Oral contraceptives and risk of endometriosis: a systematic review and meta-analysis

Paolo Vercellini1,2*, Brenda Eskenazi3, Dario Consonni4, Edgardo Somigliana1,2, Fabio Parazzini1, Annalisa Abbiati1,2, and Luigi Fedele1

1Clinica Ostetrica e Ginecologica I, Istituto ‘Luigi Mangiagalli’, Università Statale di Milano, Fondazione IRCCS ‘Ca’ Granda’ - Ospedale Maggiore Policlinico, Via Commenda 12, Milan 20122, Italy
2Centre for Research in Obstetrics and Gynaecology (C.R.O.G.), Milan, Italy
3School of Public Health, Division of Epidemiology, University of California, Berkeley CA, USA
4Unit of Epidemiology, Fondazione IRCCS ‘Ca’ Granda’ - Ospedale Maggiore Policlinico, Milan, Italy

*Correspondence address. Tel: +39-02-5503-2917; Fax: +39-02-5503-2331; E-mail: paolo.vercellini@unimi.it

Submitted on April 19, 2010; resubmitted on July 27, 2010; accepted on August 4, 2010

TABLE OF CONTENTS

• Introduction
• Methods
  • Sources
  • Study selection
  • Data extraction and analysis
• Results
• Discussion

BACKGROUND: Combined oral contraceptives (OCs) inhibit ovulation, substantially reduce the volume of menstrual flow and may hypothetically interfere with implantation of refluxed endometrial cells. The aim of this review is to establish if OC use influences the risk of endometriosis.

METHODS: We performed a MEDLINE search to identify all studies published in the last four decades (January 1970 to January 2010) in the English language on the relationship between OC exposure and risk of endometriosis. Two authors abstracted data on standardized forms.

RESULTS: We identified 608 potentially relevant studies and 18 studies (6 cross-sectional, 7 case–control and 5 cohort) were selected. Pooling of the results derived from all the included reports independently from study design, yielded a common relative risk of 0.63 [95% confidence interval (CI), 0.47–0.85] for current OC users, 1.21 (95% CI, 0.94–1.56) for past users and 1.19 (95% CI, 0.89–1.60) for ever users. Methodological drawbacks, such as uncertain temporal relationship between exposure and outcome in cross-sectional studies and suboptimal selection of controls in case–control studies, limit the quality of the available evidence.

CONCLUSIONS: The risk of endometriosis appears reduced during OC use. However, it is not possible to exclude the possibility that the apparent protective effect of OC against endometriosis is the result of postponement of surgical evaluation due to temporary suppression of pain symptoms. Confounding by selection and indication biases may explain the trend towards an increase in risk of endometriosis observed after discontinuation, but further clarification is needed. To date, the hypothesis of recommending OCs for primary prevention of endometriosis does not seem sufficiently substantiated.

Key words: endometriosis / oral contraceptives / observational studies / systematic review / meta-analysis
Introduction

Retrograde menstruation seems the most probable pathogenic mechanism for the development of endometriosis (Eskenazy and Warner, 1997; Missmer and Cramer, 2003; Viganò et al., 2004; Chapron et al., 2006; Farquhar, 2007). If this is true, the likelihood of implantation of regurgitated endometrium could be influenced by the recent major increase in the number of menstrual cycles (Thomas, 1993; Eaton et al., 1994; Evers et al., 1995; Vercellini et al., 2010). Indeed, the risk of endometriosis appears to rise with greater lifetime number of ovulatory cycles (Missmer et al., 2004). Moreover, the role of ovulation in the genesis of ovarian endometriosis has been confirmed demonstrating that endometriomas develop from follicles immediately after ovulation (Jain and Dalton, 1999), and by direct transition from haemorrhagic corpora lutea (Vercellini et al., 2009a).

Combined oral contraceptives (OCs) suppress ovulation and substantially reduce the amount of monthly uterine blood flow. Progestins may prevent implantation and growth of regurgitated endometrium, inhibiting expression of matrix metalloproteinases and angiogenesis (Vercellini et al., 2003a). Moreover, progestins have several anti-inflammatory in vitro and in vivo effects that may reduce the inflammatory state generated by the metabolic activity of the ectopic endometrium, and the consequent immune response (Vercellini et al., 2003a). In addition, OCs increase the abnormally low apoptotic activity of the endometrium of women with endometriosis (Meresman et al., 2002).

The possibility of reducing the risk of endometriosis by prescribing OCs has been repeatedly hypothesized (Anonimus, 1997; Fraser and Kovacs, 2003; Missmer et al., 2004). However, there is no consensus on the potential protective role of OCs (Vercellini et al., 1993), and it has been suggested that they might even facilitate implantation and growth of refluxed endometrial cells, thus increasing, instead of reducing, the risk of the disease (Parazzini et al., 1994; Italian Endometriosis Study Group, 1999).

Given the high frequency of OC use especially in young women needing contraception, as well as the worrying prevalence of endometriosis in this population, a relationship between the two factors, either towards a reduction or an increase in risk, would have major consequences (Egger et al., 1998). Moreover, OCs are used as a treatment for symptomatic endometriosis (Vercellini et al., 1997, 2003a, b, 2008, 2009b), and clarification of the potential effects of this type of therapy on the natural history of the disease would be extremely important. Finally, if a protective effect of OCs is demonstrated, their use could be hypothesized as a primary preventive measure in high-risk groups (e.g. first-degree relatives of women with endometriosis).

