

# Genetic factors contribute to the risk of developing endometriosis

H.Stefansson<sup>1,4</sup>, R.T.Geirsson<sup>2</sup>, V.Steinthorsdottir<sup>1</sup>, H.Jonsson<sup>1</sup>, A.Manolescu<sup>1</sup>, A.Kong<sup>1</sup>, G.Ingadottir<sup>3</sup>, J.Gulcher<sup>1</sup> and K.Stefansson<sup>1</sup>

<sup>1</sup>DeCode Genetics, Lynghals 1, Reykjavik, IS-110, <sup>2</sup>Department of Obstetrics and Gynecology, Landspítalinn University Hospital, Reykjavík, IS-101 and <sup>3</sup>Genetic Research Service Center, Nóatún 17, Reykjavik, IS-110, Iceland

<sup>4</sup>To whom correspondence should be addressed. E-mail: hreinn@decode.is

**BACKGROUND:** Endometriosis is known to cluster within nuclear families. The extent of familial clustering can be evaluated in Iceland with its large population-based genealogical database. **METHODS AND RESULTS:** Applying several measures of familiarity we demonstrated that 750 women with endometriosis were significantly more interrelated than matched control groups. The risk ratio for sisters was 5.20 ( $P < 0.001$ ) and for cousins 1.56 ( $P = 0.003$ ). The average kinship coefficient for the patients was significantly higher than that calculated for 1000 sets of 750 matched controls ( $P < 0.001$ ) and this remained significant when contribution from first-degree relatives was excluded ( $P < 0.05$ ). The minimum number of ancestors required to account for the group of patients was compared with the minimum number of ancestors required to account for the control groups at different time points in the past. The minimum number of founders for the group of patients was significantly smaller than for the control groups. Affected cousin pairs were as likely to be paternally connected as maternally connected. **CONCLUSIONS:** This is the first population-based study using an extensive genealogy database to examine the genetic contribution to endometriosis. A genetic factor is present, with a raised risk in close and more distant relatives, and a definite kinship factor with maternal and paternal inheritance contributing.

*Key words:* endometriosis/familial clustering/genetic susceptibility/heritability/population genetics

## Introduction

There is a growing body of evidence for a role for genetic factors in the pathogenesis of endometriosis, based largely on reports of familial clustering and increased prevalence among first-degree relatives (Simpson *et al.*, 1980; Lamb *et al.*, 1986; Coxhead and Thomas, 1993; Moen and Magnus, 1993; Olive, 1993; Kennedy *et al.*, 1995; dos Reis *et al.*, 1999; Kennedy, 1999). The risk to first-degree relatives of patients for developing endometriosis has been reported to be 4–8 times that of the general population. Twin studies (Moen, 1994; Hadfield *et al.*, 1997a; Treloar *et al.*, 1999) and animal studies (Hadfield *et al.*, 1997b) also support a genetic factor in the pathogenesis of endometriosis although the mode of transmission is not understood. Complex inheritance patterns are generally assumed (Simpson *et al.*, 1980; Moen and Schei, 1997; Kennedy, 1999).

The population prevalence of endometriosis is not known exactly, but has been estimated to be between 2–10% among women during the reproductive years (Olive, 1993; Moen and Schei, 1997). Diagnosis is difficult since laparoscopy or laparotomy are required. This is often performed in the context of a different diagnostic setting, such as investigation into infertility or for the purpose of sterilization.

Advances in human genetics have resulted in the identification of genes responsible for most of the known single gene disorders using linkage analysis. Single gene disorders account for ~1% of human disease. However the vast majority of human disease is either polygenetically inherited, caused by infection or other external agents, or is a product of interactions between genetic and external factors. In this study evidence for a genetic basis for endometriosis in Icelandic women was sought with several approaches using a genealogy database covering an entire nation.

