Glutathione S-transferases M1/T1 gene polymorphisms and endometriosis: a meta-analysis of genetic association studies

Sun-Wei Guo

Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA

1To whom correspondence should be addressed at: Department of Pediatrics, Medical College of Wisconsin, 8701 Watertown Plank Road, MS 756, Milwaukee, WI 53226-0509, USA. E-mail: swguo@mail.mc.w.edu

In view of the controversies surrounding the glutathione S-transferases (GST) M1/T1–endometriosis association, a meta-analysis of the GSTM1/GSTT1 genetic association studies of endometriosis was performed. In this meta-analysis involving 14 GSTM1 studies with 1539 cases and 1805 controls and nine GSTT1 studies with 746 cases and 834 controls, respectively, substantial heterogeneities among studies were found. In addition, asymmetry in funnel plot was evident, which is likely to stem from publication bias, given no apparent indication of true heterogeneity. The bias appears to be prominent for GSTM1 studies, but is less so for GSTT1 studies. After correction for this bias, there is no evidence that women with GSTM1 null genotype have increased risk of developing endometriosis as compared with women with other genotypes. For GSTT1, the risk associated with the null genotype is 29% higher than other genotypes. However, even this estimate should be viewed with a large grain of salt, because the estimate could easily lose its statistical significance if there is a realistic 69–80% publication probability.

Key words: association/endometriosis/genetic/glutathione S-transferase/GSTM1/GSTT1/meta-analysis/polymorphism

Introduction

Endometriosis is one of the most common gynecologic disorders affecting, reportedly, 2–22% of women of reproductive age (Mahmood and Templeton, 1991; Olive and Schwartz, 1993). Despite extensive research on endometriosis, our current knowledge of its etiology and pathogenesis is quite limited (Giudice et al., 1998; D’Hooghe et al., 2003). With the completion of the first draft of the human genome and the availability of cheaper and faster genotyping technologies, there is a burgeoning interest in identifying genes and genetic polymorphisms that predispose women to increased risk of developing endometriosis (Zondervan et al., 2002a; Kennedy, 2003), based on reports that there is a heritable component in endometriosis susceptibility (Bischoff and Simpson, 2000). Among all published studies aimed at the identification of genetic susceptibility to endometriosis, glutathione S-transferases (GST) gene polymorphisms GSTM1 and GSTT1 are the most studied (see the website: http://www.well.ox.ac.uk/~S-transferase/GSTM1/GSTT1/meta-analysis/polymorphism).

GSTs are phase II enzymes involved in the detoxification of a broad range of toxic compounds and carcinogens (Hayes and Pulford, 1995), including dioxins and polycyclic aromatic hydrocarbons (PAH), which are ubiquitous and probably the most feared environmental contaminants worldwide. In humans, six classes of GST enzymes, α, μ, θ, π, ζ and η, have been identified, with each class being encoded by a separate gene or gene subfamily (Nebert and Vasiliou, 2004). The GSTM1 gene, located on chromosome 1p13.3 (Pearson et al., 1993), codes for cytosolic GST μ class enzyme, and has a deletion polymorphism that, when homozygote (GSTM1 null), results in the complete absence of a functional gene product (Seidgard et al., 1988). The GSTT1 gene, located on chromosome 22q11.2 (Webb et al., 1996), codes for GST class enzyme and also has an inactivating homozygous deletion polymorphism (Pemble et al., 1994). The frequency of the GSTM1 null genotype varies from population to population and was reported to be about 53% in Caucasians and Asians (Coton et al., 2000; Garte et al., 2001; Geisler and Olshan, 2001). The frequency of its GSTT1 counterpart is lower and was reported to range from 11 to 20% in Caucasians and about 47% in Asians (Rebeck, 1997).

The first report on the GSTM1 and endometriosis link was made by Baranov et al. in 1996 in a study of the proportion of GSTM1 null genotype in patient groups with different medical conditions and in some Slavic populations (Baranov et al., 1996). In 1999, Baranova et al. reported possible involvement of GSTT1 deletion polymorphism in the development of endometriosis. The impetus for evaluating the endometriosis patient group was apparently due to the report that exposure to dioxin increases the risk of endometriosis in rhesus monkeys, which was first reported in 1993 (Rier et al., 1993). Given the detoxification properties of the GST enzyme family, it is plausible that the lack of or reduced function of detoxification enzymes due to the deletion polymorphism may predispose women with increased risk for endometriosis.

Despite this plausibility, however, there have been conflicting reports on the GSTM1/GSTT1 and endometriosis association (see Results section). In addition, a recent comprehensive reappraisal of all published primate data on the dioxin–endometriosis link casts serious doubt about the dioxin–endometriosis link (Guo, 2004). In fact, the reappraisal exposes some major methodological deficiencies in the 1993 article that had not been recognized previously. These deficiencies seriously undermine the conclusions of the first report (Guo, 2004). Three recent publications on the same topic (Fierens et al., 2003; De Felip et al., 2004; Lim et al., 2004), not included in the reappraisal, appear to lend further support to the view that such a link does not exist.
In the light of this development, and in view of the controversies surrounding the GSTM1/GSTT1–endometriosis association, the role of GSTs, as a detoxification mechanism, in the pathogenesis of endometriosis begs for a reassessment. To provide a comprehensive and quantitative assessment of the impact of GSTM1/GSTT1 gene polymorphisms on the risk of endometriosis, a meta-analysis of all relevant published genetic association studies was performed to estimate effect size, to determine the extent of heterogeneity in the strength of associations between studies and to attempt to identify possible sources of this heterogeneity by searching for possible publication bias.

Materials and methods

Search strategy and inclusion and exclusion criteria

The PubMed database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) was searched in a systematic and diligent manner for all genetic association studies on GSTM1/GSTT1 and endometriosis from 1996, when the GSTM1 association endometriosis was first reported (Baranov et al., 1996), up until the end of 2004. The search used the following keywords association studies, endometriosis, glutathione, glutathione-S-transferase, GSTM1, GSTT1, GSTM1 polymorphisms and GSTT1 polymorphisms, or their combinations. Because there is only one published study that examined GSTP1 polymorphism and its effect on the risk of endometriosis, GSTP1 was not included in this analysis. The references of all computer-identified publications were searched for additional studies, and the PubMed option related articles was used to search for potentially relevant articles. Articles published in all languages were selected if they met the following criteria: (i) study was an association study, and (ii) study reported on the genotypic frequencies of either GSTM1 or GSTT1 in a group of unrelated patients with endometriosis and a group of unrelated individuals serving as a control group. Those retrieved studies that are review articles or that provide no new data were excluded. In the case of sequential or multiple publications of analyses of the same or overlapping data sets, the study that reported the data from the largest or most recent publication was included, as recommended by Little et al. (2002).