However, to our knowledge no specific attempt has been made to review systematically the scientific literature on the potential relationship between OC exposure and the risk of endometriosis. With the aim of disentangling this intriguing and controversial issue, we have reviewed the available epidemiological evidence published on the topic in the last four decades. Hospital-based studies were pooled in a meta-analysis with consideration given to the effect of exposure in current, past, and ever OC users.

Methods

The present literature overview was conducted according to the MOOSE guidelines for systematic reviews of observational studies (Stroup et al., 2000). As published, de-identified data were used, the present study was exempt from Institutional Review Board approval.

Sources

This review was restricted to published research articles that compared exposure to OCs in women with surgically diagnosed endometriosis with that in women without a diagnosis of endometriosis. Several different strategies were adopted to identify medical papers published on the association between OC use and diagnosis of endometriosis. We conducted a MEDLINE search from January 1970 to January 2010 using the combinations of medical subject heading terms ‘endometriosis’, ‘dysmenorrhea’, ‘oral contraceptives’ and ‘progestins’. Only those publications written in English were included. All pertinent articles were retrieved and the reference lists were systematically reviewed in order to identify further reports that could be included in the meta-analysis. Moreover, review articles, books and monographs published on endometriosis were consulted and their reference lists searched for additional potential studies. Proceedings of scientific meetings were not included. No attempt was made to identify unpublished studies.

Study selection

An initial screening of the title and abstract of all articles was performed to exclude citations deemed irrelevant by two observers (e.g. if women without endometriosis were not evaluated). Studies were excluded if ad interim results were reported in advance of an available later full report. Reports were categorized based on research design into cross-sectional, case–control and cohort studies.

The quality of studies was evaluated by means of the Newcastle–Ottawa Scale, a validated modality for assessing observational and non-randomized studies (Wells et al., 2000). The scale uses a score system based on three criteria: selection of participants, comparability of study groups and assessment of exposure. No cut-off level was set for inclusion in the meta-analysis, as the use of quality scoring for study selection is controversial. According to Greenland and O’Rourke (2001), it appears that ‘quality’ is of fairly high dimension and possibly non-additive and nonlinear, and that quality dimensions are highly application-specific. Hence, when used to directly modify weights or contributions of individual studies, implementation of quality scores in meta-analysis could produce biased estimates of effect. Nonetheless, since a formal evaluation of study methodology could be of benefit to the reader in order to allow a general overview of the robustness of the available data, we included the results of our assessment of study quality in the present report.

Data extraction and analysis

Two authors (E.S. and A.A.) independently evaluated all articles and abstracted data on standardized forms. A final abstraction form was compiled from the two evaluation forms, with correction or resolution of any discrepancies between reviewers by consensus reached after discussion. The year of publication, location, setting, study design, number and clinical characteristics of study subjects and modality of diagnosis of endometriosis were recorded. Adjusted odds ratio (OR, cross-sectional and case–control studies), or rate ratio (cohort studies) and their 95% confidence intervals (CIs) were extracted. When performing overall estimates, these measures will be collectively reported here as relative risk (RR). In one study (Whitoff et al., 2000) we could calculate adjusted ORs for past and current users as a weighted average of the ORs and 95% CI by length of OC use. In the remaining, mostly cross-sectional, studies only crude estimates were reported or calculated by us. Meta-analyses were performed by pooling the logarithms of estimates and using the DerSimonian and Laird random-effects method (DerSimonian and Laird, 1986).
The Cochrane Q and the $I^2$ statistics, which describe the proportion of the total variation of estimates across studies owing to heterogeneity rather than chance (Higgins et al., 2003), were then calculated. Negative values of $I^2$ are set equal to 0 so that $I^2$ lies between 0 and 100%. A value of 0% indicates no observed heterogeneity, whereas $I^2$ values of 25, 50 and 75% indicate low, moderate and high heterogeneity, respectively (Higgins et al., 2003).

For the association between OC use and endometriosis, we also explored sources of heterogeneity using study design as an independent explanatory variable defined a priori. A sensitivity analysis was conducted to verify if any one study unduly influenced the pooled effect size.

Begg’s funnel plots, which plot RR on a log scale (effect) against SE of log RR (precision), were generated and visually inspected to determine if the included studies were non-representative of the body of possible studies on the subject (which could result from publication or other biases). When biases and heterogeneity are not present, the variation in the estimated effect decreases with increasing sample size, and the plot resembles a symmetrical funnel. To facilitate interpretation, pseudo 95% confidence limits have been included in funnel plots. Egger’s approach to testing the significance of Begg’s funnel plot asymmetry was also used. A probability level of <5% was considered significant for publication bias (Begg and Mazumdar, 1994; Egger et al., 1997). All analyses were performed using Stata 10.0 (Stata Corp., 2007).

