Understanding and quantifying uncertainty due to multiple biases in meta-analyses of observational studies

Timothy Shin Heng Mak

Department of Epidemiology and Biostatistics
Imperial College London

Thesis submitted for the degree of Doctor of Philosophy

I hereby declare that the work in this thesis is my own contribution. Any published and unpublished work of others has been acknowledged in the text and a list of references is given.

Abstract

There has been considerable interest recently in quantifying uncertainty beyond that due to random error in meta-analyses. This is particularly relevant to meta-analyses of observational studies, since error in estimates from these studies cannot be attributed to a randomization mechanism. Typically, observational studies are also subject to error due to measurement error, non-participation, and incomplete adjustment for confounding. Errors due to these sources are often referred to as bias. To quantify uncertainty due to bias, researchers have proposed using "bias models" and giving subjectively elicited probability distributions to parameters that are not identifiable in the models.

In a typical meta-analysis, probability distributions involving tens of parameters will have to be elicited. At the same time, the resulting estimate and uncertainty interval of the overall (meta-analytic) effect measure will generally be very sensitive to this multi-dimensional subjectively-elicited distribution. To overcome some of the problems associated with the use of such a distribution, I propose an alternative method for eliciting and quantifying uncertainty due to bias. In the method of this thesis, the lower and upper bounds of bias parameters are elicited instead of probability distributions. The most extreme Bayesian posterior inference for the target parameter of interest within the specified bounds is sought through an algorithm. The resulting lower and upper bounds for the target parameter of interest have interpretation of a Robust Bayes analysis.

In this thesis, the method is applied to a meta-analysis of childhood leukaemia and exposure to electromagnetic fields. The method of this thesis was found to produce uncertainty intervals that are generally more conservative in comparison with the standard approach. It is also proposed that the method be used as a tool for sensitivity analysis, and some interesting insight is gained from the childhood leukaemia data.

Contents

List of Figures

List of Tables

Abbreviations

CCM	Cyclic Coordinate Method
EMF	Electromagnetic Fields
EURO	Estimated Uncertainty RegiOn
FE	Fixed-effects (meta-analysis model)
FPCI	Feasible Posterior (95%) Credible Interval
FPI	FPMI and FPCI collectively
FPMI	Feasible Posterior Median Interval
HEIR	Honestly Estimated Ignorance Region
MCRA	Monte Carlo Risk Assessment
MCSA	Monte Carlo Sensitivity Analysis
OOR	Observed Odds Ratio
OR	Odds Ratio
RD	Relative Difference
RE	Random-effects (meta-analysis model)
sens	sensitivity
SES	Socioeconomic Status
spec	specificity
TOR	True Odds Ratio
UI	Uncertainty Interval

Notation

Functions

$Pr(\cdot)$	The probability of ...
$Pr(\cdot\|\cdot)$	The conditional probability of ...
$p(\cdot)$	The probability density function of ...
$F(\cdot)$	The (cumulative) density function of ...
$F^{-1}(\cdot)$	The inverse of the (cumulative) density function
$E(\cdot)$	The expectation of ...
$\text{expit}(x)$	$= \text{logit}^{-1}(x)$
$Lik(\cdot)$	The likelihood of ...
$\text{Odds}(p)$	The odds transform of p, $= \frac{p}{1-p}$
$\text{logit}(p)$	The log-odds transform of p, $= \log\left(\frac{p}{1-p}\right)$
$\text{logit}^{-1}(x)$	The inverse log-odds transform of x, $= \frac{\exp x}{1+\exp x}$

Distributions

$Bin(n, p)$	Binomial distribution with total number n and probability p
$\text{Logistic}(\mu, \tau)$	Logistic distribution with mean μ and scale parameter τ
$N(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$U(a, b)$	Uniform distribution with lower limit a and upper limit b

Special parameters

N	Total number of participants (in the case/control group of a case-control study)
X	The data (generally), often refers to the set of $\{Y_{si}, N_{si}, s = 1 \dots 14, i = 0, 1\}$
Y	Number of exposed participants (in the case/control group of a case-control study)
i (in subscript)	Index denoting case/control status: $i = 1$: case; $i = 0$: control
p	Misclassified probability of exposure (as opposed to π)
s (in subscript)	The index of study in a meta-analysis
γ_s	$= \operatorname{logit} \pi_{s0}$
δ_s	$= \theta_s - \theta$, the difference between the study-specific θ_s and the meta-analytic θ
$\boldsymbol{\eta}$	The vector of parameters that are treated non-probabilistically
θ	The target parameter of inference, typically a log odds ratio
$\boldsymbol{\xi}$	The vector of parameters that are treated probabilistically excluding θ
θ^*	θ subject to bias due to incomplete control of confounding
$\hat{\theta}_L$	The 2.5%-ile of the posterior distribution of θ
$\hat{\theta}_M$	The median of the posterior distribution of θ
$\hat{\theta}_U$	The 97.5%-ile of the posterior distribution of θ
$\hat{\boldsymbol{\Theta}}$	$\hat{\theta}_L$, $\hat{\theta}_M$, $\hat{\theta}_U$ collectively
λ_s	$= \operatorname{logit} \pi_{s0} + \delta_s$
π	True probability of exposure (as opposed to p)
π^*	Probability of exposure subject to non-participation bias

Acknowledgement

First of all, I would like to thank God, who has not only given me the skills and opportunity to embark on such a project, but also gave me numerous inspired ideas over the years, without which this thesis could not have been written. Moreover, He has also blessed me with a variety of people over the years, who not only helped me finish this thesis, but also mature as a person. My sincere thanks go to:

- My supervisors Lesley and Nicky, for giving me the opportunity to work on such a topical and important project, and also devoting so much of your time to it. My thanks also for your helping me focusing my ideas and coming out with a coherent plan for the thesis.
- Cancer Research UK, for giving me such a generous stipend
- Professor Deborah Ashby, for helping me come up with a sound PhD plan during the upgrade exam
- Dr Mark Little, for standing in for Nicky as co-supervisor during her leave
- Ed Chan, for lending me your computer to run my optimization
- Professor Sander Greenland, for a lot of good advice at the beginning of the project
- Professor Berc Rustem, for advice on optimization
- Mr Jay Singh Sangha, for allowing me to stay on at my office after my writeup period finished
- Mom and dad, and Chee, and Shun Heng, for being there and helping me through the hard times
- All my friends, of which there are too many to name, for supporting me emotionally during this time

Chapter 1

Introduction

1.1 Philosophical issues in the quantification of uncertainty

In observational epidemiologic studies, estimates of exposure effects and associations of risk factors and disease are usually accompanied by an estimate of uncertainty. This usually takes the form of a standard error or confidence interval or a p value. The precise interpretation of these measures of uncertainty, however, is generally problematic (Greenland, 1990). Since observational studies involve no randomization in the allocation of exposure or risk factors, such uncertainty measures are not generally regarded as representing uncertainty of causal effects due to randomization as in randomized trials. Greenland (1990) considered two alternative interpretations of these measures of uncertainty: They could represent uncertainty due to random sampling of the study sample from an underlying large population, or they could represent an estimate of "random error" in a "stochastic modelling" exercise. However, he considered both of these interpretations still problematic. The first interpretation is problematic because many if not most observational studies in epidemiology do not involve random sampling from a large population either. Case-control studies, for example, often involve recruiting (or trying to recruit) all available incidence cases within a particular time period in a particular catchment area. Controls may be recruited from other hospital patients, or friends of the patients, or from a register of some kind, often matched collectively on age and sex. Cohort studies often involve the following-up of a convenient sample, in order to maximize sample size. Moreover, this interpretation applies only to the quantification of uncertainty of the *association* between risk factors and disease, even though the ultimate objective of the epidemiologic studies is the elucidation of causes of disease (Cochran and Chambers, 1965; Hill, 1965). The second interpretation of estimates of uncertainty as "random error" in a stochastic modelling exercise is problematic because in order that parameters from the stochastic model be identifiable from the data, often strict and unrealistic assumptions are needed and it is unclear what inference can be made from the model if the assumptions are known to be violated. Moreover, since knowledge of the causation of disease remains incomplete at best, particularly in relation to long

term causes, much uncertainty remains as to the correct model form as well as the correct confounders to include in any particular situation. Thus, in practice measures of uncertainty commonly reported in observational studies gives only very tentative information concerning the true uncertainty underlying the relationship between the risk factor and the disease. Indeed, almost always, it under-represents the true degree of uncertainty.

Our "true" uncertainty over a particular estimate in general goes far beyond what is reported in the confidence interval or p-value. It encompasses our uncertainty over the accuracy of our data in reflecting reality, uncertainty over whether we have collected the most relevant measure of exposure as well as confounding factors, uncertainty over the representativeness of our sample, uncertainty over how well the observed association approximates the causal relationship, and uncertainty over how well our model approximates reality (Greenland, 2005a). It would be ideal to have a quantitative summary of all these uncertainties and not just the standard confidence intervals and p-values. But when we consider going about achieving this, we immediately encounter a variety of problems. Many of these are philosophical. A brief review of some of the issues is given below.

1.1.1 The subjective nature of uncertainty

The widespread use of methods used to quantify uncertainty based on the assumption of randomization or random sampling is in part due to the relative lack of subjective input needed for these methods. In random sampling from a large population (population N), for example, the probability of selection for every individual is uncontroversially $1/N$ and the chance of being selected is to all intents and purposes independent. With most other kinds of uncertainty, however, no such objective standards exist. For example, consider the probability that the observed association between our exposure and disease differs from the true causal relationship by more than a factor of 1.5, or the extent of error resulting from the use of our simplified model to represent reality. Uncertainty of this kind is much more difficult to quantify objectively because it depends how well one knows about the problem at hand.

Many theorists have in fact separated uncertainty into two different kinds (Ayyub and Klir, 2006; Ferson *et al.*, 2004): *Aleotory* uncertainty refers to uncertainty that cannot be reduced or eliminated by increasing our knowledge or further research. It is thus "inherently random". *Epistemic* uncertainty refers to all other kinds of uncertainty — uncertainty that can be reduced through further research or understanding. Uncertainty due to randomization or random sampling would generally belong to the aleotory category.

From a purely philosophical point of view, however, the distinction between aleotory and epistemic uncertainty is unsatisfactory since one can argue that no uncertainty is irreducible and all uncertainty simply reflects a lack of knowledge (e.g. De Finetti, 1974). Nonetheless, in practice, the distinction may help us decide whether a particular kind of uncertainty can be quantified "objectively" or not. If a

particular source of uncertainty is considered to all intents and purposes irreducible beyond a certain limit (like the outcome of a dice throw) by most people, then this "limit" in knowledge can be presented "objectively" as an agreed degree of uncertainty.

1.1.2 Classical, frequentist, and subjective probabilities

In practice, the following distinction also arises between aleotory and epistemic uncertainty: While the use of probability (and probability distributions) as a measure of aleotory uncertainty has been fairly uncontroversial, many have found it unsatisfactory to use probability to quantify epistemic uncertainty (Ferson *et al.*, 2004). One of the fundamental problems of using probability to quantify epistemic uncertainty is that the "classical" definition as well as the "frequentist" definition of probability does not lend itself to describing epistemic uncertainty. In the "classical" definition of probability, a probability of an event E is defined as the ratio of the number of "equally probable cases" for which E is true to the total number of possible cases. Thus, considering drawing from a well-shuffled pack of cards, drawing the ace of spades is considered "equally probable" to drawing any other cards, and hence the probability of drawing the ace of spade is 1/52, given 52 cards in a pack. The "frequentist" definition defines a probability as the "long term relative frequency of an event". Thus the probability of drawing the ace of spade can be estimated as precisely as one wants by drawing repeatedly from the deck (with replacement), although it can never be known exactly. These two definitions of probability are designed to work with uncertainty arising from randomization or random sampling. They do not, however, lend themselves to a clear interpretation when applied to epistemic uncertainty. Consider, for example, trying to quantify the uncertainty over the height of the Statue of Liberty. If I say it has probability 0.7 of being taller than 60m, it does not really mean I have 10 cases in mind, 7 of which has a Statue of Liberty with greater than 60m! When dealing with epistemic uncertainty, therefore, we need to embrace a subjective definition of probability, which is simply a measure on a continuous scale from 0 to 1 reflecting the subjective judgement of uncertainty of the person toward a particular *proposition*, and which also obeys the axioms of probability (see e.g. Good, 1950; Fishburn, 1986; Greenland, 1998). Subjective probabilities, therefore, can be expected to vary from person to person. Although in some situations, it may be possible to define an "ideal" probability judgement (the so-called "logical probability" or "credibility" (see e.g. Good, 1988)), this may not be possible or even desirable in all situations.

Traditionally, however, *frequentist* statisticians reject the whole idea of subjective probabilities altogether and therefore only apply probability methods to situations where probabilities can be interpreted in terms of long term frequency. For example, if θ represents the population mean income which we are uncertain of, and a sample of incomes from the population is obtained, instead of quantifying uncertainty over θ directly, frequentists instead quantify uncertainty over the estimated *confidence interval* of θ, because uncertainty over θ is epistemic, whereas uncertainty over the confidence interval of θ is

aleotory if the sample is random. One problem with the frequentist approach is that we are left with no numerical summary of uncertainty over θ at all, and this can be especially problematic in a decision making scenario, because in these situations, we generally want to assign probabilities to θ, e.g. if we want to find the action which maximize expected utility.

1.1.3 Quantifying subjective uncertainty

If we embrace a subjective interpretation of probability, we have the option of quantifying uncertainty over our parameters of interest directly. The use of Bayes' Theorem then allows us to "update" our probability judgement in the light of data (see e.g. Berger, 1985; Robert, 2007; Gelman *et al.*, 2004). Thus we have a *prior* probability $p(E)$ over some proposition E, reflecting our uncertainty over the truth of E, before we observe some data X. Our probability judgement after seeing X is called the *posterior* probability, and is usually denoted $p(E|X)$. The change from $p(E)$ to $p(E|X)$ may then be seen as a gain in understanding. This *Bayesian* method of quantifying as well as updating uncertainty fits well with decision making, since $p(E|X)$ can be used directly for that purpose. Moreover, De Finetti (1974) proved that representing uncertainty in terms of probability is optimal in that if we were to gamble based on any other numerical measure of uncertainty that is not a probability (in that they do not conform to the probability axioms), it is possible to find a situation where we can be guaranteed to make a loss simply by examining our judgements.

However, apart from the problem that it is by its very nature subjective, there is another major limitation with the use of subjective probability. A probability is by definition a precise number on a real scale, whereas humans do not reason in precise numbers. For example, a precise characterization of a person's probability on a continuous scale involves making an infinite number of probability judgements (e.g. $p(\theta = 0.5)$, $p(\theta = 0.500001)$, $p(\theta = 0.500002)$, etc.). In practice, we either have to discretize the space into a manageable number of categories or else use a mathematical equation to describe what we believe, in which case probabilities are expressed as integrals of a probability distribution function. The problem that this creates escalates when the number of parameters increases, since then we have to consider joint distributions of parameters unless there are good reasons why our subjective judgement of probabilities are independent. These measures mean that we necessarily sacrifice a degree of accuracy and it must be hoped that such loss of accuracies does not affect our results too much.

As a result of these practical problems with probability as a representation of uncertainty, there has been much research on overcoming this problem. The use of hierarchical prior distributions is likely to reduce the inaccuracy arising when assigning probability distributions to a large number of similar parameters (Good, 1980; Berger, 1985). Koopman (1940) introduced some calculus for working with *imprecise probabilities*, which are interval-valued probabilities (i.e., instead of being precise numbers, each probability has a lower and and upper limit). The Robust Bayes approach (Berger, 1984, 1990)

addresses the difficulties in specifying precise prior probability distributions to parameters by allowing the consideration of *classes* of distributions. Other authors avoided probabilistic measures of uncertainty altogether and introduced other measures such as "possibility" (Zadeh, 1978), and proposed axioms and rules for working with these measures. In this thesis, I examine the use of a combined probabilistic/non-probabilistic approach to quantifying uncertainty, which is also a kind of Robust Bayes analysis.

1.2 Probabilistic and non-probabilistic approaches to quantifying uncertainty

Because quantifying uncertainty with the assumption of some underlying randomization or random sampling mechanisms is often not satisfactory, increasingly researchers are looking toward statistical techniques that try to quantify other sources of errors too. In the epidemiologic literature, errors in estimation that result from model assumptions not being met in reality is often referred to as *bias*, since estimators based on the wrong assumptions are believed to produce *biased* estimates of the effect of interest (Greenland, 2005a; Greenland and Lash, 2008; Lash *et al.*, 2009). Statistical models or techniques that aim to quantify or estimate the effect of bias on the target estimate are often referred to as bias models. Bias models have been used to quantify bias due to measurement error (e.g. Tenenbein, 1970; Hui and Walter, 1980; Selen, 1986; Espeland and Hui, 1987; Armstrong *et al.*, 1989; Greenland, 1989; Rosner *et al.*, 1989; Carroll *et al.*, 1995; Paulino *et al.*, 2003; Greenland, 2008), unmeasured confounding (e.g. Rosenbaum and Rubin, 1983b; Robins, 1988; Angrist *et al.*, 1996; Balke and Pearl, 1997; Lin *et al.*, 1998; Tian and Pearl, 2000; MacLehose *et al.*, 2005; Sturmer *et al.*, 2005; Gustafson *et al.*, 2010), and non-participation bias (Hansen and Hurwitz, 1946; Politz and Simmons, 1949; Little, 1982, 1993; Copas and Li, 1997; Geneletti *et al.*, 2009). Generally speaking these techniques involve expanding standard statistical models to involve additional parameters that characterize the biases (c.f. Gustafson, 2005). The expansion of the model will generally cause the parameter of interest to be unidentifiable from the existing data (i.e. it will not be possible to find a consistent estimator of the parameter of interest). However, this can be addressed in one of several ways. One strategy is to introduce additional data to help us identify these parameters. Two such techniques have received particular attention in recent years: The Hui-Walter paradigm (Hui and Walter, 1980) allows the use of more than one independent measurements of an exposure to identify bias due to misclassification; the use of "instrumental variables" (variables that are related to the exposure but are otherwise unlikely to be associated with the disease in any other ways) can be used to reduce bias due to confounding from unmeasured confounders (Angrist *et al.*, 1996). However, these models generally still rely on certain assumptions, which may or may not be met in practice and there is often no way of checking. For example, the Hui-Walter paradigm requires that misclassification errors are independent between the two or more exposure measurements, as well as

the assumption that misclassification probabilities do not differ between cases and controls (the diseased and the healthy). The use of "instrumental variables" relies on the assumption that the instrumental variable is not otherwise related to the disease other than through the exposure. In any case, these techniques are not applicable in the absence of "additional data". An alternative strategy to deal with these expanded models with unidentifiable parameters is to carry out sensitivity analyses (Rosenbaum, 2005). In sensitivity analyses, some of the parameters of the models are given *assumed* values, in order that the parameter of interest can be identified by the data. These assumed values are varied and the researcher examines how sensitive the estimates of the target parameter are to changes in the assumed values. Sensitivity analyses have a relatively long history of application in epidemiology. One commonly cited early example is the investigation of Cornfield *et al.* (1959) to find out if the observed association between lung cancer and smoking might be explained by an unknown binary confounder. However, one major limitation of traditional sensitivity analyses is that it is difficult to deal with a model with many unidentifiable parameters, since in such a case there are very many scenarios (assumed values) that need to be investigated. Even if it is computationally feasible to obtain estimates for all of the scenarios, it may be difficult to derive a summary measure of uncertainty from all the different estimates. Several strategies can be adopted to overcome these limitations in traditional sensitivity analyses. Here I describe a non-probabilistic approach and a probabilistic approach in the following two sections, before discussing their relative advantages. The strategy to be adopted in this thesis will involve ideas from both approaches.

1.2.1 A non-probabilistic approach to draw inference in models with unidentifiable parameters

To illustrate these different approaches, I now introduce some notation. Let us assume that we are interested in estimating the parameter θ, which is related to the data X, possibly through a number of other parameters $\boldsymbol{\eta} = \{\eta_1, \eta_2, \ldots, \eta_n\}$, i.e.:

$$X \sim p(\theta, \boldsymbol{\eta}) \tag{1.1}$$

where $p(\theta, \boldsymbol{\eta})$ denotes an arbitrary probability distribution involving parameters $\theta, \boldsymbol{\eta}$. If θ is identifiable, it means there is a one-to-one correspondence between the function $p(\theta, \boldsymbol{\eta})$ and the parameters $(\theta, \boldsymbol{\eta})$, and hence observing the distribution of X (which gives us an estimate of $p(\theta, \boldsymbol{\eta})$) allows us to make inference to θ. However, frequently, bias models are not identifiable, although it is often the case that if $\boldsymbol{\eta}$ or a subset of $\boldsymbol{\eta}$ are known then θ becomes identifiable given X. If this is the case, we can form an estimator of θ of the form:

$$\hat{\theta} = f(X, \boldsymbol{\eta}_U) \tag{1.2}$$

where $\boldsymbol{\eta}_U \subseteq \boldsymbol{\eta}$. One non-probabilistic approach is to define ranges $\mathcal{E}$ for our unknown parameters $\boldsymbol{\eta}_U$ and observe the range of possible $\hat{\theta}$ when $\boldsymbol{\eta}_U$ takes values within this range, i.e., we want to find:

$$[\min_{\boldsymbol{\eta}_U \in \mathcal{E}} \hat{\theta}, \max_{\boldsymbol{\eta}_U \in \mathcal{E}} \hat{\theta}] \tag{1.3}$$

Intervals derived this way have been called Honestly Estimated Ignorance Regions (HEIRs) by Vansteelandt *et al.* (2006). This approach has been used for estimating the prevalence of disease or the proportion of a particular response from surveys with missing data (Hansen and Hurwitz, 1946; Birnbaum and Sirken, 1950). It has also been applied to the causal inference literature to obtain bounds to causal risk differences, causal risk ratios, and causal odds ratios (Balke and Pearl, 1997; Chiba *et al.*, 2007; MacLehose *et al.*, 2005; VanderWeele, 2008; Kuroki *et al.*, 2010).

1.2.2 A probabilistic approach to draw inference in models with unidentifiable parameters

A commonly used probabilistic approach to requires us to specify a probability distribution $p(\boldsymbol{\eta}_U)$ to $\boldsymbol{\eta}_U$. Within this approach, there are two variants: Monte Carlo Sensitivity Analysis (MCSA) and Bayesian. In MCSA, we sample from $p(\boldsymbol{\eta}_U)$ and for each sample $\boldsymbol{\eta}_{U,i}$, we obtain $\hat{\theta}_i = f(X, \boldsymbol{\eta}_{U,i})$. The mean or median of $\hat{\theta}_i$ over the sample can then form an estimate for θ, and an uncertainty interval obtained by, e.g., the 2.5% and 97.5%-ile. MCSA has been used mainly in the risk assessment literature, where it is also referred to as Monte Carlo Risk Assessment (MCRA) (Greenland, 2001). In general, the uncertainty interval derived from MCSA may be difficult to interpret. Unlike in the non-probabilistic approach, they certainly do not correspond to limits of estimates. They also do not have the interpretation of posterior intervals from a Bayesian analysis, although with a minor modification, they may approximate them in certain situations (Greenland, 2005a).

For the Bayesian approach, we not only need to assign a prior distribution to $\boldsymbol{\eta}_U$, but to the entire set $\boldsymbol{\eta}$ as well as to θ, possibly as a joint distribution: $p(\theta, \boldsymbol{\eta})$. Given the likelihood (i.e. $p(X|\theta, \boldsymbol{\eta})$), $p(\theta, \boldsymbol{\eta})$ implies the posterior distribution $p(\theta|X)$ through Bayes' theorem, and an estimate of θ can be obtained by noting its median or mean and an uncertainty interval obtained through the 2.5%, 97.5%-ile. Examples of the Bayesian approach to inference in unidentifiable models in the literature include e.g. Gustafson *et al.* (2001), McCandless *et al.* (2007), Stamey *et al.* (2008).

1.3 Relative merits of the probabilistic and non-probabilistic approaches

The probabilistic and the non-probabilistic approaches as described in the previous section each have their merits and limitations, which are discussed below.

1.3.1 Advantages of the non-probabilistic over the probabilistic approach

1. A natural definition for a "conservativeness" of uncertainty intervals From the very definition of the lower and upper bounds of $\hat{\theta}$ given in (1.3), it is clear that if we have parameter space $\mathcal{F}$ for $\boldsymbol{\eta}_U$ that is a subset of $\mathcal{E}$, i.e. $\mathcal{F} \subseteq \mathcal{E}$, then necessarily, the lower limit of the possible $\hat{\theta}$ within $\mathcal{F}$ must be greater than or equal to that within $\mathcal{E}$, and the upper limit under $\mathcal{F}$ must be less than or equal to that under $\mathcal{E}$. Hence, we have a natural definition for the "conservativeness" of the uncertainty intervals (UI): If $\mathcal{F}$ is a subset of $\mathcal{E}$, then the UI derived using $\mathcal{F}$ as limits for $\boldsymbol{\eta}_U$ is less conservative than those derived with $\mathcal{E}$ as limits. On the other hand, defining conservativeness of probability distributions $p(\boldsymbol{\eta}_U)$ is more difficult since the relationship between $p(\boldsymbol{\eta}_U)$ and the posterior median/credible interval limits of θ can be complicated. The possibility to compare "conservativeness" of intervals simply by comparing the limits of parameter space in the non-probabilistic approach helps resolve one of the fundamental problems in inference with subjective inputs. A reader who disagrees with the limits given to $\boldsymbol{\eta}_U$ used in a particular analysis can check if his/her limits are within those used by the researcher. If so, he/she can be sure that his/her UI will also fall within the researcher's reported UI.

2. Limits are easier to elicit than probability distributions For a single uncertain parameter, eliciting upper and lower limits require the elicitation of only 2 numbers. A probability distribution, on the other hand, is an infinite dimension object, and in practice can only be elicited either by discretizing or by adoption of a mathematical form. This can result in inaccuracies of the posterior distribution (Berger, 1984). The problem is made more complicated when the number of unknown parameters increases. Often, we need to make the assumption that our probabilistic judgement on the parameters are independent of one another, simply because finding a suitable distribution to model the dependence among several parameters can be extremely difficult. Eliciting limits for multiple parameters, on the other hand, presents few additional difficulties above the one parameter case.

3. It does not suffer from complications due to different possibilities of parameterization In many statistical models, there are multiple ways of assigning prior distributions to the parameters.

For example, in a simple case-control study model, we have:

$$Y_i \sim Bin(N_i, \pi_i) \qquad i = 0, 1 \tag{1.4}$$

$$\text{logit}\, \pi_1 = \text{logit}\, \pi_0 + \beta \tag{1.5}$$

where Y_1 and N_1 denote the number of exposed and the total number of participants among the cases and Y_0 and N_0 the same among the controls. If we want to assign prior distributions to the parameters, it is not always clear whether we should assign to π_0 and π_1, or π_0 and β, or π_1 and β. This has caused concern in some Bayesian models in the literature (e.g. Rice, 2005). Moreover, sometimes we may have prior information on all three parameters π_0, π_1, and β. However, it may be difficult to specify a prior distribution that satisfy all of our beliefs. For example, say we want π_0 and π_1 to have a marginal distribution of Uniform$(0, 0.2)$, and β to have a marginal distribution of Normal$(0, 1)$. Assigning a Uniform$(0, 0.2)$ prior to π_0 and an independent normal distribution to β will not lead to a prior of Uniform$(0, 0.2)$ for π_1. Indeed, finding a joint distribution that does have these marginals can be a challenging mathematical problem.

The non-probabilistic approach, however, avoids these problems. It is no contradiction to require, for example, that $0 \leq \pi_0 \leq 0.2$, $0 \leq \pi_1 \leq 0.2$, and $-2 \leq \beta \leq 2$.

1.3.2 Disadvantages of the non-probabilistic approach compared to the probabilistic approach

1. No point estimate available Whereas in probabilistic approaches, one can use the median or mean of the distribution of $\hat{\theta}$ (in MCSA) or the posterior distribution $p(\theta|X)$ (in Bayesian analysis) as a point estimate of θ, there is no intuitive point estimate when we use the non-probabilistic approach.