Data extraction

Following the meta-analysis of observational studies in epidemiology (MOOSE) guidelines for reporting meta-analysis of observational studies (Stroup et al., 2000), the following data were extracted from eligible studies: authors’ names, region/country where the study was conducted, year of publication, number of cases/patients and controls, stages of endometriosis and their percentages in cases/patients, diagnostic criteria, mean age and standard deviation or age range in both control and case groups, how the controls were selected, and numbers of subjects with GSTM1/GSTT1 null/null genotype in both cases and controls. Information on whether the study made any attempt to test for Hardy–Weinberg proportion (HWP), to check for and correcting genotyping errors, and to control for confounding risk factors were also noted.

The search resulted in 15 published articles on genetic association of GSTM1 and endometriosis, 11 in English (Baranova et al., 1999; Baranova et al., 1997, 1999; Arvanitis et al., 2001, 2003; Baxter et al., 2001; Hadfield et al., 2001; Lin et al., 2003; Hsieh et al., 2004; Morizane et al., 2004; Hur et al., 2005), two each in Chinese (Ding et al., 2004) and in Russian (Ivashchenko et al., 2004). Of the eight published articles on GSTT1–endometriosis association, 6 were published in English (Baranova et al., 1999; Hadfield et al., 2001; Arvanitis et al., 2003; Lin et al., 2003; Morizane et al., 2004; Hur et al., 2005), and one each in Chinese (Ding et al., 2004) and in Russian (Ivashchenko et al., 2003). Again, one article by Ding et al. (2004) presented data on two ethnic populations and was thus counted as two studies, resulting in nine studies.

Seven studies examined, individually, the frequencies of both GSTM1 and GSTT1 null genotypes in cases and controls. However, only four studies (Baranova et al., 1999; Arvanitis et al., 2003; Morizane et al., 2004; Hur et al., 2005) examined the risk of combined GSTM1-GSTT1 null genotype. One study (Arvanitis et al., 2003) also examined the effect of combined genotype of GSTM1, GSTT1, CYP1A1, and CYP19. Data on the combined GSTM1-GSTT1 null genotype were extracted except (Baranova et al., 1999), which only mentioned in passing that the combined genotype did not increase the risk of endometriosis.

Statistical analysis

The goals of this analysis were to pool crude odds ratio (OR) estimates from included studies, to identify any heterogeneity, and attempt to identify its sources. For each included study, the OR and its 95% confidence interval (CI) and its standard error (SE) of the log OR were computed. The choice of log OR was simply due to the fact that, in contrast to the OR, its SE is unaffected by the magnitude of the log OR.

For both GSTM1 and GSTT1 polymorphisms, a summary OR incorporating both within- and between-study variations was calculated using a random-effects model proposed by DerSimonian and Laird (DerSimonian, 1996). This model also provides a means to test for heterogeneity in OR estimates across studies. In contrast to the fixed-effects model which assumes that all studies aim at evaluating a common truth and results differ purely as a result of sampling errors (Mantel and Haenszel, 1959), the random-effects model stipulates that the studies may have genuine variations in their results, in addition to sampling errors. These genuine variations may be attributable to some unknown variables not accounted for by all studies. The fixed-effects model is not appropriate when there is genuine heterogeneity (due, most likely, to population heterogeneity or heterogeneity in selecting controls etc.) in the size of genetic effects across subpopulations, in which case the random-effects model is more preferable.

An array of statistical/graphical methods were employed to identify heterogeneity and, if present, to identify its sources. The use of different methods is necessary because different methods typically entail different assumptions.

Funnel plots were used to examine asymmetry, in which the ORs were plotted on a logarithmic scale against the inverse of their corresponding standard errors, a measure of precision (Light and Pillemer, 1984). If bias is absent, small studies will have ORs that are widely scattered but symmetric about the OR estimates provided by larger, more precise studies. In this case, the plot would resemble an inverted funnel with the tip pointing roughly towards the true log OR. If publication bias is present, the plot will be asymmetric because some negative studies are not published.

The asymmetry of the funnel plot was visually evaluated by the Galbraith plot (Galbraith, 1988), in which the standard normal deviate (SND), defined as the log OR divided by its standard error, is plotted against the estimate’s precision (i.e. 1/SE), and by the test of significance of the intercept coefficient of the linear regression model of Galbraith plot (Egger et al., 1997). An intercept coefficient that statistically significantly deviates from 0 indicates asymmetry (Egger et al., 1997). Begg’s rank correlation test was also used to test for asymmetry (Begg and Mazumdar, 1994).

The sources of heterogeneity were further evaluated by a graphical method proposed by Baujat et al. (2002). In Baujat’s plot, the X-axis represents the contribution of the study in question to the overall heterogeneity, as measured by Cochran’s Q, whereas the Y-axis represents the influence of the study on the overall risk effect, defined as the square of the difference between the risk effect estimated from the study and that of the rest of the studies, weighted by the inverse of the estimated variance of the risk effect after exclusion of the study.

In addition, the sources of heterogeneity were evaluated by an inconsistency measure proposed by Higgins and Thompson (2004). This measure, I², defined to be 100% × (Q – df)/Q, where Q is Cochran’s heterogeneity statistic and df is the degrees of freedom, describes the percentage of total variation across studies that is due to heterogeneity rather than chance. In addition, the presence of
asymmetry in funnel plots, often an indication of publication bias, a nonparametric ‘trim and fill’ method proposed by Duval and Tweedie (2000) was used to estimate the number of missing studies and to estimate an ‘adjusted’ OR and its CI. This method identifies and estimates the number of asymmetric studies in the right side of the funnel. These studies are then ‘trimmed’ from the funnel, revealing a more or less symmetric but smaller funnel from which the true center of the funnel is estimated by standard meta-analysis methods. The trimmed studies are then replaced and their missing counterparts, i.e. those studies with negative findings, are statistically imputed or ‘filled’. The imputed OR estimates and their standard errors are then combined with the whole studies to come up with an adjusted ‘true’ pooled estimate of OR and its standard errors.

Finally, given strong evidence for publication bias, I investigated the effect of publication bias on the pooled OR estimate using a method by Copas and Jackson (2004). This method, under the only and plausible assumption that the chance that a study gets published cannot get smaller as the study size, as measured by precision, gets larger, estimates the maximum bias in estimating the pooled log OR under all possible publication bias mechanisms as a function of the number of unpublished studies, \( m \). This analysis was necessary because characterization of the bias mechanism can be difficult, if not impossible, and hence a sensitivity analysis of some sort will be of help in gauging the value of \( m \) which gives a bias just large enough to cast doubt on the overall conclusion. In addition, as the number of included studies for meta-analysis can be small, the lack of significant evidence for publication bias does not necessarily mean there is no publication bias at all. Hence, again, a sensitivity analysis is helpful.

All computations were carried out in R (version 2.0.0), a language and environment for statistical computing and graphics (www.r-project.org).

Role of the funding source
The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of this report.