Results

Figure 1 shows the flow diagram of the literature search results. The MEDLINE search identified 608 studies of which there were 14 abstracts (Buttram, 1979; Karnaky 1979; Sensky and Liu, 1980; Strathy et al., 1982; Kirshon and Poindexter, 1988; Parazzini et al., 1989, 1994; Mahmood and Templeton, 1991; Parazzini and Ferraroni, 1993; Vessey et al., 1993; Vercellini et al., 1993; Italian Endometriosis Study Group, 1999; Westhoff et al., 2000; Templeman et al., 2008) that reported findings on the relationship between OC exposure and risk of endometriosis; these articles were retrieved for detailed
assessment. After reviewing the relevant reference lists, we identified another eight publications (Royal College of General Practitioners, 1974; Brown, 1981; Moen, 1987, 1991; Darrow et al., 1993, 1994; Matorras et al., 1995; Sangi-Haghpeykar and Poindexter, 1995). Three additional articles were identified from searching the reference list of review articles (Anonimous, 1997; Hemmings et al., 2004; Missmer et al., 2004). No other studies were found from a search of reference lists of recent books and monographs. Of the 25 studies found, we excluded three (Karnaky 1979; Vercellini et al., 1993; Anonimous, 1997) because no original data were included, two (Buttram, 1979; Sensky and Lui, 1980) because no control group was considered and two (Darrow et al., 1993; Parazzini and Ferraroni, 1993) that reported ad interim analyses.

Data on the association between OC use and risk of endometriosis were extracted from the remaining 18 articles, all published in full in peer-review journals between 1974 and 2008 (Royal College of General Practitioners, 1974; Brown, 1981; Strathy et al., 1982; Moen, 1987, 1991; Kirshon and Poindexter, 1988; Parazzini et al., 1989, 1994; Mahmood and Templeton, 1991; Vessey et al., 1993; Darrow et al., 1994; Matorras et al., 1995; Sangi-Haghpeykar and Poindexter, 1995; Italian Endometriosis Study Group, 1999; Westhoff et al., 2000; Hemmings et al., 2004; Missmer et al., 2004; Templeman et al., 2008). Six were cross-sectional studies (Moen, 1987, 1991; Kirshon and Poindexter, 1988; Mahmood and Templeton, 1991; Italian Endometriosis Study Group, 1999; Hemmings et al., 2004), seven were case–control studies (Strathy et al., 1982; Parazzini et al., 1989, 1994; Darrow et al., 1994; Matorras et al., 1995; Sangi-Haghpeykar and Poindexter, 1995; Westhoff et al., 2000) and five were cohort studies (Royal College of General Practitioners, 1974; Brown, 1981; Vessey et al., 1993; Missmer et al., 2004; Templeman et al., 2008).

Disagreements between abstractors concerned categorization of two studies (Strathy et al., 1982; Kirshon and Poindexter, 1988). After mutual discussion, the study by Kirshon and Poindexter (1988) was categorized as cross-sectional because OC exposure and outcome were both determined at the same point in time, whereas the study by Strathy et al. (1982) was categorized as case–control because subjects were investigated in a backward direction from outcome to OC exposure.

Details of the characteristics of the selected studies are shown in Table I. Eight studies were conducted in the USA (Brown, 1981; Strathy et al., 1982; Kirshon and Poindexter, 1988; Darrow et al., 1994; Sangi-Haghpeykar and Poindexter, 1995; Westhoff et al., 2000; Missmer et al., 2004; Templeman et al., 2008), three studies each in UK (Royal College of General Practitioners, 1974; Mahmood and Templeton, 1991; Vessey et al., 1993) and Italy (Parazzini et al., 1989, 1994; Italian Endometriosis Study Group, 1999), two in Norway (Moen, 1987, 1991), one in Spain (Matorras et al., 1995) and one in Canada (Hemmings et al., 2004). A total of 4802 cases of endometriosis were found, 1565 in cross-sectional studies, 1231 in case–control studies and 2235 in cohort studies. Women were generally post-adolescent and premenopausal, although two studies included subjects up to 64 (Brown, 1981) and 69 years (Parazzini et al., 1989). Patients recruited in case–control studies were mostly infertile or with severe pelvic pain, whereas in three cross-sectional studies (Moen, 1987, 1991; Kirshon and Poindexter, 1988) only fertile, asymptomatic subjects undergoing tubal sterilization were considered. Possible additional indications for OC use besides contraception (e.g. dysmenorrhea) were never specified. Endometriosis was surgically diagnosed at laparoscopy in nine studies (Strathy et al., 1982; Moen, 1987, 1991; Kirshon and Poindexter, 1988; Darrow et al., 1994; Matorras et al., 1995; Sangi-Haghpeykar and Poindexter, 1995; Italian Endometriosis Study Group, 1999; Missmer et al., 2004), at laparotomy in two (Royal College of General Practitioners, 1974; Brown, 1981) and at laparoscopy or laparotomy in five studies (Mahmood and Templeton, 1991; Vessey et al., 1993; Parazzini et al., 1994; Hemmings et al., 2004; Templeman et al., 2008), whereas in two studies (Parazzini et al., 1989; Westhoff et al., 2000) the surgical approach was not specified. Histological confirmation was reported in five studies (Parazzini et al., 1989; Moen, 1991; Matorras et al., 1995; Westhoff et al., 2000; Missmer et al., 2004).

