The association between the combined oral contraceptive pill and insulin resistance, dysglycemia and dyslipidemia in women with polycystic ovary syndrome: a systematic review and meta-analysis of observational studies

Ilana J. Halperin1,*, Shoba Sujana Kumar1, Donna F. Stroup2, and Sheila E. Laredo1

1Department of Medicine, Women’s College Hospital, University of Toronto, 1650-111 Elizabeth St. Toronto, Toronto, Ontario, Canada M5S 1B6
2Data for Solutions, Inc., PO Box 894, Decatur, GA 30031-0894, USA

*Correspondence address. Tel: +1-647-339-5251; E-mail: ilana.halperin@utoronto.ca

Submitted on March 1, 2010; resubmitted on September 29, 2010; accepted on October 1, 2010

BACKGROUND: Polycystic ovary syndrome (PCOS) is the most common endocrine disorder of young women. First-line treatment is often the oral contraceptive pill (OC), but evidence suggests that OC may worsen metabolic outcomes in this population. We undertook this meta-analysis of observational studies and cohorts from within randomized controlled studies to investigate the association between OC use and dysglycemia, dyslipidemia and insulin resistance (IR) in women with PCOS.

METHODS: We searched MEDLINE (1966–April 2010), EMBASE (1980–April 2010) and All EBM Reviews. We included prospective cohorts and RCTs that treated women, aged 13–44, with PCOS with OC for at least 3 months. Blinded quality assessment and data extraction were conducted on 35 included studies by two independent reviewers. We used random effects methods to calculate weighted mean differences as the effect size. We investigated heterogeneity using sequential removal of studies, subgroup analysis and meta-regression.

RESULTS: OC use was significantly associated with an increase in high-density lipoprotein cholesterol (HDL-C) (P = 0.004) and triglycerides (P = 0.004). Significant heterogeneity was found in glucose, cholesterol, HDL-C, low-density lipoprotein cholesterol triglycerides, fasting glucose to insulin ratios and homeostatic model assessments-IR. Study characteristics such as mean BMI, mean age and duration of study could explain some of the heterogeneity.

CONCLUSIONS: Use of OC was not associated with clinically significant adverse metabolic consequences. Because of limitations of the underlying studies, further research including rigorously designed randomized trials would more definitively confirm our findings.

Key words: polycystic ovary syndrome / oral contraceptive pill (OC) / dysglycemia / lipids / insulin resistance

Introduction

Polycystic ovary syndrome (PCOS) is the leading cause of anovulatory infertility in developed countries (ESHRE/ASRM, 2004). The clinical features of the syndrome include oligomenorrhea, acne, hirsutism, obesity and insulin resistance (IR). IR is present in the majority of women with PCOS, regardless of BMI (Dunaif et al., 1989). Women with PCOS have an increased risk for impaired glucose tolerance (IGT) (Legro et al., 1999). IGT or type two diabetes mellitus (T2DM) develops in >40% of obese women with PCOS by the age of 30 (Ehrmann et al., 1999). Decreased high-density lipoprotein cholesterol (HDL-C) levels, and increased total cholesterol, low-density lipoprotein cholesterol (LDL-C) and triglyceride levels are also associated with PCOS (Talbott et al., 1995).
The oral contraceptive pill (OC) has been recommended for >30 years to treat women with PCOS (Ehrmann, 2005). Its beneficial effects include reducing acne, improving hirsutism and correcting oligomenorrhea (Ehrmann, 2005). Less well understood is the effect of OCs on the metabolic risk profile in PCOS.

Studies in the general population have linked OCs, composed of 30–50 µg of ethinyl estradiol (EE), to reduced glucose tolerance and IR (Rimm et al., 1992; Watanabe et al., 1994). The link between OC and dysglycemia is unclear with low-dose OCs (Chazen-Taber et al., 1997), but the risks reported in the general population may underestimate the risk for women with PCOS. Dyslipidemia has been linked to the progestogen component of OCs (Ball et al., 1990; Van Rooijen et al., 2002). If OC use can induce IR, dysglycemia and dyslipidemia in the general population, then women with PCOS may have a greater risk for adverse metabolic outcomes when exposed to OCs. Until now, published studies examining the effect of the OC in women with PCOS have been small and yielded conflicting results. Thus, clinicians lack definitive data on the metabolic changes associated with OC treatment in women with PCOS.

To investigate the association between metabolic changes and OC use in women with PCOS, we undertook a systematic review and meta-analysis. We hypothesized that OC use is associated with worsening of IR, lipid profiles and glucose metabolism. We expected that differences in effect size may be explained by differences in patient age, duration of OC use, BMI, EE dose and progestogen type.