Results
Characteristics of included studies
Fourteen studies on GSTM1-endometriosis involved 1539 cases/patients and 1805 controls, whereas nine studies on GSTT1 involved 1021 cases/patients and 1180 controls. Detailed characteristics of each study are described in Table I for GSTM1 and Table II for GSTT1. The frequency of GSTM1 0/0 genotype in controls ranges from 5.0% in Chinese people from Taiwan to 56.0% in Koreans. With the exception of a study conducted in Taiwan (Hsieh et al., 2004), the GSTM1 null genotype frequency is about 50% in all controls of included studies. The frequency of GSTT1 0/0 genotype in controls ranges from 9.7% in French to 48.3% in Koreans. The three studies, AK03, MY04 and HL04, from which data on combined GSTM1/GSTT1 null genotype could be deduced, involved 577 cases and 778 controls (Table III).

Qualitative assessment of included studies
The first publication that reported the GST–endometriosis association was based on a relatively small study involving 42 women with endometriosis (Baranov et al., 1996). Using DNA samples taken from nondescript ‘healthy male and female donors’ as a reference group, the study investigated disease associations for six different diseases, including endometriosis, with two of the six diseases each having two different subtypes. No information was given regarding the ages of the patients or the quality assurance procedures in genotyping, and no attempt was made to control for known risk factors for endometriosis, such as smoking status. The inclusion of males in the control sample may be also questionable (Zondervan et al., 2002b). Therefore, from a methodological standpoint, this study was not a methodologically rigorous one. As a result, any finding that came out of the study should be viewed as preliminary.

Of the 14 studies on GSTM1, half of them were conducted in Europe, and the other half, in East Asia. Three studies (Baranov et al., 1996, 1999; Ivashchenko et al., 2003), including the first report on the GSTM1–endometriosis association, were conducted in Russia by the same research group. Four studies were conducted in China. Among seven studies on GSTT1, three were conducted in China.

The selection of cases or patients varies among the included studies. Some selected prevalent cases, whereas others selected incident cases. Some studies selected cases/patients with a mixture of various revised American Fertility Society (rAFS) stages, or exclusively advanced rAFS stages (rAFS III/IV), whereas others merely selected ‘ovarian endometriosis’ or ‘extragenital endometriosis’ or simply did not report the type of endometriosis. Most studies did not report the ages of the recruited cases/patients. The earlier studies tended to pay less attention to methodological details, such as the criteria of case/patient selection, staging and age.

The selection of controls also varied substantially. Some were population based (but not necessarily from the same source population from which the cases arise), whereas others were hospital-based. Both population- and hospital-based controls appeared to be selected based on convenience, and further variations exist within each category. For population-based controls, for example, it ranged from ‘donors’, apparently involving both men and women, to healthy newborn females. For hospital-based studies, it ranged from women undergoing induced abortions to women with gynecologic complaints but without endometriosis.

Despite such tremendous heterogeneities, there are several conspicuous commonalities shared by all studies. First, no information was provided in any included study regarding the assessment of HWP in controls. Deviation from the HWP could be due to chance, but more likely it signals some potential biases or problems in control selection or genotyping. It also tends to inflate artificially the error of a false-positive association (Little et al., 2002). Because the evaluation of HWP is regarded as an important criterion in the evaluation of genetic association studies (Little et al., 2002), caution should be exercised in evaluating these studies. Second, the quality measures used for genotyping or the degree of reproducibility between quality control and replicates were not specified in almost all of studies. Third, no study made any serious attempt to match controls with the cases on factors that may confound the result. For example, age is the only sociodemographic risk factor consistently reported for endometriosis (Eskenazi and Warner, 1997), yet no study attempted to match, either individually or by frequency, controls in age with the cases. In several studies, in fact, the controls, as a group, were apparently younger than the cases, leaving the possibility that some of them may have endometriosis later in their life. Lastly, all studies made little, or, in most cases, no serious attempt to control for some known risk factors for endometriosis, such as smoking and body mass index (Eskenazi and Warner, 1997). In fact, in one earlier study (Baranova et al., 1997), not included in this meta-analysis but reported the bulk of the data later in Baranov et al. (1999), controls were significantly younger than the cases and had significantly more smokers. These glaring omissions suggest that the overall quality of the GST–endometriosis association studies, especially those of earlier ones, leaves much to be desired.

GSTM1/GSTT1 and risk for endometriosis
The OR for endometriosis in women with GSTM1 0/0 genotype as compared with other genotypes ranged from 0.8 to 32.6 (Figure 1A). The pooled summary OR based on a random effect model was 1.96 (95% CI = 1.29–2.98), suggesting that women with the GSTM1 0/0 genotype have nearly twice the risk of developing endometriosis as compared with women with other genotypes. However, there was strong indication of a tremendous heterogeneity among studies (\( \chi^2_{13} = 97.25, P < 5.7 \times 10^{-5} \), and the estimated variance of the random effect, \( \tau^2 = 0.53 \)).
Table I. Characteristics of included studies of glutathione S-transferases (GST) M1 and endometriosis