2. Difficulties in eliciting limits In some cases, bounds for parameters are fairly uncontroversial. For example, bounds for prevalences and proportions must be between 0 and 1. However, the use of such bounds often lead to uncertainty bounds for $\hat{\theta}$ that are meaninglessly wide. In practice, we need tighter bounds that are based on expert knowledge. Still, expert knowledge on limits of the possible range of parameter values might be hard to elicit, because elicitation on tails of distributions has proved difficult and unreliable (Garthwaite *et al.*, 2005).

3. Intervals tend to be overly conservative The non-probabilistic approach examines the *extreme* possible inference. Often, these can be highly unrealistic situations, and thus lead to very wide and uninformative uncertainty intervals. This problem exacerbates when there are more parameters, because unless there are obvious deterministic relations between parameters, it becomes more and more unlikely that parameters are all at their extreme at the same time.

1.3.3 Combining probabilistic and non-probabilistic approaches

Given the trade-offs between probabilistic and non-probabilistic approaches to quantifying uncertainty over unknown parameters, it may not be surprising that many have combined both approaches in tackling a particular problem. In the simplest case, we might combine a non-probabilistic approach to quantifying uncertain parameters in a model with a probabilistic approach to quantifying uncertainty due to random sampling. In our notation $\hat{\theta}$ is merely an estimate of θ. The remaining error between $\hat{\theta}$ and θ is assumed to be due to sampling error:

$$\hat{\theta} = \theta + \delta(\boldsymbol{\eta}) \tag{1.6}$$

$$\delta(\boldsymbol{\eta}) \sim f(\mu(\boldsymbol{\eta}), \sigma^2(\boldsymbol{\eta})) \tag{1.7}$$

where $\delta(\boldsymbol{\eta})$ is used to denote the deviance between θ and its estimate, and $f(a, b)$ denotes a general distribution with mean a and variance b. Often, given a particular set of unknown parameter values $\boldsymbol{\eta}_U$, we can derive not only $\hat{\theta}$, but also a confidence interval $(\hat{\theta}_L(\boldsymbol{\eta}_U), \hat{\theta}_U(\boldsymbol{\eta}_U))$ for θ. Given this confidence interval, we can quantify the uncertainty in θ due to both uncertainty in $\boldsymbol{\eta}$ and random sampling $\delta(\boldsymbol{\eta})$. Hence, we derive an overall uncertainty interval as:

$$[\min_{\boldsymbol{\eta}_U \in \mathcal{E}} \hat{\theta}_L, \max_{\boldsymbol{\eta}_U \in \mathcal{E}} \hat{\theta}_U] \tag{1.8}$$

(Note that the difference between (1.8) and (1.3) is in the subscript of $\hat{\theta}$.) Vansteelandt *et al.* (2006) provided some theories on coverage for uncertainty intervals derived in this way and called them Estimated Uncertainty RegiOns (EUROs). It is possible to use the same principle to construct uncertainty for estimates for which some parameters within $\boldsymbol{\eta}_U$ are quantified via non-probabilistic means and some parameters are quantified by probabilistic means.

There are other ways where non-probabilistic and probabilistic approaches can be combined in quantifying uncertainty. I mentioned in the above subsection that one of the difficulties in the non-probabilistic approach may be the specifying of limits to parameters' ranges. It has been proposed in the literature that this problem be overcome by assigning probabilistic distributions to the limits (Sahinidis, 2004). Thus, instead of finding limits to $\hat{\theta}$ subject to the straightforward constraint that $\boldsymbol{\eta}_U \in \mathcal{E}$, the constraint becomes $\Pr((\boldsymbol{\eta}_U \in \mathcal{E}) > 1 - \epsilon)$, i.e. the probability that the constraint is violated must be less than a certain proportion ϵ. There have been few, if any, applications of this kind of approach in epidemiology, although techniques for finding solutions to these problems are developed in stochastic programming literature (Prekopa, 1995).

1.4 Quantifying uncertainty in meta-analyses

The subject of this thesis is the quantification of uncertainty in meta-analyses of observational studies. Meta-analyses are studies which combine estimates collected across different studies with the aim of synthesizing information and producing a summary estimate of a particular effect of interest (Deeks *et al.*, 2001; Borenstein *et al.*, 2009). In the backdrop of the Evidence Based Practice movement, meta-analysis and quantitative synthesis have been promoted as the best source of information in informing policies (Greenhalgh, 1997; Cordray and Morphy, 2009). It is therefore arguably more important that uncertainty in estimates be quantified reliably in meta-analyses than in single studies. Traditionally, however, methods in meta-analyses make the same assumptions as in most single studies — that uncertainty is due entirely to the individual studies' "random" error. Furthermore, the extent of this "random" error is generally quantified by the reported standard error or confidence interval associated with the study-specific estimate. In the case of a meta-analysis of randomized controlled trials, this is justified to the extent that "random" error reflects error due to randomization of the individual trials. Although meta-analyses of observational studies are widespread, the statistical underpinning of modelling study-specific error as "random" is far weaker (Stroup and Thacker, 2005). The standard practice of using standard errors and confidence intervals for quantifying uncertainty is usually far from satisfactory, because of the numerous "biases" that afflict observational studies, as noted in the previous section (Greenland, 2005a).

Although there have been numerous proposals to adjust for biases in the epidemiologic literature (see section 1.2), it was not until relatively recently that systematic adjustment for biases was proposed for use in a meta-analysis (Wolpert and Mengersen, 2004; Greenland, 2005a; Greenland and Kheifets, 2006; Turner *et al.*, 2009; Welton *et al.*, 2009; Thompson *et al.*, 2011).[1] These meta-analyses combine study-specific estimates or raw data (in the case where the estimates were derived from simple 2-by-2 cross tabulation of counts) with subjectively elicited, study-specific, information on the likely extent of biases, to obtain "bias-adjusted" estimates and uncertainty intervals for a meta-analytic (overall) estimate of the effect of interest.[2] A suitable model is needed to combine the two.

In the choice of such a model, there is generally a trade-off between model complexity and availability of information. To keep the model simple, we can assume all biases are additive or multiplicative on the

[1] Previously, biases in meta-analyses were mainly addressed qualitatively, although sometimes quantitatively summarized as a "quality score". However, the use of this score in deriving statistical estimates has been severely criticized (Greenland, 1994)

[2] In the literature, attempts to quantitatively combine information from a variety of sources for the purpose of estimating a particular effect of interest is sometimes called evidence synthesis (Ades and Sutton, 2006; Turner *et al.*, 2009). However, here I distinguish between the attempts to estimate an effect of interest through a "chain of evidence" (e.g. Ades, 2003; Molitor *et al.*, 2009) and those that do so through a meta-analysis. The former estimates a parameter θ by formulating it as a function of other parameters, e.g. $\theta = f(a, b, c)$, and different sources of information may contribute to estimation by providing estimates for one or more components of the function, e.g. Study 1 is used for estimating a and study 2 for b and c, etc. This differs from a meta-analysis in which all studies estimate θ, though with different biases and error.

parameter of interest. For example, if θ_s is the parameter of interest for study s, we assume our biased estimate θ_s^{biased} is:

$$\theta_s^{\text{biased}} = \theta_s + c_s \tag{1.9}$$

or

$$\theta_s^{\text{biased}} = k_s \theta_s \tag{1.10}$$

and ask experts to give their subjective estimates of k_s or c_s either as interval estimates or as a probability distribution. When considering more than one source of bias, we simply chain together these different biases to form a more elaborate equation such as:

$$\theta_s^{\text{biased}} = k_s^{(K)}(\cdots(k_s^{(2)}(k_s^{(1)}\theta_s + c_s^{(11)} + c_s^{(12)} + \cdots) + c_s^{(21)} + \cdots)\cdots) \tag{1.11}$$

This is essentially the approach taken by Turner *et al.* (2009) and Thompson *et al.* (2011). The particular difficulty of this approach, however, is in the elicitation of suitable values or distributions for the parameters c_s and k_s — there can be very little information to go by. In rare circumstances, however, one can employ estimates of c_s based on so-called meta-epidemiologic studies (Welton *et al.*, 2009), which aimed to estimate biases by pooling over a large number of meta-analyses.

On the other hand, one can employ more elaborate *bias models* relating the true parameter value to the data (and/or to the biased parameters). These are used in this thesis, as well as in Wolpert and Mengersen (2004), Greenland (2005a) and Greenland and Kheifets (2006). One example is the relationship between the true odds ratio (TOR) and the observed odds ratio (OOR) in a case-control study with exposure misclassification, which are related by the formulae:

$$Y_i \sim Bin(N_i, p_i) \qquad i = 0, 1 \tag{1.12}$$

$$p_i = \pi_i sens_i + (1 - \pi_i)(1 - spec_i) \tag{1.13}$$

$$TOR = \frac{\pi_1(1 - \pi_0)}{\pi_0(1 - \pi_1)} \tag{1.14}$$

$$OOR = \frac{p_1(1 - p_0)}{p_0(1 - p_1)} \tag{1.15}$$

Here, Y_1 and N_1 denote the number of exposed and the total number of cases and Y_0 and N_0 denote the number of exposed and total number of controls. In this approach, we require experts to elicit intervals or prior distributions for the parameters $sens_i$ and $spec_i$ (denoting the *sensitivity* and *specificity* respectively, see equations 4.3 and 4.4 of this thesis). Based on these elicited distributions/values, we then derive an interval/posterior distribution of the TOR, or alternatively, the bias factor TOR/OOR. The advantage of this approach is that there is often more information to go by for such specific parameters as *sens* and *spec*. However, the model also becomes more complicated. What is a single (exposure

misclassification) bias is now modeled by 6 parameters ($\pi_1, \pi_0, sens_1, sens_0, spec_1, spec_0$). Specification of a (joint) prior distribution true to the experts' belief can be a considerable challenge. Moreover, this approach has only been applied in situations where the outcome is categorical and where the analyses do not depend critically on the adjustment of confounders (Wolpert and Mengersen, 2004; Greenland, 2005a; Greenland and Kheifets, 2006). While a proposal is made in chapter 6 of this thesis to extend its applicability to the case where confounders are adjusted for, it appears more research is needed before researchers can confidently apply these techniques to more general meta-analysis situations.

1.4.1 Critique of existing attempts to quantifying uncertainty due to biases in meta-analyses

So far, all previous attempts to incorporate bias adjustment in meta-analysis (Wolpert and Mengersen, 2004; Greenland, 2005a; Greenland and Kheifets, 2006; Welton *et al.*, 2009; Turner *et al.*, 2009; Thompson *et al.*, 2011) are essentially probabilistic and subjectivist, that is, they have used prior distributions to quantify subjective uncertainty over various perceived biases. This means that one of the main problems is that if the readers do not agree with the prior distributions that are used for the bias parameters, it is unsure what conclusions they can draw from the results. The authors in the above papers recognized this potential problem and did not recommend that one presents the results from only one set of prior distributions. Turner *et al.* reported the results from prior distributions elicited from several experts and both Greenland (2005a) and Greenland and Kheifets (2006) performed tabular sensitivity analyses (TSA) to see how sensitive results are to disturbance in some of the parameters. Nonetheless, given that the number of bias parameters in both cases can easily be over 100, it is understandable that many readers can still be left unsatisfied by the very limited number of scenarios that are covered in the sensitivity analyses, as some of the discussants appeared to be in Greenland's (2005a) paper.

Several authors have suggested that disagreement over prior distributions are likely to be less material than disagreement over which exact value the unknown parameters should take (Good, 1962; Berger, 1985; O'Hagan and Oakley, 2004). This, however, does not mean that disagreement over prior distributions are always immaterial. For example, the standard analysis in Greenland and Kheifets (2006) leads to a posterior median and 95% credible interval for the average study-specific odds ratio of 1.55 (0.40, 4.79).[3] This is based on a $N(0, 0.5)$ prior distribution for the parameter α_{TY}, which is the log odds ratio between true exposure and disease among those who are observed to be unexposed. If this prior distribution instead has double the variance, the posterior median and 95% credible interval becomes 2.46 (0.38, 9.16). This is a result of the change of only one out of over 100 parameters used in

[3]This posterior odds ratio is based on a reanalysis of the data using MCMC in WinBUGS 1.4 (Lunn *et al.*, 2000), as are all other results in this section. Greenland and Kheifets in fact used an approximation strategy for the estimation of this posterior distribution (pers. comm.). Thus, their reported posterior odds ratio and 95% posterior interval for the standard analysis was given as 2.8 (0.97, 8.4) in Table 3. The difference between their results and what I computed is believed to be due to inaccuracies in the approximation strategy.

this study.

Moreover, sometimes it may be the case that while differences in prior distributions lead to little difference in results when only a single study was considered, the same differences, when applied to all studies in a meta-analysis, leads to considerably greater difference in the combined (pooled) estimate than when only one study is considered. Again, considering the Greenland and Kheifets (2006) dataset, the posterior median and 95% credible interval of the study-specific odds ratio for the Coghill *et al.* (1996) study was 0.85 (0.19, 3.69), if this study was considered independently of other studies, using the standard setup for the bias parameters. When the mean of one of its parameters (α_{TX}) (which represents the ratio of true-positive odds to false-positive odds[4]), is changed from $\ln 16$ to $\ln 4$ (indicating a belief for greater possible misclassification error), the posterior distribution for the odds ratio for this study did not change materially – it still had median 0.85 and similar 95% credible interval (0.20, 3.61). However, in a meta-analysis of 15 studies, changing the mean of the prior distribution for α_{TX} from $\ln 16$ to $\ln 4$ for all studies simultaneously changes the posterior median and 95% credible interval of the average study-specific odds ratio from 1.55 (0.40, 4.79) to 1.18 (0.31, 4.31), reflecting a noticeable movement of the distribution towards 0.

The examples above considered disagreement over prior distribution. Furthermore, there can be disagreement over modelling strategies. I noted above that there is a trade-off between model simplicity and the availability of information. In the literature, many models have been developed for adjustment of biases due to exposure misclassification, non-participation bias, and incomplete control of confounding. Many of these models will also be applicable in a meta-analysis setting. It may not always be clear which model enables us to best represent our uncertainty. Consider, again, the Greenland and Kheifets (2006) model. The model the authors used for exposure misclassification was a loglinear model, described also in Greenland (2009b), and parameterized in terms of true-positive to false positive odds, ratio of case vs. control receiver operator characteristics, etc. Another possible model for exposure misclassification is described in chapter 4 of this thesis, and is parameterized in terms of *sensitivity* and *specificity* of the exposure measure in the cases and controls. Some researchers may prefer one model over another, but it does not mean that their preferred method is necessarily the most suitable.

Another problem concerns the correlation of the bias parameters in such a meta-analysis. When so many bias parameters are involved, it is natural to want to simplify analyses and assume that the prior distribution of the parameters are uncorrelated. However, bias judgement is hardly uncorrelated. If one's judgement for a particular bias is wrong in one direction for one study, it is likely to be wrong in the same direction for another study also. Furthermore, treating bias parameters as independent means that our error in judgement is essentially "random", and has all the advantages conferred by true randomization. For example, it means that if we make enough estimates, the average of our estimates will have no uncertainty left. This clearly is not reasonable in the context of subjectively elicited bias estimates.

[4] i.e. Odds(Truly exposed|Observed exposed)/Odds(Truly exposed|Observed unexposed)

Greenland (2005a) and Greenland and Kheifets (2006) recognized this and their bias parameters are correlated, to different extent, across studies. However, in general correlation of judgement is difficult to elicit. Again, results can be sensitive to the correlation structure that is assumed, leading to more problems in interpreting the results.

1.5 Aim and structure of thesis

The aim of this thesis is to explore the use of a new method – a combined probabilistic/non-probabilistic approach to quantifying uncertainty in a meta-analytic setting. In this approach, instead of giving bias parameters subjectively elicited probability distribution as in the previous approaches, we instead give these parameters *feasible ranges*, and we seek the most extreme inference for our target parameter θ that is possible within these subjectively specified ranges. As noted before, a foreseeable advantage of this approach is that there is a natural definition of conservativeness – feasible regions that are wider are always more "conservative" in that they always lead to an uncertainty region for θ that is wider. Furthermore, there is no need to specify a joint distribution for the bias parameters. It is shown that extreme inference (as defined in the next chapter) is always achieved with prior distributions with point mass at specific values of the bias parameters. In addition to being an alternative method for quantifying uncertainty, the method can also be exploited as a useful tool for carrying out sensitivity analyses, particularly in situations where a large number of unknown parameters are involved. Traditional sensitivity analyses require us to allocate precise values to the unknown parameters, and therefore only a limited number of scenarios can be investigated and reported at any one time. The method introduced in this thesis allows us to specify feasible *ranges* to the unknown parameters, thereby investigating the sensitivity of the analyses to a wide range of parameter values at the same time.

Throughout this thesis, I illustrate the techniques involved in this method and explore its strengths and weaknesses by applying it to Greenland's (2005a) meta-analysis of 14 case-control studies, the data of which are available directly from the paper. I also consider several different possible bias models for each type of bias considered, although I have chosen to focus on those that enable the best use of information that is typically available in a meta-analysis setting, which are not necessarily the ones that Greenland employed.

The structure of the thesis is as follows. Chapter 2 gives the theory and the technical details of the methods and algorithms used. Chapter 3 applies the method to a meta-analysis with simple adjustment of bias. Chapter 4 applies the method to a meta-analysis with exposure misclassification modelling. Chapter 5 considers non-participation bias. Chapter 6 considers bias due to incomplete control of confounding. Chapter 7 considers a model that models exposure misclassification, non-participation bias, and incomplete control of confounding at the same time. The findings are discussed in chapter 8 together with recommendation for future research work.

Summary:

- Although observational studies typically do not involve randomization or random sampling, standard statistical practice is to quantify uncertainty in effect estimates using techniques developed for randomized experiments. This is partly because uncertainty in such estimates is inherently subjective and the use of subjective probabilities for the quantification of uncertainty has been controversial.

- Uncertainty can be separated into two types – *aleatory* and *epistemic*, where the former refers to uncertainty that can generally be considered random and the latter to all other kinds of uncertainty. The use of probabilistic methods is uncontroversial for the quantification of aleatory uncertainty, but not so for epistemic uncertainty.

- This chapter discusses a simple non-probabilistic method and a probabilistic method for quantifying uncertainty and their respective advantages over one another. It then outlines a combined probabilistic/non-probabilistic approach which is an amalgam of the two.

- There has been increasing interest in quantifying study-specific biases in meta-analyses. Probabilistic approaches have been used in all such meta-analyses. These meta-analyses involve an enormous amount of subjective input in terms of prior distributions of bias parameters, and inference can be very sensitive to these prior distributions.

- The aim of the study is to propose, apply, and examine the use of a combined non-probabilistic/probabilistic approach to quantifying uncertainty due to biases in meta-analysis.

Chapter 2

Introduction to the methods used for quantifying bias in this thesis

In the standard approach to meta-analysis, we assume that the study-specific estimate of effect $\hat{\theta}_s$ is distributed with mean θ_s and variance $\hat{\sigma}_s^2$, with the goal of estimating θ, where

$$\theta_s = \theta \tag{2.1}$$

in a fixed-effects model and

$$\theta_s \sim p(\theta) \tag{2.2}$$

in a random-effects model, with $p(\theta)$ denoting an arbitrary distribution with mean θ (e.g. Petitti, 2000; Deeks *et al.*, 2001; Borenstein *et al.*, 2009). While this formulation is general in that it does not matter what effect measure is being used and whether the effect estimate has been adjusted for confounders, it is arguably sub-optimal in the case of a meta-analysis of studies with binary outcomes and exposure with no adjustment for confounders (Stijnen *et al.*, 2010), since in such a scenario, the raw data are typically available, and we can model them using the Binomial distribution. For a meta-analysis of case-control studies without adjustment for confounders, the (retrospective) Binomial model would be:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \tag{2.3}$$

$$\operatorname{logit} \pi_{s1} = \operatorname{logit} \pi_{s0} + \theta_s \tag{2.4}$$

where Y_{s1} denote the number of exposed cases and Y_{s0} the number of exposed controls in study s. N_{si} represent the total number of cases and controls in study s, for $i = 1, 0$. π_{si} is the corresponding prevalence of exposure in the population from which the participants are sampled and $\exp \theta_s$ is the odds ratio of disease given exposure. In this thesis, Model (2.3-2.4) serves as the basic meta-analysis model upon which extensions are made to allow for study-specific biases. Although the model in this form

does not appear applicable in a meta-analysis of binary data with adjustment of measured covariates, a proposal is made in chapter 6 to overcome this limitation.

Throughout the thesis, the way various study-specific biases are accounted for is by replacing one of the parameters in (2.3) and (2.4) with its "biased version", which is linked to the "true version" by some deterministic formula. For example, when accounting for exposure misclassification, I replace π_{si} with p_{si} in (2.3), with $p_{si} = f(\pi_{si}, \boldsymbol{\eta}^{\text{miscl}})$, where $\boldsymbol{\eta}^{\text{miscl}}$ are the bias parameters associated with exposure misclassification. When accounting for non-participation bias, I replace π_{si} with π^*_{si} in (2.3), where $\pi^*_{si} = f(\pi_{si}, \boldsymbol{\eta}^{\text{sel}})$. When accounting for bias due to incomplete control of confounding, I replace θ_s in (2.4) with θ^*_s, where $\theta^*_s = f(\theta_s, \boldsymbol{\eta}^{\text{conf}})$. The details of the formulae used are given in chapters 4, 5, and 6. In addition, I shall be giving feasible ranges (or regions) to the parameters $\boldsymbol{\eta}^{\text{miscl}}$, $\boldsymbol{\eta}^{\text{sel}}$, $\boldsymbol{\eta}^{\text{conf}}$, and seek the maximum and minimum estimate or uncertainty range for θ given these ranges.

2.1 A Robust Bayesian approach to quantify uncertainty

The approach of this thesis is thus similar to Vansteelandt *et al.* (2006). Vansteelandt *et al.* defined the Honestly Estimated Ignorance Region (HEIR) as

$$[\min_{\boldsymbol{\eta}\in\mathcal{E}}\tilde{\theta}, \max_{\boldsymbol{\eta}\in\mathcal{E}}\tilde{\theta}] \tag{2.5}$$

where $\tilde{\theta}$ represents an estimate of θ and $\mathcal{E}$ represents the feasible region of $\boldsymbol{\eta}$. They also defined the Estimated Uncertainty RegiOn (EURO) as:

$$[\min_{\boldsymbol{\eta}\in\mathcal{E}}\tilde{\theta}_L, \max_{\boldsymbol{\eta}\in\mathcal{E}}\tilde{\theta}_U] \tag{2.6}$$

where $\tilde{\theta}_L$ and $\tilde{\theta}_U$ represent limits of particular confidence intervals. The approach of this thesis uses a Bayesian estimate of θ and a Bayesian posterior interval limits for θ_L and θ_U instead of confidence intervals limits. This is because in addition to specifying a feasible region for the bias parameters $\boldsymbol{\eta}$, in many situations it would be useful to be able to specify subjectively-elicited prior distributions to θ as well as other nuisance parameters which are not part of $\boldsymbol{\eta}$. An example of this is given in section 4.2.1. Furthermore, Bayesian procedures can often be made to have good frequentist properties such as consistency of estimates and good coverage of posterior intervals (Bayarri and Berger, 2004). In the rest of this thesis, I denote by $\hat{\theta}_M$ the median of the posterior distribution of θ, and $\hat{\theta}_L$ and $\hat{\theta}_U$ are the 2.5% and 97.5% limits of the posterior distribution. I also define the feasible range of the posterior median as:

$$\text{Feasible Posterior Median Interval} = \text{FPMI} = [\min \hat{\theta}_M, \max \hat{\theta}_M] \tag{2.7}$$

and the feasible range of the posterior 95% credible interval as:

$$\text{Feasible Posterior Credible Interval} = \text{FPCI} = [\min \hat{\theta}_L, \max \hat{\theta}_U] \tag{2.8}$$

Moreover, I use $\hat{\boldsymbol{\Theta}}$ to denote $(\hat{\theta}_M, \hat{\theta}_L, \hat{\theta}_U)$ generally, when the discussion applies equally to all of $\hat{\theta}_M$, $\hat{\theta}_L$, and $\hat{\theta}_U$, and *Feasible Posterior Interval* (FPI) to denote FPMI and FPCI generally.

It can be noted that the FPIs defined above can have the interpretation of uncertainty intervals from a Robust Bayes analysis (Berger, 1990). In Robust Bayes analysis, we seek to summarize the many possible posterior inferences arising from a *class* of prior distributions. Denote by $\boldsymbol{\eta}$ the set of parameters whose uncertainty are quantified non-probabilistically, and $(\theta, \boldsymbol{\xi})$ those whose uncertainty are quantified probabilistically, where θ is the parameter of interest and $\boldsymbol{\xi}$ are other nuisance parameters (not belonging to $\boldsymbol{\eta}$). Seeking the minimum and maximum of $\hat{\theta}_M$ can be thought of as seeking the minimum and maximum posterior median among the class of prior distributions which have zero density outside the feasible region of $\boldsymbol{\eta}$. To see this, note that the posterior density of θ can be written as:

$$p(\theta|X) = \int_{\boldsymbol{\eta}} \int_{\boldsymbol{\xi}} p(\theta, \boldsymbol{\xi}, \boldsymbol{\eta}|X) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\eta} \tag{2.9}$$

Denoting the cumulative distribution of θ given X by $F_{\theta|X} : \theta \to p$, which maps θ onto the percentiles p, we have:

$$\begin{aligned}
F_{\theta|X}(\theta) &= \int_{\theta'=-\infty}^{\theta} \int_{\boldsymbol{\eta}} \int_{\boldsymbol{\xi}} p(\theta', \boldsymbol{\xi}, \boldsymbol{\eta}|X) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\eta} \mathrm{d}\theta' && (2.10) \\
&= \int_{\theta'=-\infty}^{\theta} \int_{\boldsymbol{\eta}} \int_{\boldsymbol{\xi}} \frac{p(X|\theta', \boldsymbol{\xi}, \boldsymbol{\eta}) p(\theta', \boldsymbol{\xi}) p(\boldsymbol{\eta}) p(X|\boldsymbol{\eta})}{p(X|\boldsymbol{\eta}) p(X)} \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\eta} \mathrm{d}\theta' \quad \text{(by Bayes' Theorem)} && (2.11) \\
&= \int_{\boldsymbol{\eta}} \int_{\theta'=-\infty}^{\theta} p(\theta'|X, \boldsymbol{\eta}) \mathrm{d}\theta p(\boldsymbol{\eta}|X) \mathrm{d}\boldsymbol{\eta} && (2.12) \\
&= \int_{\boldsymbol{\eta}} F_{\theta|X,\boldsymbol{\eta}}(\theta) p(\boldsymbol{\eta}|X) \mathrm{d}\boldsymbol{\eta} && (2.13)
\end{aligned}$$

Hence, we see that the cumulative distribution of θ given X is a weighted average of the cumulative distribution of θ given X and $\boldsymbol{\eta}$. Now, if our prior distribution of $\boldsymbol{\eta}$ belongs to a class that has zero mass for values outside the feasible region of $\boldsymbol{\eta}$, denoted $\mathcal{E}$, i.e.:

$$p(\boldsymbol{\eta}) = 0 \quad \forall \boldsymbol{\eta} \notin \mathcal{E}$$

then

$$p(\boldsymbol{\eta}|X) = 0 \quad \forall \boldsymbol{\eta} \notin \mathcal{E}$$

and because averages cannot be greater than the maximum or less than the minimum,

$$\min_{\boldsymbol{\eta}\in\mathcal{E}} F_{\theta|X,\boldsymbol{\eta}}(\theta) \leq F_{\theta|X}(\theta) \leq \max_{\boldsymbol{\eta}\in\mathcal{E}} F_{\theta|X,\boldsymbol{\eta}}(\theta)$$

Now our Bayesian estimates $\hat{\theta}_M, \hat{\theta}_L, \hat{\theta}_U$ are defined by the percentile function $F^{-1} : p \to \theta$, which is the inverse of the cumulative distribution function. Since the cumulative distribution function $F(\theta)$ is necessarily a monotonically increasing function, we have:

$$\min_{\boldsymbol{\eta}\in\mathcal{E}} \hat{\theta} = \min_{\boldsymbol{\eta}\in\mathcal{E}} F^{-1}_{\theta|X,\boldsymbol{\eta}}(p) \leq F^{-1}_{\theta|X}(p) \leq \max_{\boldsymbol{\eta}\in\mathcal{E}} F^{-1}_{\theta|X,\boldsymbol{\eta}}(p) = \max_{\boldsymbol{\eta}\in\mathcal{E}} \hat{\theta}$$

Thus, by finding $\min_{\boldsymbol{\eta}} \hat{\theta}$ and $\max_{\boldsymbol{\eta}} \hat{\theta}$, we give bounds to $F^{-1}_{\theta|X}(p)$. Note that when we give bounds to $F^{-1}_{\theta|X}(p)$, we are assuming that the prior distribution of $\boldsymbol{\theta}$ is the same as the prior distribution we use to calculate the bounds (i.e. $F^{-1}_{\theta|X}(p)$ and $F^{-1}_{\theta|X,\boldsymbol{\eta}}(p)$ share the same prior distribution for $\boldsymbol{\theta}$). For this to be possible, the prior distribution of $\boldsymbol{\theta}$ must not depend on $\boldsymbol{\eta}$.