### Materials and Methods

#### Search strategy

We conducted searches with the assistance of a professional librarian. We used the following MeSH terms: ‘polycystic ovary syndrome’, ‘hyperandrogenism’, ‘contraceptives, oral’, ‘estrogen’, ‘lipids’, ‘dysglycemia’ and ‘insulin resistance’ (Supplementary data, Appendix I). We searched all EBM Reviews (a compilation database through the University of Toronto that includes: Cochrane Database of Systematic Reviews, ACP Journal Club, Database of abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Health Technology Assessment and National Health Service Economic Evaluation), Medline (1966–April 2010) and EMBASE (1980–April 2010). We searched abstracts from the conference proceedings of the European Society of Human Reproduction and Embryology, the American Society of Reproductive Medicine and the Endocrine Society (2005–2009) for relevant unpublished literature and contacted the authors for their data. Experts in the field were contacted for possible unpublished research on the topic.

#### Eligibility of relevant studies

We included studies that used the NIH (Zawadzki and Dunai, 1992) or Rotterdam (ASRM/ESHRE, 2004) criteria for diagnosis of PCOS. We also included studies that did not specify the terms NIH or Rotterdam criteria but described the presence of at least two of the following three clinical criteria in their study cohort, consistent with NIH or Rotterdam criteria: oligoovulation (nine or fewer menses per year), hyperandrogenism and polycystic ovaries on ultrasound. We excluded studies of women with pre-existing diabetes. The intervention of interest was contraceptive dose EE, combined with any type of progestogen. Studies involving women aged 13–44 years were included. The minimum follow-up period required for inclusion was 3 months to allow the menstrual rate, sex hormone binding globulin and androgens time to normalize. In studies with more than one end-point, the follow-up closest to 12 months was used.

Initial inclusion criteria were restricted to RCTs that compared either OC use to non-use or two different types of OCs. The first search using these criteria resulted in six possible studies (Creatsas et al., 2000; Mastorakos et al., 2002; Cagnacci et al., 2003; Vrbikova et al., 2004; Mastorakos et al., 2006; Chantidou et al., 2008). None of the RCTs compared OC with placebo and thus the restriction to RCT design was insufficient to address our research question. Therefore, we expanded eligibility criteria to include prospective cohort studies of women with PCOS comparing our outcomes of interest before and after OC administration. This review was conducted according to guidelines for the meta-analysis of observational studies (Stroup et al., 2000).

In studies that assigned women to OCs or OCs and another intervention such as an insulin sensitizing agent (MorinPapunen et al., 2000; Harborne et al., 2003; Allen et al., 2005; Cibula et al., 2005; Lv et al., 2005; Luque-Ramirez et al., 2007; Hoeger et al., 2008; Wu et al., 2008; Bilgir et al., 2009; Essah et al., 2009; Teede et al., 2010), a lipid lowering agent (Duleba et al., 2006) or an adjuvant hormonal treatment (Gokmen et al., 1996; Acien et al., 1997; Chantidou et al., 2008; Ozdemir et al., 2008), the OC only arm was included as a prospective cohort. We excluded studies using hypocaloric diet as a co-intervention since weight loss may be insulin sensitizing. When both groups of a two arm study met inclusion criteria, we treated the data as two separate cohorts.

If data from the same cohort were reported in more than one study (Creatsas et al., 2000; Mastorakos et al., 2002, 2006), we included the outcomes from only the most recent study (Mastorakos et al., 2002), unless different outcomes were reported in the separate papers (Mastorakos et al., 2002, 2006).

#### Data extraction and quality assessment

Two reviewers (S.S.K. and I.J.H.) blinded to study authors, institutions, journal name, volume and page numbers completed data abstraction independently. When data were not available from the published papers, repeated efforts were made to contact the authors. A validated quality score (Wells et al., 2000) was modified to assign a score to each paper based on three components: selection of groups for study, quality of the adjustment for confounding and ascertainment of the outcomes of interest (Supplementary data, Appendix I). After quality assessment was completed but before analysis, we defined studies with a score of eight or nine as ‘good’, seven as ‘fair’ and six or less as ‘poor’ quality. Following published guidelines (Stroup et al., 2000), we used the quality score in subgroup analysis. Abstrator agreement was measured by kappa (K) for concordance. Disagreement was resolved by consensus and adjudication by the senior reviewer (S.E.L.).

#### Outcome measures

The primary outcomes of interest were fasting insulin, glucose, total cholesterol, triglycerides, HDL-C and LDL-C. All reported measurements of IR were extracted, but because of the variability in methods of measurement, only glucose disposal rates as measured by the hyperinsulinemic euglycemic clamp (M-value), fasting glucose to insulin ratios (G/I) and homeostatic model assessments (HOMA-IR) were pooled.

#### Statistical methods

We standardized all units to SI measurements. When reported units were not SI units, we standardized using these criteria: oligoovulation (nine or fewer menses per year), hyperandrogenism and polycystic ovaries on ultrasound. We excluded studies with more than one end-point, the follow-up closest to 12 months was used.

Initial inclusion criteria were restricted to RCTs that compared either OC use to non-use or two different types of OCs. The first search using these criteria resulted in six possible studies (Creatsas et al., 2000; Mastorakos et al., 2002; Cagnacci et al., 2003; Vrbikova et al., 2004; Mastorakos et al., 2006; Chantidou et al., 2008). None of the RCTs compared OC with placebo and thus the restriction to RCT design was insufficient to address our research question. Therefore, we expanded eligibility criteria to include prospective cohort studies of women with PCOS comparing our outcomes of interest before and after OC administration. This review was conducted according to guidelines for the meta-analysis of observational studies (Stroup et al., 2000).