The isolation of the Icelandic population for over 1100 years has limited somewhat the genetic diversity among the population (Gudmundsson *et al.*, 2000). There is a keen interest in genealogy in the population and a single payer healthcare system that ensures universal access facilitates ascertainment. A national identity number system and a population-based patient list of admissions and diagnoses are further advantages for genetic studies. A computerized genealogy database has now been developed which uses the extensive genealogical information collected over the last 10 centuries by the Icelandic people (Gudmundsson *et al.*, 2000). It includes all 283 000 presently-living Icelanders in addition to most of their ancestors since the time of settlement in the late 9th century. Over 630 000 individuals are currently registered in the database

out of the estimated 750 000 Icelanders ever living to adulthood in the country since it was settled. Those not recorded may largely represent those who never reproduced.

Using this population-based genealogical database, together with a list of women diagnosed with endometriosis during 1981–1993 the contribution of genetic factors in endometriosis has been estimated. This was done by calculating the relative risk for sisters and cousins, by comparing average kinship coefficients for the endometriosis patients and matched control groups and by estimating the minimum number of founders for the patient group compared with matched control groups.

## Material and methods

### Study population

All women with pelvic endometriosis diagnosed at laparoscopy or laparotomy between 1981–1993 at gynaecology departments in Iceland were identified by searching the national computerized database of hospital admissions for the relevant ICD 9 diagnostic codes, 617.1–617.9. Two residents in gynaecology with and under the supervision of one of us (RTG) examined individual operation notes for each patient to confirm the diagnosis and stage disease severity, and to obtain histological confirmation—available for 56% of cases. Disease severity was assessed using the revised American Fertility Society classification of endometriosis from 1985, classification system stages I–IV (American Fertility Society, 1985). A total of 750 women were on the diagnostic files and this patient group could be used to calculate relative risk, kinship coefficients and number of founders for the patient group at given time points in the past.

The genealogy database at DeCode Genetics was used to identify family connections. This database stores all available genealogy data for the last eleven centuries in Iceland including genealogy manuscripts, censuses and church books. The genealogy database used for the work described here was encrypted by the Icelandic Data Protection Commission in the same way as the national identity numbers which are used for patient identification. The genealogy database has been described in detail (Gudmundsson *et al.*, 2000; Gulcher *et al.*, 2000). The study was approved by the National Bioethics Committee and the Icelandic Data Protection Commission.

### Genealogy database and cluster function

The genealogy database is stored and maintained within a relational database. Each record in the database consists of a personal identifier, identifiers to parents, gender, dates of birth and death rounded off to the nearest decade or half-decade. All algorithms that work on family relations are implemented outside the database, in memory-based programs, since relational databases are not well suited for implementing dynamic programming algorithms. Recursive pedigree algorithms have been developed that find all ancestors in the database who are related to each member of the input list within a given number of generations back in time. Using these groups the cluster function then searches for ancestors who are common to any two or more members of the input lists.

### Relative risk calculations

The relative risk was calculated as follows:  $N_y$  was the number of individuals born in time interval (or years)  $y$ , and  $N_{ay}$  the number of affected individuals born in that time interval. The general risk for individuals born in period  $y$  was then  $N_{ay}/N_y$ . For an affected individual  $i$ , the number of relatives was  $n_{yi}$  and the number of affected relatives born in the same time interval was  $n_{ayi}$ . The relative risk of finding

the disease among relatives of affected individuals, born in period  $y$ , could be calculated as:

$$\lambda_y = \frac{\sum_i n_{ayi} \left| \sum_i n_{yi} \right.}{N_{ay}/N_y}$$

Relationships studied included sisters and first cousins of affected women. Relative risk calculations were carried out for 610 women born between 1940–1960 and 1000 matched and randomly chosen control sets of 610 women mimicking the patient group with respect to age, sex and connectivity in the database. A  $P$ -value of 0.005 for the relative risk was taken to indicate that five of the matched control groups had values as large, or larger, than those of the patients. When none of the values computed for the control groups was larger than the value for the patients, the  $P$ -value was estimated to be smaller than 0.001.

