

ORIGINAL ARTICLE

Diagnosis of pelvic inflammatory disease: time for a rethink

I Simms, F Warburton, L Weström

Sex Transm Infect 2003;**79**:491–494

Objectives: To critically evaluate the available evidence base concerned with the diagnosis of pelvic inflammatory disease (PID) based on clinical presentation, and to investigate the relation between signs and symptoms and the presence of laparoscopically diagnosed PID using the largest available dataset.

Methods: The evidence base was critically evaluated and data collected by Lund University between 1960 and 1969 were used to compare clinical presentation with the results of laparoscopic investigation. Three techniques were used in this investigation—sensitivity and specificity, likelihood ratios, and discriminant analysis.

Results: None of the variables (abnormal vaginal discharge, fever $>38^{\circ}\text{C}$, vomiting, menstrual irregularity, ongoing bleeding, symptoms of urethritis, rectal temperature $>38^{\circ}\text{C}$, marked tenderness of pelvic organs on bimanual examination, adnexal mass, and erythrocyte sedimentation rate ≥ 15 mm in the first hour) had both high specificity and sensitivity—most had low specificity and sensitivity. There was little variation in either the likelihood ratios or the post-test probabilities between the variables. The lowest likelihood ratio (0.97) produced a post-test probability of 78% (95% CI: 74% to 81%) whereas the highest (1.73) had a post-test probability of 84% (95% CI: 81% to 87%). The pretest probability of having PID based on the presence of lower abdominal pain was 79% (95% CI: 76% to 82%). The discriminant analysis indicated that three variables significantly influenced the prediction of the presence of PID: erythrocyte sedimentation rate ($p < 0.0001$), fever ($p < 0.0001$), and adnexal tenderness ($p < 0.0001$). These variables correctly classified 65% of patients with laparoscopically diagnosed PID (95% CI: 61% to 69%).

Conclusion: There is insufficient evidence to support existing diagnostic criteria, which have been based on a combination of empirical data and expert opinion. A new evidence base is urgently needed but this will require either a new investigation of the association between clinical presentation and PID based on a laparoscopic “gold standard” or the development of new diagnostic techniques.

See end of article for authors' affiliations

Correspondence to: Ian Simms, HPA Communicable Disease Surveillance Centre, 61 Colindale Avenue, London NW9 5EQ, UK; ian.simms@hpa.org.uk

Accepted for publication 12 May 2003

The treatment of pelvic inflammatory disease (PID) has recently been discussed but the evidence base supporting PID diagnosis has largely been ignored.¹ In the United Kingdom, diagnosis in primary care, sexually transmitted disease, and obstetrics and gynaecology (O&G) clinics is focused on syndromic diagnosis and the exclusion of competing diagnoses. Recommended diagnostic criteria are based on the definition proposed by Hager which, in turn, is based on a combination of empirical data and expert opinion.² The problem with the definition is that, although signs and symptoms may be diagnostic markers, none is pathognomonic. The accuracy with which signs and symptoms predict the presence of PID has been evaluated using a laparoscopic “gold standard.” However, interpretation of the evidence base has been flawed and needs to be re-evaluated so that new directions in diagnosis can be formulated objectively. Here we critically evaluate the evidence base and discuss problems of interpretation.

ASSESSMENT OF EVIDENCE BASE

A Medline search was undertaken. Seven studies were found where laparoscopy had been used as the gold standard, most of which had been included in a previous review.^{3–10} The limited evidence base is not surprising as such studies are difficult to undertake owing to the high cost, associated risks, and infrequent use of laparoscopy. Weaknesses can be seen if the studies are compared against a sample size calculation. Assuming a 5% level of significance, 80% power, and a minimum detectable difference of 5%, the number of positive (laparoscopically diagnosed PID) and negative (non-laparoscopically diagnosed PID) patients would be required at the following sensitivities: 70% (323 each of positives and negatives, total = 646), 80% (246 each, total = 492),