In studies that assigned women to OCs or OCs and another intervention such as an insulin sensitizing agent (MorinPapunen et al., 2000; Harborne et al., 2003; Allen et al., 2005; Cibula et al., 2005; Lv et al., 2005; Luque-Ramirez et al., 2007; Hoeger et al., 2008; Wu et al., 2008; Bilgir et al., 2009; Essah et al., 2009; Teede et al., 2010), a lipid lowering agent (Duleba et al., 2006) or an adjuvant hormonal treatment (Gokmen et al., 1996; Acien et al., 1997; Chantidou et al., 2008; Ozdemir et al., 2008), the OC only arm was included as a prospective cohort. We excluded studies using hypocaloric diet as a co-intervention since weight loss may be insulin sensitizing. When both groups of a two arm study met inclusion criteria, we treated the data as two separate cohorts.

If data from the same cohort were reported in more than one study (Creatsas et al., 2000; Mastorakos et al., 2002, 2006), we included the outcomes from only the most recent study (Mastorakos et al., 2002), unless different outcomes were reported in the separate papers (Mastorakos et al., 2002, 2006).

#### Data extraction and quality assessment

Two reviewers (S.S.K. and I.J.H.) blinded to study authors, institutions, journal name, volume and page numbers completed data abstraction independently. When data were not available from the published papers, repeated efforts were made to contact the authors. A validated quality score (Wells et al., 2000) was modified to assign a score to each paper based on three components: selection of groups for study, quality of the adjustment for confounding and ascertainment of the outcomes of interest (Supplementary data, Appendix I). After quality assessment was completed but before analysis, we defined studies with a score of eight or nine as ‘good’, seven as ‘fair’ and six or less as ‘poor’ quality. Following published guidelines (Stroup et al., 2000), we used the quality score in subgroup analysis. Abstrator agreement was measured by kappa (K) for concordance. Disagreement was resolved by consensus and adjudication by the senior reviewer (S.E.L.).

#### Outcome measures

The primary outcomes of interest were fasting insulin, glucose, total cholesterol, triglycerides, HDL-C and LDL-C. All reported measurements of IR were extracted, but because of the variability in methods of measurement, only glucose disposal rates as measured by the hyperinsulinemic euglycemic clamp (M-value), fasting glucose to insulin ratios (G/I) and homeostatic model assessments (HOMA-IR) were pooled.

#### Statistical methods

We standardized all units to SI measurements. When reported units were clinically improbable, that is, differing from expected clinical values by a factor greater than 10, we contacted the authors for clarification; if no response was received, we excluded the inconsistent variable from the analysis. Effect size for each outcome in each cohort was calculated as the standardized mean difference (Borenstein et al., 2009). For studies
where standard errors were not reported, we calculated relevant standard errors from reported confidence intervals (CIs). We excluded studies where CIs were not reported or were not able to be calculated (Higgins et al., 2003). Effect sizes were weighted by the inverse of the within study variance (Hedges and Olkin, 1985). We calculated both fixed and random effects combined estimates; however, given the high risk for heterogeneity, we elected to report only the more conservative random effects models. We assessed heterogeneity using a $\chi^2$ Q statistic with a significance level of $\alpha \leq 0.10$, and by calculating an $I^2$ (using values >50% as indicating important inconsistency), (Borenstein et al., 2009).

We explored heterogeneity extensively. First, we re-ran the analyses using only ‘good’ quality studies. Second, using the forest plot for each outcome, we removed cohorts sequentially in order of largest residuals (Borenstein et al., 2009). Third, a priori subgroups were used to assess the impact of age, BMI (in categories of obese ($\geq 30$ kg/m$^2$), overweight (25–29.9 kg/m$^2$) and normal), duration of exposure, EE dose, progestogen type (cyproterone acetate (CPA), drospirenone (DRS) or conventional) (Roberts and Stanley, 2005). We looked for significant differences between the subgroups using the $\chi^2$ Q statistic. Fourth, we used a meta-regression with age and BMI as continuous variables using a conservative mixed maximum likelihood (Roberts and Stanley, 2005). Finally, for outcomes where the effect size was significant, we conducted a multivariate meta-regression to assess the simultaneous contribution of age and BMI with duration of use, estrogen dose or progestin type (Borenstein et al., 2009). We evaluated publication bias in significant outcomes using Duval and Tweedie non-parametric ‘trim and fill’ method (Duval and Tweedie, 2000), Egger’s test (Egger et al., 1997), the Fail-safe N test (Rosenthal, 1979) and Kendall’s Tau (Begg and Berlin, 1988). Calculations were performed using Comprehensive Meta Analysis™ and confirmed using SAS version 9.2.