<table>
<thead>
<tr>
<th>Study (first author and code)</th>
<th>Region (province, country)</th>
<th>n1/number of cases, n2/number of controls</th>
<th>rAFS stages</th>
<th>Source of control</th>
<th>Age (years) (range) of cases</th>
<th>Age (years) (range) of controls</th>
<th>Crude odds ratio (OR) estimate (95% CI)</th>
<th>Standard deviation of log OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baranov (Baranov et al., 1996) BI96</td>
<td>St. Petersburg, Russia</td>
<td>34/42, 26/67</td>
<td>NR</td>
<td>Population</td>
<td>NR</td>
<td>18–40</td>
<td>6.70 (2.69–16.71)</td>
<td>0.47</td>
</tr>
<tr>
<td>Baranov (Baranov et al., 1999) BI99</td>
<td>St. Petersburg, Russia</td>
<td>88/150, 42/99</td>
<td>NR</td>
<td>‘Donors’</td>
<td>NR</td>
<td>NR</td>
<td>1.93 (1.15–3.22)</td>
<td>0.26</td>
</tr>
<tr>
<td>Baranova (Baranova et al., 1999) BC99</td>
<td>Clermont-Ferrand, France</td>
<td>50/65, 33/72</td>
<td>I/II, 55%; III/IV, 45%</td>
<td>Women having induced abortions</td>
<td>NR</td>
<td>(but older than controls)</td>
<td>3.94 (1.88–8.26)</td>
<td>0.38</td>
</tr>
<tr>
<td>Baxter (Baxter et al., 2001) BT01</td>
<td>South east England</td>
<td>40/84, 107/219</td>
<td>NR ‘ovarian endometriosis’</td>
<td>‘Healthy, white, Caucasian female volunteers’</td>
<td>NR</td>
<td>NR</td>
<td>0.95 (0.58–1.57)</td>
<td>0.26</td>
</tr>
<tr>
<td>Hadfield (Hadfield et al., 2001) HM01</td>
<td>England</td>
<td>59/132, 27/52</td>
<td>FH+ III/IV, 100%; FH– I–IV</td>
<td>Women with a normal pelvic at hysterectomy</td>
<td>NR</td>
<td>40–50</td>
<td>0.75 (0.39–1.42)</td>
<td>0.33</td>
</tr>
<tr>
<td>Arvanitis (Arvanitis et al., 2003) AK03</td>
<td>Crete, Greek</td>
<td>161/275, 181/346</td>
<td>NR FH–</td>
<td>Hospital: fertile, premenopausal women without endometriosis</td>
<td>NR</td>
<td>27.2 ± 3.2 (21–37)</td>
<td>34.5 ± 7.4 (26–53)</td>
<td>1.29 (0.94–1.77)</td>
</tr>
<tr>
<td>Lin (Lin et al., 2003) LZ03</td>
<td>Zhejiang, China (Han nationality)</td>
<td>49/68, 12/28</td>
<td>I/II, 24%; III/IV, 76%</td>
<td>Hospital: women of reproductive age without endometriosis who underwent operation due to ectopic pregnancy</td>
<td>NR (older than controls)</td>
<td>NR</td>
<td>3.44 (1.37–8.60)</td>
<td>0.47</td>
</tr>
<tr>
<td>Peng (Peng et al., 2003) PH03</td>
<td>Guangdong, China (Han nationality)</td>
<td>50/76, 37/80</td>
<td>I/II, 47%; III/IV, 53%</td>
<td>Hospital: women receiving pelvic surgeries due to ectopic pregnancy and sterilization, but without endometriosis</td>
<td>21–45</td>
<td>23–41</td>
<td>2.23 (1.17–4.27)</td>
<td>0.33</td>
</tr>
<tr>
<td>Morizane (Morizane et al., 2004) MY04</td>
<td>Kobe, Japan</td>
<td>57/108, 89/173</td>
<td>III/IV, 100%</td>
<td>Population (female newborns)</td>
<td>NR</td>
<td>5 min</td>
<td>1.05 (0.65–1.71)</td>
<td>0.25</td>
</tr>
<tr>
<td>Hsieh (Hsieh et al., 2004) HC04</td>
<td>Taiwan, China</td>
<td>95/150, 8/159</td>
<td>III/IV, 100%</td>
<td>Hospital</td>
<td>NR</td>
<td>NR</td>
<td>32.60 (14.87–71.46)</td>
<td>0.37</td>
</tr>
<tr>
<td>Ding (Ding et al., 2004) DCU04</td>
<td>Xinjian, China (Uygur nationality)</td>
<td>21/41, 57/107</td>
<td>I/II, 44%; III/IV, 56%</td>
<td>Hospital: women with a normal pelvic during gynecological screening (through examination, ultrasound, and serum CA-125)</td>
<td>35∀10</td>
<td>33∀12</td>
<td>0.92 (0.45–1.89)</td>
<td>0.37</td>
</tr>
<tr>
<td>Ding (Ding et al., 2004) DCH04</td>
<td>Xinjian, China (Han nationality)</td>
<td>46/80, 55/105</td>
<td>I/II, 26%; III/IV, 74%</td>
<td>Hospital: women with a normal pelvic during gynecological screening (through examination, ultrasound and serum CA-125)</td>
<td>35∀11</td>
<td>33∀13</td>
<td>1.22 (0.68–2.21)</td>
<td>0.30</td>
</tr>
<tr>
<td>Hur (Hur et al., 2005) HL04</td>
<td>Seoul, South Korea</td>
<td>112/194, 145/259</td>
<td>III/IV, 100%</td>
<td>Hospital: women undergone surgeries for nonmalignant lesions</td>
<td>NR</td>
<td>NR</td>
<td>1.07 (0.74–1.56)</td>
<td>0.19</td>
</tr>
</tbody>
</table>

FH+ (−), positive (negative) family history of endometriosis; n1, n2, number of women with the null genotype in the corresponding group; NR, not reported.
Table II. Characteristics of included studies of glutathione S-transferases (GST) T1 and endometriosis

<table>
<thead>
<tr>
<th>Study (first author)</th>
<th>Region (province, country)</th>
<th>( n_1/n_2 ) number of cases, controls</th>
<th>rAFS stages</th>
<th>Source of control</th>
<th>Age (years) (range) of cases</th>
<th>Age (years) (range) of controls</th>
<th>Crude odds ratio (OR) estimate (95% CI)</th>
<th>Standard deviation of log OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baranova (Baranova et al., 1999)</td>
<td>Clermont-Ferrand, France England</td>
<td>13/65, 7/72</td>
<td>I/II, 55%; II/IV, 45%</td>
<td>Women having induced abortions</td>
<td>NR (but older than controls)</td>
<td>NR</td>
<td>2.32 (0.86–6.24)</td>
<td>0.50</td>
</tr>
<tr>
<td>Hadfield (Hadfield et al., 2001)</td>
<td>England</td>
<td>29/116, 14/50</td>
<td>FH+ III/IV, 100%; FH– I–IV</td>
<td>Women with a normal pelvic at hysterectomy</td>
<td>NR</td>
<td>40–50</td>
<td>0.86 (0.41–1.81)</td>
<td>0.38</td>
</tr>
<tr>
<td>Ivashchenko Arvanitis (Arvanitis et al., 2003)</td>
<td>Russia Crete, Greece</td>
<td>27/74/6/40 24/275/31/346</td>
<td>NR NR FH–</td>
<td>Hospital, fertile women; fertile, premenopausal women without endometriosis</td>
<td>27.2 ± 3.2 (21–37)</td>
<td>34.5 ± 7.4 (26–53)</td>
<td>3.26 (1.21–8.75)</td>
<td>0.50</td>
</tr>
<tr>
<td>Lin (Lin et al., 2003)</td>
<td>Zhejiang, China (Han nationality)</td>
<td>53/68, 9/28</td>
<td>I/II, 24%; III/IV, 76%</td>
<td>Women having reproductive age without endometriosis who underwent operation due to ectopic pregnancy</td>
<td>NR (older than controls)</td>
<td>NR</td>
<td>7.46 (2.80–19.85)</td>
<td>0.50</td>
</tr>
<tr>
<td>Morizane (Morizane et al., 2004)</td>
<td>Kobe, Japan</td>
<td>52/108, 71/173</td>
<td>III/IV, 100%</td>
<td>Population (female newborns)</td>
<td>NR</td>
<td>5 min</td>
<td>1.33 (0.82–2.16)</td>
<td>0.25</td>
</tr>
<tr>
<td>Ding (Ding et al., 2004)</td>
<td>Xinjiang, China (Uygur nationality)</td>
<td>15/41, 32/107</td>
<td>I/II, 44%; III/IV, 56%</td>
<td>Hospital; women with a normal pelvic during gynecological screening (through examination, ultrasound, and serum CA 125)</td>
<td>35 V10</td>
<td>33 V12</td>
<td>1.35 (0.63–2.89)</td>
<td>0.39</td>
</tr>
<tr>
<td>Ding (Ding et al., 2004)</td>
<td>Xinjiang, China (Han nationality)</td>
<td>59/80, 47/105</td>
<td>I/II, 26%; III/IV, 74%</td>
<td>Hospital; women with a normal pelvic during gynecological screening (through examination, ultrasound and serum CA 125)</td>
<td>35 V11</td>
<td>33 V13</td>
<td>3.47 (1.85–6.51)</td>
<td>0.32</td>
</tr>
<tr>
<td>Hur (Hur et al., 2005)</td>
<td>Seoul, South Korea</td>
<td>104/194, 125/259</td>
<td>III/IV, 100%</td>
<td>Hospital; women undergone surgeries for nonmalignant lesions</td>
<td>NR</td>
<td>NR</td>
<td>1.24 (0.85–1.80)</td>
<td>0.19</td>
</tr>
</tbody>
</table>

