

The International Journal of Biostatistics

Volume 8, Issue 2

2012

Article 5

CAUSAL INFERENCE IN HEALTH RESEARCH

Meta-Analysis of Observational Studies with Unmeasured Confounders

Lawrence C. McCandless, *Simon Fraser University*

Recommended Citation:

McCandless, Lawrence C. (2012) "Meta-Analysis of Observational Studies with Unmeasured Confounders," *The International Journal of Biostatistics*: Vol. 8: Iss. 2, Article 5.

DOI: 10.2202/1557-4679.1350

Meta-Analysis of Observational Studies with Unmeasured Confounders

Lawrence C. McCandless

Abstract

Meta-analysis of observational studies is an exciting new area of innovation in statistical science. Unlike randomized controlled trials, which are the gold standard for proving causation, observational studies are prone to biases including confounding. In this article, we describe a novel Bayesian procedure to control for a confounder that is missing across the sequence of studies in a meta-analysis. We motivate the discussion with the example of a meta-analysis of cohort, case-control and cross-sectional studies examining the relationship between oral contraceptives and endometriosis. An important unmeasured confounder is dysmenorrhoea, which is an indication for oral contraceptive use. To adjust for unmeasured confounding, we combine random effects models with probabilistic sensitivity analysis techniques. Information about the unmeasured confounder is incorporated into the analysis via prior distributions, and we use MCMC to sample from posterior.

KEYWORDS: causal inference, bias, sensitivity analysis, Bayesian statistics

Author Notes: This work was supported by an a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. The author appreciates the helpful comments of two anonymous referees.

1. An Example of Meta-Analysis with Unmeasured Confounders

1.1 Oral Contraceptives and Risk of Endometriosis

Endometriosis is a gynecological medical condition in which endometrial tissue grows outside of the uterine cavity, typically on the ovaries. It is a leading cause of gynecologic hospitalization, with an incidence rate of roughly 250 cases per 100,000 person years in North America (Missmer, Hankinson, Spiegelman et al. 2004). Endometriosis has a long period of disease progression, and it is thought to occur in roughly 5% to 10% of women at some point in their lifetime. It is associated with pain, scarring, adhesions and infertility.

The causes of endometriosis are not well known. One important question is understanding the role of oral contraceptive use. Endometriosis is related to menstruation and the total number of ovulations in a woman's lifetime. Because oral contraceptives suppress ovulation, it has been hypothesized that they may interfere with the development of endometrial tissue. However the epidemiological evidence is mostly inconclusive. Some authors argue that oral contraceptives reduce the risk of endometriosis, whereas others argue that they may accelerate disease development and progression (see Vercellini et al. (2011) for review).

Vercellini et al. (2011) conducted a systematic review and meta-analysis comparing the incidence of endometriosis in former users of oral contraceptives versus women who have never used oral contraceptives. The meta-analysis included 11 studies (4 cross-sectional, 2 case-control and 5 cohort studies) published between 1970 and 2010. In Figure 1, we reproduce a forest plot of their results that depicts relative risk estimates for each of the studies. The authors conducted a random effects meta-analysis using the method of DerSimonian-Laird, and they obtained a pooled relative risk estimate equal to 1.21, with 95% confidence interval (0.94, 1.56). The results indicates a modest (albeit not statistically significant) increased risk of endometriosis among former users of oral contraceptives. We refer the reader to Vercellini et al. (2011) for information about each of the studies including the study population, country, year, and quality of information.

The relative risks in Figure 1 are heterogenous. The Q statistic test for heterogeneity is equal to 51.9 on 10 degrees of freedom ($p < 0.001$), which leads us to reject the null hypothesis that the true effect of oral contraceptives on endometriosis does not differ between studies. The I^2 statistic, which Higgins and Thompson (2002) define as the proportion of the variation in the point estimates that reflects genuine differences in effect size, is equal to $(Q - 10)/Q = (51.9 - 10/51.9) = 81\%$. There are many possible sources of heterogeneity, including differences between study populations, differences in methods of exposure and outcome assessment, and differences in analysis techniques and study designs. On the righthand side of Figure 1, we give the percentage weighting of each study for calculating the pooled estimate. The percentages are equal to the DerSimonian-Laird weights, which have been normalized to add up to 100% (Sutton and Higgins, 2008).

1.2 True Causation or Confounded Association?

Table 1 describes the effect measures, risk adjustment techniques and covariates that were available in each of the 11 studies. Because the incidence of endometriosis is low, the authors followed convention in meta-analysis of observational studies and they assumed that odds ratio and hazard ratio estimates can be used to approximate the relative risk. Thus all effect measures in Figure 1 are reported as relative risks. See Etminan et al. (2009) and Toh and Hernández-Díaz (2007) for other examples.

Figure 1 suggests that there may be a harmful association between oral contraceptive use and endometriosis. However, Vercellini et al. (2011) disputed this conclusion. They hypothesized that unmeasured confounding could explain the association. One important confounder is *dysmenorrhea*, which means pain during menstruation. Dysmenorrhea is a known risk factor for endometriosis. It was not measured in 10 of the 11 studies in the meta-analysis. In the study of Hemmings et al. (see Table 1), the authors measured pain during intercourse and menstruation, however this variable was not controlled for in the logistic regression analysis.

Vercellini et al. (2011) argue that oral contraceptives are prescribed to reduce menstrual pain. Women who have used oral contraceptives in the past are more likely to have dysmenorrhea compared to non-users. Because dysmenorrhea is a risk factor for endometriosis, it means that oral contraceptive users will have elevated risk of endometriosis compared to non-users. Failure to measure and adjust for dysmenorrhea will induce unmeasured confounding

that biases the relative risk estimates in Figure 1 away from zero. Vercellini et al. (2011) called this *confounding by indication*, because dysmenorrhea is a clinical indication for oral contraceptive use.

When thinking about the epidemiology of endometriosis, it is important to distinguish between two types of dysmenorrhea. *Primary* dysmenorrhea is pain during menstruation in the absence of underlying disease or pathology. It is caused by elevated endometrial prostoglandins and may occur in up to 50% of reproductive aged women (Proctor and Farquhar 2007). In contrast, *secondary* dysmenorrhea occurs when symptoms are attributable to an underlying disease, such as endometriosis. In this article, the unmeasured confounder is primary dysmenorrhea.

Although Vercellini et al. (2011) acknowledge the possibility of unmeasured confounding in Figure 1, they did not provide a quantitative assessment of bias. This makes it difficult to judge the impact of a missing confounder on the results. A further challenge is that magnitude of bias is likely to vary from one study to the next. For example, the strength of the association between oral contraceptive use and dysmenorrhea will vary between populations depending on physician prescribing preferences. Sutton and Higgins (2008) call this *heterogeneity due to bias*. Heterogeneity due to bias complicates the process of exploring sensitivity to realistic departures from no unmeasured confounding.

Briefly, we emphasize that there are several additional biases at play in Figure 1. For example, measurement error and selection bias are common in epidemiologic studies of endometriosis due to the difficulties of case ascertainment. An additional issue discussed by Vercellini et al. (2011) is publication bias. In this article, we focus on exploring sensitivity to unmeasured confounding, and we defer discussion of other biases to Section 5.

Figure 1: Relative risk estimates, with 95% confidence intervals for the association between oral contraceptives and endometriosis in a meta-analysis of 11 observational studies. The solid diamond indicates the pooled estimate with 95% confidence interval, calculated using the method of DerSimonian-Laird.

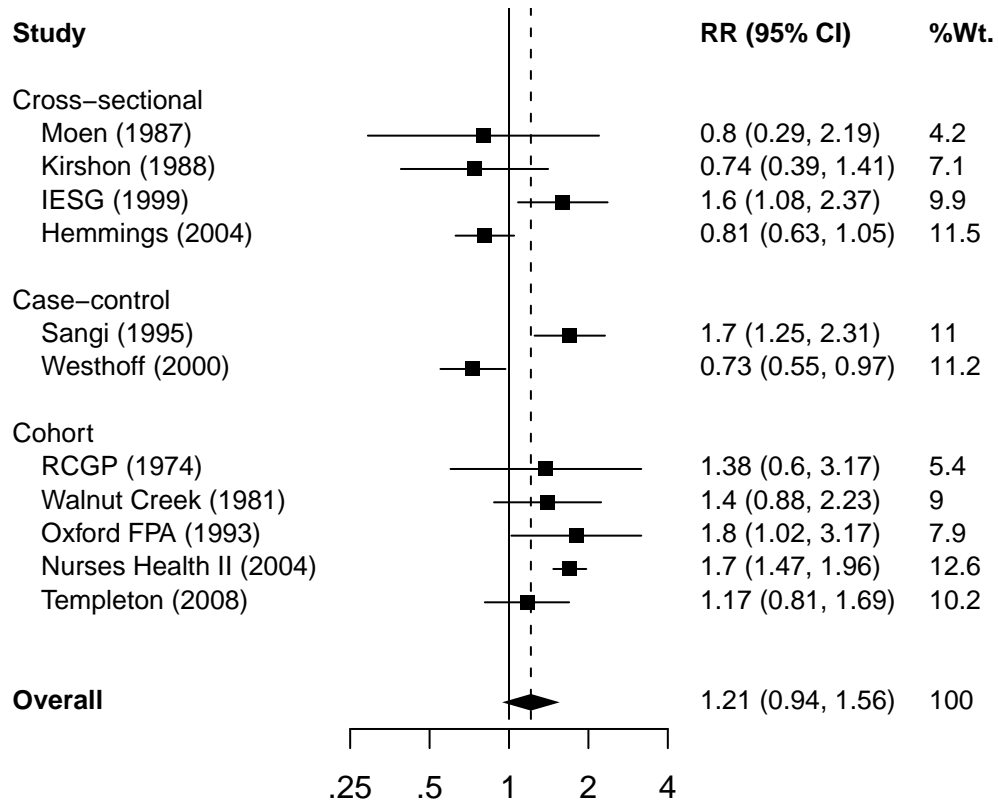


Table 1: Effect measure and risk adjustment technique used in 11 studies from the meta-analysis of oral contraceptives and endometriosis.