90% (138 each, total = 276). Consequently, most of the published studies are too small to accurately detect real differences within the data (a type two statistical error) (table 1). The insufficient sample size is reflected in the wide confidence intervals (CI) that surround estimates of specificity and sensitivity derived from these studies. For example, in a study of women with clinically and laparoscopically diagnosed PID, erythrocyte sedimentation rate (ESR) had a sensitivity of 83% (95% CI: 52% to 98%) and a specificity of 45% (95% CI: 24% to 68%).⁴ Most studies are too small to be used as an evidence base for the formulation of diagnostic criteria and meta-analysis cannot be used because of difficulties in accessing data, reconciling selection criteria and data collection methods, together with differences in diagnostic methodologies and intraobserver error. Nevertheless, it is unlikely that large-scale studies will be undertaken for the foreseeable future and this emphasises the importance of the existing evidence base. Here, three analytical techniques were used to compare the accuracy with which clinical presentation predicted the presence of laparoscopically diagnosed PID.

PATIENTS AND METHODS

The anonymised dataset included women who attended the department of O&G, Lund University Hospital, with suspected PID between 1960 and 1984.¹⁰ This analysis was confined to first episodes of suspected PID collected between 1960 and 1969—the period for which the largest number of clinical parameters was available. All patients included in the study had an initial diagnosis based on clinical presentation (signs and symptoms). The minimum criteria were lower quadrant bilateral abdominal or pelvic pain of less than 3 weeks' duration, together with two or more of the

Table 1 Summary of studies that have examined the relation between clinical and laparoscopic findings in suspected PID cases

Clinic population	Country	Sample size (laparoscopically diagnosed PID/non-laparoscopically diagnosed PID)	Author
O&G	Sweden	716 (532/184)	Jacobson* 1969 ³
O&G	Finland	35 (26/9)	Lehtinen 1986 ⁷
O&G	Sweden	552 (414/138)	Hadgu* 1986 ⁵
O&G/A&E/STD	USA	36 (22/14)	Wasserheit 1986 ⁴
O&G	Finland	41 (31/10)	Paavonen 1989 ⁶
O&G	USA	176 (134/42)	Morcos 1993 ⁸
O&G	Czech Republic	141 (43/98)	Cibula 2001 ⁹

*Both studies use data collected by Lund University.¹¹

A&E=accident and emergency; O&G=obstetrics and gynaecology; STD=sexually transmitted disease.

following: abnormal vaginal discharge, fever $>38^{\circ}\text{C}$, vomiting, menstrual irregularity, ongoing bleeding, symptoms of urethritis, rectal temperature $>38^{\circ}\text{C}$, marked tenderness of pelvic organs on bimanual examination, adnexal mass, and ESR ≥ 15 mm in the first hour. Laparoscopy was used to verify clinical diagnosis.¹¹ For the purposes of this analysis, the data were divided in two: (1) laparoscopically diagnosed PID, and (2) non-laparoscopically diagnosed PID. These groups were compared in terms of age using the *t* test, whereas the number of births before index laparoscopy, and whether an IUD had ever been used, were compared using the χ^2 test.

Three methods were used to explore the relation between clinical presentation and presence of laparoscopically diagnosed PID. Firstly, the specificity and sensitivity of individual variables were assessed together with 95% CIs. Secondly, likelihood ratios were used to assess whether the presence of individual variables altered the index of suspicion based on the pretest probability.¹² Thirdly, forward stepwise discriminant analysis, a method of finding the combination of variables that most effectively separate populations, was used to determine which of the variables best predicted the presence of laparoscopically proved PID (SPSS PC).¹³

RESULTS

A total of 623 patients were included in the analysis; 494 patients were laparoscopically confirmed as having PID and 129 were not. There was no statistically significant difference

between these groups in terms of age ($p = 0.649$), number of pregnancies ($p = 0.447$), births ($p = 0.375$), and whether an IUD had been either used ($p = 0.675$) or inserted within 6 weeks of the index laparoscopy ($p = 0.100$).