FH+ (–), positive (negative) family history of endometriosis; \( n_1, n_2 \), number of women with the null genotype in the corresponding group; NR, not reported.
For GSTT1, the OR ranged from 0.9 to 7.5 (Figure 1B). The random-effects model yielded a pooled OR of 1.77 (95% CI = 1.19–2.63), suggesting that the risk of developing endometriosis in women with the at-risk genotype is increased by nearly 80% as compared with women with other GSTT1 genotypes. Again, there was sign of substantial heterogeneity ($\chi^2 = 25.98$, $P = 0.0011$, and the estimated variance of random effect $\tau^2 = 0.24$).

For the combined GSTM1/GSTT1 null genotype, the three studies, AK03, MY04 and HL04, gave an OR estimate of 0.86 (95% CI = 0.39–1.88), 1.16 (95% CI = 0.67–2.00) and 1.45 (95% CI = 0.98–2.15), respectively. The pooled OR based on the random-effects model is 1.26 (95% CI = 0.94–1.70), which is consistent with the conclusion reached by BC99. There was no indication for heterogeneity ($\tau^2 = 0.0$, $P = 0.472$). Because all OR CI include 1, and there was no sign of heterogeneity, no further analysis was performed.

Tables IV and V summarize the ORs for endometriosis in different subgroups for GSTM1 and GSTT1 studies, respectively. Several observations are in order. First, for both GSTM1 and GSTT1, there was no difference in OR estimate between the continents where the study was conducted. Second, for GSTM1, the OR estimate from pooled studies published in English was higher than that from pooled studies published in other languages, but the two estimates overlap in CIs. For GSTT1, the magnitudes of the OR estimates were reversed. Lastly, although the OR estimate for GSTM1 from small studies, defined to be the one with overall sample size less than 200, was almost identical to that from large studies (overall sample size >200), the CI of latter OR includes 1, which indicates no significant deviation from the null effect. For GSTT1, the OR from small studies was higher than that from large studies. Similarly, the CI of the pooled OR from larger GSTT1 studies also included 1.

### Assessment of heterogeneity

In view of clear evidence of heterogeneity among included studies, I used Baujat’s plot to identify studies that account for most of the heterogeneity without having to explore all possible sources of heterogeneity by subgroup analyses. Figure 2 (upper panel) shows the Baujat’s plot for both GSTM1 and GSTT1 studies. For GSTM1, it can be seen that the study HC04 stands out prominently as being the study that contributes most to the overall heterogeneity (upper left panel of Figure 2). For GSTT1, it is LZ03 that was the most influential. The $\hat{F}$ values were 86.6 and 69.5% for GSTM1 and GSTT1, respectively, suggesting that most of the variability across studies is due to heterogeneity rather than chance. After removing HC04, the $\hat{F}$ only reduced very moderately to 67.6% and the heterogeneity test remains highly significant ($P = 0.00022$). After further removing GSTM1 studies BI96, BC99 and LZ03, the $\hat{F}$ finally dipped below 39–23.3% and the heterogeneity test was no longer significant at 0.1 level (Figure 2, lower left panel).

After removing LZ03 from GSTT1 studies, the $\hat{F}$ increased to 87%, and the Cochran’s heterogeneity test was still significant ($P = 1.9 \times 10^{-14}$). After further removing DC04H, the $\hat{F}$ reduced to 13.6%, and the heterogeneity test was no longer significant ($P = 0.326$).

That the study HC04 contributes apparently a disproportional share to the overall heterogeneity in GSTM1, studies can also be viewed from a different angle. Figure 3 is the plot of log OR versus genotype GSTM1 0/0 frequency in the cases. With the only exception of HC04, there was a near-perfect linear relationship between the two variables ($\hat{R}^2_{ad} = 0.93$, $F_{1,11} = 160.8$, $P = 6.6 \times 10^{-8}$). This suggests that HC04 was evidently an outlier and somehow different from the rest of the studies.
A meta-analysis of GSTM1/GSTT1 and endometriosis association

Evaluation of publication bias

Figure 4 is the funnel plot for GSTM1 and GSTT1 studies. As can be seen from the figure, both GSTM1 studies and GSTT1 studies showed some signs of asymmetry, especially in GSTM1 studies. Begg’s rank test of asymmetry yielded rank correlation coefficient of 0.52 ($P = 0.0098$) for GSTM1 studies and of 0.39 ($P = 0.180$) for GSTT1 studies, respectively. Egger’s test of bias yielded a $P$ value of 0.042 for GSTM1 studies and a $P$ value of 0.114 for GSTT1 studies. Both tests suggest that although there was no statistically significant evidence for bias or asymmetry for GSTT1 studies, there is strong evidence for bias for GSTM1 studies. However, given the rather small number of GSTT1 studies, the lack of statistical evidence may well be attributed to the lack of statistical power. The moderate rank correlation coefficient appeared to suggest that such bias may exist.