## Results

### Search results

The literature search revealed 1218 studies (Fig. 1). Of these, we excluded 1126 based on title and 54 based on detailed review because they did not meet the pre-specified inclusion criteria, did not report on our outcome of interest or were duplicate publications. Three studies were excluded during data analysis (Cibula et al., 2002;...
Harborne et al., 2003; Meyer et al., 2007) as they did not report any measures of standard error or CIs. Seven of the included studies (Mastorakos et al., 2002, 2006, Cagnacci et al., 2003; Vrbikova et al., 2004; Chantidou et al., 2008; Wu et al., 2008; Chen et al., 2010) contributed two cohorts each, and therefore the final analyses were conducted using 42 cohorts (Table I).

### Quality assessment

There were 17 studies classified as good, 9 as fair and 16 as poor quality. The initial $K$ for agreement was 0.63 (Table II); all differences were resolved by consensus and adjudication.

### Included cohorts

Of the 35 studies, 24 were conducted in Europe, 6 in North America, 1 in South America, 4 in China and 1 in Australia. The publication dates ranged from 1995 to 2010. The median duration of follow up was 6 months (range 3–36). The median age was 24.3 years (range 15.3–33.5). The median BMI was 24.7 (range 21.8–40.1). There were eight cohorts with a mean BMI $>30$ kg/m². Of the cohorts, 24 were exposed to EE 35 mg and CPA 2 mg, 5 cohorts were exposed to EE 30 mg and DRS 3 mg, 12 cohorts were exposed to varying doses of EE and conventional progestogens (desogestrel, nor-ethindrone and norgestimate), and one cohort was treated with trans-

