each follow-up visit, a standard gynecologic examination and a transvaginal ultrasound scan were performed, the occurrence of pregnancy was recorded and any pain symptoms were evaluated on the two scales. Disease recurrence was recorded in terms of both ultrasonographic evidence of ovarian endometriomas (round-shaped cystic masses, with thick walls, regular margins, homogeneous low echogenic fluid content with scattered internal echoes and without papillary proliferations) and reoperation with histological confirmation.

Data management
Data were archived using Access 97 and then exported into SAS 6.12 (SAS Institute Inc., Cary, NC, USA) for analysis. The Kaplan–Meier
method was used to calculate the cumulative proportion of women who became pregnant, reported moderate or severe pain symptoms recurrence or presented endometriosis relapse. The curves obtained after subgrouping according to disease stage were compared by the log-rank test. The event dates used in the computation of time to pregnancy were the date of surgery and the date of the last menstrual period or follow-up visit. The event dates used in computing time to pelvic pain recurrence or disease relapse were the date of surgery and the date of recurrence of symptoms or of ultrasonographic demonstration of an endometrioma or reoperation or last follow-up visit or last menstrual period in case of conception, respectively. Because the Kaplan–Meier method allows comparison of subjects with different lengths of follow-up, the women who became pregnant were not excluded from the pain and disease recurrence evaluations.

The effect of various covariates on the time to pregnancy and the time to pain or disease recurrence was estimated using Cox’s proportional hazards model. Included in the regression equations were terms for age at surgery, BMI, revised AFS classification stage, revised AFS classification score for implants and for adhesions separately, peritoneal lesions of the anterior or posterior cul-de-sac, diameter of right and left ovarian endometriomas and post-operative medical treatment. Age at surgery, endometriosis stage, classification score and lesion diameter were treated as continuous variables to avoid over-parametrization. The same applied to diameter of lesions after the initial scoring. As a consequence, risk ratios (RR) for these variables relate to change in each specific unity. RR and the corresponding 95% confidence intervals (CI) were computed separately for the considered factors from data stratified for quinquennia of age. The significance of the main effects was tested by the Wald $\chi^2$ test (two-sided).

### Results

#### Fertility

Of the 537 women included in the analysis, 222 had no infertility factors other than endometriosis (median follow-up, 38 months), whereas 315 had additional minor infertility factors (median follow-up, 41 months). Post-operative medical treatment was used in 171/537 subjects (estrogen–progestogen combinations, 94 and gonadotrophin-releasing hormone agonists, 60; others, 17) in both the former (74/222) and the latter group (97/315). The crude pregnancy rate in women without other infertility factors in addition to endometriosis was 104/222, 47% (37/89, 42%, at stage I; 14/35, 40%, at stage II; 25/44, 57%, at stage III and 28/54, 52%, at stage IV). The crude pregnancy rate in those with other infertility factors was 141/315, 45% (49/118, 42%, at stage I; 19/50, 38%, at stage II; 36/68, 53%, at stage III and 37/79, 47%, at stage IV). The distribution of conceptions according to lesion type is reported in Table II.

#### Disease relapse

The overall cumulative probability of disease relapse at 3 years from surgery was 91/425, 21% (33/117, 28%, at stage I; 15/66, 23%, at stage II; 22/113, 20%, at stage III and 21/129, 16%, at stage IV). The overall cumulative probability of moderate or severe dysmenorrhoea recurrence at 3 years from surgery was 24% (32% at stage I, 24% at stage II, 21% at stage III and 19% at stage IV; log-rank test, $\chi^2 = 6.39$, $P = 0.094$). Cumulative dysmenorrhoea recurrence curves by stages are shown in Figure 2.

At multivariate proportional hazards regression analysis, the only covariate significantly associated with recurrence of dysmenorrhoea was age at surgery (Wald $\chi^2$ test = 3.86, $P = 0.049$; RR, 0.95, 95% CI, 0.91–1.00 and Wald $\chi^2$ test = 8.09, $P = 0.004$; RR, 0.94, 95% CI, 0.90–0.98, respectively). No other significant correlation was observed in either study group. Specifically, endometriosis stage was not associated with reproductive performance (no other infertility factors, Wald $\chi^2$ test = 0.16, $P = 0.68$; RR, 1.09; other infertility factors, Wald $\chi^2$ test = 0.12, $P = 0.73$; RR, 1.06). The same was true with regard to diameter of endometriotic ovarian cysts (data not shown) and use of post-operative medical treatment (Wald $\chi^2$ test = 0.18, $P = 0.67$; RR, 1.10 and Wald $\chi^2$ test = 0.27, $P = 0.60$; RR, 1.11, respectively).