| Source | Effect measure | Risk adjustment | Available covariates |
|-----------------------------|-------------------------|--------------------------|--|
| Cross-sectional | | | |
| Moen et al. (1987) | Odds ratio [†] | None | Age*, parity*, age at menarch*, age at sterilization*, age at first pregnancy*, parity*, # abortions*, IUD use*, menstrual cycle characteristics*, family history of endometriosis*. |
| Kirshon & Poindexter (1988) | Odds ratio [†] | None | Age*, parity*, miscarriage*, IUD use*. |
| IESG (1999) | Odds ratio | Logistic regression | Age, parity, education, abortion*, menstrual cycle length*. |
| Hemmings et al. (2004) | Odds ratio | Logistic regression | Age*, parity*, BMI, education, ethnicity*, menstrual cycle characteristics, IUD*, age at menarch*, Age at first pregnancy, pain during intercourse or menstruation*, smoking*, alcohol*, # abortions*. |
| Case-control | | | |
| Parazzini et al. (1994) | Odds ratio | Mantel-Haenszel | Age, parity*, education, # miscarriages*, # abortions*, menstrual irregularity*, age at menarche*, marital status*, age at 1 st birth*, smoking*, BMI*. |
| Westoff et al. (2000) | Odds ratio | Logistic regression | age, parity*, race*, usual source of care*, hospital. |
| Cohort | | | |
| RCGP (1974) | Risk Ratio | Indirect Standardization | Age*, parity*, social class*, smoking*, medical history, recently pregnant, tubal ligation, hysterectomy, contraception, marital status. |
| Walnut Creek (1981) | Risk ratio | Indirect Standardization | Age, parity, education, marital status, contraception. |
| Oxford FPA (1993) | Risk ratio | Indirect Standardization | Age, parity, social class*, smoking*, obesity*. |
| Nurses Health II (2004) | Hazard ratio | Cox regression | Age, parity*, race*, calendar time, age at menarche*, menstrual cycle length*, duration of lactation*, BMI, alcohol*, smoking*, birthweight*, breastfed as infant?*, multiple gestation?*. In previous 2 years: physical exam*, pap smear*, pelvic exam*, or breast exam*. |
| Templeman et al. (2008) | Odds ratio | Logistic regression | Age, parity, race, family history, menstrual irregularity, difficulty becoming pregnant*, fertility drugs*, breast fed*, tubal ligation*, menopause*, hormone use*, BMI*. |

* Not included in risk adjustment when computing the effect measure.

† Effect measure was derived by Vercillini et al. (2011).

1.3 Review: Meta-Analysis of Observational Studies with Unmeasured Confounders

Statistical methods for unmeasured confounding in meta-analysis are scarce. The approach of The Fibrinogen Studies Collaboration (2009) incorporates studies with complete information on all potential confounders in order to obtain both full and partially adjusted exposure effect estimates. The authors use a bivariate normal model to capture the relationship between the two different types of estimators. This allows one to estimate the magnitude of bias from unmeasured confounding. A different approach due to Greenland (2003, 2005) uses loglinear models for the joint distribution of the exposure, outcome and a dichotomous unmeasured confounder. Both analytic approaches require individual-level data. This is a potentially serious limitation because in meta-analysis we often only have access to adjusted relative risk estimates from each of the studies. Other Bayesian meta-analysis techniques that model bias include the papers of Welton et al. (2009), Turner et al. (2009), and Dominici et al. (1999). Wolpert and Mengerson (2004) describe a framework for modelling biases, chiefly measurement error, in cohort and case-control studies. For an overview of Bayesian meta-analysis techniques, see the review papers of Smith, Spiegelhalter and Thomas (1995) and Sutton and Higgins (2008).

In contrast, there is a vast literature on sensitivity analysis and Bayesian techniques to handle unmeasured confounding within individual studies (i.e. outside of the context of meta-analysis). See Greenland (2003, 2005) for review. The Bayesian approach takes a model for the relationship between exposure and disease that has been expanded to incorporate *bias parameters* that model unmeasured confounding. We assign prior distributions to bias parameters using external information taken from the literature. The posterior distribution of the exposure effect incorporates uncertainty about unmeasured confounding, and it can be used to compute bias-corrected point and interval estimates.

It seems natural to combine meta-analysis techniques with Bayesian sensitivity analysis for unmeasured confounding. Nonetheless, there are complicating factors. In the endometriosis data we do not have individual-level data, and standard analysis techniques (e.g. Lin, Psaty and Kronmal 1998) are not directly applicable. Additionally, it is difficult to formulate prior distributions for bias parameters in light of the possible heterogeneity due to bias. We must characterize the bias in each of the $k = 11$ different studies.

A further issue is model nonidentifiability. As argued by Greenland (2005), in Bayesian bias analysis of observational data, the posterior distribution is greatly affected the choice of prior distribution.

1.4 Plan of the Paper

In this article, we propose a general methodology to adjust for unmeasured confounding in meta-analysis of observational studies. We focus on the setting where the data are exposure effect estimates, specifically log relative risks, that have been adjusted for a set of confounders. These estimates could be obtained from cohort, case-control or cross-sectional studies. Thus we treat the effect estimates as the basic unit of analysis that are assumed to be roughly normally distributed with known standard error. See Gelman et al. (2003), Smith et al. (1995), Welton et al. (2009) and Carlin (1992) for similar examples from Bayesian meta-analysis. In Section 2, we describe the proposed method that builds upon the sensitivity analysis framework of Lin et al. (1998). We obtain results for the endometriosis data in Section 3. To better understand the performance of our method, Section 4 gives simulation results comparing to a standard Bayesian random effects meta-analysis (Carlin 1992). We illustrate that if the investigator is able to correctly glean the true distribution of bias from unmeasured confounding, then our Bayesian procedure will give interval estimates for the exposure effect with roughly nominal 95% coverage probability.

2 Meta-Analysis of Observational Studies with Unmeasured Confounders

2.1 Model for Unmeasured Confounding

Our method combines the bias modelling framework of Welton et al. (2009) with the model for unmeasured confounding due to Lin et al. (1998). Let (y_j, σ_j) for $j \in 1 : k$ denote the log relative risk estimates with standard errors, for the association between a dichotomous exposure and outcome in k different observational studies. These could be any combination of cohort studies, case-control studies or cross-sectional studies. We assume that each estimate y_j has been adjusted for a set of covariates that were measured and available for analysis in the j^{th} study. So for example, in Table 1 we can

see that in the cross-sectional study of Kirson and Poindexter there were four variables (age, parity, miscarriage and IUD use) that were available for risk adjustment.

Following convention in Bayesian meta-analysis (Gelman et al. (2003), Smith et al. (1995), Welton et al. (2009) and Carlin (1992)), we assume that y_j is approximately normally distributed

$$y_j \sim N(\theta_j^*, \sigma_j^2) \quad \text{for } j \in 1 : k, \quad (1)$$

with mean θ_j^* and variance σ_j^2 , where σ_j is the standard error of y_j and is assumed to be known. The quantity θ_j^* is the logarithm of the relative risk for the association between the dichotomous exposure and dichotomous outcome in study j , conditional the set of covariates that are measured in study j . Equation (1) is based on a large sample normal approximation for the distribution of the log relative risk estimate y_j .

The incidence of endometriosis is approximately 250 cases per 100,000 person years (Missmer et al. 2004). When the incidence is low, then the relative risk is approximated by the odds ratio and hazard ratio (Greenland 2005). Thus following convention in meta-analysis of observational studies (e.g. Vercellini et al. (2011), Toh and Hernández-Díaz (2007) and Ethminan et al. (2009)), we allow for the possibility that y_j is an adjusted log odds ratio, or alternatively, an adjusted log hazard ratio. The bias from using odds ratios or hazard ratios to estimate the relative risk is small provided that the cumulative incidence of disease is less than 10% (Greenland 2005).

Our objective is to model the bias in y_j and θ_j^* due to the missing confounder dysmenorrhea. We use the algebraic adjustment formula described by Lin et al. (1998), which is applicable to observational studies where there is a single binary unmeasured confounder. The adjustment formula assumes that the causal relative risk for the effect of the exposure on disease can be represented by a regression model that includes the exposure indicator as well as the measured covariates, and additionally, a single binary unmeasured confounder. The regression model is loglinear in the case where y_j and θ_j^* are relative risks. Alternatively, if y_j is a log odds ratio or log hazard ratio, then the regression model is logistic or a cox proportional hazards model, respectively.