---

**Table I Characteristics of included studies.**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Location (country)</th>
<th>Study design</th>
<th>n (used in analysis)</th>
<th>Mean age</th>
<th>Mean BMI</th>
<th>Duration of follow-up (months)</th>
<th>Contraceptive intervention (mg)</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acien et al. (1997)</td>
<td>Spain</td>
<td>RCT</td>
<td>12</td>
<td>23.70</td>
<td>26.90</td>
<td>3–7</td>
<td>EE 0.035/CPA 2</td>
<td>6</td>
</tr>
<tr>
<td>Allen et al. (2005)</td>
<td>USA</td>
<td>RCT</td>
<td>15</td>
<td>15.30</td>
<td>40.10</td>
<td>6</td>
<td>EE 0.035/NRG 0.25</td>
<td>9</td>
</tr>
<tr>
<td>Armstrong et al. (2001)</td>
<td>UK</td>
<td>PC</td>
<td>11</td>
<td>28.00</td>
<td>27.30</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>5</td>
</tr>
<tr>
<td>Bligir et al. (2009)</td>
<td>Turkey</td>
<td>RCT</td>
<td>20</td>
<td>24.30</td>
<td>28.2</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Cagnacci et al. (2003)</td>
<td>Italy</td>
<td>RCT</td>
<td>10</td>
<td>21.80</td>
<td>22.60</td>
<td>6</td>
<td>EE 0.040–0.030/DSG 0.25–0.125</td>
<td>9</td>
</tr>
<tr>
<td>Cagnacci et al. (2003)</td>
<td>Italy</td>
<td>RCT</td>
<td>10</td>
<td>22.70</td>
<td>23.50</td>
<td>6</td>
<td>EE 0.035/CPA 2</td>
<td>9</td>
</tr>
<tr>
<td>Chantidou et al. (2008)</td>
<td>Greece</td>
<td>RCT</td>
<td>20</td>
<td>20.98</td>
<td>22.77</td>
<td>12</td>
<td>EE 0.035/CPA 2</td>
<td>6</td>
</tr>
<tr>
<td>Chantidou et al. (2008)</td>
<td>Greece</td>
<td>RCT</td>
<td>23</td>
<td>21.74</td>
<td>22.5</td>
<td>12</td>
<td>EE 0.03–0.035/DSG 0.05–0.150</td>
<td>6</td>
</tr>
<tr>
<td>Chen et al. (2010)</td>
<td>China</td>
<td>PC</td>
<td>21</td>
<td>24.80</td>
<td>30.65</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Chen et al. (2010)</td>
<td>China</td>
<td>PC</td>
<td>35</td>
<td>24.80</td>
<td>20.53</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Cibula et al. (2005)</td>
<td>Czech Republic</td>
<td>RCT</td>
<td>15</td>
<td>23.20</td>
<td>22.10</td>
<td>6</td>
<td>EE 0.035/NRG 0.03/DSG 0.25</td>
<td>6</td>
</tr>
<tr>
<td>Duleba et al. (2006)</td>
<td>Poland</td>
<td>RCT</td>
<td>24</td>
<td>23.80</td>
<td>22.80</td>
<td>3</td>
<td>EE 0.020/DSG 0.15</td>
<td>8</td>
</tr>
<tr>
<td>Elter et al. (2002)</td>
<td>Turkey</td>
<td>RCT</td>
<td>20</td>
<td>23.45</td>
<td>21.83</td>
<td>4</td>
<td>EE 0.035/CPA 2</td>
<td>9</td>
</tr>
<tr>
<td>Essah et al. (2010)</td>
<td>US</td>
<td>RCT</td>
<td>10</td>
<td>24.30</td>
<td>32.60</td>
<td>3</td>
<td>EE 0.035/NRG 0.18/0.215</td>
<td>7</td>
</tr>
<tr>
<td>Falsetti et al. (1995)</td>
<td>Italy</td>
<td>PC</td>
<td>72</td>
<td>23.00</td>
<td>22.00</td>
<td>12</td>
<td>EE 0.035/CPA 2</td>
<td>4</td>
</tr>
<tr>
<td>Gokmen et al. (1996)</td>
<td>Turkey</td>
<td>PC</td>
<td>48</td>
<td>24.70</td>
<td>23.90</td>
<td>6</td>
<td>EE 0.035/CPA 2</td>
<td>7</td>
</tr>
<tr>
<td>Guido et al. (2004)</td>
<td>Italy</td>
<td>PC</td>
<td>15</td>
<td>24.68</td>
<td>24.68</td>
<td>12</td>
<td>EE 0.03/DRS 3</td>
<td>6</td>
</tr>
<tr>
<td>Gul et al. (2008)</td>
<td>Turkey</td>
<td>PC</td>
<td>30</td>
<td>25.1</td>
<td>23.85</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>6</td>
</tr>
<tr>
<td>Hoeger et al. (2008)</td>
<td>USA</td>
<td>RCT</td>
<td>10</td>
<td>15.40</td>
<td>37.8</td>
<td>6</td>
<td>EE 0.030/DSG 0.15</td>
<td>7</td>
</tr>
<tr>
<td>Korykowski et al. (1995)</td>
<td>USA</td>
<td>PC</td>
<td>9</td>
<td>24.00</td>
<td>28.00</td>
<td>3</td>
<td>EE 0.035/NITA 0.25–0.75</td>
<td>7</td>
</tr>
<tr>
<td>Luque-Ramirez et al. (2007)</td>
<td>Spain</td>
<td>RCT</td>
<td>15</td>
<td>23.4</td>
<td>29.2</td>
<td>6</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Lv et al. (2005)</td>
<td>China</td>
<td>RCT</td>
<td>25</td>
<td>24.35</td>
<td>21.81</td>
<td>6</td>
<td>EE 0.035/CPA 2</td>
<td>7</td>
</tr>
<tr>
<td>Mastorakos et al. (2002)</td>
<td>Greece</td>
<td>RCT</td>
<td>14</td>
<td>17.53</td>
<td>25.50</td>
<td>12</td>
<td>EE 0.035/DSG 0.15</td>
<td>8</td>
</tr>
<tr>
<td>Mastorakos et al. (2002)</td>
<td>Greece</td>
<td>RCT</td>
<td>14</td>
<td>17.46</td>
<td>24.84</td>
<td>12</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Mastorakos et al. (2006)</td>
<td>Greece</td>
<td>RCT</td>
<td>18</td>
<td>17.16</td>
<td>25.40</td>
<td>12</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Mastorakos et al. (2006)</td>
<td>Greece</td>
<td>RCT</td>
<td>18</td>
<td>17.01</td>
<td>25.80</td>
<td>12</td>
<td>EE 0.030/DSG 0.15</td>
<td>8</td>
</tr>
<tr>
<td>Morin-Papunen et al. (2000)</td>
<td>Finland</td>
<td>RCT</td>
<td>10</td>
<td>29.80</td>
<td>37.20</td>
<td>6</td>
<td>EE 0.035/CPA 2</td>
<td>7</td>
</tr>
<tr>
<td>Morin-Papunen et al. (2003)</td>
<td>Finland</td>
<td>RCT</td>
<td>9</td>
<td>28.50</td>
<td>21.80</td>
<td>6</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Nader et al. (1997)</td>
<td>USA</td>
<td>PC</td>
<td>16</td>
<td>29.00</td>
<td>36.80</td>
<td>6</td>
<td>EE 0.03/DSG 0.15</td>
<td>6</td>
</tr>
</tbody>
</table>

Continued
Meta-analysis of outcomes

The combined effect sizes for various outcomes are presented in Table III.

### Fasting glucose

OC use was not associated with change in fasting glucose ($P = 0.69$). The 29 cohorts had significant heterogeneity ($Q < 0.001, I^2 = 96.7\%$). Removal of fair and poor-quality studies or those with the largest residuals did not explain the heterogeneity. Subgroup analysis reveals significant differences in the studies across BMI categories ($P = 0.01$).

### Fasting insulin

OC use was not associated with fasting insulin ($P = 0.07$). The 33 cohorts were homogenous. When fair and poor-quality studies were removed, the lack of association persisted. Subgroup and meta-regression analyses were not significant.

### Total cholesterol

OC use was not associated with changes in cholesterol ($P = 0.36$). The 29 cohorts had significant heterogeneity ($Q < 0.001, I^2 = 99.7\%$). Subgroup analysis, meta-regression or removal of fair and poor-quality studies or those with the largest residuals did not explain the heterogeneity.