Figure 1. Individual and pooled odds ratio estimates and their 95% confidence intervals (CI) for glutathione S-transferases (GST) M1 (A) and GSTT1 (B). The size of the square is proportional to the percent weight of each study in the random effect meta-analysis. Horizontal line represents the 95% CI. The summary pooled OR and its 95% CI are indicated by the shaded diamond.
Meta-analysis of the effect of the glutathione S-transferases (GST) M1 null genotype on risk of endometriosis, according to potential sources of heterogeneity

<table>
<thead>
<tr>
<th>Studies (patients/controls</th>
<th>Pooled OR (95% CI)</th>
<th>(\tau^2) Estimate</th>
<th>(P) value for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies included</td>
<td>14 (1539/1805)</td>
<td>1.96 (1.29–2.98)</td>
<td>0.53</td>
</tr>
<tr>
<td>First versus subsequent studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First study</td>
<td>1 (42/67)</td>
<td>6.7 (2.69–12.71)</td>
<td>—</td>
</tr>
<tr>
<td>First study excluded</td>
<td>13 (1497/1738)</td>
<td>1.80 (1.19–2.74)</td>
<td>0.45</td>
</tr>
<tr>
<td>Four studies excluded</td>
<td>10 (1214/1479)</td>
<td>1.22 (1.02–1.47)</td>
<td>0.02</td>
</tr>
<tr>
<td>Continent of origin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>7 (822/894)</td>
<td>1.74 (1.10–2.75)</td>
<td>0.28</td>
</tr>
<tr>
<td>Asia</td>
<td>7 (717/911)</td>
<td>2.20 (1.02–4.78)</td>
<td>0.98</td>
</tr>
<tr>
<td>Publication language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>9 (1118/1375)</td>
<td>2.30 (1.23–4.29)</td>
<td>0.80</td>
</tr>
<tr>
<td>Non-English</td>
<td>5 (421/430)</td>
<td>1.56 (1.16–2.12)</td>
<td>0.01</td>
</tr>
<tr>
<td>Size of study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>8 (579/550)</td>
<td>1.95 (1.18–3.22)</td>
<td>0.38</td>
</tr>
<tr>
<td>Large</td>
<td>5 (961/1255)</td>
<td>1.98 (0.99–3.98)</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Table V. Meta-analysis of the effect of the glutathione S-transferases (GST) T1 null genotype on risk of endometriosis, according to potential sources of heterogeneity

<table>
<thead>
<tr>
<th>Studies (patients/controls</th>
<th>Pooled OR (95% CI)</th>
<th>(\tau^2) Estimate</th>
<th>(P) value for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies included</td>
<td>9 (1021/1180)</td>
<td>1.77 (1.19–2.63)</td>
<td>0.24</td>
</tr>
<tr>
<td>First versus subsequent studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First study</td>
<td>1 (65/72)</td>
<td>2.32 (0.86–6.24)</td>
<td>—</td>
</tr>
<tr>
<td>First study excluded</td>
<td>8 (956/1108)</td>
<td>1.73 (1.13–2.65)</td>
<td>0.26</td>
</tr>
<tr>
<td>Studies LZ03 and DC04H excluded</td>
<td>7 (873/1047)</td>
<td>1.29 (1.01–1.65)</td>
<td>0.02</td>
</tr>
<tr>
<td>Continent of origin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>4 (530/508)</td>
<td>1.44 (0.79–2.62)</td>
<td>0.21</td>
</tr>
<tr>
<td>Asia</td>
<td>5 (491/672)</td>
<td>2.06 (1.18–3.60)</td>
<td>0.30</td>
</tr>
<tr>
<td>Publication language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>5 (761/856)</td>
<td>1.43 (0.87–2.34)</td>
<td>0.22</td>
</tr>
<tr>
<td>Non-English</td>
<td>4 (260/324)</td>
<td>2.46 (1.55–3.89)</td>
<td>0.05</td>
</tr>
<tr>
<td>Study size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>6 (444/402)</td>
<td>2.39 (1.30–4.40)</td>
<td>0.39</td>
</tr>
<tr>
<td>Large</td>
<td>3 (577/778)</td>
<td>1.20 (0.92–1.56)</td>
<td>0</td>
</tr>
</tbody>
</table>

**Assessment of the effect of publication bias on risk estimate**

I further used the ‘trim and fill’ method to evaluate publication bias and its impact on pooled estimate of OR. For GSTM1 studies, the estimated number of missing studies is four (SE = 3.16), and the adjusted pooled OR estimate is 1.23 (95% CI = 0.95–1.59) (Figure 4A). It is interesting to note that the trim and fill method trimmed the four studies (B196, BC99, LZ03 and HC04) that are identical to those identified by the successive use of Baujat’s plot (see above). It is also of note that the 95% CI of adjusted pooled log OR now included 0, indicating that the effect of the putative risk genotype was not significantly different from the null. For GSTT1 studies, the estimated number of missing studies was 0 (SE = 1.41).

Although the pooled OR estimate was still greater than 1 for GSTT1 after removing two studies that contribute disproportionately to the overall heterogeneity, the funnel plot (Figure 4B) appeared to indicate that publication bias may have influenced these findings. In fact, higher log OR estimates tended to be those with higher standard errors (the correlation coefficient was 0.65, \(P = 0.16\), or 0.57, \(P = 0.14\) if the two studies were removed). Figure 5A shows the maximum bias in estimating the pooled log OR in the presence and the absence of heterogeneity for GSTM1, based on the method of Copas and Jackson (2004). For GSTM1 studies, the pooled log OR estimate was 0.67 (95% CI = 0.25–1.09) if all studies were included, or 0.20 (95% CI = 0.02–0.39) when the four studies were excluded (Table IV). In the former case, there was a clear indication for heterogeneity (\(\tau^2 = 0.53\), \(P < 0.00001\)), whereas in the latter heterogeneity appeared to be absent or negligible (\(\tau^2 = 0.02\), \(P = 0.23\)) (Table IV). For any given number of unpublished studies, Figure 5A shows the estimated largest possible bias that could result from publication bias. In the presence of heterogeneity, it can be seen that when the publication probability is 14/(14+4) = 78%, where 4 is the number of unpublished studies, the maximum bias can be as high as 0.31, which reduces the log OR from 0.67 to 0.36 and could effectively render the CI to include 0, i.e. the null effect. In the absence of heterogeneity, a bias of 0.04 would be enough to wipe out the statistical significance of the estimated log OR, which can happen when there is only one missing study, or 8/(8+1) = 89% publication probability.

For GSTT1, the pooled log OR estimate was 0.57 (95% CI = 0.17–0.97) if all studies were included, or 0.25 (95% CI = 0.01–0.50) when the two studies were excluded (Table V). In the former case, there was a clear indication for heterogeneity (\(\tau^2 = 0.24\), \(P = 0.0011\)), whereas in the latter heterogeneity appeared to be absent (\(\tau^2 = 0.02\), \(P = 0.33\)) (Table V). The log OR could easily lose its statistical significance when the publication probability is 9/(9+4) = 69% in the presence of
heterogeneity or is $7/(7+1) = 88\%$ in the absence of heterogeneity (Figure 5B). These results suggest that, with a fairly realistic publication probability ranging from 69 to 88\%, the pooled OR from the eight published studies is very vulnerable to publication bias and should thus be viewed as volatile.