Suppose that the quantities θ_j , for $j \in 1 : k$, are log relative risks for the association between the dichotomous exposure and disease in study j , adjusted for measured covariates in study j , and *additionally*, adjusted for a single binary unmeasured confounder. Lin et al. (1998) showed that if the

disease incidence is low and the binary unmeasured confounder is independent of the measured covariates within levels of the exposure, then there exists an algebraic relationship between θ_j from the full model and θ_j^* from the reduced model that omits the unmeasured confounder. Lin et al. (1998) prove that

$$\theta_j^* \approx \theta_j + \Omega(RR_j, p_{1j}, p_{0j}) \quad \text{for } j \in 1 : k, \quad (2)$$

where

$$\Omega(RR_j, p_{1j}, p_{0j}) = \log \frac{RR_j \times p_{1j} + (1 - p_{1j})}{RR_j \times p_{0j} + (1 - p_{0j})}.$$

If the disease incidence is low, then this relationship also holds for log odds ratios and log hazard ratios obtained via logistic or Cox proportional hazards regression. See Lin et al. (1998) for a complete justification of the bias formula in equation (2), and see also Arah, Chiba and Greenland (2008) for an overview of different models for unmeasured confounding.

Following Greenland (2005), we call the quantities (RR_j, p_{1j}, p_{0j}) *bias parameters* because they describe the amount of unmeasured confounding bias in study j . The quantity RR_j is the relative risk for the association between the dichotomous unmeasured confounder and the outcome in study j , conditional on the exposure and measured covariates. The quantities p_{1j} and p_{0j} are the prevalences of the unmeasured confounder in the exposed group and the unexposed group, respectively. Using the bias model, we re-write equation (1) as

$$y_j \sim N\{\theta_j + \Omega(RR_j, p_{1j}, p_{0j}), \sigma_j^2\} \quad \text{for } j \in 1 : k, \quad (3)$$

where σ_j^2 is known. If either $RR_j = 1$ or $p_{1j} = p_{0j}$, then $\Omega(RR_j, p_{1j}, p_{0j}) = 0$, and there is no unmeasured confounding.

A crucial point is that the standard error of the bias-corrected point estimator is not affected by unmeasured confounding. In other words, if you introduce unmeasured confounding into an observational study, then the location of the exposure effect estimate will shift, however its standard error will remain constant. This is proven by Lin et al. (1998, page 951), and it seems paradoxical because it is well known that covariate adjustment typically affects the precision of regression coefficients (Pocock et al. 2002). For example, The Fibrinogen Studies Collaboration (2009) compared fully adjusted versus and partially adjusted exposure effects in a meta-analysis of observational studies. The authors showed that adjusting for missing confounders does indeed

affect the standard error. However, it is important to realize that the bias correction formula due to Lin et al. (1998) incorporates perfect knowledge about the magnitude of unmeasured confounding. In contrast, for the case of The Fibrinogen Studies Collaboration (2009), the magnitude of unmeasured confounding is estimated from the data.

To complete a Bayesian analysis, it remains to assign prior distributions to model parameters. An important point is that the parameter θ_j in equation (3) is nonidentifiable unless the other parameters (RR_j, p_{0j}, p_{1j}) are known. For each j , the distribution of the single data point is y_j is governed by four parameters $(\theta_j, RR_j, p_{0j}, p_{1j})$. Even if the standard error σ_j was roughly zero (e.g. if we had infinitely large amounts of data in study j), then we would still be unable to obtain asymptotically unbiased parameter estimates. Thus estimating θ_j is hopeless without prior information about bias.

One possible remedy is to substitute fixed values for the bias parameters (RR_j, p_{1j}, p_{0j}) within the framework of a frequentist sensitivity analysis. For example, one might assume that $RR_1 = RR_2 = \dots = RR_k$, $p_{0,1} = p_{0,2} = \dots = p_{0,k}$ and $p_{1,1} = p_{1,2} = \dots = p_{1,k}$ and then explore sensitivity over a range of values for bias parameters. This presumes a fixed amount of bias is at play in each study, and then shifts the relative risk estimates accordingly. In fact, this is a special case of a Bayesian analysis using a point-mass prior on (RR_j, p_{0j}, p_{1j}) .

Additionally, we note that the bias-corrected quantities $\theta_1, \dots, \theta_k$ do not necessarily have a causal interpretation. In the j^{th} study, provided there are no *additional* unmeasured confounders, θ_j is the causal log relative risk for the effect of the exposure on the outcome, conditional on the set of measured and unmeasured confounders. However, in the endometriosis data this assumption is tenuous because of the diversity of covariates that are available for analytic adjustment in each study. Therefore, we view the quantities $\theta_1, \dots, \theta_k$ as merely a conceptual tool for obtaining biased-corrected inferences that explore sensitivity to a binary unmeasured confounder.

2.2 Prior Distributions

Write $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, $\mathbf{RR} = (RR_1, \dots, RR_k)$, $\mathbf{p}_1 = (p_{1,1}, \dots, p_{1,k})$, $\mathbf{p}_0 = (p_{0,1}, \dots, p_{0,k})$, $\mathbf{y} = (y_1, \dots, y_k)$ and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k)$. We assign priors so that $\boldsymbol{\theta}$ and $(\mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0)$ are marginally independent. This reflects the belief that the amount of unmeasured confounding in any particular study is independent of the exposure effect.

We model the individual exposure effect parameters $\theta_1, \dots, \theta_k$ as exchangeable, and following Carlin (1992), we assign a conventional hierarchical random effects prior distribution

$$\theta_j \stackrel{\text{iID}}{\sim} N(\mu, \tau^2) \quad \text{for } j \in 1 : k,$$

where the parameter μ is the mean of the distribution of exposure effects, and τ is the standard deviation. The pooled log relative risk μ is the primary target of inference, whereas the quantity τ governs the heterogeneity of exposure effects across studies.

For the hyperparameters (μ, τ^2) we assign

$$\begin{aligned} \mu &\sim N(0, 10^3) \\ \tau &\sim \text{Uniform}(0, 10^3). \end{aligned}$$

Lambert, Sutton, Burton et al. (2005) investigated the effect of using different vague prior distributions for the between study variance in Bayesian meta-analysis. They found that when the number of studies is small (≤ 5) then the prior distribution tends to greatly affect precision of the posterior distribution of μ . For example, the conventional inverse gamma prior distribution has the convenience of conditionally conjugacy, but it tends to perform poorly, particularly if the true between study variance is small.

For **RR**, we treat the quantities RR_1, \dots, RR_k as exchangeable and assign

$$\log RR_j \stackrel{\text{iID}}{\sim} N(\mu_{RR}, \tau_{RR}^2) \quad \text{for } j \in 1 : k. \quad (4)$$

We model $\log RR_j$ for $j \in 1 : k$ as independent and identically normally distributed with mean μ_{RR} and standard deviation τ_{RR} . Thus the strength of the association between the unmeasured confounder and outcome can vary from one study to the next, with heterogeneity that is governed by τ_{RR} . Because the data reveal nothing about the unmeasured confounder, the user must specify values for the hyperparameters (μ_{RR}, τ_{RR}) using external information. In Section 3.2.1 we give an illustration of prior elicitation for bias parameters, along with general advice for practitioners.

Next we assign priors for the bias parameters \mathbf{p}_1 and \mathbf{p}_0 . We assume exchangeability and assign

$$\text{logit}(p_{0j}) \stackrel{\text{iID}}{\sim} N(\mu_{p_0}, \tau_{p_0}^2) \quad (5)$$

$$\text{logit}(p_{1j}) \stackrel{\text{iID}}{\sim} N(\mu_{p_1}, \tau_{p_1}^2), \quad (6)$$

for $j \in 1 : k$. The quantities p_{0j} and p_{1j} are the prevalences of the unmeasured confounder in unexposed versus exposed individuals in the j^{th} study. They govern the strength of the association between the confounder and exposure. The quantities μ_{p_1} and μ_{p_0} are means, whereas τ_{p_1} and τ_{p_0} are standard deviations. If the quantities τ_{p_1} and τ_{p_0} are non-zero, then this means that we allow for heterogeneity wherein the strength of the association between the unmeasured confounder and exposure differs from one study population to the next. The hyperparameters must be specified by the analyst, and an illustration is given in Section 3.2.1.

Our method assumes that the partially adjusted estimates may be treated in the same manner (exchangeably), irrespectively of the extent and type of adjustment made. Given the very different types and amounts of adjustments in Table 1, this is not very plausible. There are two issues: First, the assumed exchangeability of $\theta_1, \dots, \theta_k$, and second, the exchangeability of individual parameters in each of the sets $\{RR_1, \dots, RR_k\}$, $\{p_{0,1}, \dots, p_{0,k}\}$, $\{p_{1,1}, \dots, p_{1,k}\}$. We assume exchangeability of $\theta_1, \dots, \theta_k$ because this is the assumption that underlies the systematic review and meta-analysis of Vercellini et al. (2011). This is not ideal because there could be systematic differences between parameter that are not attributable to confounding. Further, the assumed exchangeability of bias parameters is also tenuous. For example, the Nurse Health II study (Table 1) has 17 measured confounders. The magnitude of unmeasured confounding from dysmenorrhoea could be quite minimal. Additionally, one could argue that that case-control studies are more prone to bias than cohort studies. Thus one could imagine extending the prior distribution in equations (4), (5) and (6) to incorporate study-level covariates in the spirit of meta-regression (Sutton and Higgins 2008). We explore this possibility in Section 5.