#### Pain recurrence

A total of 425 subjects had moderate or severe dysmenorrhoea before surgery. The crude pain recurrence rate in these women was 91/425, 21% (33/117, 28%, at stage I; 15/66, 23%, at stage II; 22/113, 20%, at stage III and 21/129, 16%, at stage IV). The overall cumulative probability of moderate or severe dysmenorrhoea recurrence at 3 years from surgery was 24% (32% at stage I, 24% at stage II, 21% at stage III and 19% at stage IV; log-rank test, $\chi^2 = 6.39$, $P = 0.094$). Cumulative dysmenorrhoea recurrence curves by stages are shown in Figure 2.
was 12% (3% at stage I, 11% at stage II and 11% at stage III, 23% at stage IV; log-rank test, \( \chi^2 = 24.95, P = 0.0001 \)). Cumulative disease relapse curves by stages are shown in Figure 3.

At multivariate proportional hazards regression analysis, the only covariate significantly associated with recurrence of endometriosis was post-operative medical treatment (Wald \( \chi^2 \) test = 11.75, \( P = 0.0006 \); RR, 2.58, 95% CI, 1.50–4.42). No other significant correlation was observed. Again, endometriosis stage was not associated with the risk of disease recurrence (Wald \( \chi^2 \) test = 1.56, \( P = 0.212 \); RR, 1.33). This result was confirmed after removing post-operative medical treatment from the equation, with the hypothesis that it could have masked the effect of stage (Wald \( \chi^2 \) test = 3.21, \( P = 0.073 \); RR, 1.50).

**Discussion**

The current classification of endometriosis [AFS, 1985; American Society for Reproductive Medicine (ASRM), 1997] is based on visual findings, is simple and concise for easy implementation in routine practice and is sufficiently analytical and descriptive to allow clear comparison of anatomicopathological modifications with time or following treatment as well as unbiased circulation of clinical and scientific information with limited intra- and inter-observer variability (Buttram, 1985; Candiani, 1986; Canis et al., 1993, 1995; Hornstein et al., 1993; Rock, 1995; Lin et al., 1998). However, in addition to being an essential tool to give an accurate and immediate picture of the pelvic condition, such a scheme must possess intrinsic prognostic properties in terms of consistent and predictable reproductive outcomes according to the various stages of the system and in relation to specific treatments or expectant management (Rock et al., 1981; Adamson et al., 1982; Guzick et al., 1982; Palmisano et al., 1993; Schenken and Guzick, 1997). Moreover, the likelihood of symptomatic recurrence and disease relapse associated with the available therapeutic options should be reliably...
estimated (Buttram, 1987; Weitzman and Buttram, 1989; Groff, 1991; Olive, 1992; Damario and Rock, 1997; Hoeger and Guzick, 1997, 1999; Stovall et al., 1997).

There are several major unsolved problems relating to the design of a useful staging system for endometriosis. It is unclear whether endometriosis is a single disease with multiple pathological manifestations, that is, peritoneal, ovarian and deep (Brosens, 1994; Brosens et al., 1994a, 1994b) or if these lesions are expressions of different aetiologies (Nisolle and Donnez, 1997). In the latter case, it would be erroneous to devise a single instrument for what would be diverse disorders. Furthermore, what constitutes ‘less’ and ‘more’ disease is not straightforward. All the classifications are based on visual findings, but what we see is not all or always active endometriosis (Sturgis and Call, 1954; Brosens et al., 1993b; Dubuisson and Chapron, 1994; Brosens, 1997; Evers et al., 2005; Marchino et al., 2005). Many lesions are consequences of previous implants now healed (Brosens et al., 1985; Brosens, 1993) or, as in the case of ovarian cysts, their diameter may not be proportional to the amount of active endometrium present (Brosens, 1994; Brosens et al., 1994b).

Moreover, the behaviour of endometriosis is unpredictable, sometimes self-limiting its spread or even regressing (Brosens et al., 1994a; Bergqvist, 1995; Moen, 1995). If and when the disease progresses, the sequence may not necessarily correspond to the gross lesions that form the basis for defining stage gradient. Furthermore, it is very difficult to stage a disorder with reference to more than one outcome, that is, fertility, pain and recurrence. This would require the demonstration that different lesions carry different but consistent prognoses. Such a supposition may be unfounded, because the foci that cause pain may not necessarily also affect fertility and vice versa (AFS, 1993; Rock, 1993; Vercellini, 1997). Finally, if scores are attributed arbitrarily and not empirically derived (Guzick et al., 1982), the concept of a point system itself may be criticized, because the same bias could have a systematic and repeated effect on diagnosis and classification of the disorder.