2.2 Finding $\hat{\theta} = F^{-1}_{\theta|X,\boldsymbol{\eta}}(p)$ by numerical integration

As discussed in the previous section, our goal is to find $\min_{\boldsymbol{\eta}\in\mathcal{E}} \hat{\boldsymbol{\Theta}}$ and $\max_{\boldsymbol{\eta}\in\mathcal{E}} \hat{\boldsymbol{\Theta}}$ ($\hat{\boldsymbol{\Theta}} = (\hat{\theta}_L, \hat{\theta}_M, \hat{\theta}_U)$) subject to $\boldsymbol{\eta}$ being in a certain space $\mathcal{E}$. Generally, we cannot determine these values analytically, and hence the approach of the thesis is to use a search algorithm. Searching through the space of $\mathcal{E}$ can be computationally very time-consuming if the evaluation of $F^{-1}_{\theta|X,\boldsymbol{\eta}}(p)$ is slow. To overcome this problem, I considered using the posterior mode as an approximate estimate for $\hat{\theta}_M$ and $\hat{\theta}_M \pm 1.96\hat{\sigma}$ as approximation for $\hat{\theta}_U$ and $\hat{\theta}_L$, where $\hat{\sigma}$ is the square root of the inverse of the Observed Information, in accordance with standard theory (see e.g. Gelman *et al.*, 2004). However, I found that sometimes, the inverse of the Observed Information can be a very poor approximation of the posterior variance. (This happens when the posterior mode is in a very flat region of the log posterior density, rendering the Observed Information close to 0.) Moreover, there was the possibility that the posterior density is not unimodal, such that convergence may not be to the true mode. In this thesis, I have chosen to evaluate $F_{\theta|X,\boldsymbol{\eta}}(\theta)$ through numerical integration for various values of θ, and then interpolate to find $F^{-1}_{\theta|X,\boldsymbol{\eta}}$. At first this may seem computationally infeasible, especially if $\boldsymbol{\xi}$ is of a high dimension. However, in a meta-analytic situation, all parameters except the common parameters and (and possible hyperparameters in a hierarchical model) are study-specific, and that given the common parameters and the hyperparameters, the posterior distribution of the study-specific parameters are independent. This feature can be used to factorize the overall posterior distribution, such that even if the posterior density has to be integrated over, the

number of dimensions can be vastly reduced. This is particularly the case when bias parameters take on fixed values rather than distributions, as this obviates the need to integrate over the bias parameters. This approach is described more fully in the next subsection. Other options for evaluating $F^{-1}_{\theta|X,\boldsymbol{\eta}}$ are available. For example, we can make use of Laplace's approximation for $p(\theta|X,\boldsymbol{\eta})$ (Rue *et al.*, 2009; Leonard and Hsu, 1999, p.191), which is discussed further in the future discussion section of Chapter 8.

2.2.1 The Bayesian meta-analysis model

First, let us consider a standard Bayesian model of a meta-analysis of case-control studies without consideration of biases. Such a model is already given in (2.3) and (2.4), and is re-presented here:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{2.14}$$

$$\operatorname{logit} \pi_{s1} = \operatorname{logit} \pi_{s0} + \theta_s \tag{2.15}$$

Here, we further note that in a *fixed-effects* (FE) meta-analysis model, θ_s are assumed to be the same for all s, whereas in the *random-effects* (RE) model, θ_s are assumed to "be drawn" from the same distribution. In a RE model, attention is often focused on the mean of the study-specific θ_s, or the mean of the distribution of θ_s. Denoting this mean by θ, we can write:

$$\theta_s = \theta + \delta_s \tag{2.16}$$

where δ_s represents the departure of the study s-specific log odds ratio from the average log odds ratio among the population of studies. Written in this way, instead of assuming θ_s being drawn from some distribution, we assume δ_s is drawn from some distribution with mean 0:

$$\delta_s \sim f(0, \sigma^2_\delta) \tag{2.17}$$

It is readily seen that if the distribution for δ_s is a point-mass at 0, the RE model reduces to the FE model and hence the FE model is a special case of the RE model. Here, I show how we can obtain posterior inference for θ in the RE model.

2.2.2 The posterior distribution for θ in a standard random effects meta-analysis of case-control studies

Denoting the entire dataset $\{Y_{si}, N_{si} : i = 1, 2; s = 1 \dots k\}$ by X, $\boldsymbol{\pi}_0 = (\pi_{10}, \pi_{20}, \dots, \pi_{k0})$, $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_k)$, where k is the number of studies, our posterior distribution of interest is:

$$p(\theta|X) = \int_{\boldsymbol{\pi}_0 \in \mathbb{R}^k} \int_{\boldsymbol{\delta} \in \mathbb{R}^k} p(\theta, \boldsymbol{\delta}, \boldsymbol{\pi}_0|X) \, \mathrm{d}\, \boldsymbol{\delta} \, \mathrm{d}\, \boldsymbol{\pi}_0 = \int_{\boldsymbol{\pi}_0 \in \mathbb{R}^k} \int_{\boldsymbol{\delta} \in \mathbb{R}^k} \frac{Lik(X|\theta, \boldsymbol{\delta}, \boldsymbol{\pi}_0) p(\boldsymbol{\delta}, \boldsymbol{\pi_0}, \theta)}{p(X)} \, \mathrm{d}\, \boldsymbol{\delta} \, \mathrm{d}\, \boldsymbol{\pi}_0 \tag{2.18}$$

(by Bayes' Theorem) given suitable priors for θ, $\boldsymbol{\pi}_0$ and $\boldsymbol{\delta}$. In a meta-analysis, the likelihood $Lik(X|\theta, \boldsymbol{\delta}, \boldsymbol{\pi}_0)$ is a product of individual study likelihoods:

$$Lik(X|\theta, \boldsymbol{\delta}, \boldsymbol{\pi}_0) = Lik(X_1|\theta, \delta_1, \pi_{10}) Lik(X_2|\theta, \delta_2, \pi_{20}) \cdots Lik(X_k|\theta, \delta_k, \pi_{k0}) \tag{2.19}$$

Each study-specific likelihood depends only on its study-specific δ_s and π_{s0}, and not on the others. If the prior distributions of δ_s and π_{s0} are also independent across studies, the multivariate integral can be written as a product of bivariate integrals:

$$p(\theta|X) = \frac{p(\theta)}{p(X)} \prod_s \int_{\pi_{s0}} \int_{\delta_s} Lik(X_s|\theta, \delta_s, \pi_{s0}) p(\delta_s, \pi_{s0}) \, \mathrm{d}\, \delta_s \, \mathrm{d}\, \pi_{s0} \tag{2.20}$$

In a typical Bayesian random-effects model, however, δ_s are not independent across studies, but are rather *exchangeable* (e.g. Sutton and Abrams, 2001; Higgins *et al.*, 2009), i.e. in addition to (2.17), we have:

$$\sigma_\delta^2 \sim p(.) \tag{2.21}$$

This considerably complicates the calculation as it does not enable the factorization of (2.20). In this thesis, therefore, we treat δ_s as independent. This is equivalent to fixing σ_δ^2. As explained in the following chapter, there are also conceptual reasons why this might in fact be desirable.

Although we have reduced a multi-dimensional integral to a product of bivariate integrals, the computational burden is still great. Moreover, because it is the inverse cumulative distribution of $p(\theta|X)$ that is needed (i.e. $F^{-1}_{\theta|X}$), we further need to integrate over θ. As a result, we need to integrate over 3 dimensions: θ, δ_s, and π_{s0}. However, a reparameterization of the above formula can make the computational burden considerably less. Reparameterizing (π_{s0}, δ_s) as (γ_s, λ_s), where $\gamma_s = \operatorname{logit} \pi_{s0}$ and $\lambda_s = \gamma_s + \delta_s$, equation (2.20) becomes:

$$p(\theta|X) = \frac{p(\theta)}{p(X)} \prod_s \int_{\gamma_s} \int_{\lambda_s} Lik(X_s|\theta, \lambda_s, \gamma_s) p(\gamma_s, \lambda_s) \, \mathrm{d}\, \lambda_s \, \mathrm{d}\, \gamma_s \tag{2.22}$$

Writing X_s as $\{X_{s1}, X_{s0}\}$, where X_{s1} denote the case and X_{s0} the control data, (2.22) factorizes to:

$$p(\theta|X) = \frac{p(\theta)}{p(X)} \prod_s \int_{\lambda_s} Lik(X_{s1}|\theta, \lambda_s) \int_{\gamma_s} Lik(X_{s0}|\gamma_s) p(\gamma_s, \lambda_s) \,\mathrm{d}\,\gamma_s \,\mathrm{d}\,\lambda_s \tag{2.23}$$

While equation (2.23) still involves a double integral, it is of note that the integral

$$\int_{\gamma_s} Lik(X_{s0}|\gamma_s) p(\gamma_s, \lambda_s) \,\mathrm{d}\,\gamma_s \tag{2.24}$$

($= p(X_{s0}, \lambda_s)$) no longer involves θ. Therefore, it only needs to be evaluated once when we integrate over θ. This can greatly reduce the computational burden.

In case-control studies, it has been recommended that prior distributions be given independently for π_{s0} and θ_s (Greenland, 2001, 2005b). This translates to independent γ_s and δ_s (with $\gamma_s = \operatorname{logit} \pi_{s0}$ and $\lambda_s = \gamma_s + \delta_s$). For this reason, it is useful to rewrite the joint distribution $p(\gamma_s, \lambda_s)$ as:

$$p(\gamma_s, \lambda_s) = p_{\gamma_s}(\gamma_s) p_{\delta_s}(\lambda_s - \gamma_s) \tag{2.25}$$

(by change of variables). In this way, the joint distribution is simply the product of two (generally) known univariate distribution, and complications in prior specification is avoided.

2.2.3 Further details on integration

Our goal is to evaluate the median, 2.5%-ile, and 97.5%-ile of the posterior distribution of θ given $\boldsymbol{\eta}$, which is the inverse of the posterior cumulative distribution $F_{\theta|X,\boldsymbol{\eta}}(\theta)$. Now,

$$F_{\theta|X,\boldsymbol{\eta}}(\theta) = \int_{\theta'=-\infty}^{\theta} p(\theta'|X, \boldsymbol{\eta}) \,\mathrm{d}\,\theta' \tag{2.26}$$

where $p(\theta|X, \boldsymbol{\eta})$ is defined in equation (2.23) except here we have added the $\boldsymbol{\eta}$ as part of the condition, to highlight the dependence of the posterior distribution on $\boldsymbol{\eta}$. From formula (2.23) and (2.26), it can

be seen that $F_{\theta|X,\boldsymbol{\eta}}(\theta)$ is a triple integral. It is helpful to separate the three levels of integral as:

$$F_{\theta|X,\boldsymbol{\eta}}(\theta) = \frac{1}{p(X|\boldsymbol{\eta})} \int_{\theta'=-\infty}^{\theta} p(\theta') \prod_s p(X_s|\theta', \boldsymbol{\eta}) \, \mathrm{d}\, \theta' \tag{A1}$$

$$p(X|\boldsymbol{\eta}) = \int_{\theta'=-\infty}^{\infty} p(\theta') \prod_s p(X_s|\theta', \boldsymbol{\eta}) \, \mathrm{d}\, \theta' \tag{A2}$$

$$p(X_s|\theta, \boldsymbol{\eta}) = \int_{\lambda_s=-\infty}^{\infty} Lik(X_{s1}|\theta, \lambda_s, \boldsymbol{\eta}) p(X_{s0}, \lambda_s|\boldsymbol{\eta}) \, \mathrm{d}\, \lambda_s \tag{B}$$

$$p(X_{s0}, \lambda_s|\boldsymbol{\eta}) = \int_{\gamma_s=-\infty}^{\infty} Lik(X_{s0}|\gamma_s, \boldsymbol{\eta}) p(\gamma_s, \lambda_s) \, \mathrm{d}\, \gamma_s \tag{C}$$

It is clear from the above integrals that C is nested in B and B in A1 and A2. Thus, to evaluate integral A1 and A2, we first evaluate integrals C and then B, and finally A1 and A2 simultaneously. Moreover, it is sufficient to consider integration techniques for one-dimension only. Techniques for integration over one dimension are well known for over a century. These generally aim to approximate an integral by a weighted sum of the function evaluated at a finite number of abscissas (Press *et al.*, 2007):

$$\int_a^b f(\theta) \, \mathrm{d}\, \theta \approx \sum_{i=1}^{n} w_i f(\theta_i) \tag{2.27}$$

Generally, the most efficient integration algorithms would choose the number (n) and location (θ_i) of the abscissas to evaluate the integral in order to achieve a required level of precision. In general, when only one integral needs be evaluated, these algorithms are to be preferred over schemes which fix the number and locations of the abscissas in advance (Press *et al.*, 2007). In this thesis, we are searching over the feasible space of $\boldsymbol{\eta}$ for the extreme $\hat{\boldsymbol{\Theta}} = F^{-1}_{\theta|X,\boldsymbol{\eta}}(p), p = 0.025, 0.5, 0.975$. As $\boldsymbol{\eta}$ changes, so do some of the components of the integrals A1, A2, B, C. In this thesis, it will be the case that some of the components of A1, A2, B, C will stay the same, and hence it will be efficient to reuse those components without having to re-calculate them every time $\boldsymbol{\eta}$ changes (see Figure 2.1). To achieve this, we also need to fix the locations of the abscissas in advance (Krommer and Ueberhuber, 1998, p.346).

Moreover, in order that the algorithm can be accurate and efficient at the same time, we need to adapt the domain of integration to the distribution (e.g. for a distribution which has 99.99% of its mass from -2 to 4, our integration domain would probably be from around -2 to 4, whereas for a distribution with 99.99% mass from 4 to 6, our integration domain would probably be around 4 to 6). For this thesis, I have developed some *ad hoc* algorithms to achieve this and these are given in Appendix A. In order to reuse the abscissas even as the domain changes, the abscissas need to be equally spaced (see Figure

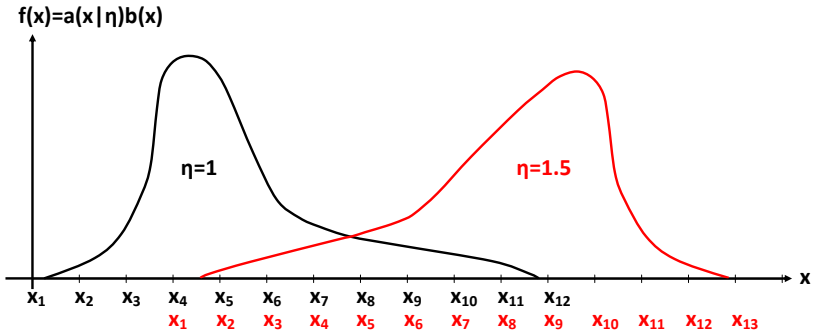


Figure 2.1: Suppose we want to integrate the function: $f(x) = a(x|\eta)b(x)$. When $\eta = 1$, it would be sufficient to use the set of points $(x_1, x_2, x_3, \ldots, x_{12})$ for integration. When η changes from 1 to 1.5, we would want a different set of points for integration. Here, we denote these by $(x_1, x_2, x_3, \ldots, x_{13})$. It would be advantageous to choose $(x_1, x_2, x_3, \ldots, x_{13})$ such that $(x_1, x_2, x_3, \ldots, x_9)$ coincides with $(x_4, x_5, x_6, \ldots, x_{12})$. This is because although $a(x|\eta)$ still needs to be evaluated for $(x_1, x_2, x_3, \ldots, x_{13})$, $b(x)$ only needs to be evaluated for $(x_{10}, x_{11}, x_{12}, x_{13})$, and not for $(x_1, x_2, x_3, \ldots, x_9)$, since these have been evaluated. This scheme is easy to apply only if the abscissas are equally spaced.

2.1). Integration with equally spaced abscissas for smooth functions is best done using Newton-Cotes formulae (Press *et al.*, 2007), and these are employed in this thesis. Appendix A.4 gives further details.

Once we have evaluated the integrand of integral A1 (and A2) at a number of equally-spaced abscissas, we can then work out the approximated cumulative distributions $\hat{F}(\theta_1)$, $\hat{F}(\theta_2)$, $\hat{F}(\theta_3)$, ..., $\hat{F}(\theta_n)$ using the Newton-Cotes formula, where:

$$\hat{F}(\theta_i) = \sum_{j=1}^{n} w_j^{(i)} p(\theta_j|X, \boldsymbol{\eta}) \approx F_{\theta|X,\boldsymbol{\eta}}(\theta_i) = \int_{\theta'=-\infty}^{\theta_i} p(\theta'|X, \boldsymbol{\eta})\, \mathrm{d}\,\theta' \tag{2.28}$$

and $w_j^{(i)}$ are the weights used in the Newton-Cotes formula (see Appendix A.4). Since we seek the inverse of the cumulative distribution function, we then plot $\hat{F}(\theta_i)$ against θ_i, and interpolate to locate θ_i (see Figure 2.2). In fact, accuracy is lost if we simply use the linear interpolation scheme as depicted in Figure 2.2, since the true cumulative distribution must be smooth whereas Figure 2.2 represents it as piecewise linear. In this thesis, I instead use linear interpolation after first transforming $\hat{F}(\theta_i)$ on the inverse Normal scale.

2.3 Searching through $\mathcal{E}$ for the most extreme $\hat{\theta}$

In the above section, I described the techniques for evaluating $\hat{\boldsymbol{\Theta}} = F^{-1}_{\theta|X,\boldsymbol{\eta}}(p), p = 0.025, 0.5, 0.975$. Here, I describe the algorithm for searching through the feasible space of $\boldsymbol{\eta}$ (denoted $\mathcal{E}$) to find min/max $\hat{\boldsymbol{\Theta}}$.

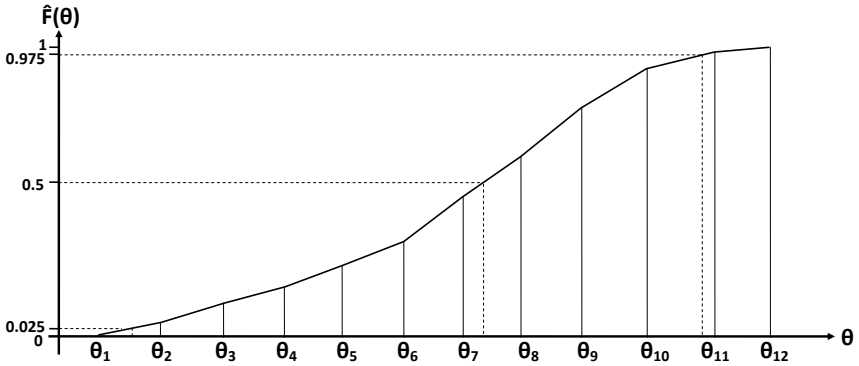


Figure 2.2: Finding the 2.5%, 50%, and 97.5%-ile of $p(\theta|X, \boldsymbol{\eta})$ by interpolation

First, let us note that if $\mathcal{E}$ is defined arbitrarily, then optimizing over this space can be quite difficult. In this thesis, however, the space of $\mathcal{E}$ is always regular, with boundaries that are linear. This means that optimization can be achieved by a simple iterative algorithm. In the case where the feasible space is "box-shaped", i.e. where the feasible region of any $\eta_i \in \{\eta_1, \eta_2, \ldots, \eta_R\}$ can be described by an inequality of the form:

$$l_i \leq \eta_i \leq u_i \tag{2.29}$$

where l_i and u_i denote the lower and upper limit of the feasible region, an algorithm known as the cyclic coordinate method (Vassiliadis and Conejeros, 2009) that allows us to search through the feasible space is given in Box 2.1.

It may be worth noting that the algorithm as presented in Box 2.1 is quite inefficient for general optimization problems (Press *et al.*, 2007, p. 509). In the problems of this thesis, however, it presents two advantages:

1. In every iteration of the algorithm, only one η_i out of $(\eta_1, \eta_2, \ldots, \eta_R)$ is changed. This means that in the evaluation of the integrals A1, A2, B, and C, many components retain their value from iteration to iteration, saving the need to re-evaluate all components every time.

2. It is easily extensible to deal with the more complicated constraints given in chapter 4 and 5

It is important to note that the algorithm only aims to seek a *local* maximum/minimum whereas the aim of the thesis is to find the *global* minimum/maximum.[1] If a problem contains multiple local maxima or minima, then the point of convergence may not be the global maximum/minimum that we seek. In general, it can be difficult to verify whether a local maximum/minimum is in fact also the global. Indeed, often the only guaranteed solution is to explore the entire parameter space, which, however, is often infeasible, particularly as the number of dimensions is large. We can, however, repeat

[1]Briefly, the local maximum/minimum is the maximum/minimum within a neighbourhood of the minimum/maximum, whereas the global maximum/minimum is the absolute maximum/minimum over the entire parameter space.

1. Permute the bias parameter $\eta_1, \eta_2, \ldots, \eta_R$ into a random sequence $\eta_{(1)}, \eta_{(2)}, \eta_{(3)}, \ldots, \eta_{(R)}$. For example, when $R = 5$, one possible sequence is: $\eta_{(1)} = \eta_2, \eta_{(2)} = \eta_1, \eta_{(3)} = \eta_4, \eta_{(4)} = \eta_5, \eta_{(5)} = \eta_3$.

2. Start with any feasible $\boldsymbol{\eta}^{[0]} = (\eta^0_{(1)}, \eta^0_{(2)}, \ldots, \eta^0_{(R)})$. In my program, these starting values are generated randomly.

3. Holding $\eta_{(2)}, \eta_{(3)}, \ldots, \eta_{(R)}$ fixed, i.e. $\eta_{(2)} = \eta^0_{(2)}, \eta_{(3)} = \eta^0_{(3)}, \ldots, \eta_{(R)} = \eta^0_{(R)}$, optimize the objective function with respect to $\eta_{(1)}$ alone (See Appendix B). Replace $\eta_{(1)}$ by the new, optimized value $\eta^1_{(1)}$. The new vector of $\boldsymbol{\eta}$ now would be: $\boldsymbol{\eta}^{[1]} = (\eta^1_{(1)}, \eta^0_{(2)}, \ldots, \eta^0_{(R)})$.

4. Repeat 3 above for $\eta_{(2)}$, holding $\eta_{(1)}, \eta_{(3)}, \ldots, \eta_{(R)}$ fixed, generating $\boldsymbol{\eta}^{[2]} = (\eta^1_{(1)}, \eta^1_{(2)}, \ldots, \eta^0_{(R)})$ and so on. When we reach the end of the sequence, repeat from the beginning. For example: $\boldsymbol{\eta}^{[R+1]} = (\eta^2_{(1)}, \eta^1_{(2)}, \ldots, \eta^1_{(R)})$

5. After every evaluation, compare $\boldsymbol{\eta}^{[n]}$ with $\boldsymbol{\eta}^{[n-R]}$. This is done by a standard relative difference (RD) formula in numerical analysis:

$$RD(\boldsymbol{\eta}^{[n]}, \boldsymbol{\eta}^{[n-R]}) = \max_r \left\{ RD_r(\eta^{p_r}_{(r)}, \eta^{p_r-1}_{(r)}) \right\} \tag{2.30}$$

$$RD_r(\eta^{p_r}_{(r)}, \eta^{p_r-1}_{(r)}) = \frac{|\eta^{p_r}_{(r)} - \eta^{p_r-1}_{(r)}|}{|\eta^{p_r-1}_{(r)}| + 1} \tag{2.31}$$

   If RD is less than a certain tolerance tol, then declare convergence, and report $\boldsymbol{\eta}^{[n]}$ as the optimized vector of $\boldsymbol{\eta}$.

6. In theory, it is possible that the maximum/minimum is in a flat region. In this case, the sequence of $\boldsymbol{\eta}^{[n]}$ may not converge. Declare that a flat region is encountered if $RD(f(\boldsymbol{\eta})^{[n]}, f(\boldsymbol{\eta})^{[n-R]})$ is less than a certain tolerance tol'. Typically, this tolerance level will be much less than tol. In my program, I used 0.0001 for tol and 1×10^{-10} for tol'.

Box 2.1: The cyclic coordinate method for searching through $\mathcal{E}$ for max/min $\hat{\boldsymbol{\Theta}}$

the optimization several times from different starting points and use different optimization schedules, i.e. different sequences of $\eta_{(1)}, \eta_{(2)}, \ldots, \eta_{(R)}$, to see if they converge into different maxima/minima (see Box 2.1 for notations). This is why I used randomly generated starting points, and randomly permuted parameter sequences, even though computational time can be significantly reduced by having well-chosen starting values. If optimization from different starting points using different parameter sequences all converge to the same point, then we may have more confidence that the maximum/minimum achieved is the global maximum/minimum. This is the approach that is followed in this thesis. For all of the problems in this thesis, I repeated the optimization 4 times. When more than one optima is encountered, I repeated the optimization a further 4 times until no more optima are found. Discussion then focuses on this overall maximum/minimum as if the global optimum.[2]

A note on programming

All computer programs used to implement the algorithms were written in Stata 10.1 (StataCorp, 2007), with much use made of the Mata programming language that is part of Stata. The Mata programming language has syntax similar to C but has syntax similar to R or Matlab for the manipulation of matrices. Programs are compiled before running (rather than interpreted as in R or Java). The codes for all programs used in this thesis are found in the supplementary CD-ROM, and instructions are also given in the CD-ROM for reproduction of all results. During the development of the programs, a selection of results were checked against implementation using WinBUGS 1.4 (Lunn *et al.*, 2000) to check for possible programming error in the implementation of the numerical integration.

A note on results

To avoid having too many tables in the thesis, results of the extreme locations found in all of the optimization done in this thesis are given in the "Location tables.pdf" file in the supplementary CD-

[2]Because of inaccuracies in the evaluation of $\hat{\boldsymbol{\Theta}} = F^{-1}_{\theta|X,\boldsymbol{\eta}}(p)$ (due to limitation in computer time), the values of $\boldsymbol{\eta}$ upon convergence would be slightly different even if they were all from the same mode. Thus, an arbitrary cut-off needs to be used to distinguish between "true" multi-modality and "apparent" multimodality due to computational inaccuracies. In this thesis, in order that two $\boldsymbol{\eta}$'s be classified as being from different modes, they must have a *relative difference* of more than $e^{-5} = 0.0067$, where *relative difference* (RD) is defined by:

$$RD(\boldsymbol{\eta}^{[a]}, \boldsymbol{\eta}^{[b]}) = RD((\eta_1^{[a]}, \eta_2^{[a]}, \ldots, \eta_R^{[a]}), (\eta_1^{[b]}, \eta_2^{[b]}, \ldots, \eta_R^{[b]})) \tag{2.32}$$

$$= \max_r \left\{ RD_r(\eta_r^{[a]}, \eta_r^{[b]}) \right\} \tag{2.33}$$

$$RD_r(\eta_r^{[a]}, \eta_r^{[b]}) = \frac{|\eta_r^{[a]} - \eta_r^{[b]}|}{|\eta_r^{[a]}| + 1} \tag{2.34}$$

The choice of this cut-off point is decided by the examination of the relative differences calculated in the various problems of this thesis. In general, $\boldsymbol{\eta}$'s from the same mode have log(RD) of around $-\infty$ to -7, and $\boldsymbol{\eta}$ from different modes have log(RD) of around -4 to 0. Thus -5 is around the midpoint between the two distributions.