### Kinship coefficients

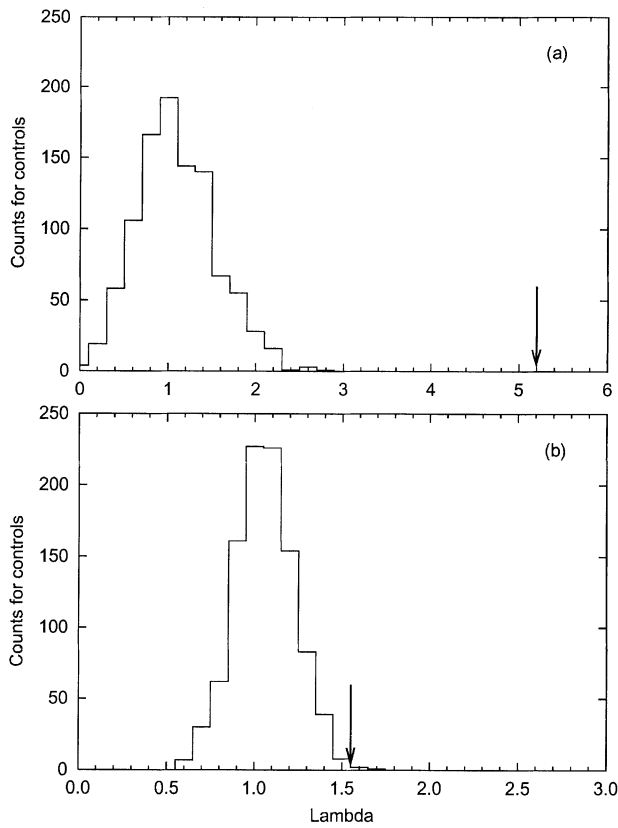
Kinship coefficients (KC) were calculated for the patients and their spouses. The KC is the probability that a randomly selected allele of a person is identical-by-descent to a randomly selected allele of another person (Lange, 1997). Average KC values were calculated for the group of patients and similarly for the 1000 control groups by averaging the KCs of every pair in a given group. Because of the size of the pedigrees, the pair-wise KC could not be calculated exactly. Monte Carlo simulations were used to evaluate the average KC by randomly transferring distinct alleles from the founders of the pedigrees to the next generations. In the Monte Carlo simulations we have taken into account six ancestral generations from the patient and control groups representing 35 000–38 000 individuals for each pedigree.

### Minimum Founder Test

The Minimum Founder Test was used to estimate whether the trait had a familial component (Gudmundsson *et al.*, 2000). It determines the minimum number of founders at a given time point in the past that is required to account for the test group and compares it with the minimum number of founders required to account for a set of control groups of the same size and with a similar structure. Given a set  $S$  of people,  $F(S)$  is set as the size of the smallest set  $A$  of ancestors, such that all the people in  $S$  were descendants of the ancestors in  $A$ . The problem of calculating  $F(S)$  can be reduced to a set covering problem (SCP). A branch and bound type of optimization algorithm was used to solve the SCP.  $F_y$  is defined as  $F$  before, except with the added constraint that all the ancestors were born before year  $y$ .  $F_y$  was calculated for the list of endometriosis patients ( $P$ ), and for 1000 control groups, each consisting of 750 individuals, for years  $y$  ranging from 1600–1960. For each  $y$ ,  $\bar{F}_y$  was the mean of  $F_y$  for the controls, and  $\sigma_y$  the standard deviation. Each control group  $C$  was built in such a way that every person in  $P$  corresponded to one person in  $C$ , born in the same year as  $P$  and having the same number of parents and grandparents documented in the genealogy database as  $P$ . Apart from this constraint, the control subjects were drawn at random from the genealogy database, irrespective of their disease status.