2.3 Model Fitting and Computation

According to Bayes theorem, the posterior distribution is proportional to the likelihood function multiplied by the prior distribution. We have

$$P(\boldsymbol{\theta}, \mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0, \mu, \tau^2 | \mathbf{y}, \boldsymbol{\sigma}) \propto \prod_{j=1}^k \exp \left[-\frac{1}{2\sigma_i^2} \{y_i - (\theta_j + \Omega(RR_j, p_{1j}, p_{0j}))\}^2 \right] \times (7)$$

$$P(\theta_i | \mu, \tau) P(RR_j) P(p_{1j}) P(p_{0j}) \times (8)$$

$$P(\mu) P(\tau), (9)$$

where equation (7) is the likelihood function, equation (8) is the prior distributions on $(\theta_j, RR_j, p_{0j}, p_{1j})$, and equation (9) is the prior on the hyperparameters (μ, τ) . We sample from the posterior distribution using Markov chain Monte Carlo. We update sequentially from the following three conditional distributions, which are written to reflect the conditional independence in the posterior distribution,

$$P(\boldsymbol{\theta} | \mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0, \mu, \tau^2, \mathbf{y}, \boldsymbol{\sigma}) \quad P(\mu, \tau^2 | \boldsymbol{\theta}) \quad P(\mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0 | \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\sigma}).$$

Computer code written using the software R (R Development Core Team 2010) is available from the author's website. Sampling from $P(\mu, \tau^2 | \boldsymbol{\theta})$ can be done using standard computational techniques for hierarchical Bayesian models. We sample first from the marginal distribution $P(\tau^2 | \boldsymbol{\theta})$ and then from the conditional distribution $P(\mu | \tau^2, \boldsymbol{\theta})$ by using the algorithm described in Section 3.3 of Gelman et al. (2003) for semi-conjugate prior distributions.

Standard calculations based on completing the square for Bayesian conditionally conjugate Gaussian models show that the conditional distribution $P(\boldsymbol{\theta} | \mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0, \mathbf{y}, \boldsymbol{\sigma})$ is multivariate normal with diagonal covariance matrix. We have

$$\theta_j | RR_j, p_{1j}, p_{0j}, y_j, \sigma_j \sim N \left(\frac{\tau^2 \{y_i - \Omega(RR_j, p_{1j}, p_{0j})\} + \sigma_j^2 \mu}{\sigma_j^2 + \tau^2}, \frac{\sigma_j^2 \tau^2}{\sigma_j^2 + \tau^2} \right). \quad (10)$$

This equation says that, given the bias parameters (RR_j, p_{0j}, p_{1j}) and data (y_j, σ_j) , the posterior distribution of θ_j is normal with mean that is a weighted average of the prior mean of θ_j , which is μ , and the bias-corrected exposure effect estimate, which is $y_i - \Omega(RR_j, p_{1j}, p_{0j})$. We sample from $P(\boldsymbol{\theta} | \mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0, \mu, \tau^2, \mathbf{y}, \boldsymbol{\sigma})$ by simulating from a multivariate normal with mean and variance from equation (10).

Updating from $P(\mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0 | \mathbf{y}, \boldsymbol{\theta})$ is accomplished using Metropolis Hastings steps because of the cumbersome non-linear dependence of $\Omega(RR_j, p_{1j}, p_{0j})$ on the parameters (RR_j, p_{0j}, p_{1j}) . We update the vectors \mathbf{RR} , \mathbf{p}_1 and \mathbf{p}_0 using a random walk Metropolis Hastings step with proposal distribution that is multivariate normal with mean zero and covariance matrix equal to the identity matrix multiplied by a tuning parameter. The tuning parameter is adjusted using trial MCMC simulation runs to ensure satisfactory convergence. We refer the reader to Gelman et al. (2003) for further discussion of MCMC.

3. Application to Endometriosis Data

3.1 Preliminary Bayesian Analysis that Ignores Unmeasured Confounding

Before applying our proposed methodology, it is informative to do a preliminary analysis using a Bayesian random-effect model that ignores unmeasured confounding. We use the method of Carlin (1992), which henceforth we denote by the acronym CARLIN. To implement CARLIN, we sample from the posterior distribution in equations (7)-(9), while forcing $\Omega(RR_j, p_{1j}, p_{0j}) = 0$ for $j \in 1 : k$. We update from the conditional distributions $P(\boldsymbol{\theta}|\mu, \tau^2, \mathbf{y}, \boldsymbol{\sigma})$ and $P(\mu, \tau^2|\boldsymbol{\theta})$ using the MCMC procedure described in Section 2.3. We need not update the bias parameters $(\mathbf{RR}, \mathbf{p}_1, \mathbf{p}_0)$ because we are assuming that there is no unmeasured confounding.

The results of applying CARLIN are presented in Figure 2, which plots the posterior means of the relative risks $\exp(\theta_1), \dots, \exp(\theta_{11})$ for the 11 studies, with equi-tailed 95% posterior credible intervals. The diamond interval at the base of the figure gives the posterior mean and credible interval of the pooled relative risk $\exp(\mu)$. Contrasting Figures 1 and 2, we see that the pooled relative risk estimates are identical, but there are larger differences in the estimated study-specific effects. In Figure 1, we have 1.21 with 95% interval estimate (0.94, 1.56), versus 1.21 (0.89, 1.62) for CARLIN. As expected, the Bayesian interval estimate is wider because CARLIN admits uncertainty in the variance parameter τ^2 (Sutton and Higgins, 2008). The method of moments point estimate of τ via DerSimonian and Laird is equal to 0.35, whereas the posterior mean of τ via CARLIN is 0.43 with 95% equi-tailed credible interval is (0.22, 0.67).

We see large differences between Figures 1 and 2 for the study-specific effects because of shrinkage towards the common mean $\exp(\mu)$. The credible intervals for $\exp(\theta_1), \dots, \exp(\theta_k)$ are more narrow because the Bayesian approach borrows information from all 11 studies in order to make inferences about a single study. See Carlin (1992) and Smith et al. (1995) for discussion.

Our Bayesian analysis confirms that there is a great deal of heterogeneity in the effect estimates. We calculate the I^2 statistic from the MCMC output using the method of Higgins and Thompson (2002), and we obtain $I^2 = 64\%$, (versus 81% using DerSimonian and Laird in Section 1.1). Figure 2 also present weights for each study for computing the posterior mean of μ . In

the j^{th} study, the weight is the reciprocal of the marginal posterior variance of θ_j , which is $\{Var(\theta_j|\mathbf{y}, \boldsymbol{\sigma}) + E(\tau^2|\mathbf{y}, \boldsymbol{\sigma})\}^{-1}$ for $j \in 1 : k$. The weights are normalized across k studies to add up to 100%.

3.2 Adjustment for Unmeasured Confounding Due to Dysmenorrhea

3.2.1 An Example of Prior Elicitation and Recommendations for Practitioners

We now extend the analysis of Section 3.1 by incorporating uncertainty about unmeasured confounding. We use the Bayesian method of Section 2 (denoted henceforth by the acronym BAYES) where we allow that $\Omega(RR_j, p_{1j}, p_{0j}) \neq 0$. First, we must specify the hyperparameters (μ_{RR}, τ_{RR}) , (μ_{p_0}, τ_{p_0}) and (μ_{p_1}, τ_{p_1}) , which determine the prior distribution for bias parameters in equations (4)-(6).

Choosing the prior distribution is a delicate matter because it can greatly influence the results. We must guess plausible values of the bias parameters. This can be accomplished by looking to the literature to identify studies that investigate the pattern and distribution of dysmenorrhea in reproductive aged women. We require information about how dysmenorrhea affects oral contraceptive use and the risk of endometriosis.

To illustrate, we begin by guessing values for the hyperparameters (μ_{RR}, τ_{RR}) from equation (4), which govern the effect of dysmenorrhea on risk of endometriosis. Cramer et al. (1986) conducted a case-control study of 4062 American women and obtained an estimated relative risk for the association between moderate menstrual pain and endometriosis equal to 3.4 with 95% CI (2.2, 5.2). Another cross-sectional study of American women yielded a relative risk of 2.89 with 95% CI (1.17, 7.12) (Kresch et al. 1984). Using these two different studies, we set μ_{RR} equal to average $1.142 = (\log(3.4) + \log(2.89))/2$. The quantity τ_{RR} is the standard deviation of RR_1, \dots, RR_k across the sequence of studies, and eliciting a value is more speculative. Thus we conservatively set $\tau_{RR} = \log(3.4) - \log(2.89) = 0.162$ as the range of the log relative risks.