ROM. Where more than one local minima are found in any of the problems, the table also presents the location of all the local minima.

Summary:

- The basic meta-analysis model used in this thesis is the Binomial model of (2.3), (2.4). To account for bias, some of the parameters are replaced by their biased counterparts, which are related to the unbiased parameters through a function with bias parameters $\boldsymbol{\eta}$.
- Bias parameters $\boldsymbol{\eta}$ are given the *feasible region* $\mathcal{E}$, and the combined probabilistic/non-probabilistic approach of this thesis is to use a search algorithm to search through $\mathcal{E}$ for the most extreme posterior median or 95% credible interval for θ, the effect of interest (which depends on $\boldsymbol{\eta}$).
- An algorithm was devised for evaluating the posterior median and credible interval quickly. The cyclic coordinate method is used to search through $\mathcal{E}$ for the most extreme posterior inference for θ.

Chapter 3

Quantifying uncertainty in a meta-analysis with simple bias adjustments

In this chapter, I apply the method of this thesis to a simple bias model and describe several features of the methodology, especially in the context of the example data of this thesis — the meta-analysis of the Greenland (2005a) data, which is a meta-analysis of 14 case-control studies of childhood leukaemia in relation to Extremely Low Frequency Electromagnetic Fields (henceforth shortened to EMF). Below I give a few more details on this dataset, and some background of the epidemiology of childhood leukaemia in relation to EMF.

3.1 The Greenland (2005a) data

The data of this meta-analysis are given in Table 3.1. As mentioned above, the data are a collation of individual study data by Greenland of 14 case-control studies of childhood leukaemia in relation to exposure to high levels of EMF. In these studies, EMF exposure was assessed in different ways, but for each study, Greenland chose the measure that in his judgement corresponded most closely to the average exposure prior to the onset of disease. He then classified exposure to high levels of EMF as having measurement (often a 24-hour measurement in the bedroom of a child, or calculations based on distance from the child's residence to the nearby external sources) of at least 0.3 μT. There has, however, been no clear understanding of any biological causal mechanism that relates high EMF exposure to risk of childhood leukaemia (WHO, 2007). In the laboratory, exposure to EMF of even a magnitude greater than 0.3μT has not been demonstrated to cause any long-term changes to bodily function in the human body (WHO, 2007). However, because of epidemiological evidence, especially from the meta-analyses of Greenland *et al.* (2000) and Ahlbom *et al.* (2000), the International Agency for Research on Cancer (IARC) classifies EMF has a *possible* carcinogen to the human body (IARC, 2002). Bias remains a strong possibility for the explanation of the observed association in these meta-analyses. For this

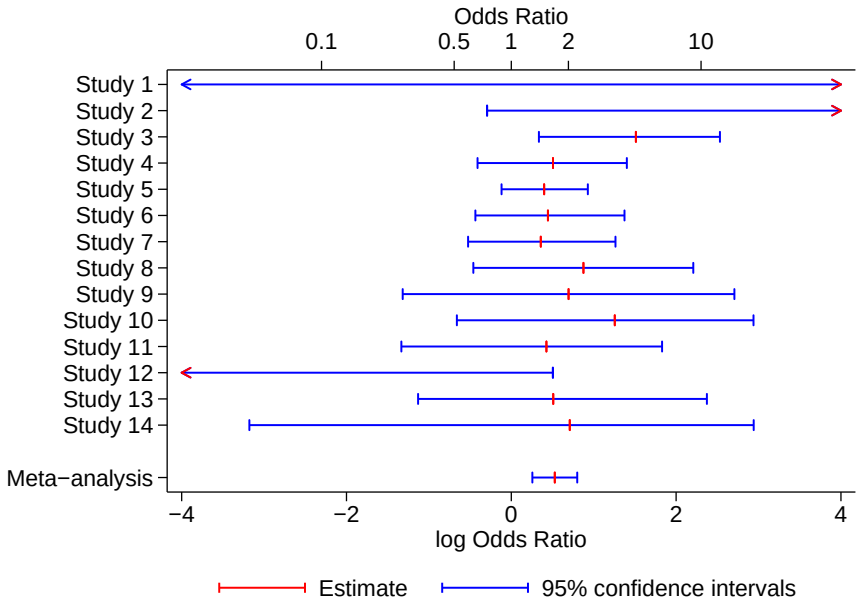


Figure 3.1: A standard non-Bayesian meta-analysis using the Mantel-Haenszel method.

reason, and also to demonstrate techniques for multiple bias modelling, Greenland (2005a) re-analyzed his meta-analysis of 2000 with additional data, and took into account of additional uncertainty due to biases.

The results of two non-Bayesian methods for analyzing this dataset are given at the bottom of Table 3.1. The Dersimonian-Laird estimate of the between-study variance for the data is 0, which means that the random-effects (RE) model is equivalent to the fixed-effects (FE) model using the Dersimonian-Laird method (Dersimonian and Laird, 1986). These results are also plotted in Figure 3.1. As can be seen, the confidence intervals from these meta-analytic summaries suggest that the odds ratio of childhood leukaemia due to high levels of EMF is significantly different from 1.

3.1.1 A Bayesian analysis for a FE meta-analysis

As mentioned in section 2.1, in this thesis we are primarily interested in *Bayesian* estimates of meta-analytic summary odds ratios and their credible intervals, in particular the median $\hat{\theta}_M$, the 2.5%-ile $\hat{\theta}_L$, and the 97.5%-ile $\hat{\theta}_U$. We want to find max/min $\hat{\boldsymbol{\Theta}} = \{\hat{\theta}_M, \hat{\theta}_L, \hat{\theta}_U\}$, subject to the bias parameters $\boldsymbol{\eta}$ being within a certain feasible region $\mathcal{E}$. To obtain Bayesian inference, we need to give prior distributions to the non-bias parameters. Recall that in the standard meta-analysis model of this thesis, we have:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad s = 1 \ldots 14, i = 0, 1 \tag{3.1}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{3.2}$$

Index	Study	Exposed case (Y_{s1})	Total case (N_{s1})	Exposed controls (Y_{s0})	Total controls (N_{s0})	Odds Ratio
1	Coghill *et al.* (1996)	1	56	0	56	∞
2	Dockerty *et al.* (1998)	3	87	0	82	∞
3	Feychting and Ahlbom (1993)	6	38	22	554	4.53(1.72, 12.0)
4	Kabuto *et al.* (2006)	11	312	13	603	1.66 (0.73, 3.75)
5	Linet *et al.* (1997)	42	638	28	620	1.49(0.91, 2.44)
6	London *et al.* (1991)	17	162	10	143	1.56 (0.69, 3.53)
7	McBride *et al.* (1999)	14	297	11	329	1.43 (0.64, 3.20)
8	Michaelis *et al.* (1998)	6	176	6	414	2.40 (0.76, 7.55)
9	Olsen *et al.* (1993)	3	833	3	1666	2.00 (0.40, 9.95)
10	Savitz *et al.* (1988)	3	36	5	198	3.51 (0.80, 15.4)
11	Tomenius (1986)	3	153	9	698	1.53 (0.41, 5.72)
12	Tynes and Haldorsen (1997)	0	148	31	2004	0
13	UKCCS (1999)	5	1057	3	1053	1.66 (0.40, 6.98)
14	Verkasalo *et al.* (1993)	1	32	5	320	2.03 (0.23, 18.0)
	Pooled OR (Mantel-Haenszel method)					1.68 (1.27, 2.22)
	Pooled OR (Maximum Likelihood)					1.69 (1.28, 2.23)
	Bayesian FE model					1.52 (1.16, 1.99)
	Bayesian RE model: $\delta_s \sim N(0, 0.15)$					1.54 (1.06, 2.24)

Table 3.1: The Greenland (2005a) childhood leukaemia-EMF data

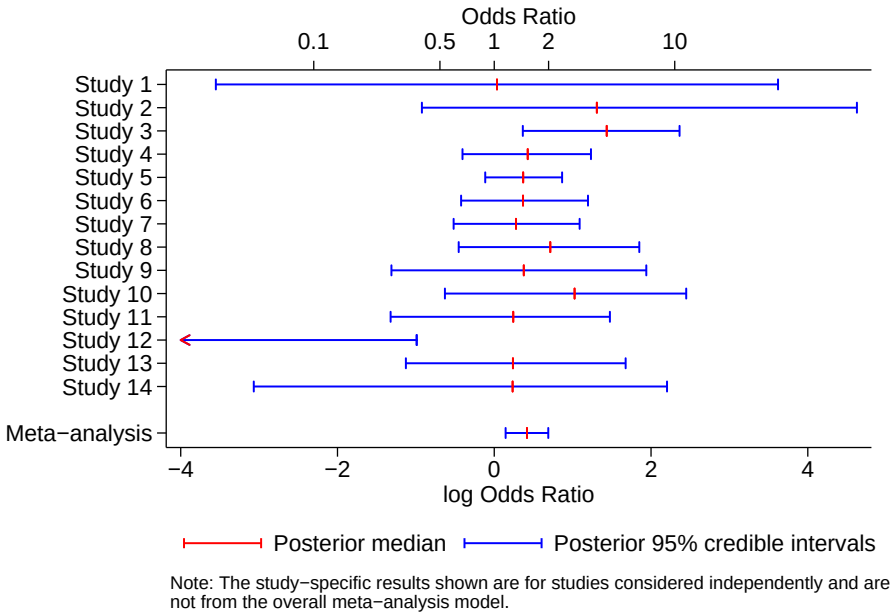


Figure 3.2: A standard Bayesian FE meta-analysis model with $\theta \sim N(0, 100)$.

In a FE model, we have: $\theta_s = \theta$, and hence we have to assign prior distribution to θ and π_{s0}, or alternatively, to θ and $\gamma_s = \text{logit}\,\pi_{s0}$. The RE model is considered in the next subsection.

In this thesis, I consider the following 'nearly flat' prior distributions for θ and γ_s:[1]

$$\gamma_s \sim \text{Logistic}(0, 1) \qquad s = 1 \ldots 14 \tag{3.3}$$

$$(\text{equivalent to } \pi_{s0} \sim U(0, 1)) \tag{3.4}$$

$$\theta \sim N(0, 100) \tag{3.5}$$

Results from these analyses are given in Table 3.1 and Figure 3.2.

The posterior median and 95% credible interval for the odds ratio is 1.52 (1.16, 1.99) for this method. Compared to the non-Bayesian results (1.68, 95% CI=[1.27, 2.22]), we see that this appears to be somewhat closer to the null. (This is because the prior distribution we give to $\gamma_s = \text{logit}\,\pi_{s0}$ is not completely flat. γ_s, being a Logistic(0,1) distribution, has disproportionate mass around $\gamma_s = 0$. If a completely flat prior (i.e. $\gamma_s \sim U(-\infty, \infty)$) was used for γ_s, the results would be much closer to the non-Bayesian results.

[1]Note that I have avoided using improper priors such as:

$$\gamma_s \sim U(-\infty, \infty)$$

or

$$\theta \sim U(-\infty, \infty)$$

This is to avoid having improper posterior distributions associated with having zero counts in some of the studies, e.g. Coghill *et al.* (1996) and Tynes and Haldorsen (1997)

3.1.2 A Bayesian RE meta-analysis

For a RE model, instead of $\theta_s = \theta$, we have $\theta_s = \theta + \delta_s$, and we need to assign a prior distribution to δ_s as well. Often, in Bayesian RE meta-analyses, this parameter is given a hierarchical prior distribution, e.g.:

$$\delta_s \sim N(0, \sigma_\delta^2) \tag{3.6}$$

with σ_δ^2 given a second-level distribution (Smith *et al.*, 1995; Sutton and Abrams, 2001; Higgins *et al.*, 2009). However, the specification of this second-level distribution can be a contentious issue, particularly when the number of studies is not large (Sutton and Abrams, 2001; Congdon, 2006, p.155). Non-informative priors for σ_δ^2 can lead to *over-smoothing*, resulting in little information being available in the data to inform θ (Congdon, 2006, p.155) in meta-analyses that are moderate or small in size. Most attempts to overcome the issue generally suggest replacing this prior distribution with an informative one (Sutton and Abrams, 2001; Congdon, 2006; Gustafson *et al.*, 2006), possibly with the parameters estimated from data (e.g. DuMouchel, 1996).

Alternatively, we can fix σ_δ^2 at suitable values. In the previous chapter, we have noted that the evaluation of our Bayesian estimates $\hat{\boldsymbol{\Theta}}$ is made considerably easier if σ_δ^2 were fixed. In view of the difficulties in specifying a suitable distribution for the parameter, fixing it at reasonable values has additional appeal. Indeed, a number of the explicitly Bayesian meta-analyses with consideration of biases have preferred to use fixed values for σ_δ^2 instead of giving it a second-level distribution, presumably for similar reasons (Greenland and Kheifets, 2006; Welton *et al.*, 2009). I therefore also follow this approach in this thesis.

σ_δ^2 represents the variance of the study-specific effects θ_s. Assuming the absence of bias due to exposure misclassification, non-participation bias, and bias due to incomplete control of confounding, study-specific effects can still differ from one another due to differences in design and differences in the underlying populations. Greenland and Kheifets (2006) considered a prior variance of approximately 0.5 for θ_s, the study-specific log odds ratio, and that 70% of this variance of might be common to all studies, leaving 30% due to study-specific features.[2] If the variance of θ_s is 0.5, then $\sigma_\delta^2 = 0.15$.

Together with the nearly flat prior of $N(0, 100)$ for θ, the posterior median and 95% credible interval for θ is 1.54 (1.06, 2.24). This result is also given in Table 3.1 and Figure 3.3. Compared with the FE model, the posterior 95% intervals are wider, both for the study-specific odds ratio and the overall average odds ratio. Because the information from each study is split between informing θ and δ_s (whereas in a FE model, all information is used for θ alone), there is now less information for informing θ, resulting in a wider posterior credible interval.

[2]This is derived from the prior distribution Greenland and Kheifets gave to the parameter α_{TY} in the model, which equals the exposure-disease log odds ratio among those that are classified as unexposed. Although not technically the study-specific log odds ratio, its distribution is almost identical to the study-specific log odds ratio based on simulation.

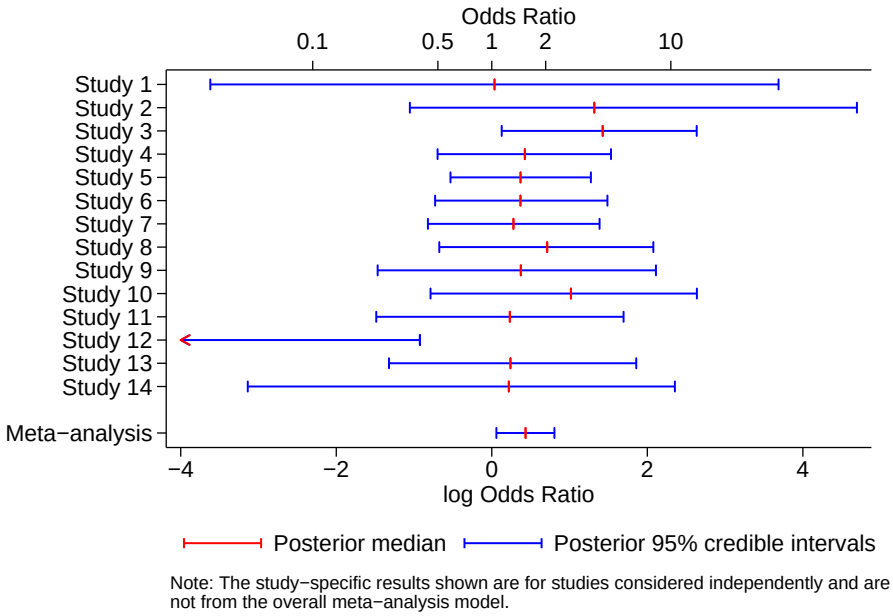


Figure 3.3: A standard Bayesian RE meta-analysis model with $\theta \sim N(0, 100)$ and $\delta_s \sim N(0, 0.15)$.

3.2 Applying the method of this thesis to the Greenland data with a simple bias model

Now, let us consider extending the standard Bayesian meta-analysis model to allow for a simple bias adjustment in each of the study. Recall that our standard meta-analysis model has the form:

$$Y_{si} \sim Bin(N_{si}, \pi_{s0}) \tag{3.7}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{3.8}$$

Here, we replace (3.8) with

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s^* \tag{3.9}$$

$$\theta_s^* = \theta_s + \eta_s \tag{3.10}$$

As an initial illustration, let us consider the following feasible region for η_s:

$$-0.5 \leq \eta_s \leq 0.5 \qquad s = 1 \ldots 14 \tag{3.11}$$

This feasible range for η_s implies that the true odds ratio may be biased by a factor of $\exp(0.5) = 1.65$ in either direction. Using the methods described in the previous chapter, we now search through the

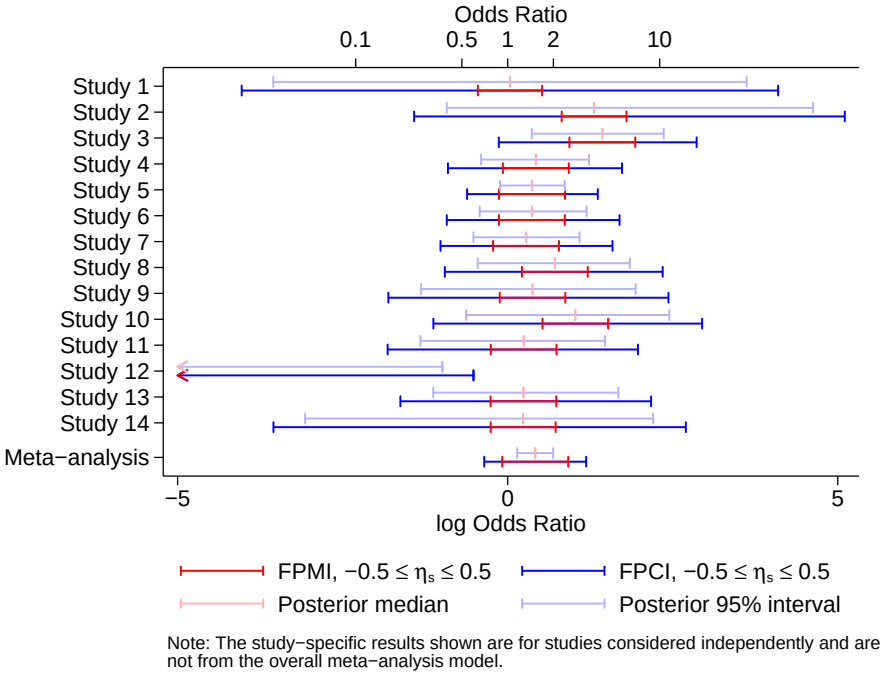


Figure 3.4: FE meta-analysis of Greenland data with simple bias: $-0.5 \leq \eta_s \leq 0.5$.

feasible region of $\eta_s, s = 1 \ldots 14$ for the most extreme $\hat{\boldsymbol{\Theta}}$. Figure 3.4 and 3.5 present the Feasible Posterior Intervals (FPI, c.f. equations 2.7 and 2.8) for θ with this feasible region for $\boldsymbol{\eta}$ for a FE and a RE meta-analysis, respectively. As before, $\theta_s = \theta$ for the FE model, and $\theta_s = \theta + \delta_s, \delta_s \sim N(0, 0.15)$ for the RE model. Optimization took 2 to 15 seconds to run for the single studies and 60 to 130 seconds to run for the meta-analysis on a Intel Pentium Core 2 Quad computer (using only 1 core). As might be expected, for $\min \hat{\boldsymbol{\Theta}}$, $\eta_s = 0.5$ for all s, and for $\max \hat{\boldsymbol{\Theta}}$, $\eta_s = -0.5$ for all s, and there was no sign of multimodality.[3]

We see in figures 3.4 and 3.5 that in comparison with the meta-analyses without bias, the Feasible Posterior Median Interval (FPMI) is almost exactly $[\hat{\theta}_M - 0.5, \hat{\theta}_M + 0.5]$, and the 95% Feasible Posterior Credible Interval (FPCI) for θ with bias is almost the same as $[\hat{\theta}_L - 0.5, \hat{\theta}_U + 0.5]$, where I have used $\hat{\theta}_M$, $\hat{\theta}_L$, and $\hat{\theta}_L$ to denote the posterior median and 95% limits in the meta-analysis without bias. Importantly, it can be seen that the FPMI in the meta-analysis setting is not shorter than that in the individual study setting, although the FPCI is shorter in the meta-analysis setting. Roughly speaking, we can

[3]It may be wondered whether in such a simple scheme, an analytical solution exists for $(\eta_1, \eta_2, \ldots, \eta_s)$ that maximizes/minimizes $\hat{\boldsymbol{\Theta}}$, since intuition suggests that if the posterior median for θ_s^* is around k, then the posterior median for θ_s would be around $k - \eta_s$, such that the posterior mean/median of θ_s might be minimized when η_s is at its most positive and maximized when η_s is at its most negative. And since $\theta = \theta_s - \delta_s$, the maximum/minimum posterior median of θ is also likely to be located with η_s at its most negative/positive. However, I have not been able to derive such association analytically, although if the likelihood of θ is Normal (through the Central Limit Theorem for large samples), then an analytical solution exists.

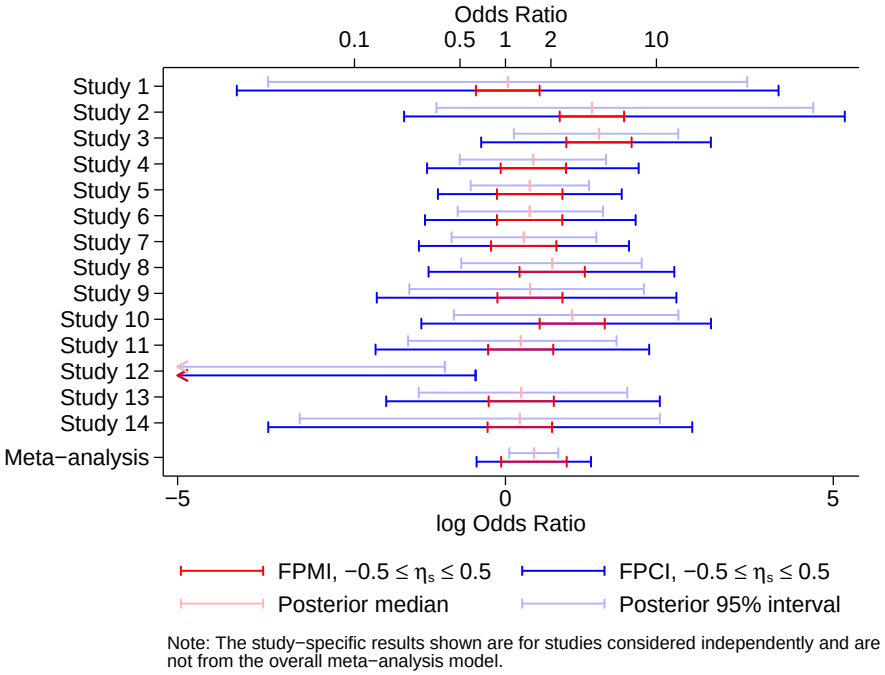


Figure 3.5: RE meta-analysis of Greenland data with simple bias: $-0.5 \leq \eta_s \leq 0.5$.

consider the FPMI to be due to uncertainty in bias alone, whereas the FPCI reflects uncertainty due to both bias and random sampling (or residual randomness, if randomness is defined in the subjective sense).[4] We therefore see that the effect of pooling studies in a meta-analysis is to reduce uncertainty due to random sampling, but not uncertainty due to bias. This reflects the very essence of the method in this thesis, which is to allow the pooling of data to reduce aleatoric but not epistemic uncertainty.

3.2.1 Some observations

This feature of the methodology, however, can lead to some counterintuitive results. Consider, for example, a meta-analysis of the above 14 studies, except that study 5 is known to have no bias. The FPIs for this scenario are given in Figure 3.6.

We see that the uncertainty interval for study 5 is in fact narrower than the overall meta-analytic uncertainty interval. If study 5 also happens to be our most relevant study, the analysis would suggest that there is no gain in our estimation of θ by including the other studies also in the meta-analysis (i.e. we have a less precise estimate from the meta-analysis than from the estimate from study 5

[4]This follows from standard large sample theory, which states that in large samples, as the likelihood becomes Normal and the variance small, Bayesian inference (with sensible choices of prior) becomes almost identical to maximum likelihood inference. The median becomes the same as the mode, and can be considered a location estimate. The credible interval becomes the same as the confidence interval, reflecting uncertainty due to random sampling. See Gelman *et al.* (2004, p.106) for more details.

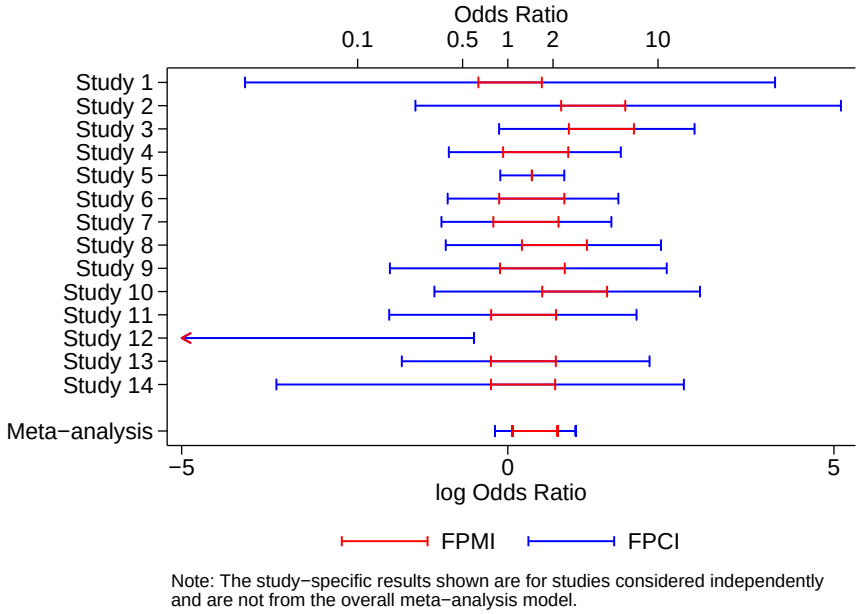


Figure 3.6: FE meta-analysis of Greenland data with simple bias: $-0.5 \leq \eta_s \leq 0.5$ for all studies except study 5

alone). This is certainly an important feature of the methodology that has to be borne in mind. The combined probabilistic/non-probabilistic method of this thesis is inherently conservative. Although it may be very unlikely that all study-specific biases η_s in fact equal their lower or upper bound, i.e. $\eta_s = -0.5$ or $0.5, s = 1 \ldots 14$, this scenario is considered "possible" by this method, and the posterior inference for this rather extreme scenario is not weighted down in any way. When we are considering only one study, the extreme scenario concerns an extreme value for only one parameter, and is thus in some sense much less "extreme" than the extreme scenario considered in a meta-analysis.