### HDL cholesterol

OC use was significantly associated with an increase in HDL-C ($P = 0.004$; Fig. 2). This association remained significant when fair and poor studies were removed ($P = 0.01$). The 27 cohorts exhibited significant heterogeneity ($Q < 0.001, I^2 = 99.4\%$). Removal of fair and poor-quality studies or those with the largest residuals did not explain the heterogeneity.

Subgroup analysis did not reveal any significant differences in the studies across BMI category. However, when studies with a BMI of $30\,\text{kg/m}^2$ or higher were combined, there was a significant association ($P = 0.01$), whereas those studies with a BMI $<30\,\text{kg/m}^2$ did not show a significant association between OC use and HDL-C.

### LDL cholesterol

OC use was not associated with increased LDL-C ($P = 0.14$). The 22 cohorts showed significant heterogeneity ($Q < 0.001, I^2 = 99\%$) which was not explained by quality of study, sequential study removal or subgroup analysis. In studies where the average age of dermal EE $50\,\mu\text{g}$ and $2\,\text{mg}$ of CPA orally (Table I).

### Table I

<table>
<thead>
<tr>
<th>Reference</th>
<th>Location (country)</th>
<th>Study design</th>
<th>n (used in analysis)</th>
<th>Mean age</th>
<th>Mean BMI</th>
<th>Duration of follow-up (months)</th>
<th>Contraceptive intervention (mg)</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ozdemir et al. (2008)</td>
<td>Turkey</td>
<td>RCT</td>
<td>40</td>
<td>22.70</td>
<td>24.3</td>
<td>6</td>
<td>EE 0.030/DRS 3</td>
<td>8</td>
</tr>
<tr>
<td>Palep-Singh et al. (2004)</td>
<td>UK</td>
<td>PC</td>
<td>13</td>
<td>26.40</td>
<td>29.20</td>
<td>6</td>
<td>EE 0.030/DRS 3</td>
<td>4</td>
</tr>
<tr>
<td>Pehlivanov and Mitkov (2007)</td>
<td>Bulgaria</td>
<td>PC</td>
<td>20</td>
<td>26.45</td>
<td>26.45</td>
<td>6</td>
<td>EE 0.035/DRS 3</td>
<td>5</td>
</tr>
<tr>
<td>Sagsoz et al. (2009)</td>
<td>Turkey</td>
<td>PC</td>
<td>20</td>
<td>25.70</td>
<td>22.60</td>
<td>3</td>
<td>EE 0.030/DRS 3</td>
<td>6</td>
</tr>
<tr>
<td>Teede et al. (2010)</td>
<td>Australia</td>
<td>RCT</td>
<td>26</td>
<td>33.50</td>
<td>35.80</td>
<td>6</td>
<td>EE 0.035/CPA 2</td>
<td>6</td>
</tr>
<tr>
<td>Villaseca et al. (2004)</td>
<td>Chile</td>
<td>RCT</td>
<td>16</td>
<td>22.40</td>
<td>32.80</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>7</td>
</tr>
<tr>
<td>Vrbikova et al. (2006)</td>
<td>Czech Republic</td>
<td>PC</td>
<td>19</td>
<td>24.30</td>
<td>7 $+/−$ 3.8</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>2</td>
</tr>
<tr>
<td>Vrbikova et al. (2004)</td>
<td>Czech Republic</td>
<td>RCT</td>
<td>12</td>
<td>26.10</td>
<td>24.70</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>7</td>
</tr>
<tr>
<td>Vrbikova et al. (2004)</td>
<td>Czech Republic</td>
<td>RCT</td>
<td>12</td>
<td>25.50</td>
<td>24.40</td>
<td>3</td>
<td>Transdermal EE 0.05/CPA 2</td>
<td>7</td>
</tr>
<tr>
<td>Wu et al. (2008)</td>
<td>China</td>
<td>RCT</td>
<td>8</td>
<td>25.30</td>
<td>25.00</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Wu et al. (2008)</td>
<td>China</td>
<td>RCT</td>
<td>12</td>
<td>21.40</td>
<td>24.70</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>8</td>
</tr>
<tr>
<td>Ye et al. (2003)</td>
<td>China</td>
<td>PC</td>
<td>15</td>
<td>26.80</td>
<td>7 $+/−$ 3.8</td>
<td>3</td>
<td>EE 0.035/CPA 2</td>
<td>5</td>
</tr>
</tbody>
</table>

RCT, randomized controlled trial; PC, prospective cohort; EE, ethinyl estradiol; CPA, cyproterone acetate; DRS, drospirenone; DSG, desogestrel; CE, conjugated estrogen; NRG, norgestimate; NTA, norethindrone acetate.
the women was older, the association between LDL-C and OC use was less pronounced ($P = 0.02$).