Based on the present results, we cannot reject the hypothesis that the AFS/ASRM classification system has inadequate power in discriminating between clinical conditions with different long-term outcomes and hence has a limited role in the formulation of a reliable prognosis. Most of the surgical procedures in this study were performed under the supervision of a senior author, and there were no exclusions apart from those specified in the study protocol. One expert research fellow (G.A.) reviewed all surgical records and staging forms. Only the revised AFS system (1985) was adopted, because the more recent ‘revised’ ASRM scheme (1997) is identical with regard to attributable points as well as threshold scores between stages, the only difference being the identification of different morphological lesion types (red, white and black) which, however, are not given specific scores. Accordingly, diagnostic and classification biases are unlikely. Moreover, the surgeons who recorded the clinical data were unaware of the hypothesis of the study. Admittedly, the exclusion of patients with incomplete follow-up information could have introduced a selection bias, because it is not possible to exclude that these subjects have different characteristics or different outcomes.

The current classification of endometriosis was not originally validated for surgical difficulty, and it is a clinical tenet that it has been specifically devised for infertile women (Garry, 2004). However, we did not observe differences in pregnancy rates across the four disease stages. This is in agreement with the findings of previous studies (Guzick et al., 1997; Fujishita et al., 2002) and is most probably the result of a lack of effect of the various components of the scoring system on the likelihood of pregnancy. Specifically, the presence of endometriotic ovarian cysts on one or both ovaries did not influence reproductive performance independently of their diameter. This may partly explain the negligible predictive value of the classification. In fact, endometrioma diameter is decisive in stage attribution according to the present system (AFS, 1985; ASRM, 1997).

Patient age at surgery was the only independent factor significantly associated with a reduction in the likelihood of
conception. Unexpectedly, cumulative pregnancy rates were similar in women with or without other infertility factors in addition to endometriosis. This is at odds with the findings of Guzick et al. (1997) who did not demonstrate an effect of age but observed a negative impact of a male factor. Our data confirm that post-operative medical treatment is of no benefit in terms of reproductive outcome (Vercellini et al., 2003).

Women who underwent IVF were excluded from the analysis because our aim was to assess the effect of stage and surgery specifically in women seeking conception spontaneously. If the worst cases were shifted more readily to IVF, potential differences in reproductive prognosis could have been reduced. However, IVF was never suggested in the post-operative period unless major tubal damage was observed at laparoscopy.

The classification of endometriosis was not even predictive of moderate or severe dysmenorrhoea recurrence. This prospective evidence is in line with the results of several studies that could not demonstrate retrospectively an association between different endometriosis stages, scores or lesions and frequency and severity of pain (Fedele et al., 1990, 1992; Vercellini et al., 1991, 1996; Parazzini et al., 2001). The limited number of post-operative events prevented meaningful multivariate analyses of both deep dyspareunia and non-menstrual pain recurrence. Interestingly, the de-novo appearance of any pelvic pain symptom was rare.

Data on disease relapse were somewhat contradictory. Survival analysis demonstrated significant differences in endometriosis recurrence rate between stages. However, neither stage nor separate scores for active lesions and adhesions had any impact on multivariate Cox's proportional hazard analysis. The substantial risk increase, observed with the use of post-operative medical therapy, should be considered with caution as it is most likely because, given the non-experimental setting, adjuvant treatment was used in patients with the most extensive and invasive disease, a condition which could be per se prone to relapse. Generally speaking, the incidence of disease persistence/recurrence observed in our series compares favourably with literature data (Wheeler and Malinak, 1987; Redwine, 1996).

The most probable explanation of the results of our studies, as well as of previous studies, lies in the hypothetical biometrical drawbacks inherent to the system itself, that is, the predictive value of the classification would be acceptable only if different lesions are considered (Wiegierinck et al., 1993), different points are assigned and different threshold scores are identified. However, there are alternative hypotheses. We, like other authors, did not evaluate the impact on fertility of the presence of various lesions per se, which constitutes the basis on which staging is based, but the impact of removal of those lesions, that is, response to treatment only. Owing to the absence of an untreated control group, it cannot even be excluded that surgery for endometriosis, in contrast with surgery for cancer, is more beneficial in severe than in minimal or mild disease (Milingos et al., 2006). Alternatively, the hypothesised gradient of effect of lesions based on their type, site and dimension may be an erroneous belief.

One fact stands out amidst the uncertainties surrounding endometriosis and its staging: a substantial modification of the currently adopted classification system based on robust epidemiological evidence is warranted, with inclusion of not only organic consequences of disease invasiveness (Chapron et al., 2003) but also functional ones (Garry, 2004; Vercellini et al., 2004).

References


Submitted on June 5, 2003; resubmitted on April 4, 2006; accepted on April 26, 2006