**Discussion**

In this meta-analysis on GSTM1/T1–endometriosis association studies, I found substantial heterogeneities among studies, which are likely to result from publication bias. The publication bias appears to be prominent and significant for GSTM1 studies, but is less so for GSTT1 studies. After correction for this bias, I found no evidence that women with GSTM1 null genotype have increased risk of developing endometriosis as compared with women with other genotypes. For GSTT1, the risk associated with the null genotype is 29\% higher than other genotypes (the pooled OR = 1.29, 95\% CI = 1.01–1.65, after removing two studies contributing disproportionately heterogeneity). However, even this estimate should be viewed with a pinch of salt, because the estimate could easily lose its statistical significance if there is a realistic 69–80\% publication probability. It should be noted that asymmetry of funnel plots can be caused by biases other than publication bias, such as true heterogeneity (effect size differs according to study size), data irregularities (poor methodological design of small studies, inadequate analysis and fraud) and chance (Sterne et al., 2001). In the absence of clear evidence for true heterogeneity and data irregularity and in view of a well-known ‘tendency on the part of investigators, reviewers and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings’ (Dickersin, 1990), publication bias becomes the primary suspect for the asymmetry. This suspicion appears to be reinforced...
and somewhat justified by analyses using a range of statistical tests that are designed to test for the existence of this bias, such as the trim and fill method (Figure 4A and B) and the Copas and Jackson method (Figure 5A and B).

The findings of this report appear to be consistent with a well-recognized perennial problem associated with genetic association studies of complex diseases (Risch and Botstein, 1996; Cardon and Bell, 2001): lack of consistent replication. In an extensive review of 166 disease–gene associations, it is reported that only six, or less than 4% of them had been consistently replicated (Hirschhorn et al., 2002). Other investigators reported slightly better percentage of 16% (Ioannidis et al., 2003). In fact, quantitative analyses of published association studies revealed that significant heterogeneity or lack of consistency is rampant, and that a small sample size of the first publication and a large number of studies are two independent predictors of reaching discrepancies (Ioannidis et al., 2001). In addition, the magnitude of the genetic effect differs significantly between large and small studies (Ioannidis et al., 2003), as evidenced also in this analysis.

There are many possible explanations for the lack of consistency that plagues genetic association studies: genuine population heterogeneity, difference in selection of cases and/or controls, false negative due to lack of statistical power, or false positive due to population stratification. For example, prevalent or incident cases or the selection of patients with different pathological stages (not necessarily equivalent to the rAFS stages) could have different genetic components. Flawed study design, faulty statistical analysis, genotyping errors or selective reporting also exacerbate the lack of consistency and often are the causes for the first positive reports.

Figure 3. Log OR versus the frequency of glutathione S-transferases (GST) M1 null genotype in cases/patients. The dashed line represents the linear regression after removing the study HC04.

Another study in point is Hsieh et al. (2004), which reported an OR of 32.6 for GSTM1 null genotype, higher than the risk of ovarian cancer with BRCA2 mutations (Rafnar et al., 2004). Yet extraordinary claims usually warrant careful examinations. Like that of BI96, no information was given regarding the ages of the patients or the quality assurance procedures in genotyping, and no attempt was made to control for known risk factors for endometriosis, such as smoking status. The selection of controls was also based on convenience: they were women admitted to the hospital with nonspecific disorders. What sets this study apart from others is its inordinately low frequency of GSTM1 null genotype (5%), far lower than reported 40–50% in different populations (Seidegard et al., 1990), than the reported 58% in the Chinese (Board, 1981) and than those in Chinese, Japanese and Korean reported in (Lin et al., 2003; Peng et al., 2003; Ding et al., 2004; Morizane et al., 2004; Hur et al., 2005). In fact, the frequency is significantly lower than any studies included in this analysis. As Figure 3 shows, this population appears to be fundamentally different than other populations. Naturally, there is question as to why this control group has a genotype frequency that is vastly different from others, and whether the selected controls were appropriate for the association study. A test of HWP in the controls may have been useful in the identification of any anomaly in this group.

Is it possible that GSTM1 or GSTT1 by itself does not exhibit appreciable marginal effect on the risk of endometriosis but interacts with other genes to confer, jointly, increased risk of endometriosis? This is possible. In fact, Hadfield et al. (2001) suggested that the combination of GSTM1 null genotype and the CYP1A1 polymorphism is associated with a small increased risk. Arvanitis et al. (Arvanitis et al.,
Figure 4. Funnel plot for glutathione S-transferases (GST) M1 (A) and GSTT1 (B) studies. Each alphabet-numeric combination represents a study. The × represents missing studies estimated and imputed by the trim and fill method. The unshaded diamond indicates the magnitude of the pooled log OR estimate based on the original data, and the shaded diamond indicates the magnitude of the adjusted pooled log OR estimate after the trim and fill procedure. The horizontal line passing the diamond represent the 95% CI for the estimate. The vertical dotted line represents a null effect.
also reported a similar finding. However, there is concern that the finding of Hadfield et al. (2001) may be a result of a post hoc analysis of data, and that the finding was not quite self-consistent (the difference between male controls and familial endometriosis cases is significant, but is not for other groups of cases and controls). This may stem from the fact that the sample sizes in (Hadfield et al., 2001) are not sufficient to permit a thorough analysis of gene–gene interactions. The finding of Arvanitis et al. (2003) shares the same problem and might also be spurious because the finding was concluded after computing over one hundred 95% CIs (Tables II and III of Arvanitis et al., 2003). This post hoc analysis, in conjunction with multiple testing, is akin to the subgroup analysis in clinical trials (Freemantle et al., 1999).

The findings are also consistent with a recent reappraisal of the link between dioxin and endometriosis, which maintains that, because of major study design flaws and faulty statistical analysis, the first report of the dioxin–endometriosis link is questionable at best (Guo, 2004). Some recent studies on the link appear to lend further support for this view (Fierens et al., 2003; De Felip et al., 2004; Lim et al., 2004).

Figure 5. The effect of publication bias on the pooled log OR: maximum bias as a function of the number of missing studies. H, presence of heterogeneity; N, no heterogeneity.
Recent laboratory findings also cast doubt on the dioxin–endometriosis link. The effects of dioxin have been shown to be mediated through high-affinity binding to the arylhydrocarbon receptor (AhR), which forms an activated heterodimer complex with the structurally related AhR nuclear translocator protein (ARNT) upon binding to dioxin (Bofinger et al., 2001). The two most studied members of the AhR gene battery are cytochrome P450 enzymes 1A1 (CYP1A1) and 1B1 (CYP1B1), which in conjunction with AhR and ARNT, form the phase I biotransformation process in dioxin detoxification, whereas GSTs play their roles in phase II. It has been recently reported that there is no significant endometriosis-related differences in CYP1A1, CYP1B1, AhR and ARNT expression in cultured human endometrial explants exposed to dioxin cultures (Bofinger et al., 2001; Pitt et al., 2001). The expression of AhR and dioxin-related genes in the endometrium do not differ in women with or without endometriosis (Igarashi et al., 1999).