3.3 Meta-analysis with an informative prior on θ

Before we leave this chapter, let us consider the situation where we replace the uninformative $N(0, 100)$ prior for θ with an informative prior. When there is little in the data to inform our target parameter (such as when uncertainty over the bias parameters is great), the influence of the prior distribution becomes much greater, and the use non-informative prior distributions to avoid prior elicitation is no longer be justifiable (see e.g. Berger, 2006). (This is why an informative prior distribution for θ is used in the next chapter.) Here, we follow Greenland and Kheifets (2006) in giving a $N(0, 0.5)$ prior for the study s-specific log odds ratio, i.e.

$$\theta_s \sim N(0, 0.5) \tag{3.12}$$

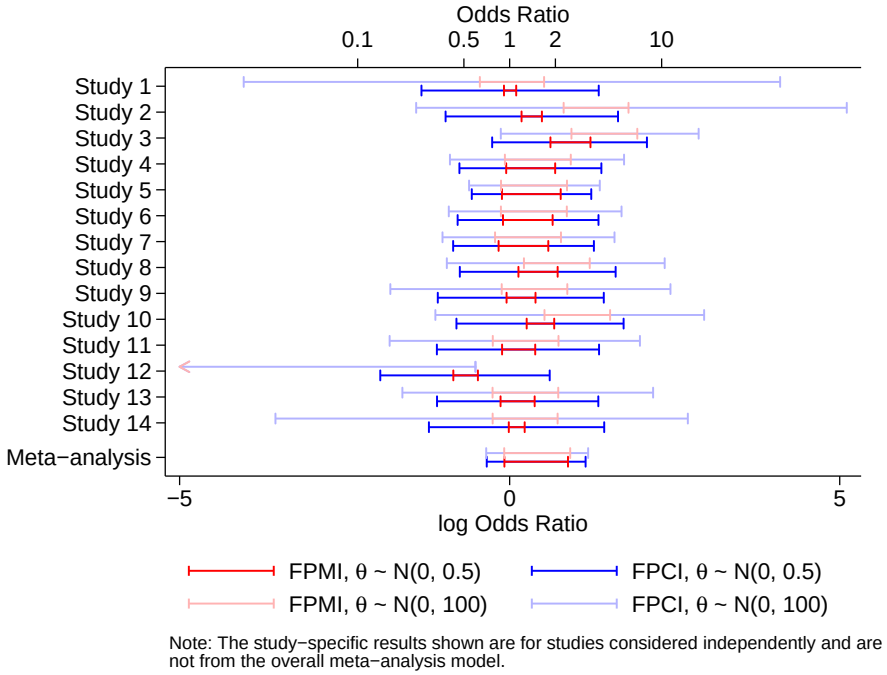


Figure 3.7: FPIs for a FE meta-analysis in simple bias model with $\theta \sim N(0, 0.5)$, $\delta_s = 0$, as compared to the model with $\theta \sim N(0, 100)$, $\delta_s = 0$.

This prior corresponds to a 95% credible interval of [0.25, 4.00] for the odds ratio (i.e. for $\exp \theta_s$). In a RE model, if we continue to have 70% of the variance of θ_s being common across studies, we have a prior distribution of:

$$\theta \sim N(0, 0.35) \tag{3.13}$$

In a FE model, we have:

$$\theta_s = \theta \sim N(0, 0.5) \tag{3.14}$$

The FPIs for these models with informative priors for θ are given in Figures 3.7 and 3.8, both in the case where studies are considered independently, and together in a meta-analysis, and are compared with the case where the 'nearly flat' prior distribution of (3.5) is used.

Perhaps the most striking feature of the FPIs when an informative prior is used for θ is that the FPMI is somewhat narrower for each of the studies when the informative prior for θ is in place, the reduction being particularly great for studies with little data, e.g. studies 1 and 2. However, the overall meta-analytic FPI is less affected, and moreover, the meta-analytic FPMI appears even wider than FPMIs of the individual studies. It appears that by pooling studies together, our uncertainty over θ due to bias has actually increased!

The reason for this is because of the influence of the prior distribution. The prior distribution has median 0 and if there were no data, the posterior median remains 0 whatever value the bias parameters

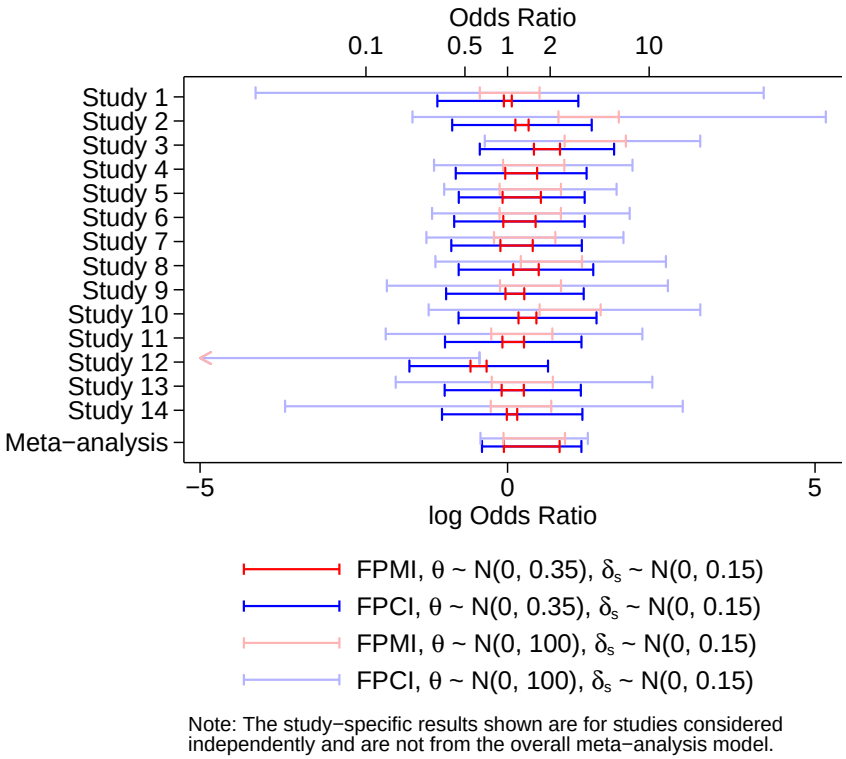
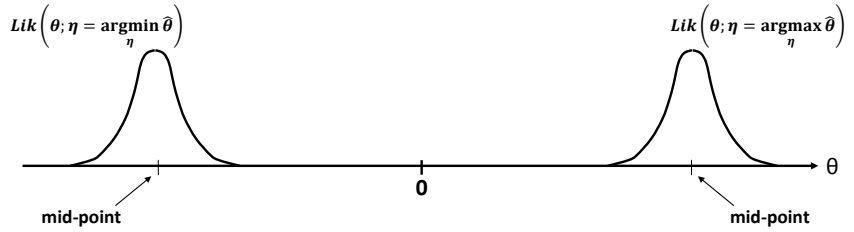


Figure 3.8: FPIs for a RE meta-analysis in simple bias model with $\theta \sim N(0, 0.5)$, $\delta_s = 0$, as compared to the model with $\theta \sim N(0, 100)$, $\delta_s = 0$.

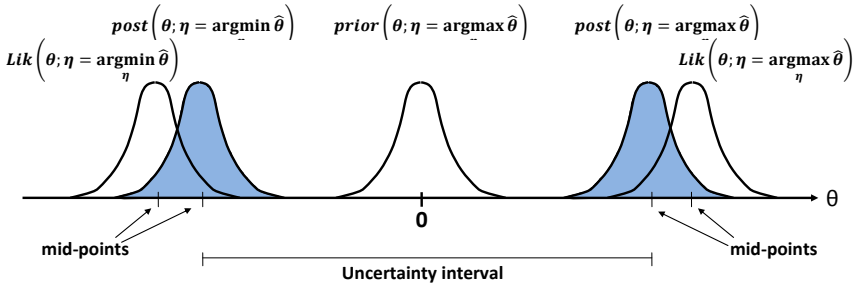
$\boldsymbol{\eta}$ takes. As the amount of data increases, the posterior distribution becomes more sensitive to changes in the bias parameters $\boldsymbol{\eta}$, leading to wider and wider FPMI. A meta-analysis has more data than any of the individual studies, and therefore can be expected to have a wider FPMI than any of the individual studies. A heuristical explanation of this phenomenon is given in Figure 3.9, although we should take note that the diagram represents a great simplification of the mathematics and the true relationship between the posterior median and the data is much more complicated.

Note that when an informative prior distribution is used, it becomes much more possible that the overall meta-analytic FPCI is wider than the study-specific one. This is because the meta-analysis contains more data than a single study, and hence the influence of the prior on the posterior of θ is much less in the meta-analysis than for a single study, leading to an increase in the width of the FPMI. The increase can be so large that it more than offsets the decrease in the width of the FPCI due to the decrease in uncertainty due to random sampling. For illustration, we may consider the scenario where $-2 \leq \eta_s \leq 2$ for all s. Figure 3.10 gives the feasible region for $\hat{\boldsymbol{\Theta}}$ for this scenario for a FE model.

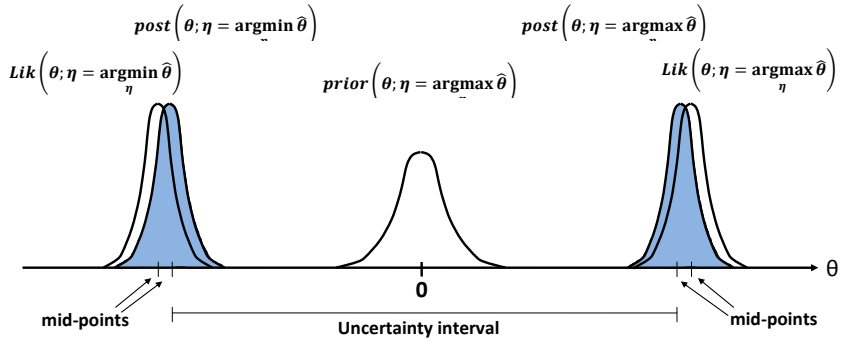
In summary, the results of this section shows that when using the combined probabilistic/non-probabilistic method of this thesis, an increase in data does not always lead to a decrease in uncertainty intervals.



(a)



(b)



(c)

Figure 3.9: Given different bias parameter values, the likelihood function with respect to θ can be very different, as seen in figure 3.9a. The influence of a zero-centred prior distribution is generally to move the posterior distribution closer to zero than the likelihood (is from 0) (figure 3.9b). The effect of increasing the amount of data is generally to increase the information available in the likelihood and thus to reduce the amount of shrinkage toward zero. Thus, this tends to lead to wider intervals (figure 3.9c).

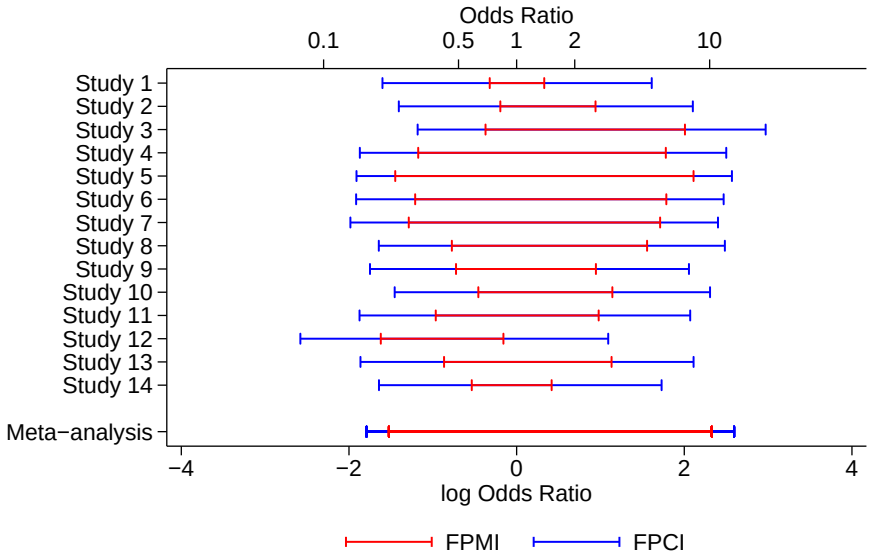


Figure 3.10: FPIs for $\hat{\boldsymbol{\Theta}}$ in simple bias model with $-2 \leq \eta_s \leq 2$, $\theta \sim N(0, 0.5)$, $\delta_s = 0$

Summary:

- In this chapter, I introduced the Greenland (2005a) meta-analysis of 14 case-control studies of childhood leukaemia and EMF, which serves as the example dataset of this thesis.
- I applied the method of this thesis to find the Feasible Posterior Intervals (FPI) for θ, the meta-analytic average log odds ratio, assuming study-specific log odds ratio are subject to a simple additive bias of $-0.5 \leq \eta_s \leq 0.5$.
- Unlike confidence intervals in conventional meta-analysis, the Feasible Posterior Median Interval (FPMI) is not shortened in the meta-analysis as compared to the study-specific FPMI. The FPCI, however, is shortened. Roughly, the FPMI reflects uncertainty due to bias alone, which is treated non-probabilistically, and the FPCI reflects uncertainty due to both bias and residual random error, which is treated probabilistically.
- When an informative prior is used for θ instead of an uninformative prior, some counterintuitive results can occur in that the meta-analytic FPMI/FPCI can be wider than the study-specific one. This is because there is more information in a meta-analysis to overcome the information contained in the prior, and posterior inference therefore becomes more sensitive to the bias parameters in a meta-analysis.

Chapter 4

Quantifying uncertainty in meta-analyses of case-control studies subject to exposure misclassification

In this chapter, I illustrate the combined probabilistic/non-probabilistic approach to quantifying uncertainty of bias due to exposure misclassification, again with application in Greenland's childhood leukaemia-EMF data. Before I discuss the details, it might be useful to consider how exposure misclassification might have arisen in these studies:

1. It is unsure what the relevant measure of EMF is, because the etiology of childhood leukaemia is unclear. In particular, the timescale is uncertain. It is also uncertain whether it is the average intensity, or the highest intensity to which the child was exposed, or whether it was the fluctuation in exposure that is relevant (ICNIRP, 2001).

2. Exposure to EMF was either measured *after* the development of the disease at the residences of the participants or estimated using calculations based on the distance of the residences from the nearest power lines or transformer stations. While these estimates may correlate positively with the real exposure levels, it is clear that they are far from accurate.

3. Moreover, the extent of error can be expected to differ from one child to another. Many children moved homes, and some were able to provide more residences for measurements than others. Some measurements were made over the course of 24 hours. Some were made for only a few seconds. Some measurements were only made outside the apartment/flat, while others were made inside the bedroom. Some children spent more time at home, particularly younger children.

4.1 Adjustment for error due to exposure misclassification

There is a large body of literature on adjustment of relative risk estimates in case-control studies with exposure misclassification (e.g. Diamond and Lilienfeld, 1962; Selen, 1986; Espeland and Hui, 1987; Drews and Greenland, 1990; Gustafson *et al.*, 2001; Fox *et al.*, 2005; Chu *et al.*, 2006; Greenland, 2008). One class of methods considers the joint distribution of the correctly classified and misclassified exposure in a loglinear model (Greenland and Kheifets, 2006; Greenland, 2009a,b). This class of methods considers the problem of misclassification in terms of parameters such as the ratio of the true-positive to false positive odds within the controls, the ratio of these two ratios between cases and controls, and other parameters. Elicitation of suitable values for these parameters, however, is not as intuitive as the others, and these methods are therefore not considered. Another class of methods deals with misclassification in a regression context by approximation methods based on methods developed for measurement error models (for continuous variables) (Kuchenhoff *et al.*, 2006). These methods do not fit easily into a Bayesian modelling framework and are also not discussed.

The vast majority of studies on misclassification models focus on methods that model the marginal distribution the true and the misclassified exposure, with the two distributions linked together by certain *misclassification probabilities*. Because the true exposure probabilities are related to the misclassified exposure probabilities through a linear transform, these methods have been termed matrix methods (Morrissey and Spiegelman, 1999), and two approaches are available.

The first approach considers the probabilities of the misclassified exposure as a linear function of the probabilities of true exposure. Denoting the vector of true exposure probabilities by $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)^T$ and the vector of misclassified exposure probabilities by $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)^T$ for an exposure with k levels, this approach considers the misclassification probability matrix $\boldsymbol{A}$, such that:

$$\boldsymbol{p} = \boldsymbol{A}\boldsymbol{\pi} \tag{4.1}$$

The second approach considers the reverse, i.e., the probability matrix $\boldsymbol{B}$, such that:

$$\boldsymbol{\pi} = \boldsymbol{B}\boldsymbol{p} \tag{4.2}$$

In both approaches, the researcher needs to obtain suitable estimates for the probabilities in either the matrix $\boldsymbol{A}$ or the matrix $\boldsymbol{B}$. In the literature on *exposure* misclassification, however, almost all studies have taken the first approach (4.1), although when diagnostic misclassification is the subject, the second approach (4.2) is usually used (e.g. Gastwirth, 1987; Paulino *et al.*, 2003). Possible reasons for this may be:

- The probabilities in $\boldsymbol{A}$ are more "transportable", in that if they are estimated in one study, they are more likely to be similar in another study (Greenland, 2008). (Note that Greenland's argument

was not based on empirical evidence.)

- It is usually more reasonable to assume the probabilities in $\boldsymbol{A}$ to be independent of disease status (i.e. non-differential exposure misclassification) than it is to assume $\boldsymbol{B}$ to be independent. This is when we suppose that apparent exposure status gives no information on disease independent of true exposure status, which is often a reasonable assumption (Diamond and Lilienfeld, 1962). When this assumption is valid, we can impose the constraint that the probabilities in $\boldsymbol{A}$ are the same for diseased and non-diseased individuals, which can reduce the need for finding estimates for $\boldsymbol{A}$.

- When applying Bayesian procedures for inference, it is easier to specify a prior distribution on $\boldsymbol{\pi}$ in the first approach than in the second approach, because in the second approach, the prior has to satisfy constraints given by $\boldsymbol{B}$.

For these reasons, in this thesis, I also follow the usual approach of modelling $\boldsymbol{p}$ as a linear function of $\boldsymbol{\pi}$.

For an exposure with only 2 levels (i.e. $k = 2$), the misclassification probability matrix $\boldsymbol{A}$ can be summarized by two probabilities: the *sensitivity* and the *specificity*. In the rest of this thesis, I denote sensitivity by *sens* and specificity by *spec*, where they are defined as follows:

$$sens = \frac{\left(\begin{array}{l}\text{Number of truly exposed people who}\\ \text{would have been categorized as exposed in the population}\end{array}\right)}{\text{Total number of truly exposed people in the population}} \tag{4.3}$$

$$spec = \frac{\left(\begin{array}{l}\text{Number of truly unexposed people who}\\ \text{would have been categorized as unexposed in the population}\end{array}\right)}{\text{Total number of truly unexposed people in the population}} \tag{4.4}$$

In a two-level exposure case, a more transparent version of equation (4.1) is:

$$p = \pi sens + (1 - \pi)(1 - spec) \tag{4.5}$$

To extend our meta-analysis model to take into account of misclassification, we modify our standard random-effects (RE) meta-analysis of case-control studies (equations 2.14, 2.15, and 2.16), i.e.:

$$Y_{si} \sim Bin(N_{si}, p_{si}) \qquad i = 0, 1 \tag{4.6}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{4.7}$$

$$\theta_s = \theta + \delta_s \tag{4.8}$$

to become:

$$Y_{si} \sim Bin(N_{si}, p_{si}) \qquad i = 0, 1 \tag{4.9}$$

$$p_{si} = \pi_{si} sens_{si} + (1 - \pi_{si})(1 - spec_{si}) \tag{4.10}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{4.11}$$

$$\theta_s = \theta + \delta_s \tag{4.12}$$

4.1.1 Some theoretical observations

In a typical Bayesian analysis, we would give prior distributions to $(sens_{s0}, spec_{s0}, sens_{s1}, spec_{s1})$ for each of the studies s, and examine the posterior distribution of θ. (For single case-control studies, this approach has been used in Gustafson *et al.* (2001), Chu *et al.* (2006, 2010), among others.) In the approach of this thesis, however, we rather assign a feasible region to ($sens_{s0}$, $spec_{s0}$, $sens_{s1}$, and $spec_{s1}$) and search through this region for the most extreme Bayesian posterior inference. Before we examine this with an example, let us consider where we might expect to find these extreme inferences.

Rearranging equation (4.5) as:

$$\pi = \frac{p - (1 - spec)}{sens + spec - 1} \tag{4.13}$$

and differentiating π with respect to $sens$, we find that:

$$\frac{\partial \pi}{\partial sens} = \frac{1 - spec - p}{(sens + spec - 1)^2} \begin{cases} > 0 & \text{if } spec < 1 - p \\ = 0 & \text{if } spec = 1 - p \\ < 0 & \text{if } spec > 1 - p \end{cases} \tag{4.14}$$

Moreover, it is easy to show that when $sens + spec > 1$, $spec \geq 1 - p$.[1] The condition $sens + spec > 1$ can be translated as

$$Pr(\text{Truly exposed}|\text{Observed to be exposed}) > Pr(\text{Truly exposed}|\text{Observed to be unexposed})$$

(see Appendix C.2) i.e. that those who are observed to be exposed are more likely to be truly exposed. If this is true, π will decrease with increasing $sens$ (since $\frac{\partial \pi}{\partial sens} \leq 0$).

Rearranging equation (4.5) as:

$$\pi = 1 + \frac{p - sens}{sens + spec - 1} \tag{4.15}$$

[1] This can be seen by rearranging (4.5) as: $spec = 1 - p + \pi(sens + spec - 1)$. Because $0 \leq \pi \leq 1$, when $sens + spec > 1$, $spec \geq 1 - p$

and differentiating π with respect to *spec*, we find that:

$$\frac{\partial \pi}{\partial spec} = \frac{sens - p}{(sens + spec - 1)^2} \begin{cases} > 0 & \text{if } sens > p \\ = 0 & \text{if } sens = p \\ < 0 & \text{if } sens < p \end{cases} \tag{4.16}$$

Again, it can be shown that when $sens + spec > 1$, $sens \geq p$.[2] Therefore we would expect π to increase with increasing *sens* (since $\frac{\partial \pi}{\partial spec} \geq 0$). For large θ_s, we would expect π_{s1} to be large and π_{s0} to be small, since $\theta_s = \text{logit}\, \pi_{s1} - \text{logit}\, \pi_{s0}$. Thus, for large $\hat{\boldsymbol{\Theta}}$, we would expect $sens_{s1}$ to be near its lower limit, $sens_{s0}$ to be near its upper limit, $spec_{s1}$ to be near its upper limit and $spec_{s0}$ to be near its lower limit. For small $\hat{\boldsymbol{\Theta}}$, we would expect the reverse. However, the exact relationship between $\hat{\boldsymbol{\Theta}}$ and $\{sens_{s1}, sens_{s0}, spec_{s1}, spec_{s0} : s = 1 \ldots 14\}$ is considerably more complex than that between θ and $\{sens_{s1}, sens_{s0}, spec_{s1}, spec_{s0} : s = 1 \ldots 14\}$. Thus a search algorithm is needed to identify the exact extreme locations.

4.2 The Greenland meta-analysis example

Before we can apply the search algorithm to search through the space of $sens_{s1}, sens_{s0}, spec_{s1}$, and $spec_{s0}$ to find the maximum/minimum $\hat{\boldsymbol{\Theta}}$ in the Greenland example, a number of issues need to be addressed. First of all, we need to specify a suitable feasible region for $sens_{s1}, sens_{s0}, spec_{s1}, spec_{s0}$. Occasionally, for some exposure, studies have been conducted to estimate the values of these parameters. This is not the case for EMF exposure. However, we are not completely ignorant of the likely values of these parameters. As shown in the previous section, if those who are classified as exposed are more likely to be truly exposed than those who are classified as non-exposed, we have $sens + spec > 1$. This reduces the feasible region of $sens_{s1}, sens_{s0}, spec_{s1}, spec_{s0}$ by exactly one half. Furthermore, it is also shown in the previous section that when $sens + spec > 1$,

$$spec > 1 - p \tag{4.17}$$

$$sens > p \tag{4.18}$$

where p is the prevalence of *mis*-classified exposure, which can be estimated from the data. For exposure with low (misclassified) prevalence, this reduces the feasible region of *spec* considerably.

In my Master's thesis (Mak, 2008), I further examined whether we can make use of studies which estimated the correlation between various different measures of EMF levels to estimate the misclassifica-

[2] This can be seen by rearranging (4.5) as: $sens = p + (1 - \pi)(sens + spec - 1)$. Because $0 \leq 1 - \pi \leq 1$, when $sens + spec > 1$, $sens \geq p$

tion parameters (Appendix C). It was found that the misclassification parameters cannot be estimated to any degree of precision, although the correlation between various measurements of EMF and the true exposure levels experienced by the child is more likely to be low rather than high. From this, we may further infer that $sens + spec$ is probably not close to 2 (see Appendix C). Furthermore, since $sens + spec = 1$ corresponds to the scenario where someone who is observed to be exposed is no more likely to be truly exposed than someone who is observed to be unexposed, it is likely that $sens + spec$ would be greater than 1. For these reasons, I believe that reasonable bounds for $sens + spec$ might be:

$$1.05 \leq sens_{si} + spec_{si} \leq 1.7 \qquad i = 0, 1 \quad s = 1 \ldots 14 \tag{4.19}$$

Based on equations (4.17), (4.18), I gave lower limits to $sens_{si}$ and $spec_{si}$ using the standard 95% confidence limits for p_{si} (denoted $[p_{si}^-, p_{si}^+]$), i.e.:

$$spec_{si} > 1 - p_{si}^+ \tag{4.20}$$

$$sens_{si} > p_{si}^- \tag{4.21}$$

Together with equation (4.19), I derived the intervals for $sens_{s1}, sens_{s0}, spec_{s1}$, $spec_{s0}$, and the exact values are given in Table 4.1.

Finally, it may be expected that $sens_{s1}$ and $spec_{s1}$ do not differ very much from $sens_{s0}$ and $spec_{s0}$, since in most studies, methods for assessment of exposure are similar for cases and controls. When $sens_1 = sens_0$ and $spec_1 = spec_0$, we have the non-differential misclassification assumption, whose property was well known since at least Bross (1954). In the present situation, however, because in some studies cases provide better access for exposure measurement than controls, and also because exposure is a dichotomization of a continuous exposure measurement, non-differential misclassification is believed not to hold exactly. Nonetheless, I believe the following might be reasonable bounds for these parameters.

$$|sens_{s1} - sens_{s0}| \leq 0.05 \tag{4.22}$$

$$|spec_{s1} - spec_{s0}| \leq 0.02 \tag{4.23}$$

$$s = 1 \ldots 14 \tag{4.24}$$

Here I have given a wider possible degree of non-differentiality for $sens$ than for $spec$, because the range of $spec_{si}$ can be known more precisely than the range of $sens_{si}$ through p_1 and p_0, as shown in Table 4.1.

Index	p_0 (95% CI)	p_1 (95% CI)	$sens_0$	$spec_0$	$sens_1$	$spec_1$	Prior distribution of logit π_0	Prior 95% interval of π_0
1	0 (0, .064)	.018 (.0005, .096)	[.05, .764]	[.936, 1]	[.05, .796]	[.904, 1]	N(-6.335,1)	[.0002, .012]
2	0 (0, .044)	.034 (.0072, .097)	[.05, .744]	[.956, 1]	[.05, .797]	[.903, 1]	N(-6.335,1)	[.0002, .012]
3^H	.04 (.025, .06)	.158 (.06, .313)	[.05, .76]	[.94, 1]	[.06, 1]	[.7, 1]	N(-4.602,1)	[.0014, .066]
4^H	.022 (.012, .037)	.035 (.018, .062)	[.05, .737]	[.963, 1]	[.05, .762]	[.938, 1]	N(-5.218,1)	[.0008, .037]
5^H	.045 (.03, .065)	.066 (.048, .088)	[.05, .765]	[.935, 1]	[.05, .788]	[.912, 1]	N(-4.472,1)	[.0016, .075]
6^H	.07 (.034, .125)	.105 (.062, .163)	[.05, .825]	[.875, 1]	[.062, .863]	[.837, 1]	N(-4.029,1)	[.0025, .112]
7^H	.033 (.017, .059)	.047 (.026, .078)	[.05, .759]	[.941, 1]	[.05, .778]	[.922, 1]	N(-4.776,1)	[.0012, .056]
8	.014 (.0053, .031)	.034 (.013, .073)	[.05, .731]	[.969, 1]	[.05, .773]	[.927, 1]	N(-5.617,1)	[.0005, .025]
9	.0018 (.0004, .0053)	.0036 (.0007, .01)	[.05, .705]	[.995, 1]	[.05, .71]	[.99, 1]	N(-7.705,1)	[.0001, .0032]
10^H	.025 (.0082, .058)	.083 (.018, .225)	[.05, .758]	[.942, 1]	[.05, .925]	[.775, 1]	N(-5.059,1)	[.0009, .043]
12	.013 (.0059, .024)	.02 (.0041, .056)	[.05, .724]	[.976, 1]	[.05, .756]	[.944, 1]	N(-5.734,1)	[.0005, .022]
13	.015 (.011, .022)	0 (0, .025)	[.05, .722]	[.978, 1]	[.05, .725]	[.975, 1]	N(-5.551,1)	[.0005, .027]
14	.0028 (.0006, .0083)	.0047 (.0015, .011)	[.05, .708]	[.992, 1]	[.05, .711]	[.989, 1]	N(-7.246,1)	[.0001, .005]
15	.016 (.0051, .036)	.031 (.0008, .162)	[.05, .736]	[.964, 1]	[.05, .862]	[.838, 1]	N(-5.541,1)	[.0006, .027]

Table 4.1: Prior distribution for π_{s0}, prevalence (p_0, p_1), and ranges of *sens* and *spec* derived using the Greenland (2005a) childhood leukaemia-EMF data. The H index denotes a high-exposure study.