The quality scores of the considered studies ranged from 4 to 9 and the average score for all 18 studies was 5.9 (SD, 1.7); corresponding figures were 5.3 (1.8) for cross-sectional studies, 5.4 (1.8) for case–control studies and 7.4 (0.5) for cohort studies (Table I). There was careful selection of cases in all studies, as case definition was adequately achieved by recruiting subjects through surgery. The cases were also largely representative of the source populations, limiting the risk of selection bias. Four studies (Strathy et al., 1982; Parazzini et al., 1989, 1994; Darrow et al., 1994) used hospital or non-community controls, such as patients with acute conditions (Parazzini et al., 1989, 1994) or with gynaecological disorders other than endometriosis (Strathy et al., 1982; Darrow et al., 1994). Studies that did better on control selection employed community controls (Moen 1987, 1991; Kirshon and Poindexter, 1988; Mahmood and Templeton, 1991; Matorras et al., 1995; Sangi-Haghpeykar and Poindexter, 1995; Italian Endometriosis Study Group, 1999; Westhoff et al., 2000; Hemmings et al., 2004; Templeman et al., 2008). The overall performance of the studies on comparability of participants was adequate in 14 studies (Royal College of General Practitioners, 1974; Brown, 1981; Parazzini et al., 1989, 1994; Mahmood and Templeton, 1991; Vessey et al., 1993; Darrow et al., 1994; Matorras et al., 1995; Sangi-Haghpeykar and Poindexter, 1995; Italian Endometriosis Study Group, 1999; Westhoff et al., 2000; Hemmings et al., 2004; Missmer et al., 2004; Templeman et al., 2008) in which potential confounders were controlled. Because some studies were fairly small in size (Strathy et al., 1982; Moen 1987, 1991; Kirshon and Poindexter, 1988), the authors could not adjust for potential confounders. The modality of assessment of exposure was clearly defined in 10 studies (Royal College of General Practitioners, 1974; Brown, 1981; Strathy et al., 1982; Vessey et al., 1993; Matorras et al., 1995; Italian Endometriosis Study Group, 1999; Westhoff et al., 2000; Hemmings et al., 2004; Missmer et al., 2004; Templeman et al., 2008). In the other eight studies the assessment of exposure was suboptimal: in two studies exposure was investigated solely through self-administered questionnaires (Mahmood and Templeton, 1991; Sangi-Haghpeykar and Poindexter, 1995), in two studies the interviewer was not blinded to cases and controls status (Moen, 1987; Darrow et al., 1994), whereas in four studies it was unclear whether exposure was assessed before or after surgery (Kirshon and Poindexter, 1988; Parazzini et al., 1989, 1994; Moen 1991).
<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>Country</th>
<th>Study population</th>
<th>Study design</th>
<th>Diagnostic criteria</th>
<th>No. of cases</th>
<th>Quality of study&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Royal College of General Practitioners</td>
<td>1974</td>
<td>UK</td>
<td>Women 15 to &gt; 45 years old using oral contraceptives</td>
<td>Cohort</td>
<td>Surgical diagnosis at laparotomy</td>
<td>43</td>
<td>7</td>
</tr>
<tr>
<td>Brown; Walnut Creek Contraceptive Drug Study</td>
<td>1981</td>
<td>USA</td>
<td>Women 18–64 years old using oral contraceptives</td>
<td>Cohort</td>
<td>Surgical diagnosis at laparotomy</td>
<td>104</td>
<td>7</td>
</tr>
<tr>
<td>Strathy et al.</td>
<td>1982</td>
<td>USA</td>
<td>Women 20–39 years old, fertile and infertile</td>
<td>Case–control</td>
<td>Visualization at laparoscopy</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Moen</td>
<td>1987</td>
<td>Norway</td>
<td>Mostly fertile, asymptomatic women, 23–46 years old</td>
<td>Cross-sectional</td>
<td>Visualization at laparoscopy</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Kirshon et al.</td>
<td>1988</td>
<td>USA</td>
<td>Mostly fertile, asymptomatic women, 22–41 years old</td>
<td>Cross-sectional</td>
<td>Visualization at laparoscopy</td>
<td>42</td>
<td>4</td>
</tr>
<tr>
<td>Parazzini et al.</td>
<td>1989</td>
<td>Italy</td>
<td>Women 20–69 years old with a surgical diagnosis of endometriosis</td>
<td>Case–control</td>
<td>Histological confirmation</td>
<td>114</td>
<td>4</td>
</tr>
<tr>
<td>Mahmoud and Templeton</td>
<td>1991</td>
<td>UK</td>
<td>Women with chronic abdominal pain or with dysfunctional uterine bleeding; fertile or infertile</td>
<td>Cross-sectional</td>
<td>Visualization at laparoscopy or laparotomy</td>
<td>227</td>
<td>5</td>
</tr>
<tr>
<td>Moen</td>
<td>1991</td>
<td>Norway</td>
<td>Mostly fertile, asymptomatic women, 23–50 years old</td>
<td>Cross-sectional</td>
<td>Surgical visualization at laparoscopy or histological confirmation</td>
<td>42</td>
<td>4</td>
</tr>
<tr>
<td>Vessey et al.</td>
<td>1993</td>
<td>UK</td>
<td>Women 25–39 years old using oral contraceptives</td>
<td>Cohort</td>
<td>Visualization at laparoscopy or laparotomy</td>
<td>138</td>
<td>7</td>
</tr>
<tr>
<td>Darrow et al.</td>
<td>1994</td>
<td>USA</td>
<td>Women 29–45 years old with a surgical diagnosis of endometriosis</td>
<td>Case–control</td>
<td>Visualization at laparoscopy</td>
<td>104</td>
<td>4</td>
</tr>
<tr>
<td>Parazzini et al.</td>
<td>1994</td>
<td>Italy</td>
<td>Women &lt;30 to &gt; 40 years old with a surgical diagnosis of endometriosis</td>
<td>Case–control</td>
<td>Visualization at laparoscopy or laparotomy</td>
<td>377</td>
<td>5</td>
</tr>
<tr>
<td>Matorras et al.</td>
<td>1995</td>
<td>Spain</td>
<td>Infertile women, age limits not reported</td>
<td>Case–control</td>
<td>Visualization at laparoscopy or histological confirmation</td>
<td>174</td>
<td>6</td>
</tr>
<tr>
<td>Sangi-Haghpeykar and Poindexter</td>
<td>1995</td>
<td>USA</td>
<td>Fertile women &lt; 25 to &gt; 36 years old</td>
<td>Case–control</td>
<td>Visualization at laparoscopy</td>
<td>126</td>
<td>6</td>
</tr>
<tr>
<td>Italian Endometriosis Study Group</td>
<td>1999</td>
<td>Italy</td>
<td>Women &lt;24 to &gt;40 years old, infertile or with abdominal pain, with no previous diagnosis of endometriosis</td>
<td>Cross-sectional</td>
<td>Visualization at laparoscopy</td>
<td>339</td>
<td>7</td>
</tr>
<tr>
<td>Westhoff et al.</td>
<td>2000</td>
<td>USA</td>
<td>Women 18–74 years old with a surgical diagnosis of a benign ovarian tumours</td>
<td>Case–control</td>
<td>Histological confirmation</td>
<td>311</td>
<td>9</td>
</tr>
<tr>
<td>Hemmings et al.</td>
<td>2004</td>
<td>Canada</td>
<td>Premenopausal women with regular menstrual cycles</td>
<td>Cross-sectional</td>
<td>Visualization at laparoscopy or laparotomy</td>
<td>896</td>
<td>8</td>
</tr>
<tr>
<td>Missmer et al.</td>
<td>2004</td>
<td>USA</td>
<td>Registered nurses 25–42 years old with a surgical diagnosis of endometriosis</td>
<td>Cohort</td>
<td>Visualization at laparoscopy or histological confirmation</td>
<td>1721</td>
<td>8</td>
</tr>
<tr>
<td>Templeman et al.</td>
<td>2008</td>
<td>USA</td>
<td>Female teachers and school administrators 22 to &gt; 90 years old with a surgical diagnosis of endometriosis</td>
<td>Cohort</td>
<td>Surgical diagnosis at laparoscopy or laparotomy</td>
<td>229</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality of study&lt;sup&gt;a&lt;/sup&gt;</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>According to the Newcastle–Ottawa Scale (Wells et al., 2000).</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Literature data, 1974–2008.
We could clarify (Parazzini, personal communication) that in one study laparotomy was the standard surgical approach (Parazzini et al., 1989) and that in two studies exposure was assessed before surgery (Parazzini et al., 1989, 1994). However, these aspects appeared unclear or not specified in the published reports.