**Triglycerides**

OC use was significantly associated with an increase in triglycerides ($P = 0.004$) (Fig. 3). Removal of fair and poor-quality studies led to a loss of significance in the combined estimate ($P = 0.12$). The 25 cohorts exhibited significant heterogeneity ($Q = 0.001$, $I^2 = 99.6\%$). Removal of fair and poor-quality studies or those with the largest residuals did not explain the heterogeneity.

Subgroup analysis revealed that some heterogeneity can be explained by differences in study duration; effect size was significant at all durations, but studies longer than 6 months exhibited a stronger association between OC use and increased triglycerides (1.72 vs. 0.35, $P = 0.001$). There was no significant difference across mean BMI subgroups, or when BMI was used in the univariate or multivariate regression. From the multivariate model, mean age, mean BMI and duration of study together account for 56% of the variability in the triglyceride outcome.

This outcome showed some evidence of publication bias suggesting that studies with large effect sizes may be underrepresented [Egger’s intercept = 15.43, $P < 0.001$; Kendall’s Tau (0.39, $P = 0.009$)]. However, Fail-safe $N$ suggests 5534 studies are needed to make the result non-significant and the trim and fill adjustment showed no statistical difference from the overall effect size for triglycerides.

**Fasting glucose to insulin ratio**

OC use was not associated with $G_0/I_0$ ($P = 0.41$). The eight cohorts had significant heterogeneity ($Q < 0.001$, $I^2 = 99.99\%$) not explained by quality of study, sequential study removal, or subgroup analysis. Univariate meta-regression suggested that as average BMI of the women in the study increased, the association between $G_0/I_0$ and OC was stronger ($P = 0.009$).

**HOMA-IR**

OC use had no association with HOMA-IR ($P = 0.45$) for 11 cohorts with significant heterogeneity ($Q < 0.001$, $I^2 = 92\%$). The lack of association persisted when poor-quality studies were removed. When one study (Chen et al., 2010) was removed, statistical
Homogeneity was achieved ($Q = 0.59, I^2 = 0$) but the combined effect size remained non-significant. This study compared obese and non-obese women on 35 μg of EE and 2 mg of CPA. The effect size for HOMA-IR in the non-obese group on the OC was much higher than for any other cohort. Univariate meta-regression suggested that as average BMI of the women in the studies increased, the association between HOMA-IR and OCs was less pronounced ($P = 0.0006$).

**Euglycemic hyperinsulinemic clamp:** glucose disposal rate (M)
OC use was not associated with change in $M$-value ($P = 0.96$). The five cohorts were homogenous. Subgroup analysis and meta-regression were not significant.

**Discussion**

**Principal findings**

Our results suggest that OC use is not associated with a significant change in insulin, glucose, total cholesterol or LDL-C in women with PCOS. We found a statistically significant but clinically small increase in HDL-C and triglycerides. No significant change was found when IR was measured by HOMA-IR, $G_0/I_0$ or the glucose disposal rate. For our primary outcomes, the 95% CIs are sufficiently narrow to indicate that clinically important adverse effects on metabolic outcomes in the first 12 months of use of the OC can be excluded.

Most outcomes had significant heterogeneity; we were unable to completely explain heterogeneity using reported characteristics of the cohorts. Differences in average BMI of women in the individual cohorts explained a portion of the heterogeneity found in fasting glucose, HDL-C, HOMA-IR and $G_0/I_0$. Differences in study duration accounted for some of the heterogeneity found in triglycerides. Differences in mean age of the women in the individual cohorts contributed to the heterogeneity found in LDL-C and HDL-C. Interestingly, estrogen dose and progesterone type used by the various cohorts could not explain any of the heterogeneity or effect sizes seen, which suggests that different types of OCs used in women with PCOS do not result in different metabolic effects. The tests for publication bias in the triglyceride outcome were inconsistent. Given the limitations in statistical tests for publication bias (Sterne *et al.*, 2000), the magnitude of the Fail-safe N (>5000 studies) and the comprehensive nature of our literature search, we are comfortable that the significant association between OC use and serum triglycerides is not influenced by publication bias.

Subgroup analysis and meta-regression by BMI did not show any consistent trends. Results may differ if studies included more women with higher BMI, typical of women with PCOS. The subgroup analysis and meta-regression suggested a stronger association between OC use and HDL-C in those studies of women with a higher average BMI. The meta-regression does not suggest any modifying effect of BMI on the association of OC use and triglycerides. Although there was no significant association between OC use and IR, BMI does appear to modify the relationship between OC use and HOMA-IR: as the average BMI of a cohort increased, there was less of an association. In contrast, as the average BMI of a cohort increased the association between OC use and $G_0/I_0$ became stronger. These results are somewhat unexpected as IR usually worsens as BMI increases. We expected that BMI would modify the association between OC use and metabolic outcomes.
outcomes as obesity and the metabolic syndrome are commonly linked to low HDL-C, elevated triglycerides and increased IR (Trevisan et al., 1998). Given the observational nature of the data and the use of cohort, rather than patient level, characteristics, we caution against drawing clinical conclusions from subgroup analyses alone.