4.2.1 Prior distribution for π_{s0} and θ_s

Although considerable effort has been spent to derive suitable bounds for the bias parameters $sens_{s1}$, $sens_{s0}$, $spec_{s1}$, $spec_{s0}$, there remains considerable uncertainty in these parameters. Perhaps of more importance is the fact that severe misclassification of exposure remains a possibility (severe meaning $0.5 < sens + spec < 1.5$ or so). When exposure is severely misclassified, the data will generally contain very little information on the true probability of exposure π_{s0} and π_{s1}. This can be seen by considering equation (4.13):

$$\pi = \frac{p - (1 - spec)}{sens + spec - 1} \tag{4.25}$$

If $sens$ and $spec$ were fixed, and p and π were random and unknown, then:

$$\mathrm{Var}(\pi) = \mathrm{Var}\left(\frac{p - (1 - spec)}{sens + spec - 1}\right) \tag{4.26}$$

$$= \frac{\mathrm{Var}(p)}{(sens + spec - 1)^2} \tag{4.27}$$

Thus, we see that if $sens + spec$ is close to 1, the variance of π can be many times larger than the variance of p. Information on π is thus "diluted" relative to information on p. When this is the case, the posterior distributions π_{s0} and π_{s1} tend to be greatly influenced by their prior distributions. Using flat or nearly flat priors such as $\pi_{s0} \sim U(0, 1)$ often results in posterior distributions that suggest π_{s0} is many times that of p_{s0} when applied to the Greenland data of this thesis (because p_{s0} tends to be around 0-0.05 (see Table 4.1), whereas the distribution $U(0, 1)$ has 50% of mass between 0.25 and 0.75). A classical measurement error model, however, would imply that it is more likely that $\pi_{s0} < p_{s0}$.[3] Partly for this reason, Greenland and Kheifets (2006) assigned a prior distribution of approximately $N(\mathrm{logit}\, 0.0025, 3)$ for $\mathrm{logit}\, \pi_0$ among studies which were carried out in "high-exposure countries"[4] and a prior of $N(\mathrm{logit}\, 0.001, 3)$ for all other countries.[5] This, however, suggests that the true exposure

[3] In a classical measurement error model for measurements of EMF, we may write that the observed EMF exposure levels (Z) equals their true EMF levels (ζ) plus an error term (ϵ):

$$Z = \zeta + \epsilon$$

where ϵ is typically 0 centred, and ϵ is independent of ζ. Our observed prevalence estimate ($p = Y/N$) estimates $Pr(Z > c)$, and our true prevalence is $Pr(\zeta > c)$, where c denotes a certain threshold, 0.3μT in this particular case. If ζ and ϵ were both Normal, and ϵ has 0 mean, then $Pr(Z > c)$ will always be greater than $Pr(\zeta > c)$ if $Pr(\zeta > c) < 0.5$. Since 0.3μT is quite a high threshold for exposure, $Pr(\zeta > c) \ll 0.5$, and hence we would expect the observed prevalence to be greater than the true prevalence.

[4] This includes all studies in North America as well as the Japanese study (Kabuto *et al.*, 2006) and the Feychting and Ahlbom (1993) study, the latter because it particularly selected a high-exposure population for the study.

[5] The actual prior distribution of π_0 is in fact the distribution of E_{1+0}/E_{++0}, where E_{1+0} is the expected number of truly exposed controls and E_{++0} the number of controls. $\mathrm{logit}\, \pi_0$ is parameterized as

$$\mathrm{logit}\, \frac{E_{1+0}}{E_{++0}} = \alpha_T + \log(1 + \exp(\alpha_X + \alpha_{TX})) - \log(1 + \exp(\alpha_X))$$

prevalence is some several hundred times lower than the observed.[6] This does not seem reasonable and therefore, in this thesis, I elicited my own prior distributions for π_{s0}, where $\text{logit}\,\pi_{s0} \sim N(\text{logit}\,\mu_s, 1)$, and μ_s is taken to be 1/4 of the observed prevalence estimate. A variance of 1 for the distribution means that the probability of π_{s0} being less than the observed prevalence estimate $\hat{p}_{s0} = Y_{s0}/N_{s0}$ is around 0.91 if $\hat{p}_{s0}$ is close to 0, i.e. when we can approximate $\text{logit}\,p_{s0}$ with $\log p_{s0}$. This is because I agree with Greenland and Kheifets that the true prevalence is likely to be less than the observed prevalences (based on the classical measurement error model), but not by as much as they suggested. For studies 1 and 2, because the observed prevalence estimate is 0, I used the average of $\text{logit}\,\mu_s$ among the low-exposure studies instead for the mean. The distribution of $\text{logit}\,\pi_{s0}$ and the 95% interval for π_{s0} for each study is given in column 8 of Table 4.1.

As for the prior distribution of θ_s, I continue with:

$$\theta_s \sim N(0, 0.5) \tag{4.28}$$

as used in the previous chapter.

For a fixed-effects (FE) model, this leads to:

$$\theta \sim N(0, 0.5) \tag{4.29}$$

$$\delta_s = 0 \tag{4.30}$$

For a random-effects (RE) model, if 70% of the variance is common across studies, this leads to:

$$\theta \sim N(0, 0.35) \tag{4.31}$$

$$\delta_s \sim N(0, 0.15) \tag{4.32}$$

$$s = 1 \ldots 14 \tag{4.33}$$

as in chapter 3.

---

where $\alpha_T \sim N(\text{logit}(0.001 + h0.0015), 3)$, $\alpha_X \sim N(0, 400)$, and $\alpha_{TX} \sim N(\log 16, 2)$, $h = 0$ for low exposure studies, and $h = 1$ for high exposures studies. Following simulation, it is found that the latter terms $\log(1 + \exp(\alpha_X + \alpha_{TX}))$ and $\log(1 + \exp(\alpha_X))$ nearly cancels one another, and thus the distribution is very similar to that of α_T alone.

[6] It can be seen that these prior distributions have mean 0.0025 and 0.001 on the logit scale. The 95% prior credible intervals for these distributions are (0.000078, 0.069) and (0.000034, 0.029) for π_0. If we compare these values to the observed prevalence of exposure given in Table 4.1, we find that these values are a magnitude lower than the observed prevalence, particularly for the high exposure studies. It appears that Greenland and Kheifets consider that studies have generally overestimated the true prevalence of exposure, and possibly up to several hundred times. The extent of overestimation as suggested by Greenland and Kheifets' prior thus appears to be somewhat extreme. If the true prevalence of exposure were in the order of, say 0.0001-0.0005, then practically all of the observed exposed participants would not have been truly exposed, and we might as well discard our data completely.

4.2.2 Modification to the search algorithm

If Table 4.1 contains all of the constraints we have over the parameters $sens_{s1}, sens_{s0}, spec_{s1}$, and $spec_{s0}$, then we can apply the algorithm of chapter 2 to search through the feasible space for the most extreme posterior distribution of θ, given the prior distributions of Table 4.1 for π_{s0} and (4.31), (4.32), (4.29), (4.30) for θ and δ_s. However, since there is still much uncertainty over the parameters within this region, it is very helpful if we can incorporate the constraints that concern $sens + spec$ (4.19), and those that limit the difference between *sens* and *spec* between cases and controls (4.22), (4.23), i.e.:

$$1.05 \leq sens_{si} + spec_{si} \leq 1.7 \qquad i = 0, 1 \tag{4.34}$$

$$|sens_{s1} - sens_{s0}| \leq 0.05 \tag{4.35}$$

$$|spec_{s1} - spec_{s0}| \leq 0.02 \tag{4.36}$$

$$s = 1 \ldots 14 \tag{4.37}$$

to further narrow the feasible space.

The algorithm of chapter 2, as specified, cannot deal with these constraints that involve more than one parameter at a time. However, it appears that a simple modification can be made to the algorithm, which would enable it to search through the feasible region. This modification is described below together with a heuristic explanation. It is noted that a rigorous proof of the convergence of the algorithm to a stationary point is not yet available.

1. First, let us consider the cyclic coordinate method (CCM) as used in the previous chapter and described in Box 2.1, summarized here in vector notation. Thus, in iteration n, we seek a new vector $\boldsymbol{\eta}^{[n]}$ by updating our old vector $\boldsymbol{\eta}^{[n-1]}$ by choosing λ which minimizes $f(\boldsymbol{\eta}^{[n]})$, where:

$$\boldsymbol{\eta}^{[n]} = \boldsymbol{\eta}^{[n-1]} + \lambda \boldsymbol{\Delta} \tag{4.38}$$

where $\boldsymbol{\Delta}$ can be any one of the following *search direction vectors*:

$$\boldsymbol{\Delta} = \begin{pmatrix} \Delta[\eta_1] \\ \Delta[\eta_2] \\ \vdots \\ \Delta[\eta_{R-1}] \\ \Delta[\eta_R] \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \text{ or } \cdots \text{ or } \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{4.39}$$

In terms of parameters $sens_{s0}$, $sens_{s1}$, $spec_{s0}$, $spec_{s1}$ the search direction vectors are:

$$\boldsymbol{\Delta} = \begin{pmatrix} \Delta[sens_{11}] \\ \Delta[sens_{10}] \\ \vdots \\ \Delta[spec_{N1}] \\ \Delta[spec_{N0}] \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \text{ or } \cdots \text{ or } \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{4.40}$$

2. The presence of constraints (4.34), (4.35), and (4.36), i.e.:

$$1.05 \leq sens_{si} + spec_{si} \leq 1.7 \qquad i = 0, 1 \tag{4.41}$$

$$|sens_{s1} - sens_{s0}| \leq 0.05 \tag{4.42}$$

$$|spec_{s1} - spec_{s0}| \leq 0.02 \tag{4.43}$$

however, prevents us from searching through the feasible region of the these parameters using the directions (4.40) alone. (For an illustration, see Figure 4.1.) To overcome this problem, we include some additional search directions to optimize along such that for each s, the search direction vectors are:

$$\begin{pmatrix} \vdots \\ \Delta[spec_{s-1\,0}] \\ \Delta[sens_{s1}] \\ \Delta[sens_{s0}] \\ \Delta[spec_{s1}] \\ \Delta[spec_{s0}] \\ \Delta[sens_{s+1\,1}] \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 0 \\ \vdots \end{pmatrix}, \tag{4.44}$$

$$\begin{pmatrix} \vdots \\ 0 \\ 1 \\ 1 \\ -1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 1 \\ -1 \\ -1 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ \vdots \end{pmatrix}$$

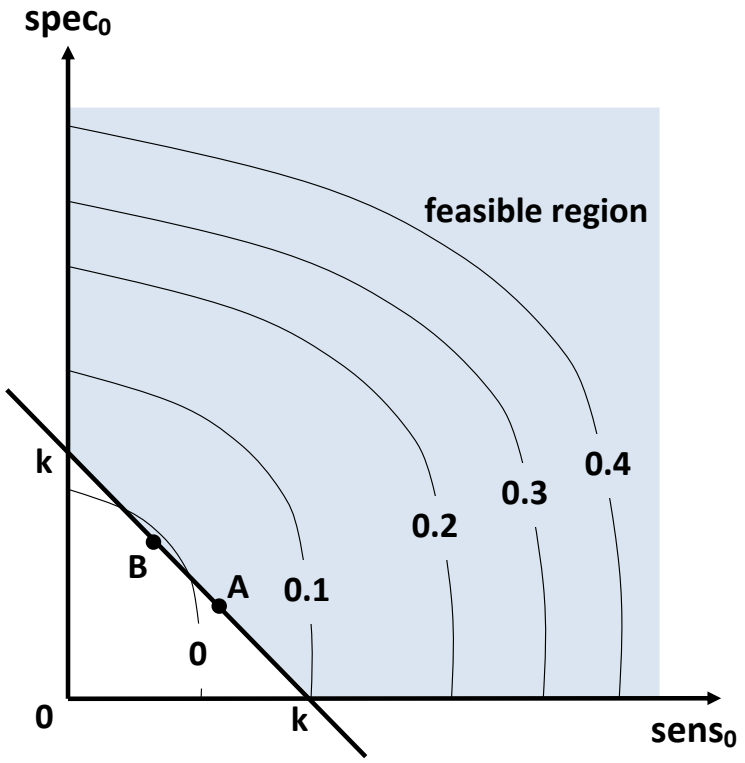


Figure 4.1: Consider the minimization of the function $f(sens_0, spec_0)$ subject to $sens_0 + spec_0 \geq k$. In the figure, the contour lines show the value of $f(sens_0, spec_0)$ and the true minimum is at B, and let us suppose that the algorithm has reached the point A. It is now not possible to improve at point A by searching in the direction of $sens_0$ and $spec_0$ alone because decreasing $sens_0$ would violate the constraint while increasing $spec_0$ does not decrease the function.

This can be thought of as an *active set* search strategy, where we try to optimize the function within the subspace defined by the active constraints (Nocedal and Wright, 1999, p.455). (A constraint $c_i(\boldsymbol{\eta}) \leq 0$ is active at $\boldsymbol{\eta}^*$ if $c_i(\boldsymbol{\eta}^*) = 0$.) For example, searching along the direction

$$\begin{pmatrix} \Delta[sens_1] \\ \Delta[sens_0] \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{4.45}$$

allows us to search within the space defined by the constraint:

$$sens_1 - sens_0 = k \tag{4.46}$$

with respect to $sens_1$ and $sens_0$ without violating it (supposing the constraint is active). Again,

searching along the direction

$$\begin{pmatrix} \Delta[sens_1] \\ \Delta[sens_0] \\ \Delta[spec_1] \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \tag{4.47}$$

allows us to search within the space defined by the constraints

$$sens_1 - sens_0 = k_1 \tag{4.48}$$

$$sens_1 + spec_1 = k_2 \tag{4.49}$$

with respect to $sens_1$, $sens_0$, and $spec_1$ without violating them (supposing the constraints are active at the minimum/maximum). We, however, do not know which of the constraints will be active in advance, and hence we iteratively search along each of the constraints defined by (4.44), which cover all possibilities imposed by (4.34), (4.35), (4.36) until no improvement can be found.

4.2.3 Results

With the adjustment made to the algorithm in section 4.2.2, we can now optimize $\hat{\boldsymbol{\Theta}}$ subject to the constraints of Table 4.1 as well as the linear inequality constraints:

$$1.05 \leq sens_{si} + spec_{si} \leq 1.7 \tag{4.50}$$

$$|sens_{s1} + sens_{s0}| \leq 0.05 \tag{4.51}$$

$$|spec_{s1} + spec_{s0}| \leq 0.02 \tag{4.52}$$

Because each study now has 4 parameters, and we have an additional 9 search directions to go through (for each study, c.f. (4.44)), optimization takes considerably longer than for the problems of the previous chapters. Single-study optimizations take around 20-160 seconds each to run, and meta-analysis optimization takes from 500 up to 10000 seconds on my Intel Pentium Core 2 Quad computer. The reason the time it takes to run the optimization varies so much is partly because it is sometimes difficult to adapt the integration algorithm used to the particular posterior distribution of θ as $sens_{s0}, sens_{s1}, spec_{s0}, spec_{s1}$ vary. Because accuracy is favoured over efficiency, there are situations where the evaluation becomes inefficient. However, it is interesting to note that on a more powerful PC with the Intel Core i7-2600 processor, optimization takes only around one quarter of the time. Thus continual development in computer technology can make these computer-intensive tasks more and more feasible.

It is also found that for some of the problems in this chapter, the algorithm converges to different local optima with different starting points/search direction schedule. For single-study problems, more than 1 local optima were found in 3% of the problems. For meta-analysis problems, 29% of the problems

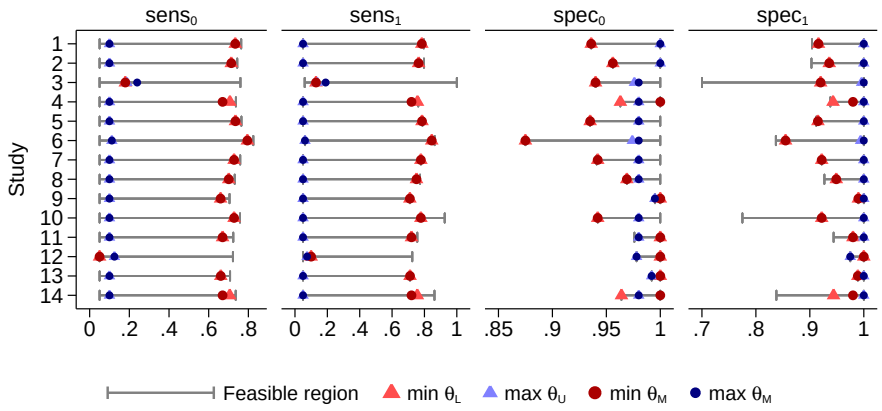


Figure 4.2: Locations of $sens_0$, $sens_1$, $spec_0$, $spec_1$ at max/min $\hat{\theta}_M$, min $\hat{\theta}_L$, and max $\hat{\theta}_U$ when studies are considered individually, with $\theta_s \sim N(0, 0.5)$

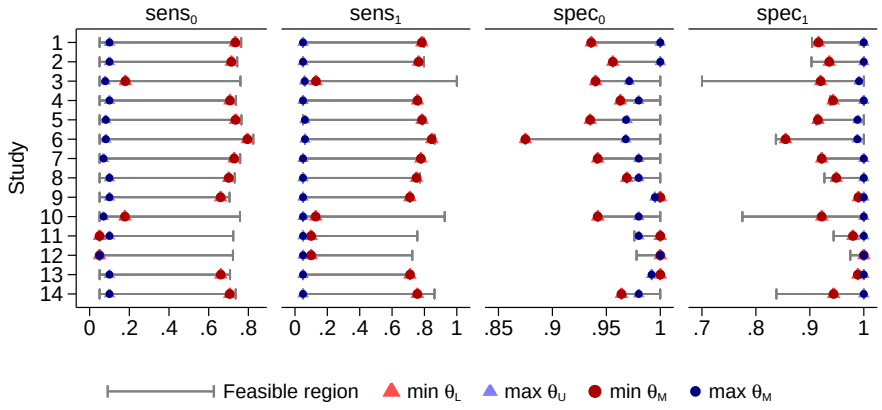


Figure 4.3: Locations of $sens_0$, $sens_1$, $spec_0$, $spec_1$ at max/min $\hat{\theta}_M$, min $\hat{\theta}_L$, and max $\hat{\theta}_U$ when studies are considered in a meta-analysis, with $\theta_s = \theta \sim N(0, 0.5)$

have more than 1 local optima, and 15% has more than 2. The largest number of local optima found for a single problem was 10. All of these local optima are tabulated in the document "Location tables.pdf" in the supplementary CD-ROM. As mentioned in chapter 2, however, it is not feasible to verify that the most extreme of the local optima are indeed the global optima, and it is conceded that the global optimum may not have been found in some of the problems. Nonetheless, for most of the local optima found, their values in $\hat{\boldsymbol{\Theta}}$ were similar. Therefore, error resulting from the failure to find the global optima is unlikely to be large.

Before we examine the FPI of this model, let us first note the locations of $sens_{s1}$, $sens_{s0}$, $spec_{s1}$, $spec_{s0}$ for the most extreme $\hat{\boldsymbol{\Theta}}$ found. When the studies are considered independently from each other, the locations are given in Figure 4.2. When in a meta-analysis, they are given in Figure 4.3. I have presented the locations for the FE model only.

In examining Figures 4.2 and 4.3, we find that min $\hat{\boldsymbol{\Theta}}$ tends to be found at $spec_0 = \min spec_0$, $sens_1$

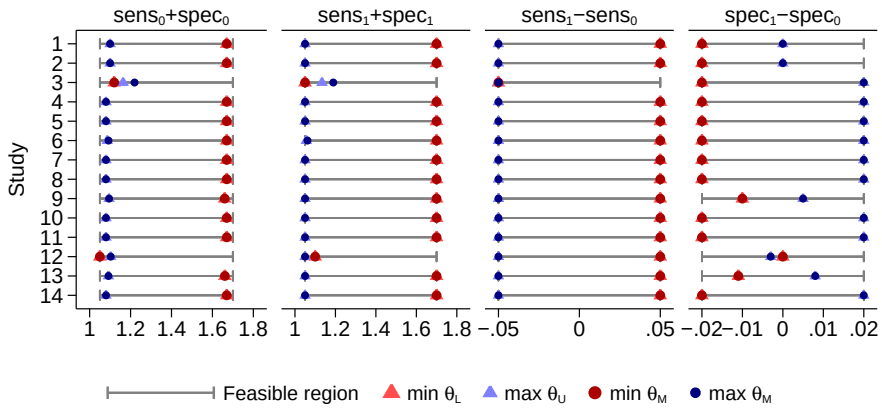


Figure 4.4: Locations of $sens_0 + spec_0$, $sens_1 + spec_1$, $sens_1 - sens_0$, $spec_1 - spec_0$ at max/min $\hat{\theta}_M$, min $\hat{\theta}_L$, and max $\hat{\theta}_U$ when studies are considered individually, with $\theta_s \sim N(0, 0.5)$

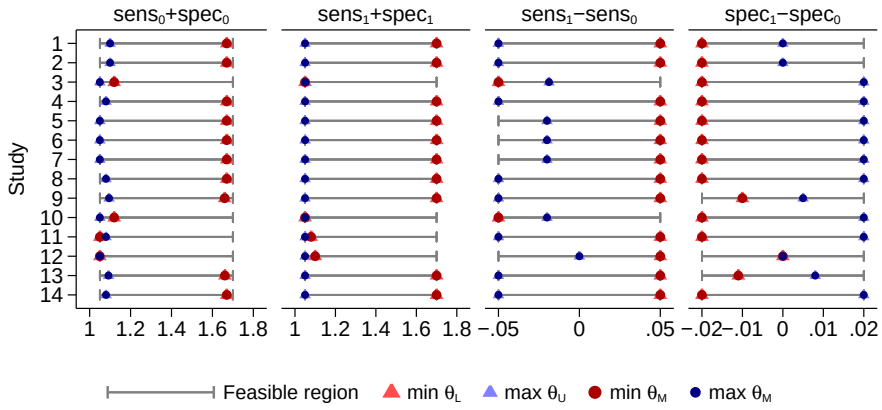


Figure 4.5: Locations of $sens_0 + spec_0$, $sens_1 + spec_1$, $sens_1 - sens_0$, $spec_1 - spec_0$ at max/min $\hat{\theta}_M$, min $\hat{\theta}_L$, and max $\hat{\theta}_U$ when studies are considered in a meta-analysis, with $\theta_s = \theta \sim N(0, 0.5)$

and $sens_0$ near their maximum, and $spec_1$ near its minimum, as would be predicted from the theories given in section 4.1.1. However, often, they are not found at these extreme points. This is partly because of the constraints between $sens_0$, $sens_1$, $spec_0$, and $spec_0$ (equations 4.50, 4.51, 4.52). If we examined the constraints, we would find that these two-variable constraints are active in almost all of the extreme points (Figures 4.4 and 4.5). The only exception appears to be $\max \hat{\theta}_M$ and $\max \hat{\theta}_U$ when study 3 is considered by itself. At these points, the location of $sens_0$ and $sens_1$ were not limited by $spec_0$ and $spec_1$, or by their own boundaries.

Occasionally, the location of $sens_0$, $sens_1$, $spec_0$, and $spec_0$ is found at the complete opposite of where we might expect them based on theories of section 4.1.1. An example is $spec_0$ for study 9. (And this was not because of a failure to locate the global maximum.) All this suggests that we are unlikely

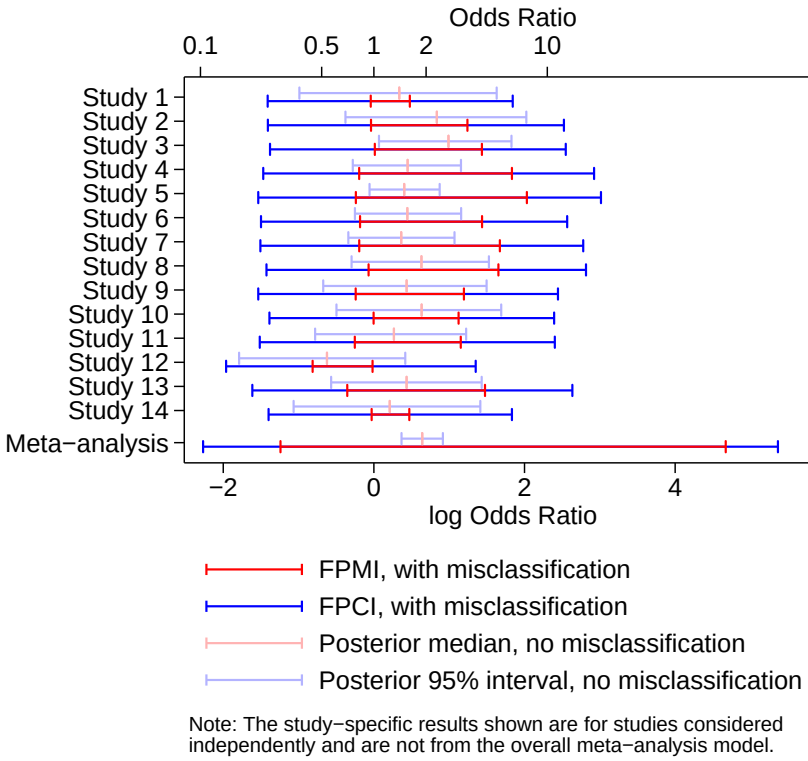


Figure 4.6: FPIs for $\hat{\boldsymbol{\Theta}}$ in a FE model with exposure misclassification adjustments.

to find a simple relationship between the location of the extreme points and the feasible region.

We turn now to the Feasible Posterior Intervals (FPIs), which are graphed in Figure 4.6 and 4.7 for the FE and the RE model, respectively. Again, FPIs are given in both the case when studies are considered independently and together in a meta-analysis. These FPIs are contrasted with the posterior credible interval when no exposure misclassification bias is assumed, i.e. when $sens_{si} = spec_{si} = 1, s = 1 \ldots 14, i = 0, 1$.