Pooling of the results derived from all the included reports independent of study design, yielded a common RR of 0.63 (95% CI, 0.47–0.85) for current users, 1.21 (95% CI, 0.94–1.56) for past users and 1.19 (95% CI, 0.89–1.60) for ever users (Figs 2–4), suggesting a potentially protective effect of OCs during periods of use, but a small increase in risk after their discontinuation.

At sensitivity analysis, the common RR for current users varied from 0.57 (95% CI, 0.44–0.73) excluding the study by the Italian Endometriosis Study Group (1999), to 0.67 (0.50–0.90) excluding the study by Westhoff et al. (2000). Likewise, the common RR for past users varied from 1.15 (95% CI, 0.90–1.48) excluding the Nurses’ Health II study (Misser et al., 2004), to 1.30 (95% CI, 1.03–1.64) excluding the study by Westhoff et al. (2000). Corresponding figures for ever users were, respectively, 1.12 (95% CI, 0.82–1.52) after exclusion of the study by Parazzini et al. (1989), and 1.33 (95% CI, 1.02–1.73) after exclusion of the study by Westhoff et al. (2000).

Analyses were conducted also separately for the three study designs. In cross-sectional studies, among current users, there was a trend towards a decrease in odds in one study (Kirshon and Poindexter, 1988), no effect in one (Moen, 1991) and an increase in odds in one (Italian Endometriosis Study Group, 1999; Fig. 2). A trend towards a decrease in odds in past OC users was observed in three studies (Moen, 1987; Kirshon and Poindexter, 1988; Hemmings et al., 2004), and a significant increase in one (Italian Endometriosis Study Group, 1999; Fig. 3). Ever users were at increased odds in two studies (Italian Endometriosis Study Group, 1999; Mahmood and Templeton, 1991; Fig. 4). Pooling of the results yielded a common RR of 1.09 (95% CI, 0.59–2.01) for current users, 0.98 (95% CI, 0.64–1.49) for past users and 1.41 (95% CI, 1.12–1.80) for ever users (Figs 2–4).