## Results

Estimates of the risk to sisters and first cousins of patients with endometriosis are shown in Figures 1a and b. The relative risk for sisters was 5.20, with 95% confidence interval (3.40,



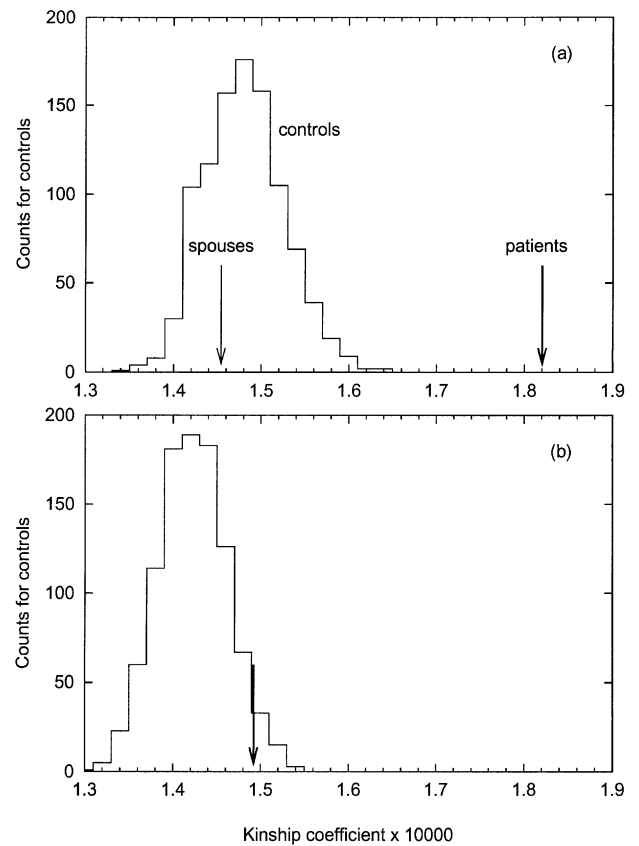
**Figure 1.** Estimates of the risk of developing endometriosis for: sisters (a) and cousins (b). The histograms show the risk distribution obtained for the control groups, and the arrows indicate the values obtained for patients. The risk ratio for sisters is 5.20 ( $P < 0.001$ ) and for cousins 1.56 ( $P = 0.003$ ).

7.16). This was significantly different from the calculated values in the control sets ( $P < 0.001$ ). For first cousins the risk ratio of 1.56 was also significantly different from the mean among the control sets ( $P = 0.003$ ) with confidence interval (1.13, 1.93).

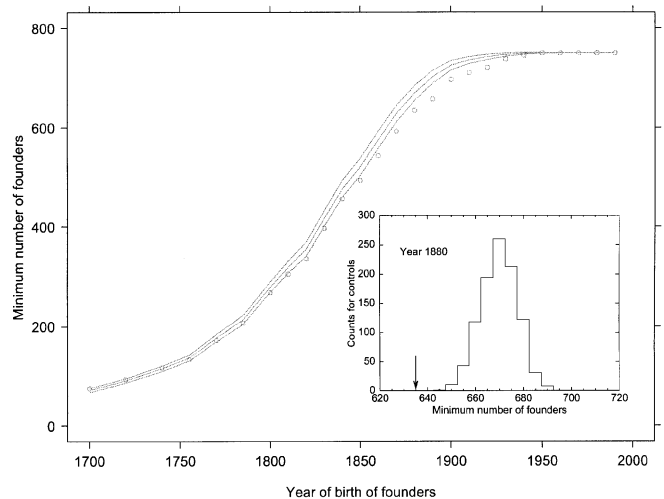
The average KC for the group of patients was  $1.82 \times 10^{-4}$ . This value is significantly higher ( $P < 0.001$ ) than all the average KC values obtained for the 1000 control groups (Figure 2a). The affected women are therefore significantly more related than the women in the control groups. The average KC for the group of spouses of patients was  $1.45 \times 10^{-4}$  and was significantly lower than that of the patient group and within the KC distribution for the control groups (Figure 2a).