Here, we note that not only was there the added uncertainty due to bias, as reflected by the wider Feasible Posterior Median Intervals (FPMIs), but the uncertainty due to randomness was also increased, as reflected by the increased length of the Feasible Posterior Credible Intervals (FPCIs) on top of the increase in the width of the FPMI, particularly in studies with more data. This can be explained by the fact that misclassification essentially "dilutes" data (c.f. equation 4.27), such that the effective amount of data is reduced. Moreover, we see that large positive values for θ are much more possible than large negative values, for all studies except study 12. This reflects the impact of the near-non-differential misclassification constraint of (4.22) and (4.23), since the effect of non-differential misclassification is to bias estimates towards the null in classical estimates (Bross, 1954), and study 12 is the only study that

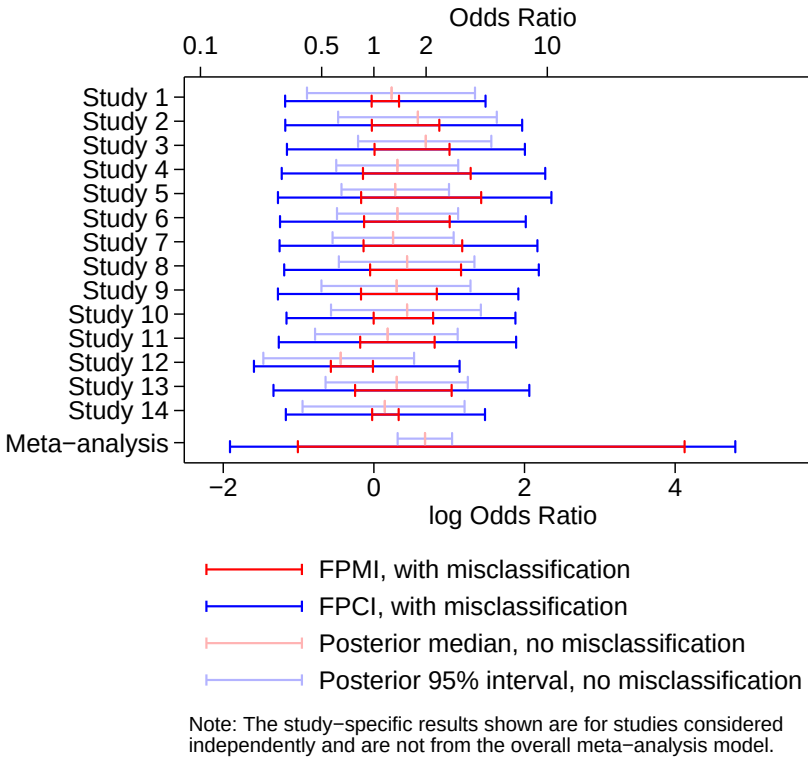


Figure 4.7: FPIs for $\hat{\boldsymbol{\Theta}}$ in a RE model with exposure misclassification adjustments

has an estimate of odds ratio less than 1.[7]

Perhaps more striking is the fact that the overall meta-analytic FPMI is now so wide that it exceeds the widths of the FPCI for all individual studies. This parallels the situation in the previous chapter where the uncertainty in bias is severe ($-2 \leq \eta_s \leq 2$) (c.f. Figure 3.10). There, I explained that two factors were at work. First, there are more data in the meta-analysis, so there is more information to overcome the information in the prior distribution. Secondly, the extremes in the meta-analytic situation are "more extreme" than the extremes in the single-study situation. I believe this is still the case here. An additional issue when faced with such a wide FPI is that there is evident conflict between the prior distribution and the likelihood when $\boldsymbol{\eta}$ takes on some of the values within $\mathcal{E}$. For example, in Figure 4.3, we see that at $\max \hat{\boldsymbol{\Theta}}$, $spec_{s1}$ was high (near 1), and $sens_{s1}$ and $sens_{s0}$ were low for all studies, with $sens_{s0} > sens_{s1}$. These misclassification probabilities together with the data (which gives estimates of p_{s1} and p_{s0} close to 0) give much weight to π_{s1} close to 1 and much weight to π_{s0} close to 0, resulting in a

[7]The reader may also observe that for the analysis *without* bias, the meta-analytic posterior median and 95% credible interval for the overall log odds ratio is considerably more to the right than study-specific posterior median and 95% credible interval. This is in fact an artifact due to the prior distribution we give for π_{s0} in this chapter, which has mean 4 times less than that of $\hat{p}_{s0}$, where $\hat{p}_{s0}$ is the observed prevalence of exposure in the controls. When there is no misclassification, this prior does not make sense, since $p_{s0} = \pi_{s0}$. However, I have decided to show these results alongside the uncertainty intervals *with* misclassification bias, to highlight the difference the bias parameters make to the posterior distribution.

likelihood which favours very positive values of $\theta_s = \text{logit}\,\pi_{s1} - \text{logit}\,\pi_{s0}$ (whereas the prior distribution of $N(0, 0.5)$ for θ places nearly 0 weight on these extremely positive values of θ). In a single-study setting, the conflict between the prior distribution and the likelihood is perhaps not great since the likelihood is probably relatively flat as compared with the prior. In a meta-analysis, this is no longer the case, and we see that even though our prior distribution ($\theta \sim N(0, 0.5)$) gives very little weight to values of θ greater than 2 or so, the posterior has so much weight on positive values of θ that the overall posterior median for θ is around 4. In view of this, it is natural to want to "re-specify" our feasible region for $\boldsymbol{\eta}$ such that it does not conflict as much with the prior distribution. However, this may be difficult, since the conflict only appears to arise in the meta-analysis situation, and not so much in the single-study situation.

4.3 Sensitivity Analyses

However, even when the approach of this thesis is unsatisfactory as a means of quantifying *overall* uncertainty about θ, it may still be used as a tool for sensitivity analysis. In the current situation, for example, we may instead ask: What values do the bias parameters have to take in order that the data may give us useful information over the possible range of θ? What if, for example, the *sens* and *spec* were in fact all towards the higher ends of our uncertainty intervals? What if non-differential misclassification were exact rather than near? What if $sens + spec$ can be known to be high? In the following subsections, I consider these scenarios for the FE model only. The RE model has very similar patterns and so its results are not presented.

4.3.1 High, medium, and low *sens* and *spec*

Here, let us consider three different levels of *sensitivity* and three different levels of *specificity*, high, medium, and low, with their definition given in Table 4.2.

The definitions of low, medium, and high sensitivity and specificity are derived by equally dividing the shared range of $sens_0$ and $sens_1$, and $spec_0$ and $spec_1$ into three parts. In Figure 4.8, I graphed the changes in FPI of θ for the meta-analysis in relation to these different sensitivity and specificity levels. Here, we see that increasing specificity tends to increase the width of the FPMI, whereas increasing sensitivity tends to decrease its width. As a result, the widest FPMI and FPCI is observed when specificity is high and sensitivity is low. The shortest uncertainty interval is observed when specificity is low, and when this is the case, varying the sensitivity does not appear to affect the FPI very much. One observation which may partly explain the pattern is that when specificity is low, the information afforded by the data is little. As a result, the prior distribution still has great influence over the posterior distribution, whatever values the bias parameters take. As a result, the posterior median cannot depart

	sens			*spec*		
Index	Low	Medium	High	Low	Medium	High
1	[.05, .288]	[.288, .526]	[.526, .764]	[.936, .957]	[.971, .985]	[.985, 1]
3	[.06, .293]	[.293, .527]	[.527, .76]	[.94, .96]	[.96, .98]	[.98, 1]
4	[.05, .279]	[.279, .508]	[.508, .737]	[.963, .975]	[.975, .988]	[.988, 1]
5	[.05, .288]	[.288, .527]	[.527, .765]	[.935, .957]	[.957, .978]	[.978, 1]
6	[.062, .316]	[.316, .571]	[.571, .825]	[.875, .917]	[.917, .958]	[.958, 1]
7	[.05, .286]	[.286, .523]	[.523, .759]	[.941, .961]	[.961, .98]	[.98, 1]
8	[.05, .277]	[.277, .504]	[.504, .731]	[.969, .979]	[.979, .99]	[.99, 1]
9	[.05, .268]	[.268, .487]	[.487, .705]	[.995, .997]	[.997, .998]	[.998, 1]
10	[.05, .286]	[.286, .522]	[.522, .758]	[.942, .961]	[.961, .981]	[.981, 1]
11	[.05, .275]	[.275, .499]	[.499, .724]	[.976, .984]	[.984, .992]	[.992, 1]
12	[.05, .274]	[.274, .498]	[.498, .722]	[.978, .985]	[.985, .993]	[.993, 1]
13	[.05, .269]	[.269, .489]	[.489, .708]	[.992, .995]	[.995, .997]	[.997, 1]
14	[.05, .279]	[.279, .507]	[.507, .736]	[.964, .976]	[.976, .988]	[.988, 1]

Table 4.2: Definition of low, medium, and high *sensitivity* and *specificity*

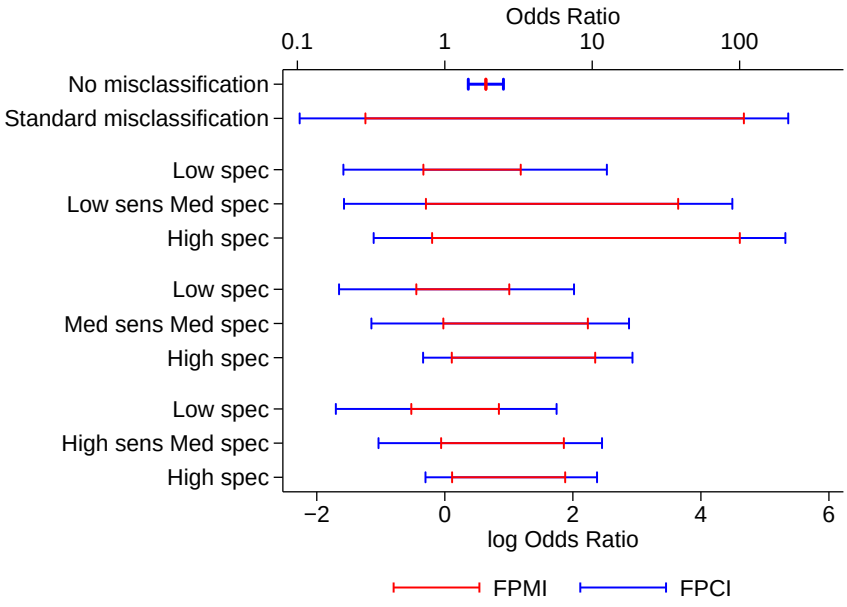


Figure 4.8: Change in the possible range of $\hat{\Theta}$ in relation to changes in sensitivity and specificity.

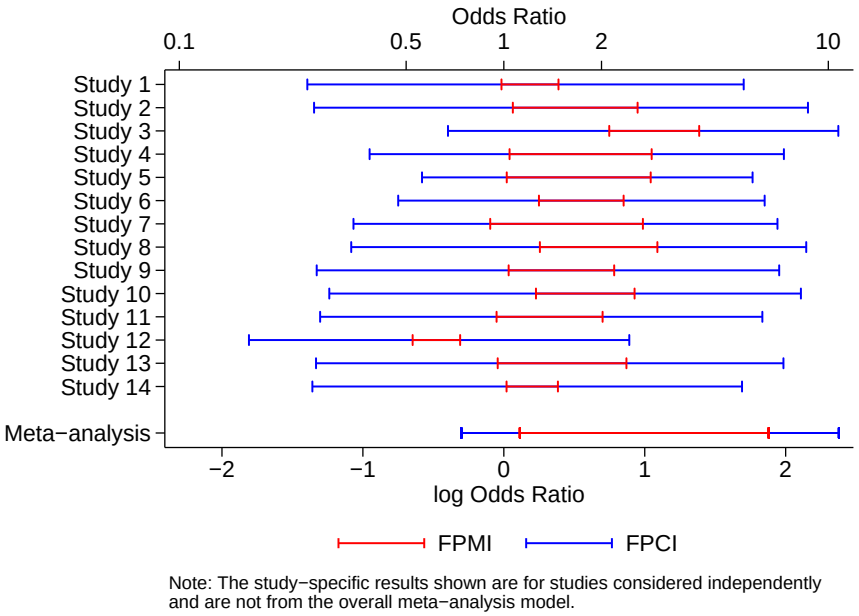


Figure 4.9: FPIs for the high *sens*, high *spec* scenario, comparing independent study inference to the meta-analytic inference for θ

greatly from the prior median, which is 0.

It may be interesting to compare the meta-analytic FPI with the individual-study FPI for the various sensitivity and specificity scenario. The comparison for the high *sens*, high *spec* scenario is given in Figure 4.9.

Here, we note that although the meta-analytic FPCI is of a similar width to the study-specific FPCI, it is altogether more positive. This is because in the meta-analysis setting, the pooling of data affords the likelihood more weight to drive the posterior away from the prior. Therefore, if *sens* and *spec* are high, the combined information from the studies gives stronger evidence for a positive θ than when studies are considered independently.

The same, however, cannot be said in a situation where *spec* is low. When this is the case, there is very little information in the data, and the increase in FPMI in the meta-analytic setting more than offsets the precision gained in the pooling of data. Figure 4.10 shows this for the case when *spec* is low and *sens* is high, but the same pattern is observed for other *sens* levels.

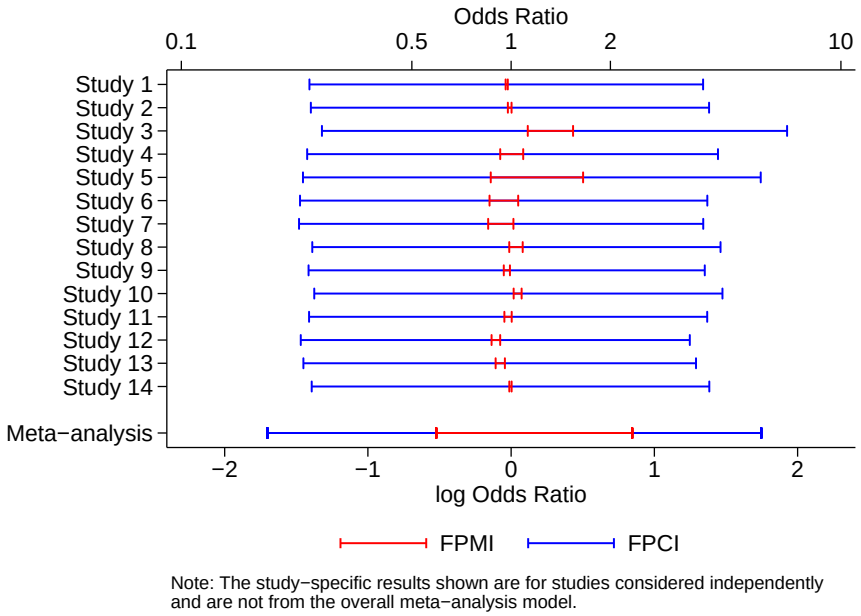


Figure 4.10: FPIs for the high *sens*, low *spec* scenario, comparing independent study inference to the meta-analytic inference for θ

4.3.2 Non-differential/differential misclassification

In this subsection, I examine whether imposing the assumption of strict non-differential misclassification affects the FPIs materially. Recall that in the above model, it was assumed that:

$$|sens_1 - sens_0| \leq 0.05 \tag{4.53}$$

$$|spec_1 - spec_0| \leq 0.02 \tag{4.54}$$

(c.f. equations 4.22, 4.23). Strict non-differential misclassification requires:

$$sens_1 = sens_0 \tag{4.55}$$

$$spec_1 = spec_0 \tag{4.56}$$

The FPIs under this assumption are given in Figure 4.11, which contrasts the FPIs under the strict non-differential misclassification assumption with that under the assumptions of (4.53) and (4.54). It can be seen that the FPMIs are shortened both in the individual study case and the meta-analysis case. The reduction in width is more dramatic in the meta-analysis situation, such that the lower limit of the FPMI is now almost 0. This indicates even small changes in the allowable degree of differential misclassification can have large effects on the range of possible posterior inference. Nonetheless, the lower limit of the FPCI remains negative, and a negative relationship between EMF and childhood leukaemia remains a possibility.

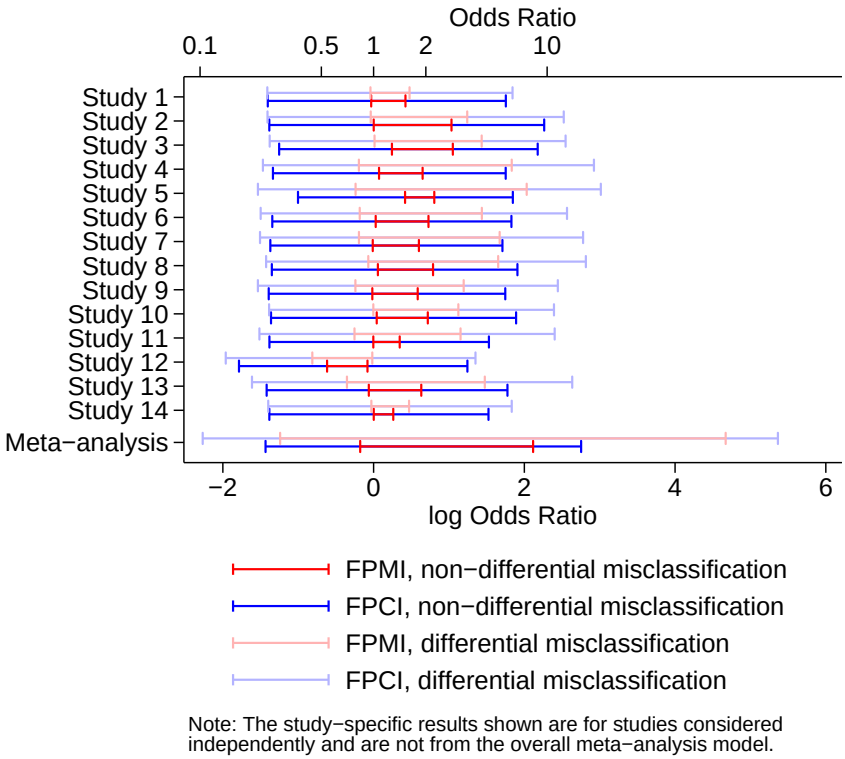


Figure 4.11: FPIs for non-differential misclassification, compared to the standard differential misclassification constraints (4.53, 4.54) of this chapter, for a FE model

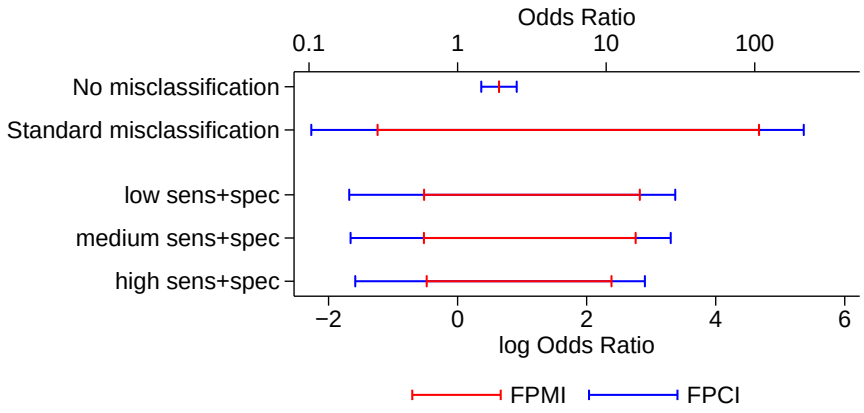


Figure 4.12: Change in the possible range of $\hat{\boldsymbol{\Theta}}$ in relation to changes $sens_1 + spec_1$ and $sens_0 + spec_0$ in the FE meta-analysis model

4.3.3 High, medium, and low $sens + spec$

Another sensitivity analysis we can conduct is with respect to $sens + spec$. As discussed in Appendix C, $sens + spec$ is related to the correlation of the raw (uncategorized) true and measured exposure levels. We are unable to estimate the correlation precisely, and therefore in the above, we use a wide range for $sens + spec$, from 1.05 to 1.7. In Figure 4.12, we can examine how $sens + spec$ affects our uncertainty over θ.

Interestingly, as $sens + spec$ becomes larger, the upper bound of $\hat{\boldsymbol{\Theta}}$ becomes smaller, although the lower bound remains roughly the same. Hence, we do not have more evidence that θ is greater than zero even if $sens + spec$ is known to be near its upper bound of 1.7.

In Figures 4.13, I compare the FPIs for the individual study scenario and the meta-analysis scenario for the case where $sens + spec$ is near the upper end, i.e. with $1.5 \leq sens_{si} + spec_{si} \leq 1.7, s = 1 \ldots 14, i = 0, 1$. It can be seen that even in this scenario the meta-analytic FPCI remain considerably wider than the study-specific FPCI. Thus a negative meta-analytic θ remains a strong possibility even when $sens + spec$ is high. This contrasts with results from section 4.3.1, which shows that if *both* sensitivity and specificity are high, the data give evidence for a positive θ.

4.3.4 General observations

In conventional sensitivity analyses for bias models, we assign specific values to the unknown parameters $\boldsymbol{\eta}$ and examine how statistical inference to the target parameter θ changes with $\boldsymbol{\eta}$. By using the methods of this thesis, however, we can assign feasible regions to $\boldsymbol{\eta}$, and examine the maximum and minimum possible statistical inference within the feasible region. This overcomes the major limitation of traditional sensitivity analyses — that if there are too many bias parameters, it becomes difficult to explore all the different possibilities.

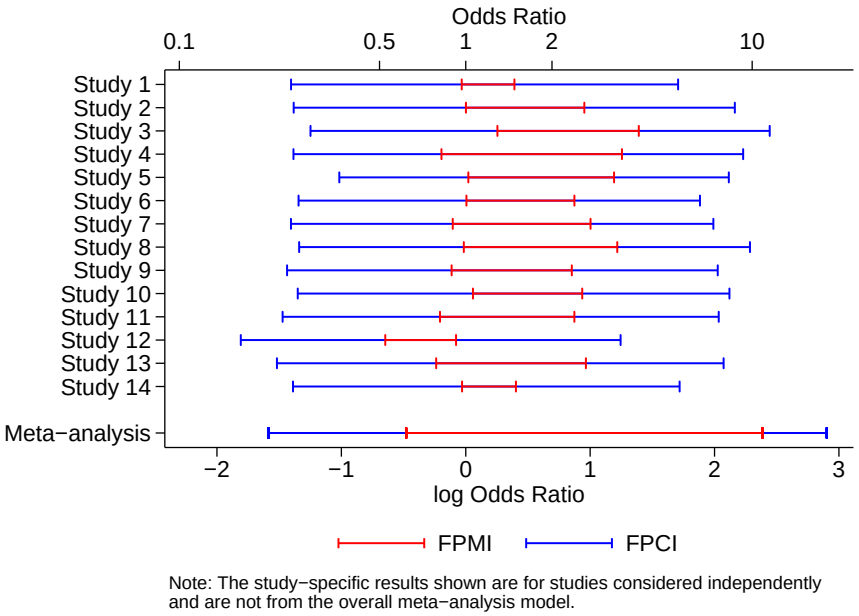


Figure 4.13: FPIs for the situation where $1.5 \leq sens_{si} + spec_{si} \leq 1.7, s = 1 \ldots 14, i = 0, 1$, in the meta-analysis setting and in the individual study setting

However, sometimes sensitivity analyses carried out in this way also display the phenomenon we saw earlier in the standard analyses — that the meta-analytic FPIs become wider than the study-specific FPIs. In general, a wide FPCI straddling across 0 is not very useful. It does not give evidence for a positive or a negative association, neither does it give evidence for a lack of association. Thus we are not able to say whether particular values for $\boldsymbol{\eta}$ are associated with positive or negative inference for θ. To overcome this problem, we may want to further restrict the range of values given to $\boldsymbol{\eta}$. However, this becomes unrealistic as it becomes more and more unlikely that all bias parameters take values within the feasible region as that region becomes more and more narrow.

Summary:

- Consideration of bias due to exposure misclassification poses a number of additional difficulties for the method of this thesis. Firstly, estimates for sensitivity and specificity are often not available. Moreover, if severe misclassification is a possibility, prior distributions for π_{s0} and θ_s must be chosen carefully.

- When misclassification probabilities are poorly known, it is useful to restrict the parameter space of $sens_{s1}$, $spec_{s1}$, $sens_{s0}$, $spec_{s0}$ through some additional constraints, such as near-non-differential misclassification, or use correlation estimates to give bounds to $sens+spec$. Using these constraints that include more than one parameter, however, requires modifying the search algorithm of chapter 2 to search through more directions.

- Multi-modality also becomes an issue searching through the feasible region of $sens_{s1}$, $spec_{s1}$, $sens_{s0}$, $spec_{s0}$ for max/min $\hat{\boldsymbol{\Theta}}$. There are typically more modes in a meta-analysis (with more parameters and hence a larger parameter space) than for single studies.

- Even when the technical aspects of the method are overcome, under severe misclassification, uncertainty due to bias is often much greater than that due to randomness. Because the method of this thesis does not allow the pooling of information to reduce uncertainty due to bias, the meta-analytic FPI for θ can remain very wide.

- Nonetheless, the method is also useful as a tool for sensitivity analysis, and we can consider different sub-space within the original feasible region of $\boldsymbol{\eta}$ and see how inference for θ changes with $\boldsymbol{\eta}$. This extends conventional sensitivity analyses in that we are able to consider *ranges* of values rather than unique values for $\boldsymbol{\eta}$ and therefore reduce the number of sensitivity analyses that need to be done.

- In this chapter, it was found that only when specificity and sensitivity are both high, the data give evidence for a positive relationship between EMF and childhood leukaemia using the Greenland data. It was also found that a small degree of differential misclassification ($|sens_1 - sens_0| < 0.5, |spec_1 - spec_0| < 0.02$) can render the FPI significantly wider than under strict non-differential misclassification.

Chapter 5

Accounting for uncertainty due to non-participation bias

Many studies require informed consent from the participants. However, typically not all invited to participate would participate. This can be due to unavailability or refusal. If those who are unavailable or refuse differ in characteristics from those who participate, then non-participation bias is likely to arise. The literature on adjustment for bias due to non-response is almost entirely dedicated to one of two models: the *pattern mixture* model (e.g. Hansen and Hurwitz, 1946; Birnbaum and Sirken, 1950; Little, 1993) and the *selection* model (e.g. Politz and Simmons, 1949; Rubin, 1977; Greenland, 2005a; Geneletti *et al.*, 2009). If $X = \{X_1, X_2, \ldots, X_n\}$ denote the data with no non-response, and $S_i = 0, 1$ denotes whether X_i is observed or not, then the pattern-mixture model considers the likelihood of X and S as:

$$p(X, S|\theta, \boldsymbol{\eta}) = p(S|\boldsymbol{\phi_1^{(pm)}})p(X|S, \boldsymbol{\phi_2^{(pm)}}) \tag{5.1}$$

while the selection model considers the likelihood as:

$$p(X, S|\theta, \boldsymbol{\eta}) = p(X|\boldsymbol{\phi_1^{(s)}})p(S|X, \boldsymbol{\phi_2^{(s)}}) \tag{5.2}$$

where $\boldsymbol{\phi_1^{(pm)}}$, $\boldsymbol{\phi_2^{(pm)}}$, $\boldsymbol{\phi_1^{(s)}}$, $\boldsymbol{\phi_2^{(s)}}$ are generally different parameters. The selection model has the advantage that the parameter of interest (i.e. θ in our case) can generally be represented as a parameter within $\boldsymbol{\phi_1^{(s)}}$, whereas in the pattern-mixture model, the parameter usually has to be written as a function of parameters from both $\boldsymbol{\phi_1^{(pm)}}$ and $\boldsymbol{\phi_2^{(pm)}}$. Moreover, in cases where X_i is multivariate, i.e. $X_i = \{X_{i1}, X_{i2}, \ldots, X_{ip}\}$, and where X_i can be partially observed, e.g. when $S_{i1} = 1, S_{i2} = 0$, it may sometimes be reasonable to assume $Pr(S|X, \boldsymbol{\phi_2^{(s)}})$ depends only on the observed data, i.e.: $Pr(S_i|X_i, \boldsymbol{\phi_2^{(s)}}) = Pr(S_i|X_i', \boldsymbol{\phi_2^{(s)}}), X_i' = \{X_{ij} : S_{ij} = 1\}$ (a condition known as missing at random (Rubin, 1976)), resulting in a marginal likelihood (i.e. integrating over $X'' = \{X_{ij} : S_{ij} = 0\}$) for the target parameter (assumed within $\boldsymbol{\phi_1^{(s)}}$) that does not depend on the auxiliary parameters $\boldsymbol{\phi_2^{(s)}}$ (see Geneletti *et al.*, 2009, for a

number of examples).

In the situation considered here, however, where a person's disease and exposure status are either both observed or both not observed, the missing at random assumption is equivalent to the assumption that response (or nonresponse) does not depend on exposure. If this assumption is not fulfilled, the auxiliary parameters $\boldsymbol{\phi_2^{(s)}}$ are needed in the selection model. In the simple case-control study with a binary exposure variable X, four parameters are inherent in $\boldsymbol{\phi_2^{(s)}}$:

$$\begin{array}{ll} Pr(S=1|X=1,\text{case}) & Pr(S=1|X=1,\text{control}) \\ Pr(S=1|X=0,\text{case}) & Pr(S=1|X=0,\text{control}) \end{array}$$

These parameters are not identifiable from the data. On the other hand, if we consider the pattern-mixture model, only two unidentifiable parameters are needed:

$$Pr(X|S=0,\text{case}) \quad Pr(X|S=0,\text{control})$$

because $Pr(S|\text{case})$ and $Pr(S|\text{control})$ are often available in study reports. For this reason, in this thesis, I adopt the pattern-mixture model.