The results of case–control studies show inconsistencies similar to those of the cross-sectional ones. In fact, among the case–control studies, current users were at significantly decreased odds in two reports (Sangi-Haghpeykar and Poindexter, 1995; Westhoff et al.,

### Table: Sources and Results

<table>
<thead>
<tr>
<th>Source (year)</th>
<th>No. of cases</th>
<th>RR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kirshon &amp; Poindexter (1988)</td>
<td>42</td>
<td>0.68 (0.35, 1.34)</td>
<td>8.55</td>
</tr>
<tr>
<td>Moen et al. (1981)</td>
<td>42</td>
<td>0.97 (0.41, 2.29)</td>
<td>6.80</td>
</tr>
<tr>
<td>Italian Endometriosis Study Group (1999)</td>
<td>339</td>
<td>1.60 (0.99, 2.37)</td>
<td>9.62</td>
</tr>
<tr>
<td>subtotal</td>
<td>(I-squared = 56.6%, p = 0.100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parazzini et al. (1994)</td>
<td>377</td>
<td>0.00 (0.46, 1.75)</td>
<td>8.77</td>
</tr>
<tr>
<td>Sangi-Haghpeykar &amp; Poindexter (1995)</td>
<td>126</td>
<td>0.40 (0.21, 0.75)</td>
<td>9.26</td>
</tr>
<tr>
<td>Westhoff et al. (2000)</td>
<td>311</td>
<td>0.33 (0.18, 0.61)</td>
<td>9.35</td>
</tr>
<tr>
<td>subtotal (I-squared = 61.1%, p = 0.076)</td>
<td></td>
<td>0.49 (0.27, 0.88)</td>
<td>27.38</td>
</tr>
<tr>
<td>Cohort</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC00 (1974)</td>
<td>43</td>
<td>0.50 (0.23, 1.07)</td>
<td>7.72</td>
</tr>
<tr>
<td>Walnut Creek (1981)</td>
<td>104</td>
<td>0.60 (0.28, 1.30)</td>
<td>7.65</td>
</tr>
<tr>
<td>Oxford F.P.A. (1993)</td>
<td>136</td>
<td>0.40 (0.21, 0.75)</td>
<td>9.26</td>
</tr>
<tr>
<td>Nurses’ Health II (2004)</td>
<td>1721</td>
<td>0.60 (0.62, 1.03)</td>
<td>14.03</td>
</tr>
<tr>
<td>Templeman et al. (2008)</td>
<td>229</td>
<td>0.41 (0.21, 0.79)</td>
<td>8.90</td>
</tr>
<tr>
<td>subtotal (I-squared = 44.6%, p = 0.125)</td>
<td></td>
<td>0.57 (0.40, 0.80)</td>
<td>47.56</td>
</tr>
<tr>
<td>Overall (I-squared = 62.8%, p = 0.003)</td>
<td></td>
<td>0.63 (0.47, 0.85)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**Figure 2** Forest plot for the relationship between OCs and endometriosis in current users. Forest plot showing individual and combined effect size estimates and 95% CIs in the studies that evaluated the relationship between current OC use and endometriosis. Horizontal lines indicate 95% CIs; boxes show the study-specific weight; rhombi represent combined effect sizes; dashed line indicates the overall estimate.
2000; Fig. 2), whereas past users were at increased odds in one study (Parazzini et al., 1994) and at decreased odds in another one (Westhoff et al., 2000; Fig. 3). Among ever users, two studies showed a significant decrease in odds (Strathy et al., 1982; Westhoff et al., 2000), whereas an increase in odds was observed in three studies (Parazzini et al., 1989, 1994; Matorras et al., 1995) and a mixed effect related to choice of controls in one (Darrow et al., 1994; Fig. 4). Pooling of the results yielded a subtotal common RR of 0.49 (95% CI, 0.27–0.88) for current users, 1.11 (95% CI, 0.49–2.55) for past users and 1.09 (95% CI, 0.72–1.64) for ever users (Figs 2–4).

The results of the five available cohort studies, i.e. the Royal College of General Practitioners Study (1974), the Walnut Creek Contraceptive Drug Study (Brown, 1981), the Oxford Family Planning Association Study (Vessey et al., 1993), the Nurses’ Health Study II (Missmer et al., 2004) and the California Teachers Study (Templeman et al., 2008), consistently demonstrated a lower risk of endometriosis among current OC users when compared with women who had never used OCs. However, a higher risk was observed in former OC users. The risk varied from 0.40 to 0.80 in current users and from 1.17 to 1.80 in those who had discontinued the pill (Figs 2 and 3). The subtotal common RR was 0.57 (95% CI, 0.40–0.80) for current users, and 1.60 (95% CI, 1.40–1.82) for past users.

We observed a considerable amount of total variation across the included studies owing to heterogeneity with regard to past users in cross-sectional (Fig. 3, $P = 0.03$), and ever users in case–control studies (Fig. 4, $P < 0.01$). Two case–control studies reported opposite findings in past users (Fig. 3). Removal of studies one at a time from the meta-analyses suggested that one study with a small sample size (Strathy et al., 1982) may have influenced the pooled subtotal estimate of cross-sectional studies. In fact, this exclusion yielded a RR of 0.80 (95% CI, 0.63–1.01), with fading of between-study heterogeneity ($I^2 = 0\%$, $P = 0.97$). Conversely, heterogeneity regarding past users in case–control studies could not be resolved by removal of any study. Of note, no significant heterogeneity was observed across cohort studies with regard to either current or past users. When only these studies were included in the meta-analysis, $I^2$ was 44.6% for current users and 1.6% for past users, suggesting that a likely source of the variation observed across cross-sectional and
Figure 4 Forest plot for the relationship between OCs and endometriosis in ever users. Forest plot showing individual and combined effect size estimates and 95% CIs in the studies that evaluated the relationship between ever OC use and endometriosis. Horizontal lines indicate 95% CIs; boxes show the study-specific weight; rhombi represent combined effect sizes; dashed line indicates the overall estimate.