None of the variables had both high specificity and sensitivity (table 2). Some achieved high sensitivity (tenderness of pelvic organs on bimanual examination, ESR) or high specificity (proctitis symptoms, vomiting), but most had low specificity and sensitivity.

The pretest probability was 79% (494/623), 95% CI: 76% to 82%. All the likelihood ratios were positive and there was little variation in either the likelihood ratios or the post-test probabilities between the variables (table 2). For example, the lowest likelihood ratio (0.97) produced a post-test probability of 78% (95% CI: 74% to 81%) whereas the highest likelihood ratio (1.73) had a post-test probability of 84% (95% CI: 81% to 87%). Consequently, for all the variables studied the post-test probability was not significantly different from the pretest probability.

The discriminant analysis indicated that three variables significantly influenced the prediction of the presence of PID: ESR (correlation value = 0.669; $p < 0.0001$), fever (CV = 0.584; $p < 0.0001$), and adnexal tenderness (CV = 0.540; $p < 0.0001$). These variables correctly classified 65% of patients with laparoscopically diagnosed PID (95% CI: 61% to 69%) (table 3). The other variables did not reach significance—that is, their presence did not increase the probability that a patient had PID.

Table 2 Prediction of laparoscopically diagnosed PID: sensitivity and specificity of signs and symptoms, likelihood ratios and post-test probabilities (pretest probability = 79%)

Signs and symptoms	Sensitivity (%)	Specificity (%)	Laparoscopically diagnosed PID			
	(95% CI)	(95% CI)	Present (n = 494)		Likelihood ratio* (positive)	Post-test probability
			No (%)	Absent (n = 129) No (%)		
Vaginal discharge	74 (69.99 to 77.90)	24 (16.95 to 32.34)	366 (74)	98 (76)	0.98	0.79
Fever	47 (42.49 to 51.47)	64 (55.43 to 72.58)	234 (47)	47 (36)	1.30	0.83
Vomiting	14 (11.03 to 17.34)	88 (81.55 to 93.34)	68 (14)	16 (12)	1.11	0.81
Menstrual irregularity	45 (40.49 to 49.45)	57 (48.36 to 66.03)	223 (45)	56 (43)	1.04	0.80
Ongoing bleeding	25 (21.34 to 29.17)	77 (68.49 to 83.73)	124 (25)	29 (22)	1.12	0.81
Urinary symptoms	35 (30.81 to 39.41)	64 (55.43 to 72.58)	173 (35)	46 (36)	0.98	0.79
Proctitis symptoms	10 (7.43 to 12.90)	92 (86.21 to 96.22)	50 (10)	10 (8)	1.31	0.83
Tenderness of pelvic organs on bimanual examination	99 (97.65 to 99.67)	0.007 (<0.001 to 2.84)	489 (99)	128 (99)	1.00	0.79
Palpable adnexal mass or swelling	52 (47.52 to 56.51)	70 (61.06 to 77.54)	258 (52)	39 (30)	1.73	0.84
Erythrocyte sedimentation rate ≥ 15 mm in 1st hour	81 (77.23 to 84.34)	33 (25.28 to 42.17)	402 (81)	86 (66)	1.22	0.82

*Likelihood ratio interpretation: >10 and <0.1 (large difference between pretest and post-test probability), 5–10 and 0.1–0.2 (moderate), 2–5 and 0.5–0.2 (small), 1–2 and 0.05–1 (small and rarely important).¹⁴

Table 3 Classification of results from forward stepwise discriminant analysis

		*Predicted group membership		
		Absent	Present	Total
Laparoscopically diagnosed PID	Absent	85	44	129
	Present	174	320	494
	Total	258	366	623

*Patients with erythrocyte sedimentation rate, adnexal tenderness, and fever.