Next, we choose values for the hyperparameters (μ_{p_0}, τ_{p_0}) from equation (5), which govern the distribution of dysmenorrhea among women who do not use oral contraceptives. Dysmenorrhea is common in reproductive aged women and it has been well studied. For example, Proctor and Farquhar (2007)

describe a systematic review of 7 different studies that reported prevalences equal to 72%, 93%, 80%, 73%, 60%, 55% and 60%. In this case, there is a wealth of available information and we set $\mu_{p_0} = 0.989$ and $\tau_{p_0} = 0.818$, which are the mean and standard deviation of these 7 quantities, computed on the log odds scale.

Finally, we set the hyperparameters (μ_{p_1}, τ_{p_1}) for equation (6), which govern the distribution of dysmenorrhea among oral contraceptive users. Unfortunately, there are few published studies investigating the patterns of oral contraceptive use in relation to primary dysmenorrhea. One example is the study of Robinson et al. (1992). The authors found that women with severe dysmenorrhea were 8.0 (1.37, 46.1) times more likely to use oral contraceptives. This gives a log odds ratio of 2.08 with standard error 0.89. In this case, we have only a single study to work with, and it is difficult to elicit plausible values for (μ_{p_1}, τ_{p_1}) . Because $\text{logit}(p_{1j}) = [\text{logit}(p_{1j}) - \text{logit}(p_{0j})] + \text{logit}(p_{0j})$, we assign a prior for $\text{logit}(p_{1j})$ where $\mu_{p_1} = 0.989 + -2.08 = -1.091$ and $\tau_{p_1} = \sqrt{(0.818^2 + 0.89^2)} = 1.21$. Note that we assume prior independence between p_{1j} and p_{0j} , which may be implausible. In some settings, it may be convenient to reparametrize the quantity $\Omega(RR_j, p_{1j}, p_{0j})$ in equation (3) in terms of the ratio p_{1j}/p_{0j} or relative odds $\frac{p_{1j}}{(1-p_{1j})} / \frac{p_{0j}}{(1-p_{0j})}$ in order to more easily elicit a prior distribution that describes the association between the exposure and the unmeasured confounder.

We recommend that practitioners use the results of systematic reviews for eliciting prior distributions for bias parameters. In the absence of systematic reviews, we recommend compiling the results of individual studies in the manner described above. If there are no published studies that describe the distribution of the unmeasured confounder then an alternative is to survey subject area experts to elicit their beliefs about the magnitude of the possible bias. See Turner et al. (2009), who describes elicitation scales for bias and a method for gauging the probability distribution.

3.2.2 Results

We apply BAYES to adjust for unmeasured confounding. The results are presented in Figure 3. The key observation is that, compared with Figures 1 and 2, we see that the estimates are shifted to the left and a null association. This makes sense intuitively. During the posterior updating, we are sampling from the posterior distribution of bias parameters and then obtaining bias-

corrected study specific effects. The analysis presumes that exposed subjects have artificially high risk of endometriosis and the relative risk estimates are shifted accordingly. The pooled relative risk is 1.00 with 95% CI (0.74, 1.35). Thus by incorporating prior beliefs about the magnitude of confounding from dysmenorrhea, the meta-analysis no longer suggests a harmful association between oral contraceptives and endometriosis.

4. Simulation Experiment

4.1 Performance of BAYES when the Investigator's Prior is Equal to Nature's Prior

A difficulty with the preceding analysis is that the results depend heavily on the prior distribution for bias parameters. If we pick the “right prior” then the interval estimates will be suitably shifted towards the truth. However many things could go wrong. If the location parameters μ_{RR} , μ_{p_0} and μ_{p_1} are poorly chosen then the bias-corrected interval estimates from BAYES will miss the truth entirely. The reason that the results are sensitive to the prior distribution is because of nonidentifiability. In equation (3), we see that the distribution of the data point y_j is a function of 4 unknown parameters. Even if the standard errors σ_j are close to zero, then we could still not hope to obtain asymptotically unbiased estimates of θ_j . Thus any conclusions that we obtain from using BAYES are largely the result of the prior distribution.

Do 95% credible intervals from BAYES have 95% frequentist coverage probability? What constitutes a “good prior”? To what extent will the performance of BAYES deteriorate relative to CARLIN through a careless choice of prior distribution? These questions are answered in part in a recent paper by Gustafson and Greenland (2009) describing interval estimation in observational studies. In the endometriosis data analysis of Section 3.2, *there are two distinct parameter distributions*. On the one hand, there is the *Investigator's Prior*, which we define as the prior distributions that are used in equations (8) and (9) to calculate the posterior distribution. On the other hand, there is *Nature's Prior*, which we define as the true frequency distribution of parameters that are used to generate data across the sequence of k studies in the meta-analysis. Gustafson and Greenland (2009) prove that if the Investigator's Prior is equal to Nature's Prior, then the 95% credible intervals will have exactly nominal 95% frequentist coverage probability when averaged over the

parameter space. This property holds regardless of sample size and regardless of whether the model is identifiable or not. See McCandless et al. (2007) for further discussion of Bayesian interval estimation in nonidentifiable models.

We study the performance of BAYES using simulations. We generate synthetic meta-analysis data using the simulation design of Lambert et al. (2005). We limit our investigation to the case of $k = 6$ observational studies, each of which has 5000 study participants. Denote the individual-level data in study j as (Y_{ji}, X_{ji}) , for $i \in 1 : n = 5000$, where Y_{ji} and X_{ji} are dichotomous outcome and exposure variables taking values 1 or 0 to denote presence or absence of the outcome or exposure. We assume that in each study there is a single binary unmeasured confounder, denoted U_{ji} , which also takes values 1 or 0. For example, setting $U_{ji} = 1$ means that the i^{th} individual in the j^{th} study has dysmenorrhea. Thus the complete data are (Y_{ji}, X_{ji}, U_{ji}) . We generate data using the model

$$\begin{aligned} Y_{ji}|X_{ji}, U_{ji} &\sim \text{Bernoulli}\{\exp(-4 + \theta_j X_{ji} + \log RR_j U_{ji})\} \\ U_{ji}|X_{ji} = 1 &\sim \text{Bernoulli}(p_{j1}) \\ U_{ji}|X_{ji} = 0 &\sim \text{Bernoulli}(p_{j0}) \\ X_{ji} &\sim \text{Bernoulli}(1/2). \end{aligned}$$

Setting the y-intercept equal to -4 to ensures that the outcome is rare in keeping with the modelling assumptions of Lin et al. (1998) for estimating relative risks. Because U is unmeasured this means that the observed data in study j consists merely of a 2×2 table tabulated over X and Y .

To generate data, we require four parameters $(\theta_j, RR_j, p_{j0}, p_{j1})$ for each $j \in 1 : k$. We simulate parameters by sampling from Nature's prior, which we specify as:

$$\theta_j \sim N(\mu = 0, \tau = 1/4) \tag{11}$$

$$\log RR_j \sim N(1, 1/4) \tag{12}$$

$$\text{logit} p_{j0} = 0 \tag{13}$$

$$\text{logit} p_{j1} \sim N(1, 1/4), \tag{14}$$

for $j \in 1 : k$. Equation (11) says that X (oral contraceptive use) and Y (endometriosis) are not associated, on average, apart from some heterogeneity in the exposure effects, which is determined by $\tau = 1/4$. The unmeasured confounder U_{ji} is dysmenorrhea, and equation (12) says that U_{ji} increases

the risk of endometriosis with a relative risk of $\exp(1) = 2.71$, on average. Equations (13) and (14) govern the prevalence of the unmeasured confounder within exposure groups. For simplicity, we fix $\text{logit}p_{j0} = 0$, which implies that the prevalence dysmenorrhea among non-users of oral contraceptives is exactly $p_{j0} = \exp(0)/(1 + \exp(0)) = 50\%$ for all individuals in each study. Finally, equation (14) controls the association between the unmeasured confounder and the exposure. It says that the prevalence of dysmenorrhea is $\exp(1)/(1 + \exp(1)) = 73\%$, on average, among users of oral contraceptives.

Our simulation proceeds as follows. For $j \in 1 : k = 6$, we generate parameters from equations (11)-(14). Next, for $i \in 1 : n = 5000$, we generate (Y_{ji}, X_{ji}, U_{ji}) . Because U_{ji} is assumed to be unmeasured, we analyze the $k = 6$ datasets using the log linear model

$$\log[P(Y_{ji} = 1|X_{ji})] = \alpha^* + \theta_j^* X_{ji}. \tag{15}$$

and we obtain 6 log relative risk estimates y_1, \dots, y_6 of $\theta_1^* \dots, \theta_6^*$ with standard errors $\sigma_1, \dots, \sigma_6$. The estimates y_1, \dots, y_6 are *unadjusted* for confounding because equation (15) does not control for U_{ji} .