Figure 5 Begg’s funnel plot for the effect of current OC exposure on risk of endometriosis. Begg’s funnel plot with pseudo 95% CI of the estimate in 11 studies which evaluated the relationship between current OC use and endometriosis. No indication of asymmetry was observed.

Figure 6 Begg’s funnel plot for the effect of ever OC exposure on risk of endometriosis. Begg’s funnel plot with pseudo 95% CI of the estimate in nine studies which evaluated the relationship between ever OC use and endometriosis. No indication of asymmetry was observed.
adequate analysis, and differences in underlying risk (Egger et al., 1997). In this regard, the within-group discrepancies may be partly explained by differences in the characteristics of the populations being investigated and in the disease definition adopted (Holt and Weiss, 2000). For instance, Moen (1987) evaluated mostly fertile and asymptomatic women, whereas in the study by Parazzini et al. (1994), cases underwent surgery because of pain, infertility or an adnexal mass.

Differences in study design appear to constitute the more likely explanation of the between-group discrepancies (Holt and Weiss, 2000; Zondervan et al., 2007). In cross-sectional studies exposure is ascertained as the same point in time as outcome. Consequently, the investigator cannot be certain that exposure preceded outcome. Actually, this is a problem for all the designs unless absence of endometriosis is confirmed a priori in the presumably non-diseased group.

This disadvantage has important implications for causal inference that rests on the unknown true temporal sequence of events (reverse causality bias). Furthermore, cross-sectional studies share with case–control studies the potential for sample distortion bias. In fact, in case–control studies the outcome is often the very reason for referrals that may be independently associated with exposure. This is most likely to occur if the cases are referred by clinicians who also prescribe the exposure agent (Eskenazy and Warner, 1997). For example, if cases are referred by gynaecologists, whereas controls are not, cases may be more likely to be taking OC pills than controls, even if there is no true association between OC exposure and the study outcome. Differential recall bias may be an even greater concern as, theoretically, cases (women with endometriosis) might be more likely to remember exposure (having taken the pill) than controls.

CoHORT studies (follow-up of non-symptomatic and presumably unaffected women until disease develops) should offer the most robust evidence, because subjects are identified prior to exposure, thereby providing a more accurate picture of the baseline state and leading to a more reliable estimate of the magnitude of the effect of considered risk factors. Data from cohort studies have consistently demonstrated a protective effect of current OC use (relative reduction, 43%; 95% CI, 20–60%), whereas previous use seems to increase the risk. However, although study subjects were asymptomatic at baseline, it is not possible to exclude the possibility that the disease was already present in some of them before exposure.

There are various biological interpretations for a possible effect of the OC pill on endometriosis. The pill may reduce risk by suppressing ovulation, as regular, ovulatory cycles increase the risk of the disease (Eskenazy and Warner, 1997; Missmer and Cramer, 2003; Missmer et al., 2004; Viganò et al., 2004). On the other hand, endometrial tissue seeded into the peritoneum of castrated female monkeys implanted even in the absence of gonadal hormones, but survived only if estradiol and/or progesterone were supplemented (DiZerega et al., 1980). In this light, the pill could be viewed as a rescue factor for regurgitated endometrial glands that would otherwise undergo necrosis and resorption during the physiologically hypo-estrogenic menstrual milieu (Parazzini et al., 1994).

However, a problem common to all the studies considered is the impossibility of knowing the exact onset of disease, as opposed to the date of diagnosis (Missmer et al., 2004). In fact, the induction period (interval between exposure and disease onset) in endometriosis is unknown and the latent period (interval between disease

---

**Figure 7** Begg’s funnel plot for the effect of past OC exposure on risk of endometriosis. Begg’s funnel plot with pseudo 95% CI of the estimate in 11 studies which evaluated the relationship between past OC use and endometriosis. There was some indication of asymmetry (Begg’s test, $P = 0.59$; Egger’s test, $P = 0.06$), owing to a large study (Missmer et al., 2004).

---

Discussion

Overall, based on the results of this systematic review and meta-analysis, OC exposure appears to decrease the risk of endometriosis in current users, but potentially to increase it in previous users. Risk reduction in current users has been consistently observed in all case–control and cohort studies (Fig. 2), whereas the increase in risk in past users is based mainly on findings from cohort studies (Fig. 3). In fact, the overall 21% RR increase calculated in past users after pooling data of all studies independent of design was not significant, as the 95% CI included unity (0.94–1.56). Conversely, combining results from cohort studies only yielded a 60% increase in risk of endometriosis (95% CI from 40 to 82%).

Heterogeneity across studies included in the same group was not strong except in past use in cross-sectional and ever use in case–control studies. Of note, 8 of the 11 studies evaluating current OC use reported a substantial decrease in risk of endometriosis.