DISCUSSION

High diagnostic accuracy is essential for effective patient management. It determines the quality of surveillance and epidemiological studies which, in turn, influence the efficiency of control and prevention strategies. "Lower abdominal pain plus two or more symptoms and signs" or "lower abdominal pain, adnexal tenderness, and cervical motion tenderness" are widely recommended diagnostic criteria for PID, but are not supported by the evidence base.¹⁵⁻¹⁶ The Lund study is the only investigation of sufficient size to act as an evidence base, and has been used as the primary source of data for the formulation of PID diagnostic guidelines.¹¹ However, a number of problems are associated with its use. Firstly, there is a temporal bias as the data were collected in Sweden between 1960 and 1967. At that time the dominant cause of PID was *Neisseria gonorrhoeae*. It is thought that gonococcal PID is more symptomatic than chlamydial PID. Consequently, where the prevalence of *Chlamydia trachomatis* is high and that of *N gonorrhoeae* low, as it is in most industrialised countries today, the specificity and sensitivity of clinical parameters are likely to be lower than in the Lund dataset. Secondly, data were based on women with acute PID attending O&G, whereas today most cases are diagnosed in primary care where cases have mild/unspecific symptoms. And, thirdly, laparoscopy may lack sensitivity and specificity when compared to fimbrial biopsy and plasma cell endometritis, as it may not identify mild intratubal inflammation and cannot detect endometriosis.

Analysis of the Lund data showed that, after exclusion of competing diagnoses, women with lower abdominal pain had a high pretest probability of having laparoscopically diagnosed PID. Analyses of the other clinical variables showed that specificity and sensitivity were inappropriate measures because of the wide CIs. The use of likelihood ratios showed that the post-test probability was not significantly different from the pretest probability. In contrast, the discriminant analysis, which took all the variables into consideration, clearly showed that only the presence of ESR, fever, and adnexal tenderness influenced pretest probability. These findings again emphasise the limitations of the Lund dataset for the formulation of diagnostic guidelines as today few clinics use ESR as a diagnostic criterion and few cases present with fever. The most effective diagnostic criteria are based on the presence of lower abdominal pain and exclusion of competing diagnosis, which justifies the high index of suspicion considered acceptable on the grounds that early intervention prevents sequelae. However, early antibiotic treatment may also be associated with elevated risk of increased antibiotic resistance, potential side effects (such as candidosis), and unnecessary patient anxiety caused by the diagnosis of a condition that is largely associated with sexually transmitted infections.

This study showed that there is insufficient evidence to support existing diagnostic guidelines. A new evidence base is urgently needed but this will require either a new investigation of the association between clinical presentation

and PID based on a laparoscopic gold standard, or the development of new diagnostic techniques. A specifically designed study of clinical presentation and PID would be very costly and time consuming. And, because signs and symptoms are not pathognomonic, it would add little to current knowledge. A variety of diagnostic techniques have been used, including pelvic imaging techniques, such as transvaginal ultrasound (with or without Power Doppler) and magnetic resonance imaging, and fimbrial and endometrial biopsy. However, the quality of evidence supporting these techniques is variable as some are based on small scale, observational studies that were only undertaken at a single location.¹⁷ In addition, they require equipment not generally available within primary care and genitourinary medicine settings in the United Kingdom. The diagnostic problem presented by PID can only be resolved by the development of a simple laboratory test that can accurately diagnose PID. Vaginal white blood cell count (WBC) has been suggested as a sensitive marker of upper genital tract infection but again this may be a marker of other pathologies.¹⁴ Other inflammatory mediators, such as cytokines and interferons, such as TNF- α and IFN- γ , need to be investigated either in the cervix or the endometrium.

PID is a leading cause of reproductive ill health in women. A substantial burden of PID is thought to exist in reproductive age women although little is known of its epidemiology in England. The chief medical officers expert advisory group on genital *C trachomatis* infection recently highlighted the urgent need for information concerning PID epidemiology.¹⁸⁻¹⁹ Accurate diagnosis is key to achieving these goals but, as this study has shown, this needs to be re-evaluated if accurate information is to be gathered.

ACKNOWLEDGEMENT

We thank Dr Susan Hillis, Centres for Disease Control (Atlanta, USA) for supplying data used in this study.