Next, we apply BAYES and CARLIN to the simulated log relative risks y_1, \dots, y_6 . For BAYES, we must specify the hyperparameters $(\mu_{RR}, \tau_{RR}, \mu_{p_0}, \tau_{p_0}, \mu_{p_1}, \tau_{p_1})$, which characterize the Investigator's Prior beliefs about bias. For the moment, we consider the idealized scenario where the Investigator's Prior is identical to Nature's Prior. Thus we set $\mu_{RR} = 1$, $\tau_{RR} = 1/4$, $\mu_{p_{j0}} = 0$, $\tau_{p_{j0}} = 0$, $\mu_{p_{j1}} = 1$, $\tau_{p_{j0}} = 1/4$. In addition, we apply a gold standard analysis technique, which we call GOLD. GOLD is implemented as follows: We analyze the $k = 6$ datasets using the log linear model

$$\log[P(Y_{ji} = 1|X_{ji})] = \alpha + \theta_j X_{ji} + \beta U_{ji}. \tag{16}$$

GOLD is defined as applying CARLIN to the log relative risk estimates y_1, \dots, y_6 that are obtained from fitting equation (16) to the individual-level data (Y_{ji}, X_{ji}, U_{ji}) . Because GOLD adjusts for missing confounder U_{ji} directly, it describes the best case scenario where there is no unmeasured confounding. Certainly one cannot expect to do better than GOLD. Thus GOLD serves as a useful benchmark for comparing the extent that BAYES succeeds to adjust for unmeasured confounding.

The results for a single meta-analysis are depicted in Figure 4. Vertical lines with symbols $\theta_1, \dots, \theta_6$ indicate the true relative risks for the relationship between X and Y in each of the $k = 6$ studies. As expected, interval estimates

Table 2: Simulation results from Section 4.1. Coverage probabilities of 95% interval estimates that are calculated by applying either GOLD, CARLIN or BAYES to ensembles of 100 different simulated meta-analysis datasets. Simulation standard errors for coverage probabilities are $< \sqrt{\frac{0.5 \times 0.5}{100}} = 5\%$.

| | Study-specific effects | | | | | | Underlying mean |
|--------|--|------------|------------|------------|------------|------------|-----------------|
| | for the 6 studies in the meta-analysis | | | | | | effect |
| | θ_1 | θ_2 | θ_3 | θ_4 | θ_5 | θ_6 | μ |
| GOLD | 95% | 96% | 96% | 95% | 95% | 96% | 96% |
| CARLIN | 72% | 74% | 72% | 76% | 74% | 74% | 66% |
| BAYES | 95% | 96% | 95% | 95% | 96% | 95% | 96% |

from CARLIN are biased to the *right* compared to GOLD because CARLIN does not adjust for the missing confounder U . In contrast, BAYES intervals are less biased and similar to GOLD.

To distill away the effect of random simulation error, Table 2 reports coverage probabilities of 95% interval estimates obtained when we apply either GOLD, CARLIN or BAYES to 100 different simulated meta-analysis datasets. The coverage probabilities are calculated by taking the average number times the intervals cover the true parameters. So, for example, in the bottom right corner of the Table 2 we can see the value 96%. This means that BAYES intervals for the pooled relative risk $\mu = 0$ covered the true value 96 times out of 100. As expected, GOLD gives perfect 95% coverage probability, whereas the coverage from CARLIN is much reduced. For BAYES, the coverage probability is also roughly 95%. This is to be expected based on the theory of Gustafson and Greenland (2009). Bayesian interval estimates give nominal level frequentist coverage, despite the fact that the model for unmeasured confounding is nonidentifiable.

4.2 Performance When the Prior is Misspecified

Nonetheless, Table 2 is far from realistic because it is unlikely that the investigator would have the good fortune to know the precise shape and location of Nature’s prior. In the endometriosis data example, our efforts to elicit the

hyperparameters $(\mu_{RR}, \tau_{RR}, \mu_{p_0}, \tau_{p_0}, \mu_{p_1}, \tau_{p_1})$ are merely speculative. Thus it is useful to understand the performance of BAYES when the investigator has chosen the wrong prior.

Accordingly, we reproduce the simulation of Section 4.1 in the setting where the Investigator's prior and Nature's prior are unequal. We consider the situation where, unbeknownst to the investigator, bias parameters are drawn from a Gaussian distribution with a different location. To generate synthetic data, we sample parameters as follows. First, we set $\text{logit}p_{j0} = 0$ and sample θ_j from equation (11) just like in Section 4.1. However, we sample the bias parameters RR_j and p_{j1} from

$$\begin{aligned} \log RR_j &\sim N(a, 1/4) \\ \text{logit}p_{j1} &\sim N(-a, 1/4), \end{aligned}$$

rather than from equation (12) and (14), where $a = 0, 1, 2$. When $a = 2$, then it means that, on average, the magnitude of unmeasured confounding is very large. The quantities $\log RR_j$ and $\text{logit}p_{j1}$ will typically be large in magnitude across the sequence of studies. In contrast, when $a = 0$ then there is no unmeasured confounding, on average.

We apply GOLD, CARLIN and BAYES to the simulated meta-analysis datasets. For BAYES, we require the hyperparameter inputs for the Investigator's Prior beliefs about bias. Following the recommendation of a reviewer, we consider two prior distributions: A *narrow prior*, with $(\mu_{RR}, \tau_{RR}, \mu_{p_{j0}}, \tau_{p_{j0}}, \mu_{p_{j1}}, \tau_{p_{j1}}) = (1, 1/4, 0, 0, -1, 1/4)$, and additionally, a *conservative prior* with $(\mu_{RR}, \tau_{RR}, \mu_{p_{j0}}, \tau_{p_{j0}}, \mu_{p_{j1}}, \tau_{p_{j1}}) = (1, 1/2, 0, 0, -1, 1/2)$. The narrow prior is exactly the same prior distribution that was used in Section 4.1. So, if it is the case that $a = 1$, the the Investigator's Prior is equal to Nature's prior. However, if it is the case that $a = 0$ or $a = 2$, then it means that the investigator has erred and picked the wrong prior in the bias analysis. The conservative prior distribution is more diffuse and it should assist BAYES to overcompensate for errors in misspecifying the location parameters μ_{RR} and $\mu_{p_{j1}}$.

The results are presented in Table 3. Each row of the table contrasts the performance of BAYES versus CARLIN and GOLD for a different values of a . In the centre row, right hand side, we see that BAYES with a narrow prior gives roughly perfect 96% coverage probability because the Investigator's Prior is equal to Nature's Prior. The other rows in the table illustrate the extent to which BAYES will collapse relative to CARLIN in the case where the Investigator's prior has been poorly chosen.

GOLD always gives perfect 95% coverage. The bottom row of the table ($a = 0$) describes the scenario where there is no unmeasured confounding, on average. CARLIN gives 96% coverage probability versus 76% coverage probability for BAYES with a narrow prior and 85% coverage for BAYES with a conservative prior. Thus the bottom row of the table tells us that if, in fact, there is truly no unmeasured confounding, then BAYES will not give nominal 95% coverage probability because it needlessly shifts the interval estimates off to the left. Using a conservative prior distribution improves the coverage, but CARLIN is still superior. The top row of Table 3 describes the reverse scenario where, unbeknownst to the investigator, there is an extremely large amount of unmeasured confounding. In this case, CARLIN fails catastrophically with 4% coverage probability. BAYES does better with coverages of 35% or 33% for narrow versus conservative priors, respectively.

The results indicate that careful diligence is required when eliciting the shape and characteristics of the Investigator's Prior. The locations of the distributions is critical because a careless choice will shift the posterior distribution away from the truth entirely. Using a more diffuse prior distribution (e.g. by choosing large values for the standard deviation parameters $\tau_{RR}, \tau_{p_{j0}}, \tau_{p_{j1}}$) is helpful, but it has the disadvantage that it will increase the length of the interval estimates. In Table 3, we see that intervals from BAYES with wide priors have greater average length than intervals with narrow priors.

Our simulations indicate that BAYES is no panacea. It is only as good as the prior distributions that are supplied by the investigator. If, in fact, there is no bias at all, then CARLIN is superior to BAYES. However if biases are present, then BAYES will do better than CARLIN. Thus our simulations echo a conclusion of Gustafson and Greenland (2009): It may be better to model uncertainty bias in observational studies, even approximately, rather than ignore it altogether. A careful choice of prior distribution will yield Bayesian interval estimates with good frequentist performance. Even a poor choice of prior distribution may be an improvement over standard methods that demand identifiability (e.g. by assuming zero bias).

Figure 2: Relative risk estimates, with 95% credible intervals for the association between oral contraceptives and endometriosis using the Bayesian random effects meta-analysis technique CARLIN. The solid diamond denotes the pooled relative risk estimate with 95% credible interval.