There are several potential sources of the limited funnel plot asymmetry observed in studies which evaluated the relationship between past OC use and endometriosis, including not only publication bias, but also poor methodological design of small studies, inadequate analysis, and differences in underlying risk (Egger et al., 1997). In this regard, the within-group discrepancies may be partly identified by the poor methodological design of small studies, including not only publication bias, but also poor methodological design of small studies, inadequate analysis, and differences in underlying risk.
onset and detection) may vary widely owing to differences in symptom severity, health care resource use and diagnostic delay (Eskenazy and Warner, 1997). Accordingly, the observed results are based on the analysis of the incidence of endometriosis diagnosis rather than incidence of disease onset. In other words, since OCs are often prescribed as a first-line treatment for dysmenorrhea, a symptom strongly associated with the presence of endometriosis, it is impossible to understand if women who have used OCs and are subsequently diagnosed with endometriosis developed the disease before or after exposure. In fact, women with endometriosis-induced dysmenorrhea might have been selectively excluded from the ‘never OC users’ category at any point, with a consequent increased risk for past users as a group. On the other hand, OC use could reduce the likelihood of diagnosis of endometriosis, as the pill reduces pelvic pain symptoms and, hence, current users tend not to be investigated for and diagnosed with endometriosis.

Women with undiagnosed endometriosis could, therefore, increase the number of controls who use OCs, leading to an apparent protective effect for current users (Chiaffarino et al., 1998). Consequently, it is probable that the diagnosis of endometriosis is only delayed in OC users once the symptoms re-emerge after pill discontinuation because of resumption of the metabolic activity of ectopic implants. This confounding by selective mechanisms and indication bias renders interpretation of data somewhat problematic. Moreover, several community and hospital-based controls in the evaluated studies did not have endometriosis excluded by direct surgical visualization of the abdomino-pelvic cavity, raising the possibility of disease misclassification. This would underestimate the relationship between OC exposure and risk of endometriosis.

A limitation of meta-analyses involving observational studies is that it is difficult to control for confounding and selection bias often associated with this type of investigation, with the inherent danger of producing tight CIs around spurious results (Egger et al., 1998). In these circumstances it has been suggested that great efforts should be put into performing a complete literature search, selecting studies and extracting data in a reproducible and objective fashion, and examining carefully possible sources of heterogeneity between different studies, rather than focusing on statistical combination of data (Egger et al., 1998). In this regard, we performed a thorough literature review adopting different modalities of article search to avoid major bias in data gathering, and these were extracted from the reports of two independent observers who admittedly were not blinded. Rejected studies and the reasons for their exclusion are described. We explored heterogeneity in funnel plots and by means of sensitivity analysis, and addressed potential sources of confounding and bias extensively. Moreover, the data included in our analysis are the only available evidence on which to base clinical understanding and treatment decision making, and so far there have been no formal attempts to review data systematically on the potential relationship between OC use and endometriosis. Even considering all the above methodological issues, it appears unlikely that OCs influence the risk of endometriosis to any great extent, also because a consistent dose–response effect for lifetime duration of use has not been observed (Parazzini et al., 1994; Chiaffarino et al., 1998). Furthermore, also the pattern of risk with time since last use does not support a causal relationship (Parazzini et al., 1994; Chiaffarino et al., 1998).

Concerning the possibility of primary prevention, Missmer et al. (2004), based on the observation that the ovulatory cycle-associated risk of endometriosis seems greatest among never users of OCs, maintain that ‘prescription of OC before disease onset is a valid health intervention’. However, considering that the relationship between past OC use and endometriosis is not yet definitely clarified, we believe that large-scale trials are warranted before suggesting OCs for primary prevention of the disease. Moreover, such a policy would imply systematic prescription of OCs to all adolescent women until they wanted to conceive, with all the associated major organizational problems, health care resources expenditure and potential association with psychological as well as organic morbidity (Davey Smith and Egger, 1994).

In conclusion, a reduction in risk of endometriosis has been consistently observed in current OC users. However, drawbacks inherent to the design of several studies, such as uncertain temporal relationship between exposure and outcome in cross-sectional studies, and suboptimal selection of controls in case–control studies, limit the quality of the available evidence. In particular, in hospital-based case–control studies, controls are often not reflective of the distribution of key characteristics of the population from which the cases arose.

A potential increase in risk of endometriosis after OC discontinuation might well be related to methodological drawbacks, but to date we do not have definitive proof that the association is spurious, and the problem of clarifying the reason for this finding remains. It appears unlikely that past OC use favours future development of endometriosis, but we still cannot exclude this hypothesis formally. At present, the role of confounding and selection as well as other biases must be considered with caution in the interpretation of the available epidemiologic findings.

In order to finally disentangle the above issues, future cohort studies must include data that specifically describe the indication for pill prescription as well as definition of gynaecologic symptoms preceding OC use. In addition, younger women should be enrolled and assessment of menstrual pain in general, and of symptoms associated with endometriosis in particular, should be indicated among the primary aims. In fact, the disease may develop and progress during the postmenarchal period, adolescence or early adulthood. Such temporal windows have not been adequately evaluated in existing cohort studies.

**Authors’ roles**

P.V. conceived and designed the study and drafted the manuscript. B.E. interpreted the data and revised the manuscript. D.C. analysed and interpreted the data. E.S. selected the articles and retrieved the data. F.P. interpreted the data and revised the manuscript. A.A. selected the articles and retrieved the data. L.F. interpreted the data and revised the manuscript. All the authors approved the final version of the manuscript.
Funding

Supported in part by the University of Milan School of Medicine ResearchGrant FIRST no. 12-01-5068118-00067.

References


Zondervan KT, Cardon LR, Kennedy SH. What makes a good case–control study? Designed issues for complex traits such as endometriosis. Hum Reprod 2007; 17:1415–1423.