Because of a possible diagnostic bias, where relatives of patients may be more likely to seek medical opinion, we calculated also the average KC for patients and control groups excluding the contribution from sisters to sisters and mothers to daughters. This lowered the average KC for both patients and controls (these pairs have the highest pair wise KC, 0.25). The average KC for the endometriosis patients was then  $1.49 \times 10^{-4}$  and the control distribution shifted also to lower average KC values (Figure 2b) but the average KC for the patients was still significantly larger than for the control groups ( $P = 0.044$ ).

The familial clustering of the endometriosis patients calculated with the Minimum Founder Test is shown in

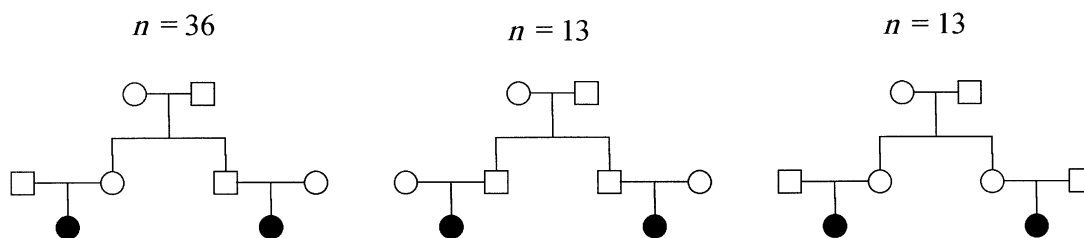


**Figure 2.** Average kinship coefficients (KC) for all patients ( $n = 750$ ) and their spouses and for the combined matched control groups (a). In (b) the contributions from mothers to daughters and between sisters were excluded in the average KC calculations. The average KC is reduced for the group of patients and for the control groups but is still significantly higher for the patient group as indicated by the arrow ( $P = 0.044$ ).



**Figure 3.** The Minimum Founder Test shows that the patient group is more related than the controls. Between the years 1830–1900 the patient group has significantly fewer founders [differs by more than 2 standard deviations from the mean (dotted lines)].

Figure 3. The continuous curves show the distribution of the minimum number of founders for the control groups for every year. The middle line corresponds to the maximum of the



**Figure 4.** Inheritance patterns for affected cousins: paternal, maternal and mixed inheritance patterns and numbers of each type.

distribution and the other two lines correspond to one standard deviation from the maximum. The circles show the results for the patients. The insert of Figure 3 is a histogram (minimum number of founders) for the controls for year 1880. The arrow points at the minimum number of founders for the patient group ( $n = 750$ ) for the year 1880 (635 founders). Between the years 1830 and 1900 the minimum number of founders behind the endometriosis patient group was significantly lower (at least one standard deviation lower) than for the control groups.

In addition, to assess possible mitochondrial transmission of a susceptibility gene, 62 affected cousin pairs were investigated. Thirteen pairs were maternally connected, 13 paternally and 36 pairs had mixed pattern (Figure 4).

**Discussion**

Epidemiological investigation of a disease like endometriosis, where under-diagnosis can be expected in the general population, can lead to biased results if only first-degree relatives are studied. The use of a population-based patient group eliminates much of the potential sampling bias and allows comparison of more distantly related patients. Using three different approaches a familial component in endometriosis was detected. Including relatives beyond the nuclear family reduces the risk of observing associations that are related to shared environmental factors (Sadovnick and Macleod, 1981). In this study, more remotely-related endometriosis patients than sisters and mother-daughter pairs were used to estimate a contribution from a genetic factor in endometriosis.