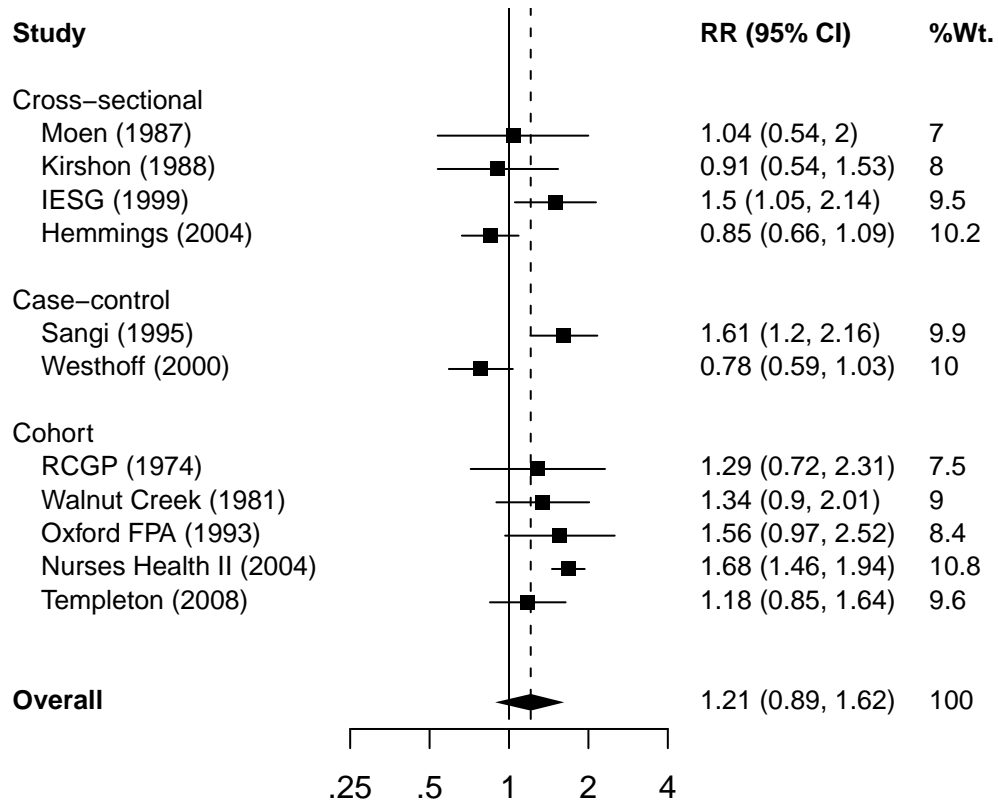


Figure 3: Relative risk estimates, with 95% credible intervals for the association between oral contraceptives and endometriosis calculated using BAYES, which adjusts for unmeasured confounding from dysmenorrhea. The solid diamond denotes the pooled relative risk estimate with 95% credible interval.

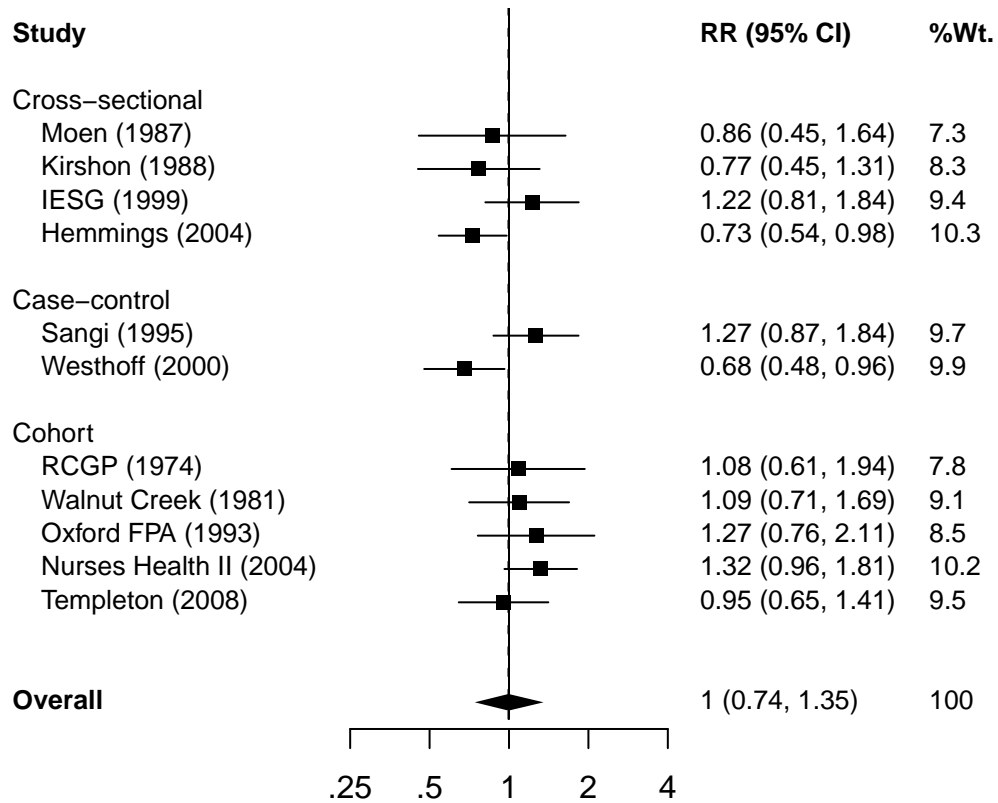


Figure 4: 95% interval estimates calculated by analyzing the data in a single synthetic meta-analysis, using the method GOLD (————), CARLIN (-----) or BAYES (— — —). The symbols $\theta_1, \dots, \theta_6$ indicate the *true* relative risks in each of the 6 studies in the meta-analysis. Diamonds indicate the pooled relative risk estimates with 95% interval.

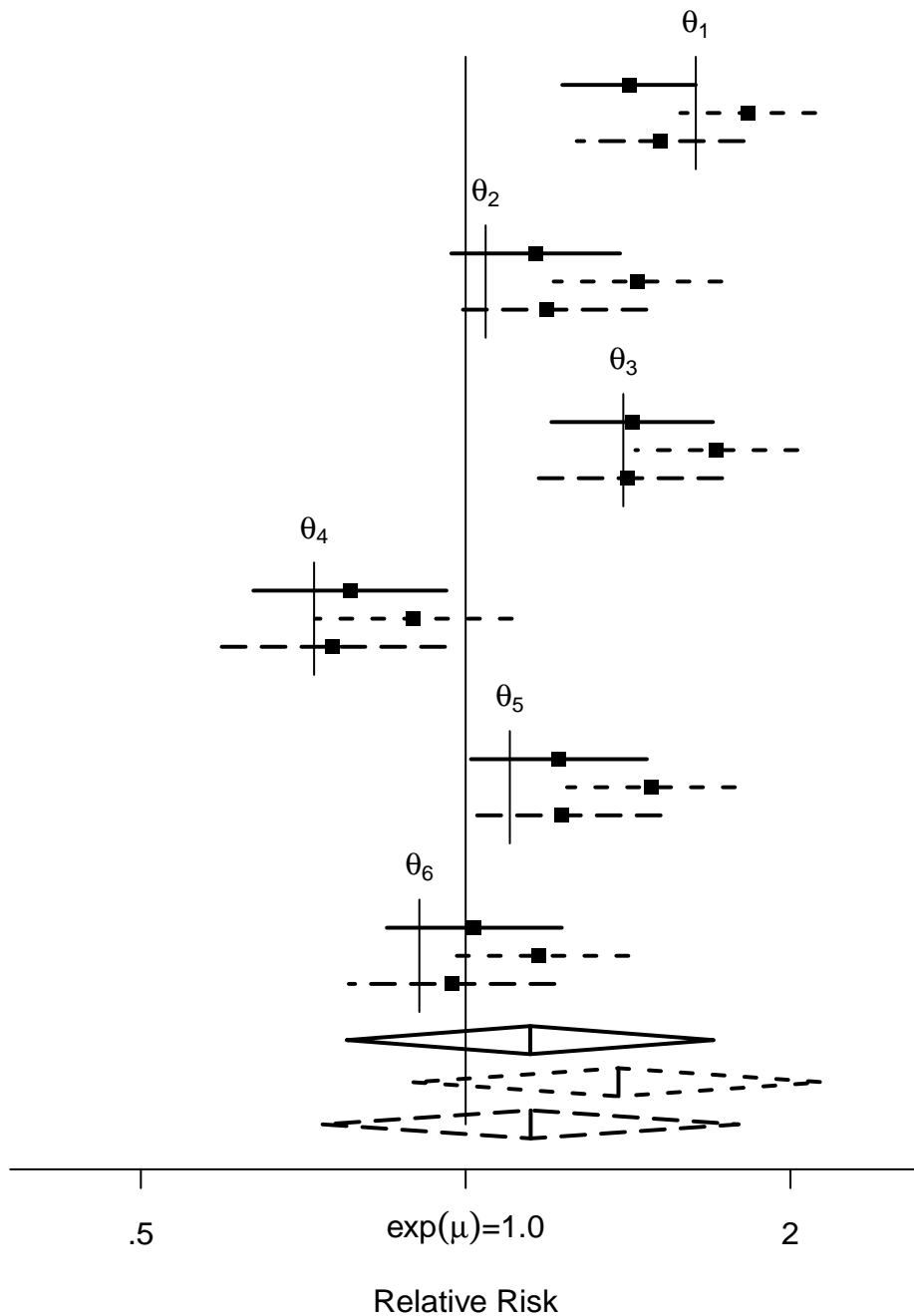


Table 3: Simulation results from Section 4.2. Coverage probabilities of 95% interval estimates (with average length in brackets) that are calculated by applying either GOLD, CARLIN or BAYES to ensembles of 100 different simulated meta-analysis datasets for $a = 0, 1$ or 2 . Simulation standard errors for coverage probabilities are $< \sqrt{\frac{0.5 \times 0.5}{100}} = 5\%$ and for average lengths are < 0.02 .

| Value for a | Bias Parameters in Nature's Prior | | | GOLD | CARLIN | BAYES | |
|------------------|--------------------------------------|----------------------|----------------------|------------|------------|---------------------------|---------------------------------|
| | $\log RR_j$ | $\text{logit}p_{j0}$ | $\text{logit}p_{j1}$ | | | Narrow Prior [†] | Conservative Prior [⊕] |
| 2 | $\sim N(2, 1/4)$ | $= 0$ | $\sim N(-2, 1/4)$ | 96% (0.46) | 4% (0.46) | 35% (0.47) | 33% (0.48) |
| 1 | $\sim N(1, 1/4)$ | $= 0$ | $\sim N(-1, 1/4)$ | 96% (0.52) | 66% (0.52) | 96%* (0.52) | 96% (0.54) |
| 0 | $\sim N(0, 1/4)$ | $= 0$ | $\sim N(0, 1/4)$ | 95% (0.59) | 96% (0.59) | 76% (0.60) | 85% (0.61) |

[†] Bias Parameters in Investigator's Prior are $\log RR_j \sim N(1, 1/4)$, $\text{logit}p_{j0} = 0$, $\text{logit}p_{j1} \sim N(-1, 1/4)$.