It is regrettable that although all included studies attempted to measure the risk of endometriosis for GST polymorphisms, no attempt has ever been made to measure the blood concentration of any dioxin or dioxin exposure levels in cases and controls, in conjunction with the genotyping. Though technically and logistically demanding, these measurements, if made, would have been more convincing if there is a belief that the role of GSTs in endometriosis pathogenesis is the detoxification of dioxins or other environmental pollutants. Moreover, no study has ever attempted to resolve the conundrum: should GST null genotype increase the risk of endometriosis which is associated with subfertility, which, in turn, apparently reduces fitness, why do GSTM1/ GSTT1 polymorphisms still exist? Although one could argue that this may be attributable to fairly recent dramatic changes in environment or reproductive behaviour, clearly the rising prevalence of endometriosis outpaces remarkably the change in any GST allele frequency.

Genetic association studies attempt to identify genes or markers that are closely linked with the genes that confer increased disease risk through identification of differential genotypic frequencies in cases and controls. Their strengths and pitfalls have been well elucidated by many authors (Kidd, 1993; Cardon and Bell, 2001). Although the growing interest in this approach to nonmendelian complex diseases such as endometriosis is driven by successes in mapping mendelian diseases and by the increasingly faster and cheaper methods for genotyping more and more genetic markers, its technical simplicity, as well as its cookie-cutter conceptual simplicity may also play a role. The unawareness or ignorance of its explicit and implicit assumptions or lack of carefully thought-out study design may be an important reason as to why nonreplication is so rampant in genetic association studies.

Because there are over 36 000 protein-coding genes (Venter et al., 2001), 1.8 million (and counting) single nucleotide polymorphisms (SNPs) (http://snp.cshl.org/), and 2.91 billion base-pairs in the human genome (at which genetic variation could occur and have influence on phenotypic variation), the prior probability that a particular gene polymorphism may increase the risk of endometriosis is minuscule. This is especially true when even astronomically more combinations of genotypes and/or genetic models are considered, let alone in combination with environmental effects. Therefore, a small P value may not be that impressive when the prior probability is incorporated. Even if there is a prior belief that a particular gene polymorphism, like GSTM1 null genotype, may be biologically plausibly involved in the pathogenesis, it is still a speculation because biological plausibility ‘is too often not based on logic or data, but only on prior beliefs’ (Rothman and Greenland, 1998). Furthermore, the GST-endometriosis link also ignores the possibility of genetic redundancy (Tautz, 1992; Nowak et al., 1997) that may exist in the detoxification pathways.

As there are several known risk factors, such as age and smoking, for endometriosis (Eskenazi and Warner, 1997), failure to control for these risk factors may possibly yield spurious results, a fact known as Simpson’s paradox. In addition, because a positive finding in an association study only means association, not necessarily a causal relationship, it is possible that a gene demonstrated to be associated with endometriosis may well be responsible for a risk factor which, in turn, increases the risk of endometriosis. For example, younger age at menarche has been identified as a risk factor for endometriosis (Moen and Schei, 1997). But ages at menarche in sisters have been shown to be highly correlated (Salces et al., 2001). Body mass index, length of menstrual cycles and the amount of menses are also risk factors for endometriosis but may, too, be familialy correlated (Salces et al., 2001, 2003). Without control for these risk factors, how can we be sure that the polymorphism of interest is an endometriosis susceptibility gene?

Failure to control for risk factors, or other genetic effects, in genetic association studies can also artificially distort the effect size of the gene polymorphism of interest when in reality it may only have a small effect on disease risk. A well-known tutorial example is college admission based on either academic achievements (‘brainy’) or athletic talents (‘sporty’). If in the general population being brainy and being sporty are assumed to be independent, then focusing on college students, who are either brainy or sporty, would lead to us to conclude, erroneously, that being brainy means less likely to be sporty (or vice versa), because being either is sufficient to be admitted to the college. This phenomenon, known as Berkson’s fallacy, could also happen when using hospital–based controls for association studies (Berkson, 1946).

As with other meta-analyses, the interpretation of this meta-analysis should be made within the context of its limitations. First, this analysis has assumed tacitly that all included studies measured the same thing. However, differences in study design and execution, in genotyping accuracy, in degree of population homogeneity, in composition of cases (percentage of stages III/IV patients), and in control selection could mean that different studies are measuring slightly different things. These differences may well attribute to the heterogeneity that has been observed. Second, this analysis, especially for GSTT1, was limited by the rather small number of studies, which could have yielded a lack of power for conducting analyses of publication bias. The small sample size also limited the ability to conduct a more meaningful subgroup analyses.

With the benefit of hindsight, of course, it is easy to be critical, because many studies were conducted without the benefit of more current molecular genetic technologies, using the cookie-cutter approach that is somewhat deceptively simple. However, the findings of these studies and their validities should be put into perspective from the vantage point of current criteria of sound genetic association studies.

Several recommendations on the future association studies of endometriosis can be made from this meta-analysis. First, a carefully thought-out study design is crucial for a successful association study. Second, it is necessary to make effort to control for known risk factors, preferably in the design stage. Third, larger sample sizes are certainly desirable. In fact, the possibility of gene–gene and gene–environment interactions and the need for control for confounding factors of environmental origins would require much larger sample sizes than studies without interaction/confounding. Fourth, special attention needs to be paid to select proper controls (see Holt and Weiss, 2000) and for controlling for population heterogeneity, such as using genomic control (Devlin and Roeder, 1999). Lastly, care should be exercised in genotyping and in checking for abnormality, such as the deviation from HWP.

In summary, this meta-analysis has found substantial heterogeneities among GSTM1/GSTT1 association studies and evidence for publication bias. The publication bias appears to be prominent and
significant for GSTM1 studies, but is less so for GSTT1 studies. After correction for this bias, there is little evidence that women with GSTM1 null genotype have increased risk of developing endometriosis as compared with women with other genotypes. For GSTT1, the risk associated with the null genotype is 29% higher than other genotypes, and this estimate is highly susceptible to publication bias and should therefore be taken with a pinch of salt.

Acknowledgements

The author thanks Dr. Elena Semina for extracting relevant data from two Russian articles. He also thanks Dr. Yiqing Song and Dr. Krina Zondervan for their helpful and constructive comments on an earlier version of the manuscript. The funding from the Children’s Hospital of Wisconsin Foundation is gratefully acknowledged.

References

Arvanitis DA, Goumenou AG, Matalliotakis IM, Koumantakis EE and Spanidios DA (2001) Low-penetrance genes are associated with increased susceptibility to endometriosis. Fertil Steril 76,1202–1206.


Baxter SW, Thomas EJ and Campbell IG (2001) GSTM1 null polymorphism and risk associated with the null genotype is 29% higher than other genotypes. For GSTT1, the risk of developing endometriosis. Hum Reprod Update 6,37–44.


Zondervan KT, Cardon LR and Kennedy SH (2002b) What makes a good meta-analysis of GSTM1/GSTT1 and endometriosis association

Submitted on March 15, 2005; accepted on July 6, 2005