CONTRIBUTORS

IS instigated the project, undertook the analyses with FW, and wrote the paper; FW advised on the statistical techniques, undertook the discriminant analysis, and wrote the statistics section; LW collected the original dataset, supplied the data, and advised on the interpretation of the data results.

Authors' affiliations

I Simms, HPA Communicable Disease Surveillance Centre, London, UK
 F Warburton, HPA Statistics Unit, London, UK
 L Weström, Department of Obstetrics and Gynaecology, University of Lund, Lund, Sweden

REFERENCES

- Ross J. Extracts from "Clinical Evidence": pelvic inflammatory disease. *BMJ* 2001;**322**:658-9.
- Hager W, Eschenbach D, Spence M, et al. Criteria for diagnosis and grading of salpingitis. *Obstet Gynecol* 1983;**61**:113-4.
- Jacobson L, Weström L. Objectivised diagnosis of acute pelvic inflammatory disease. *Am J Obstet Gynecol* 1969;**105**:1088-98.

- 4 **Wasserheit J**, Bell T, Kiviat N, *et al.* Microbial causes of proven pelvic inflammatory disease and efficacy of clindamycin and tobramycin. *Ann Intern Med* 1986;**104**:187–93.
- 5 **Hadgu A**, Weström L, Brookes C, *et al.* Predicting acute pelvic inflammatory disease: a multivariate analysis. *Am J Obstet Gynecol* 1986;**155**:954–60.
- 6 **Paavonen J**, Miettinen A, Heinonen PK, *et al.* Serum CA 125 in acute pelvic inflammatory disease. *Br J Obstet Gynaecol* 1989;**96**:574–9.
- 7 **Lehtinen M**, Laine S, Heinonen PK, *et al.* Serum C-reactive protein determination in acute pelvic inflammatory disease. *Am J Obstet Gynecol* 1986;**154**:158–9.
- 8 **Marcos R**, Frost N, Hnat M, *et al.* Laparoscopic versus clinical diagnosis of acute pelvic inflammatory disease. *J Reprod Med* 1993;**38**:53–6.
- 9 **Cibula D**, Kuzel D, Fucikova Z, *et al.* Acute exacerbation of recurrent pelvic inflammatory disease. *J Reprod Med* 2001;**46**:49–53.
- 10 **Kahn J**, Walker C, Washington A, *et al.* Diagnosing pelvic inflammatory disease. *JAMA* 1991;**266**:2594–604.
- 11 **Weström L**, Joeseof R, Reynolds G, *et al.* Pelvic inflammatory disease and fertility. *Sex Transm Dis* 1992;**19**:185–92.
- 12 **Jaeschke R**, Guyatt G, Sackett D. Users' guide to the medical literature. *JAMA* 1994;**271**:703–7.
- 13 **SPSS Inc.** *SPSS for windows, release 10.0.5.* Chicago: SPSS Inc, 1999.
- 14 **Peipert J**, Boardman L, Hogan J, *et al.* Laboratory evaluation of acute upper genital tract infection. *Obstet Gynecol* 2002;**87**:730–6.
- 15 **Clinical Effectiveness Group.** UK national guidelines on sexually transmitted infections and closely related conditions. *Sex Transm Infect* 1999;**75**:S1–88.
- 16 **Templeton A.** *The prevention of pelvic infection.* London: RCOG Press, 1996.
- 17 **Munday PE.** Pelvic inflammatory disease—an evidence-based approach to diagnosis. *J Infect* 2000;**40**:31–41.
- 18 **Department of Health.** *Chlamydia trachomatis. Summary and conclusions of CMO's expert advisory group.* London: DoH, 1998.
- 19 **Simms I**, Hughes G, Catchpole M. Screening for Chlamydia trachomatis. *BMJ* 1998;**317**:680–1.

ELECTRONIC PAGES

In practice



Sexually Transmitted Infections receives a number of articles relating to practice and operational issues that, although important locally, cannot necessarily be extrapolated into a generalised experience. For this reason we will publish these articles, after peer review, in full through eSTI. The paper edition of the journal will feature full abstracts in the “In practice” section.