[⊕] Bias Parameters in Investigator's Prior are $\log RR_j \sim N(1, 1/2)$, $\text{logit}p_{j0} = 0$, $\text{logit}p_{j1} \sim N(-1, 1/2)$.

* Nature's Prior is equal to Investigator's Prior, which gives perfect coverage theoretically.

5. Discussion

Meta-analysis of observational data is an exciting new area of innovation in statistics. In this article, we describe a novel Bayesian procedure to control for a confounder that is missing across the sequence of studies in a meta-analysis. The studies could be any combination of cohort, case-control or cross sectional studies that use adjusted relative risks, odds ratios or hazard ratios for the measure of effect. We study the relationship between oral contraceptives and endometriosis, where the unmeasured confounder is dysmenorrhea. Prior information about the confounder is obtained from a review of the literature. After adjusting for unmeasured confounding, the interval estimate for the pooled relative risk is shifted towards the null value 1.00 with 95% CI (0.74, 1.35), which indicates no definitive association. Thus our findings echo the conclusion of Vercellini et al. (2011) that the association between oral contraceptives and endometriosis is a spurious artifact of unmeasured confounding.

A concern with our analysis is that the results are sensitive to the choice of prior distribution. A careless choice of prior will adversely affect performance. In particular, if the location of the prior is misspecified then the bias-corrected interval estimate for the summary relative risk will be shifted away from the truth. However, using simulations we demonstrate that if unmeasured confounding is present and the investigator's prior roughly approximates the true sampling distribution of bias parameters across the sequence of studies, then Bayesian credible intervals will have roughly nominal 95% coverage probability.

A weakness of our analysis is that there are several additional biases at play in the endometriosis data. Measurement error is likely to be present, particularly in disease classification because endometriosis requires laparoscopy surgery for a definitive diagnosis. In addition, control selection in case-control studies of endometriosis is fraught with difficulties leading to selection bias. Vercellini et al. (2011) also raised possibility of publication bias, although a Begg's funnel plot showed no indication of asymmetry. The focus of this article is to study whether the results are sensitive to unmeasured confounding from dysmenorrhea, and in principle our method could be extended to model multiple biases (Greenland 2005). Note however that incorporating uncertainty from additional biases requires that the investigator supply additional prior distributions, and this renders the analysis more subjective.

An additional limitation (See Section 2.2) is the assumption of prior exchangeability among components of the bias parameter vectors. Figure 1 reveals that much of the heterogeneity occurs *between* study designs (cross-sectional/case-control/cohort). Therefore a more plausible model would be to extend the priors in equations (4) - (6) using meta-regression with study-level covariates. See Sutton and Higgins (2008) for discussion of heterogeneity of bias. Indeed this was done by Vercellini et al. (2011) to model heterogeneity in the relative risks. However, the problem with study-level covariates is that they complicate the process of prior elicitation. It requires that the investigator specify the extent that unmeasured confounding depends on study design.

An interesting extension of our method would be to incorporate individual-level data from studies in which dysmenorrhea was measured. One could imagine a situation in which the investigator had access to a dataset with complete information on all confounders. In this case, a Bayesian analysis could proceed by multiplying the likelihood function in equation (7) times the likelihood for the individual-level data. Such an analysis would present many methodological challenges (see for example The Fibrinogen Studies Collaboration (2009)), and it would also be less subjective because the individual-level data would inform the prior distribution for bias parameters.

References

- Arah, O. A., Chiba, Y., Greenland, S., 2008. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders, *Annals of Epidemiology* 18, 637–46.
- Carlin, J. B., 1992. Meta-analysis for 2 x 2 tables: a Bayesian approach. *Statistics in Medicine* 11, 141–58.
- Cramer, D. W., Wilson, E., Stillman, R. J., Berger, M. J., Belisle, S., Schiff, I., Albrecht, B. et al. (1986). The relation of endometriosis to menstrual characteristics, smoking, and exercise, *The Journal of the American Medical Association* 255, 1904–8.
- Dominici, F. and Parmigiani, G. and Wolpert, R.L. and Hasselblad, V., 1999. Meta-analysis of migraine headache treatments: combining information from heterogeneous designs, *Journal of the American Statistical Association* 94, 1999.

- Etminan, M., Sadatsafavi, M., Jafari, S., Doyle-Waters, M., Aminzadeh, K., FitzGerald, J. M., 2009. Acetaminophen Use and the Risk of Asthma in Children and Adults, *Chest* 136, 1316–23.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. *Bayesian Data Analysis*, 2nd Edition, Chapman Hall/CRC:New York.
- Greenland, S., 2003. The impact of prior distributions for uncontrolled confounding and response bias, *Journal of the American Statistical Association* 98, 47–54.
- Greenland S., 2005. Multiple bias modelling for analysis of observational data (with discussion). *Journal of the Royal Statistical Society Series A* 168, 267–306.
- Kresch, A. J., Seifer, D. B., Sachs, L. B., Barrese, I., 1984. Laparoscopy in 100 women with chronic pelvic pain, *Obstetrics and Gynecology* 64, 672–4.
- Higgins, J. P. T., Thompson, S. G., 2002. Quantifying heterogeneity in a meta-analysis, *Statistics in Medicine* 21, 1539–58.
- Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., Jones, D. R., 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS, *Statistics in Medicine* 24, 2401–28.
- Lin, D. Y, Psaty, B. M., Kronmal, R. A, 1998. [Assessing the sensitivity of regression results to unmeasured confounders in observational studies.](#) *Biometrics* 54, 948–963.
- McCandless L. C., Gustafson, P and Levy, A. R, 2007. Bayesian sensitivity analysis for unmeasured confounding in observational studies, *Statistics in Medicine* 26, 2331–47.
- Missmer, S. A., Hankinson, S. E., Spiegelman, D., Barbieri, R. L., Marshall, L. M., Hunter, D. J., 2004. Incidence of laparoscopically confirmed endometriosis by demographic, anthropometric, and lifestyle factors, *American Journal of Epidemiology* 160, 784–96.
- Pocock, S. J., Assmann, S. E., Enos, L. E., Kasten, L. E., 2002. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems, *Statistics in Medicine* 21, 2917–30.

- Proctor, M. L., Farquhar, C. M., 2007. Dysmenorrhoea, *Clinical Evidence* 3, 813.
- R Development Core Team, 2010. A language and environment for statistical computing, R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org>.
- Robinson, J. C., Plichta, S., Weisman, C. S., Nathanson, C. A., Ensminger, M., 1992. Dysmenorrhea and use of oral contraceptives in adolescent women attending a family planning clinic, *American Journal of Obstetrics and Gynecology* 166, 578–83.
- Smith, T.C. and Spiegelhalter, D.J. and Thomas, A., 1995. Bayesian approaches to random-effects meta-analysis: A comparative study, *Statistics in Medicine* 14, 2685–99.
- Sutton, A. J., Higgins, J., 2008. Recent developments in meta-analysis, *Statistics in Medicine*, 27, 625–50.
- Toh, S., Hernández-Díaz, S., 2007. Statins and fracture risk. A systematic review, *Pharmacoepidemiology and Drug Safety* 16, 627–40.
- The Fibrinogen Studies Collaboration (Writing committee: Jackson, D. and White, I.) 2009. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies, *Statistics in Medicine* 28, 1218–37.
- Turner, R. M., Spiegelhalter, D. J., Smith, G., Thompson, S. G., 2009. Bias modelling in evidence synthesis, *Journal of the Royal Statistical Society: Series A* 172, 21–47.
- Vercellini, P., Eskenazi, B., Consonni, D., Somigliana, E., Parazzini, F., Abbiati, A., Fedele, L., 2011. Oral contraceptives and risk of endometriosis: A systematic review and meta-analysis, *Human Reproduction Update* 17, 159–70.

- Welton, N. J., Ades, A. E., Carlin, J. B., Altman, D. G., Sterne, J. A. C., 2009. Models for potentially biased evidence in meta-analysis using empirically based priors, *Journal of the Royal Statistical Society Series A* 172, 119–136.
- Wolpert, R. L., Mengersen, K. L., 2004. Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: Effects of environmental tobacco smoke, *Statistical Science* 19, 450–71.