Comparison with previous studies
This is the first meta-analysis to examine the effect of OC alone on metabolic outcomes. A 2007 Cochrane review (Costello et al., 2007) pooled the results of four RCTs comparing OCs with metformin and showed no significant difference in the two treatments for insulin, glucose, total cholesterol or the development of T2DM, but a significant increase in triglycerides in the OC group compared with metformin. A 2008 systematic review (Jing et al., 2008) pooled nine studies comparing EE/CPA to metformin. It showed no significant difference in the prevalence of IGT, impaired fasting glucose or T2DM; only triglycerides showed a significant difference between groups. We are reassured that our results are consistent with the results of these two meta-analyses. Increased triglycerides likely reflect the estrogenic effects of OCs (Yuan et al., 2007). Reassuringly, the triglycerides remain well below the level that would increase risk for pancreatitis (Yuan et al., 2007).

Strengths and weaknesses
A strength of this analysis is the rigorous and comprehensive literature search and the careful application of the meta-analytic method. Due to the absence of relevant comparative RCTs, we modified our protocol to include observational cohort studies and individual arms of RCTs (Shrier et al., 2007). Meta-analysis of observational studies may produce equally or more precise results for a clinical question when compared with meta-analysis of RCTs alone (Shrier et al., 2007). The most important difference between RCTs and observational studies is the potential for confounding. Our findings are limited by this risk because patients receiving OC for treatment of PCOS may have also modified certain lifestyle factors that could have affected their metabolic profile.

The lack of a standardized, feasible method for reporting IR presented challenges for pooling data, and the non-significant results for these outcomes may be due to small numbers of studies. In addition, we were not able to compare IGT across studies because of non-uniform tests. Every effort was made to contact primary authors, clarify discrepancies in units and obtain CIs and missing data; however, some outcome data had to be excluded from the analysis when authors failed to respond.

We sought to explore the association of OC use with components of the metabolic profile; however, the risk of a type I error could be inflated given the large number of primary outcomes. The lack of randomization in the existing literature does not permit any conclusion about causation. The quality of the literature used in our analysis was modest; this further affects the strength of the clinical conclusions that can be drawn from our results.

Implications for future research
As PCOS is the most common endocrine disorder of young women and OC is often offered to control menstrual and hyperandrogenic...
symptoms, further research is needed to clarify the effects of this medication on long-term metabolic outcomes. Specifically, the use of an oral glucose tolerance test before and after a reasonable duration of OC use may prove beneficial in detecting subtle, clinically relevant, changes in metabolism and should be used in future studies (Legro et al., 1999).

Standardization of IR measurements will further help to clarify the effect of the OC on the metabolic profile of women with PCOS.

Conclusion

We performed a meta-analysis of observational studies to evaluate the metabolic effects of OC use in women with PCOS using cohort level variables such age, BMI and duration to explain a portion of the variability in the outcomes of interest. In summary, short-term OC use was not associated with a clinically significant worsening in metabolic outcomes for women with PCOS. Due to the observational nature of the data, the moderate quality of the underlying studies and the relatively short follow-up period, results should be considered preliminary. Further study is needed to reassure women with PCOS that the OC is safe for long-term use.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors’ roles

I.J.H. and S.E.L. were involved in the generating the protocol, the literature search, the data abstraction form, interpreting the results and drafting the manuscript. I.J.H. and S.S.K. performed the data extraction. D.F.S. refined the data abstraction form, conducted the statistical analysis and helped to draft the manuscript.

Acknowledgements

Rita Vine helped prepare the original search terms and strategy. Sharon Sadry and Brittany Rance ran the initial searches and helped with the first round of review and article selection. Jocelyn Pang assisted with the second round literature search. Virginia Collins prepared the randomly ordered blinded packages of studies for data extraction. Katherine Pittenger and Marguerite Cox provided expert statistical programming and produced the figures. Jenny Shu translated the non-English language articles.

Conflict of interest: none declared.

References

Bal Mj, Ashwell E, Jackson M. Gilmer comparison of two triphasic contraceptives with different progestogens: effects on metabolism and coagulation proteins. Contraception 1990; 41:363–376.
Charitidou C, Farmakiotis D, Zournatzi V, Pidonia I, Pegou T, Karamanis N, Hatzisilinou M, Katsikis I, Panidis D. The administration of estrogens, combined with anti-androgens, has beneficial effects on the hormonal features and asymmetric dimethyl-arginine levels, in women with the polycystic ovary syndrome. Atherosclerosis 2008; 196:958–965.


Essah PA, Arrowood JA, Cheang KI, Adawadkar SS, Stovall DW, Nestler JE. Effect of combined oral and oral contraceptive on metabolic factors and endothelial function in overweight and obese women with polycystic ovary syndrome. Scientific Program: ENDO 2009 Unpublished.


Mastorakos G, Koliopoulos C, Creatsas G. Androgen and lipid profiles in adolescents with polycystic ovary syndrome who were treated with two forms of combined oral contraceptives. Fertil Steril 2002;77:919–927.


The Rotterdam ESHE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term