A widely used measure of familial aggregation is the sibling risk ratio, which is defined as the ratio of risk of disease manifestation, given that one sibling is affected, compared with the disease prevalence in the general population. It is critical for such a study that sampling from the population is random and free of bias. Although any patient material will suffer from not including as yet undiagnosed individuals in the population, this study encompassed all women known to have the disease during a 13-year period in one population. Another consideration is that index patients may be more aware of the diagnosis in their close relatives and thus be more diligent in their search for additional cases in their families.

A different approach to assess familial aggregation of diseases is to link affected individuals in a genealogical database, estimate KCs for all possible pairs and compare the average to matched control groups derived from the general population. Concerns over the validity of estimates of the KC

prevail in this approach, including those of ascertainment bias in a disease which is likely to be under-diagnosed. Therefore cluster effects in the diagnosis cannot be excluded, even though this study is based on all women diagnosed during a certain period. In order to circumvent some of these difficulties we used an approach relying on more distant relationships, such as first cousin pairs.

Among sisters a 5.20 fold increase in the risk of being diagnosed with endometriosis was seen. Other investigators have reported similar values where sisters had between 4–8 times higher risk of developing the disease themselves (Simpson *et al.*, 1980; Moen and Magnus, 1993). The risk among first cousins was lower, but still significantly higher than in the control groups. While sisters share 50% of their genomes, cousins share 12.5% and the difference seen for the latter group may therefore more accurately reflect the genetic liability.

The cohort studied here represents all diagnosed endometriosis patients in Iceland over a period of 13 years. Due to the relatively narrow time interval there are not many affected mother-daughter pairs or aunt-niece pairs, and no grandmother-granddaughter pairs, in the patient material. Although this cohort is not well suited for estimating the mode of inheritance in endometriosis we have estimated the decline in relative risk from affected sister pairs to affected cousin pairs. For a single locus or a two-locus heterogeneity model one would expect (Lange, 1997):

$$\frac{\lambda_{sister} - 1}{\lambda_{cousin} - 1} = 4$$

For more complex models where genes may interact in an epistatic manner, for instance, different patterns and higher values would be expected (Risch *et al.*, 1990). In this study the result was:

$$\frac{\lambda_{sister} - 1}{\lambda_{cousin} - 1} = \frac{5.20 - 1}{1.56 - 1} = 7.50$$

The decline for  $\lambda_{cousin}$  is therefore more rapid than expected for a single locus or two-locus heterogeneity model and the mode of inheritance in endometriosis is therefore likely to be more complex.

To investigate possible mitochondrial transmission, we estimated the relative risk for paternal cousins, maternal cousins and mixed pairs. We did not find any significant difference in relative risk for the three different pairs (data not shown) but the observed numbers of pairs are shown in Figure 4. The inheritance pattern suggests equal transmission of susceptibility genes through mothers and fathers to the affected daughters.

Mitochondrial inheritance is therefore unlikely in endometriosis.

The kinship results confirm the familial aggregation of endometriosis (Figure 2a). The average KC for the patient group was higher than all the average KC values obtained for the control groups and the group of spouses ( $P < 0.001$ ). When a more remote relationship was tested by removing first degree relatives from the patient list (Figure 2b) the average KC for the patients was still significantly larger than the results for the control groups ( $P < 0.05$ ).

The Minimum Founder Test showed that by going back 5–6 generations, or ~150–200 years, there was a smaller number of founders for patients compared with the matched control sets. The evidence in favour of a genetic component in the aetiology of endometriosis is thus supported by a significantly increased risk ratio for sisters and for more distant relatives such as cousins, by significantly higher KCs seen among patients than among their spouses or the matched control groups representing the population, and by the fact that the patient group had fewer founders than the controls. Endometriosis lowers fertility by nature of the disease and the effect of that may tend to reduce the strength of a heritability factor.

The strength of this study is the population-based approach, utilizing a unique genealogical database from an entire nation. This allows for accurate assessment of KCs. The Minimum Founder Test is a new approach, which can be applied to a data collection reaching several generations back in time. This is the first study showing clear evidence for a genetic factor in endometriosis that is not relying on first-degree relatives.

### Acknowledgement

The contribution of G.Sverrisdottir, M.D. and K.Jonsdottir, M.D. in gathering data on the patient group is acknowledged.

### References

American Fertility Society (1985) Revised American Fertility Society Classification of Endometriosis:1985. *Fertil. Steril.*, **43**, 351–352.

- Coxhead, D. and Thomas, E.J. (1993) Familial inheritance in a British population. A case control study. *J. Obstet. Gynaecol.*, **13**, 42–44.
- Gudmundsson, H., Gudbjartsson, D.F., Frigge, M., Gulcher, J.R. and Stefansson, K. (2000) Inheritance of human longevity in Iceland. *Eur. J. Hum. Genet.*, **8**, 743–749.
- Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. and Stefansson, K. (2000) Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.*, **8**, 739–742.
- Hadfield, R.M., Mardon, H.J., Barlow, D.H. and Kennedy, S.H. (1997a) Endometriosis in monozygotic twins. *Fertil. Steril.*, **68**, 941–942.
- Hadfield, R.M., Yudkin, P.L., Coe, C.L., Scheffler, J., Uno, H., Barlow, D.H., Kemnitz, J.W. and Kennedy, S.H. (1997b) Risk factors for endometriosis in the rhesus monkey (*Macaca mulatta*): a case-control study. *Hum. Reprod. Upd.*, **3**, 109–115.
- Kennedy, S. (1999) The genetics of endometriosis. *Eur. J. Obstet. Gynecol. Reprod. Biol.*, **82**, 129–133.
- Kennedy, S., Mardon, H. and Barlow, D. (1995) Familial endometriosis. *J. Assist. Reprod. Genet.*, **12**, 32–34.
- Lamb, K., Hoffmann, R.G. and Nichols, T.R. (1986) Family trait analysis: a case-control study of 43 women with endometriosis and their best friends. *Am. J. Obstet. Gynecol.*, **154**, 596–601.
- Lange, K. (1997) *Mathematical and statistical methods for genetic analysis*. Springer Verlag, Berlin.
- Moen, M.H. (1994). Endometriosis in monozygotic twins. *Acta Obstet. Gynecol. Scand.*, **73**, 59–62.
- Moen, M.H. and Magnus, P. (1993) The familial risk of endometriosis. *Acta Obstet. Gynecol. Scand.*, **72**, 560–564.
- Moen, M.H. and Schei, B. (1997) Epidemiology of endometriosis in a Norwegian County. *Acta Obstet. Gynecol. Scand.*, **76**, 559–562.
- Olive, D.L. (1993) Endometriosis. *N. Engl. J. Med.*, **328**, 1759–1767.
- dos Reis, R.M., de Sa, M.F., de Moura, M.D., Nogueira, A.A., Ribeiro, J.U., Ramos, E.S. and Ferriani, R.A. (1999) Familial risk among patients with endometriosis. *J. Assist. Reprod. Genet.*, **16**, 500–503.
- Risch, N. (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.*, **46**, 222–228.
- Sadovnick, A.D. and Macleod, P.M. (1981) The familial nature of multiple sclerosis: empiric recurrence risks for first, second-, and third-degree relatives of patients. *Neurol.*, **31**, 1039–1041.
- Simpson, J.L., Elias, S., Malinak, L.R. and Buttram, V.C. (1980) Heritable aspects of endometriosis I. Genetic studies. *Am. J. Obstet. Gynecol.*, **137**, 327–331.
- Treloar, S.A., O'Connor, D.T., O'Connor, V.M. and Martin, N.G. (1999) Genetic influences on endometriosis in an Australian twin sample. *Fertil. Steril.*, **71**, 701–710.

Submitted on April 30, 2001; accepted on November 2, 2001