5.1 Adjusting for non-participation bias using the pattern-mixture model

It was mentioned above that there are two unidentifiable parameters in the simple case-control study with binary exposure, using a pattern-mixture model: $Pr(X|S=0,\text{case})$ and $Pr(X|S=0,\text{control})$. We can reparameterize the model such that the unidentifiable parameters represent the odds ratios between the exposure and response for the cases and for controls:

$$\frac{Pr(X|S=0,i)}{1-Pr(X|S=0,i)}R_i = \frac{Pr(X|S=1,i)}{1-Pr(X|S=1,i)} \qquad i=\text{case/control} \tag{5.3}$$

where R_i denotes the odds ratio. Rearranging equation (5.1) and denoting by:

$$\pi_i = Pr(X=1|S=1,i) \tag{5.4}$$

$$\pi_i^* = Pr(X=1|i) \tag{5.5}$$

$$Q_i = Pr(S=1|i) \tag{5.6}$$

we have:

$$\pi_i^* = Q_i\pi_i + \frac{(1-Q_i)\pi_i}{R_i(1-\pi_i)+\pi_i} \tag{5.7}$$

Thus, we have a deterministic relationship between the prevalence of exposure among those who would respond (π_i), and the prevalence of exposure of the target population (π_i^*). Applying equation (5.7) to the meta-analysis model of (2.14) (2.15) and (2.16), we have:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{5.8}$$

$$\pi_{si} = f(\pi^*_{si}) \tag{5.9}$$

$$\text{logit}\, \pi^*_{s1} = \text{logit}\, \pi^*_{s0} + \theta_s \tag{5.10}$$

$$\theta_s = \theta + \delta_s \tag{5.11}$$

where $f(\pi^*_{si})$ is the inverse of function (5.7), and has been shown by Plackett (1965) to be the solution of a quadratic equation. Mardia (1970) further showed that the solution is always the larger of the 2 solutions of the quadratic equation (given $R_{si} \neq 1, 0 < Q_{si} \leq 1, 0 \leq \pi^*_{si} \leq 1$). In symbols, we can write:

$$\begin{aligned} \pi_{si} &= \frac{-b_{si} + \sqrt{b_{si}^2 - 4a_{si}c_{si}}}{2a_{si}} \\ a_{si} &= Q_{si}(1 - R_{si}) \\ b_{si} &= (R_{si} - 1)(Q_{si} + \pi^*_{si}) + 1 \\ c_{si} &= -R_{si}p^*_{si} \end{aligned} \tag{5.12}$$

As in previous chapters, we can assign feasible ranges for $Q_{s1}, Q_{s0}, R_{s1}, R_{s0}$, and observe the range of posterior $\hat{\boldsymbol{\Theta}}$ given by these ranges.

5.1.1 Some theoretical results

In order to see where we are likely to find the maximum and minimum of $\hat{\boldsymbol{\Theta}}$ $(= \{\hat{\theta}_L, \hat{\theta}_M, \hat{\theta}_U\})$, it is instructive to examine the partial derivatives of equation (5.7). In the following, dependence on case/control status is removed to improve clarity.

$$\frac{\partial \pi^*}{\partial Q} = \pi \left(1 - \frac{1}{R(1-\pi) + \pi}\right) \tag{5.13}$$

$$\frac{\partial \pi^*}{\partial R} = \frac{-(1-Q)\pi(1-\pi)}{(R(1-\pi) + \pi)^2} \tag{5.14}$$

It can be seen that the derivative with respect to R (equation 5.14) is always non-positive (since $0 \leq \pi, Q, R \leq 1$). This is intuitive because the greater R is, the less the prevalence of the exposure among the unobserved ($Pr(X|S=0)$) compared to the observed ($Pr(X|S=1)$) ($R = \frac{\text{Odds}(X|S=1)}{\text{Odds}(X|S=0)}$). The total population exposure prevalence is a weighted average of that of the observed and the unobserved, and thus should vary with R negatively.

It can also be seen that the derivative with respect to Q (equation 5.13) is positive when $R > 1$, and negative when $R < 1$, and 0 when $R = 1$. This is also intuitive because if the unobserved ($S = 0$) has a greater exposure prevalence ($R < 1$), then increasing the proportion of unobserved (decreasing Q) will increase the true population prevalence estimate (π^*). The same logic applies when $R > 1$. When $R = 1$, then we are saying the exposure prevalence is the same in both the observed and the unobserved, and hence the proportion of unobserved does not affect our estimate of π^*.

As in previous analyses, a simple relationship between Q, R and $\hat{\boldsymbol{\Theta}}$ probably does not exist. However, we can note where extreme values of Q_1, Q_0, R_1, and R_0 are likely to be found based on the above relationships. For example, since $\theta = \operatorname{logit} \pi_1^* - \operatorname{logit} \pi_0^*$, for more positive $\hat{\boldsymbol{\Theta}}$, we would expect R_1 to be towards its minimum and R_0 to be towards its maximum. If both R_1 and R_0 are less than 1, then we would expect Q_1 to be towards its minimum and Q_0 to be towards its maximum. The reverse would be expected when R_1 and R_0 are both greater than 1. If $R_1 < 1$ and $R_0 > 1$, then we would expect both Q_1 and Q_0 to be near their minima. Again, we can derive the exact opposite for more negative θ.

5.2 The Greenland meta-analysis example

In this section, we apply the pattern-mixture model for adjustment of non-participation bias to the Greenland meta-analysis example. Going through the studies of the childhood leukaemia example we have been considering, we can find the following reported figures of rates of participation, as tabulated in Table 5.1.

5.2.1 Constraints for Q_1, Q_0

Q corresponds to the proportion of participation among those invited to participate in the studies. A number of studies (studies 3, 9, 11, 12, 14) did not require consent for exposure measurement, as exposure was assessed by the distance between the participant's residences (which can be obtained from the records) and major sources of EMF, such as power lines and electricity transformation stations. For these studies, we may assume participation is complete:

$$Q_{s0} = Q_{s1} = 1 \qquad s = 3, 9, 11, 12, 14 \tag{5.15}$$

Among studies that require consent for studies, we have the proportion of participation among the invited. These are tabulated in Table 3.1 together with their 95% confidence intervals. For these studies, we can take the 95% Confidence Intervals limits of the proportion of participation as reasonable bounds for the parameters Q_{s0} and Q_{s1}. (We assume these proportions are estimating the *population* proportion of would-be participants, since Q_{s0} and Q_{s1} represent the *population* rather than the *sample*

Index	Study	Cases			Controls		
		Participated and measured	Invited	Proportion (95% CI)	Participated and measured	Invited	Proportion (95% CI)
1	Coghill *et al.* (1996)	56	NA[a]		56	NA[a]	
2	Dockerty *et al.* (1998)	209	303	0.69 (0.63, 0.74)	121	131	0.92 (0.86, 0.96)
3	Feychting and Ahlbom (1993)	38	38	1[b]	554	554	1[b]
4	Kabuto *et al.* (2006)	2097	3833	0.55 (0.53, 0.56)	381	781	0.49 (0.45, 0.52)
5	Linet *et al.* (1997)	620	984[c]	0.63 (0.60, 0.66)	638	818[c]	0.78 (0.75, 0.81)
6	London *et al.* (1991)	144	257	0.56 (0.50, 0.62)	164	331	0.50 (0.44, 0.55)
7	McBride *et al.* (1999)	399	675	0.59 (0.55, 0.63)	399	449	0.89 (0.86, 0.92)
8	Michaelis *et al.* (1998)	414	919	0.45 (0.42, 0.48)	176	283	0.62 (0.56, 0.68)
9	Olsen *et al.* (1993)	833	833	1[b]	1666	1666	1[b]
10	Savitz *et al.* (1988)	207	282	0.73 (0.68, 0.78)	36	103	0.35 (0.26, 0.45)
11	Tomenius (1986)	153	153	1[b]	698	698	1[b]
12	Tynes and Haldorsen (1997)	148	148	1[b]	2004	2004	1[b]
13	UKCCS (1999)	2415	7629	0.32 (0.31, 0.33)	2423	3838	0.63 (0.62, 0.65)
14	Verkasalo *et al.* (1993)	32	32	1[b]	320	320	1[b]

[a] Cases were recruited by advertisement and controls were recruited as friends of cases in this study

[b] These studies do not require consent from the participants for exposure measurement

[c] These numbers were not derived directly from the paper. Different reasons for non-participation were given in the paper and 78% and 63% were given as overall participation rates for cases and controls, respectively. These numbers are simply the number of participants divided by these percentages.

Table 5.1: Number and proportion of invited participants whose exposure measurements were obtained

participation proportion.) For example, for the Linet *et al.* (1997) study (index number 5),

$$0.75 \leq Q_{5\,1} \leq 0.81 \tag{5.16}$$

$$0.60 \leq Q_{5\,0} \leq 0.66 \tag{5.17}$$

The one study that failed to provide a measure of the proportion of participation is the Coghill *et al.* (1996) study. This study recruited cases from advertisement and controls were friends of the cases. Because the potential for non-participation bias is large in comparison with the other studies, if we want to adjust for bias in this study in the same way we adjust for bias in the other studies, it will be reasonable to give this study a range for Q_1 and Q_0 that is closer to zero than the others. Here, I give these parameters a lower bound of 0.2 and an upper bound of 0.5.

$$0.2 \leq Q_{1\,1}, Q_{1\,0} \leq 0.5 \tag{5.18}$$

5.2.2 Constraints for R_1, R_0

Unlike Q_{si}, there is little to inform possible values of R_{si}. As the odds ratio of exposure between the observable and the unobservable, its value depends on how different we believe the unobservable population is to the observable population. From a closer examination of the studies, it appears that most of the non-participation in the studies were due to refusal among both the cases and the controls. Some probable, though unconfirmed reasons why participation might be related to exposure include: (1) Those being near to high exposure sources (e.g. transmission lines) may be more interested in the project; (2) Cases who are more ill are less likely to participate. If we, for example, suppose that those who are more exposed are more likely to participate, then we have $\text{Odds}(S_{si}|X_{si} = 1) > \text{Odds}(S_{si}|X_{si} = 0) \iff R_{si} > 1$. Coupled with other unknown reasons for differential propensity to participate, let us suppose that a conservative interval for R_{s1} and R_{s0} is:

$$0.75 \leq R_{si} \leq 3 \quad i = 0, 1 \quad s = 1, 2, 4, 5, 6, 7, 8, 10, 13 \tag{5.19}$$

Although we do not have much information to inform R_{si}, there may be good reasons to suppose that the difference between R_{s1} and R_{s0} may not be too much. This is because for whatever reasons $\text{Odds}(S_{si}|X_{si} = 1)$ may be different from $\text{Odds}(S_{si}|X_{si} = 0)$, the same reasons may be at work in both cases and controls, if the invitation to participate is the same for the two groups. Thus, it will usually be reasonable to restrict the ratio between R_{s1} and R_{s0}. In the following discussion, let us assume that this ratio does not exceed 1.5 or 1/1.5, for all studies except the Coghill *et al.* study, since this study

has very different mechanisms for recruiting cases and controls.

$$1/1.5 \leq R_{s1}/R_{s0} \leq 1.5 \quad s = 2 \ldots 14 \tag{5.20}$$

The constraints on Q_1, Q_0, R_1, R_0 given in the last two subsections are summarized in Table 5.2.

Unlike bias due to exposure misclassification, non-participation bias does not "dilute" the data, and hence given the values of the non-participation bias parameters, the data in a meta-analysis generally contain fairly precise information for the prevalence of true exposure, i.e. for π_{s0} and π_{s1}. Thus, we can use non-informative prior distributions for π_{s0} and $\theta_s (= \operatorname{logit} \pi_{s1} - \operatorname{logit} \pi_{s1})$, as in chapter 3:

$$\theta \sim N(0, 100) \tag{5.21}$$

$$\pi_{s0} \sim U(0, 1) \tag{5.22}$$

As for δ_s, I follow the previous chapters in having $\delta_s = 0$ for a FE model, and $\delta_s \sim N(0, 0.15)$ for a RE model.

5.2.3 Results

The optimization of this chapter took around 10-100 seconds for single study results and 30-600 seconds for meta-analyses on an Intel Core i7-2600 computer. These optimization were much faster than those in the previous chapter because (1) there are less parameters to consider (when $Q_i = 1$ or $R_i = 1$, then bias is absent), and (2) there is not the same problem with adaptation in the range of integration as was encountered in the problems of the previous chapter. However, multi-modality is a significant problem in this chapter, with 12% of single-study and 68% of meta-analysis problems found with more than 1 mode. The presence of multiple modes was almost entirely the consequence of the constraint (5.20):

$$1/1.5 \leq R_{s1}/R_{s0} \leq 1.5 \quad s = 2 \ldots 14 \tag{5.23}$$

Frequently, one mode is found with R_{s0} at its upper or lower extreme, and $R_{s1} = R_{s0} \times 1.5^{\pm 1}$, and another is found with R_{s1} at its upper or lower extreme and $R_{s0} = R_{s1} \times 1.5^{\pm 1}$. If there is only 1 study, then at most two modes are found. In a meta-analysis, more than two modes are possible because each constituent study can contribute up to 2 modes.

Figures 5.1 and 5.2 show the Feasible Posterior Intervals (FPIs) for the above setup for both the individual study setting and the meta-analysis setting, for a fixed-effects (FE) model and a random-effects (RE) model, as compared to the situation without bias. It appears that the effect of non-participation bias is roughly additive — it behaves similarly to the simple bias model of chapter 3 in that bias increases the width of the Feasible Posterior Median Intervals (FPMIs) but does not increase

Index	Study	$[Q_0]$	$[Q_1]$	$[R_0]$	$[R_1]$	$[R_1/R_0]$
1	Coghill *et al.* (1996)	[0.20, 0.50]	[0.20, 0.50]	[0.75, 3]	[0.75, 3]	(none)
2	Dockerty *et al.* (1998)	[0.63, 0.74]	[0.86, 0.96]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
3	Feychting and Ahlbom (1993)	=1	=1	(none)	(none)	(none)
4	Kabuto *et al.* (2006)	[0.45 0.52]	[0.53, 0.56]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
5	Linet *et al.* (1997)	[0.60, 0.66]	[0.75, 0.81]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
6	London *et al.* (1991)	[0.5, 0.62]	[0.44, 0.55]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
7	McBride *et al.* (1999)	[0.55, 0.63]	[0.86, 0.92]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
8	Michaelis *et al.* (1998)	[0.42, 0.48]	[0.56, 0.68]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
9	Olsen *et al.* (1993)	=1	=1	(none)	(none)	(none)
10	Savitz *et al.* (1988)	[0.68, 0.78]	[0.26, 0.45]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
11	Tomenius (1986)	=1	=1	(none)	(none)	(none)
12	Tynes and Haldorsen (1997)	=1	=1	(none)	(none)	(none)
13	UKCCS (1999)	[0.31, 0.33]	[0.62, 0.65]	[0.75, 3]	[0.75, 3]	[0.67, 1.5]
14	Verkasalo *et al.* (1993)	=1	=1	(none)	(none)	(none)

Table 5.2: Feasible range for the non-participation bias parameters in the meta-analysis example

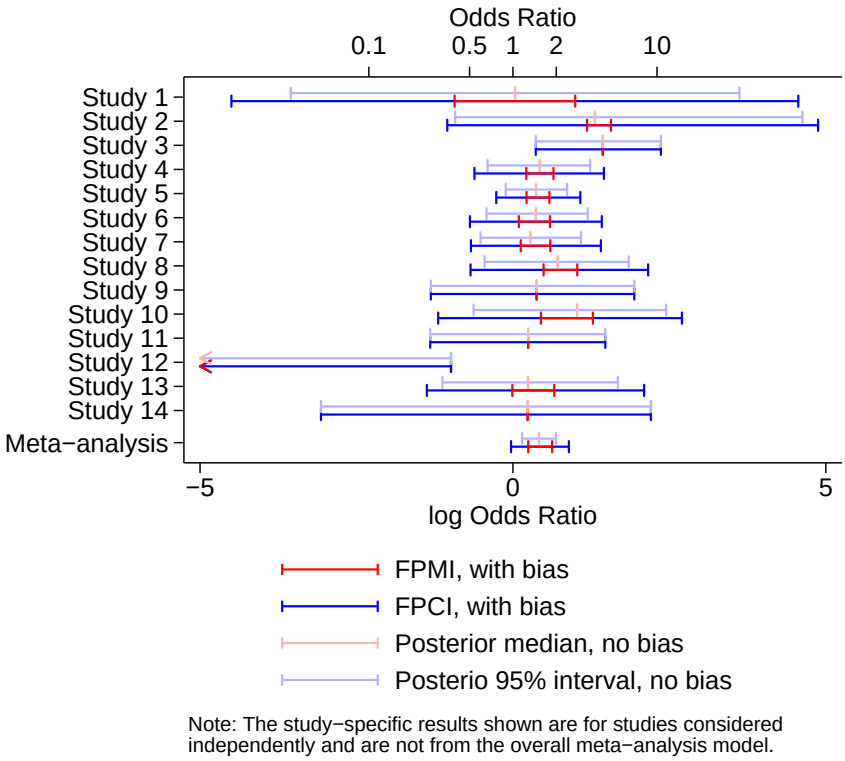


Figure 5.1: FPIs for a FE model with non-participation bias as compared to a model without bias

the Feasible Posterior Credible Intervals (FPCIs) on top of the increase in the FPMIs. The degree of non-participation bias that is possible for these studies is moderate for all studies, except perhaps for Coghill *et al.* (1996), and this is noted earlier.

The location of Q_0, Q_1, R_0, R_1 at $\max \hat{\boldsymbol{\Theta}}$ and $\min \hat{\boldsymbol{\Theta}}$ are given in Figure 5.3. Only one figure is presented for both the independent-study and the meta-analysis case, and for both the FE and the RE models, because in this example, the locations are in fact the same in all four situations. Because of the constraints (5.20), the pattern of the location of Q_0, Q_1, R_0, R_1 is quite complicated. Based on the theory of section 5.1.1, we expect R_1 to be smaller than R_0 at $\max \hat{\boldsymbol{\Theta}}$, and the reverse at $\min \hat{\boldsymbol{\Theta}}$. This can be observed in Figure 5.3. However, because of the constraints of (5.20), sometimes R_0 is found at the extreme of its feasible space while other times R_1 is. Always, one of R_0 or R_1 is found at the extreme of its feasible region. Depending on the location of R_0 and R_1, Q_0 and Q_1 are sometimes found at the lower and sometimes the upper bound. When both R_0 and R_1 are greater than or equal to 1, then one of (Q_0, Q_1) is at its minimum and the other at its maximum. When one of (R_0, R_1) is less than 1, and the other is greater than 1, then both Q_0 and Q_1 are found at their minimum. This is also in agreement with predictions from section 5.1.1.

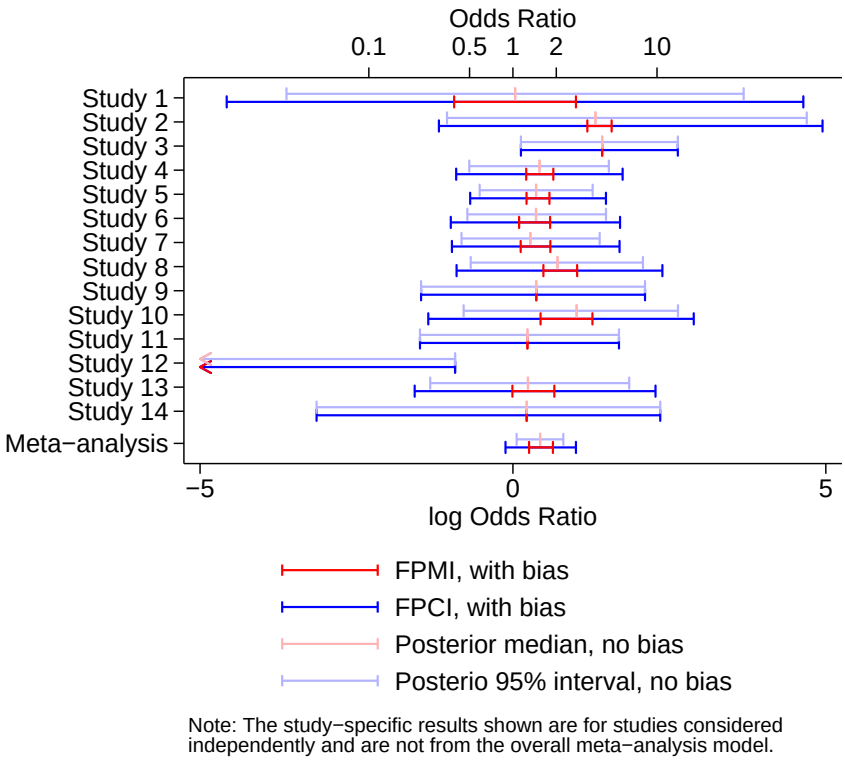


Figure 5.2: FPIs for a RE model with non-participation bias as compared to a model without bias

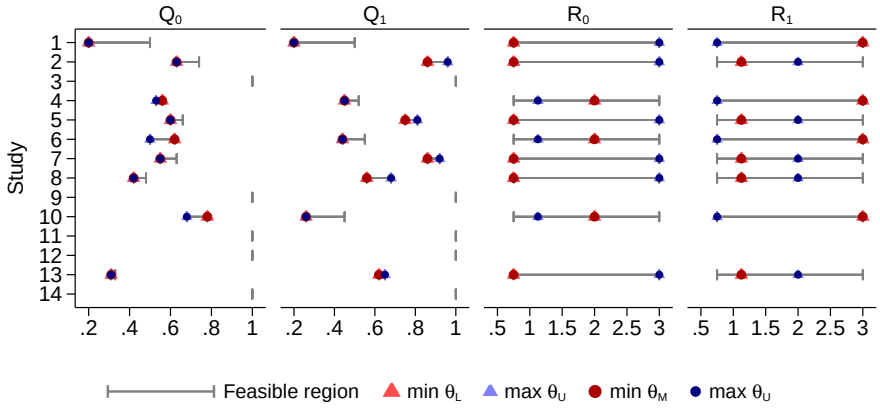


Figure 5.3: Location of (Q_0, Q_1, R_0, R_1) at max/min $\hat{\boldsymbol{\Theta}}$ when studies are considered independently as well as when considered in a meta-analysis

5.3 Sensitivity analyses

As with the previous chapter, we can conduct sensitivity analyses to see how further restricting the range of feasible non-participation bias affects the uncertainty interval of $\hat{\mathbf{\Theta}}$. Again, we consider the FPI for θ only in the FE meta-analysis setting.

As a first example, consider the scenario where those who are more exposed are certainly more likely to participate, so that we have $\text{Odds}(S|X=1) > \text{Odds}(S|X=0)$, or $R > 1$. Replacing constraints

$$0.75 \leq R_{si} \leq 3 \quad i = 0, 1 \quad s = 1, 2, 4, 5, 6, 7, 8, 10, 13 \tag{5.24}$$

by

$$1 \leq R_{si} \leq 3 \quad i = 0, 1 \quad s = 1, 2, 4, 5, 6, 7, 8, 10, 13 \tag{5.25}$$

we find that:

$$[\min \hat{\theta}_M, \max \hat{\theta}_M] = [0.29, 0.61] \tag{5.26}$$

$$[\min \hat{\theta}_L, \max \hat{\theta}_U] = [0.15, 0.88] \tag{5.27}$$

in the meta-analysis. This compares to

$$[\min \hat{\theta}_M, \max \hat{\theta}_M] = [0.24, 0.63] \tag{5.28}$$

$$[\min \hat{\theta}_L, \max \hat{\theta}_U] = [-0.03, 0.89] \tag{5.29}$$

for the standard setup (5.24), and we see that the FPIs are only slightly narrowed. Figure 5.4 shows results from further sensitivity analyses altering the feasible range of R_1 and R_0. From the theory in section 5.1.1, we know that when $R_1 = R_0 = 1$, no non-participation bias is present. As R_1 and R_0 is allowed to depart more and more from 1, the uncertainty interval also increases in width.

Also, when R_1 and R_0 becomes larger, $\hat{\mathbf{\Theta}}$ also tends to become more positive (see bottom 4 lines of Figure 5.4). This is when we have held constant the width of $\log R_1^+ - \log R_0^-$, where R_i^+ and R_i^- denote the upper and lower bound of the feasible region of R_i.

Apart from varying the ranges of R_1 and R_0, we can also examine the effect of varying the range of R_1/R_0. For example, one possible scenario is that while the control group's readiness to participate is related to the exposure (with those who are more exposed to EMF more likely to participate), the cases' readiness to participate may be less related, but rather more related to factors such as health. If this was the case, we might expect R_1 to be less than R_0. For example, if we replace the constraints:

$$1/1.5 \leq R_{s1}/R_{s0} \leq 1.5 \quad s = 1 \ldots 14 \tag{5.30}$$

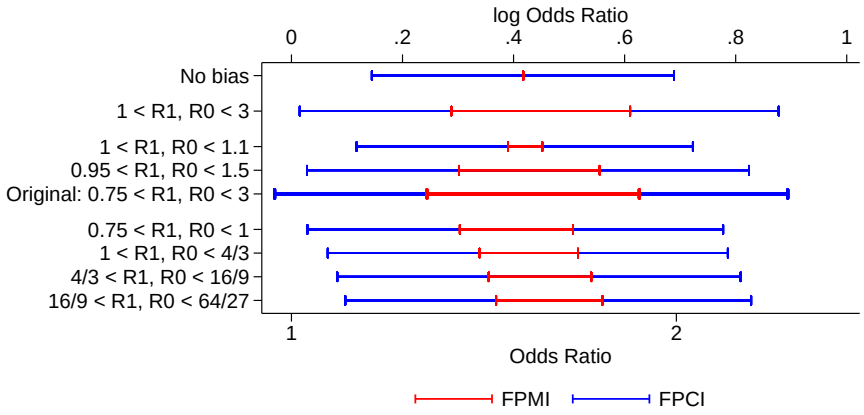


Figure 5.4: Sensitivity analysis to changes to the Feasible Posterior Intervals of θ in response to changes in the feasible range of R_1 and R_0 in the FE meta-analysis

by

$$1/1.5 \leq R_{s1}/R_{s0} \leq 1 \quad s = 1 \ldots 14 \tag{5.31}$$

we have:

$$[\min \hat{\theta}_M, \max \hat{\theta}_M] = [0.32, 0.63] \tag{5.32}$$

$$[\min \hat{\theta}_L, \max \hat{\theta}_U] = [0.07, 0.89] \tag{5.33}$$

Compared to the standard setup (5.30), where we have:

$$[\min \hat{\theta}_M, \max \hat{\theta}_M] = [0.24, 0.63] \tag{5.34}$$

$$[\min \hat{\theta}_L, \max \hat{\theta}_U] = [-0.03, 0.89] \tag{5.35}$$

we see that the FPI has a more positive lower bound, although the upper bound remains nearly the same. Figure 5.5 displays further sensitivity analyses of how varying the possible range of R_1/R_0 affects the uncertainty interval. We see that the location of the uncertainty interval of $\hat{\boldsymbol{\Theta}}$ is sensitive to the values of R_1/R_0. When R_1 is 0.667-0.8 times R_0, the interval is shifted to the right and the evidence for a positive θ is stronger. On the other hand when R_1 is greater than R_0, the interval is shifted to the left and the evidence is weaker. This is easy to explain: the greater R is, the smaller the true exposure prevalence (π^*) is compared with the observed prevalence (π) (equation 5.7). When R_1 is less than 1, π_1^* is increased relative to π_1. When R_0 is greater than 1, π_0^* is decreased relative to π_0. Since $\theta_s = \operatorname{logit} \pi_1^* - \operatorname{logit} \pi_0^*$, this will tend to increase θ. Finally, let us note that $R_1 = R_0$ is not equivalent to the absence of non-participation bias. This is because non-participation bias depends also on Q_1 and Q_0, and even when $Q_1 = Q_0$, some non-participation bias is possible because the scale by which Q and R affects π is not the log odds scale.

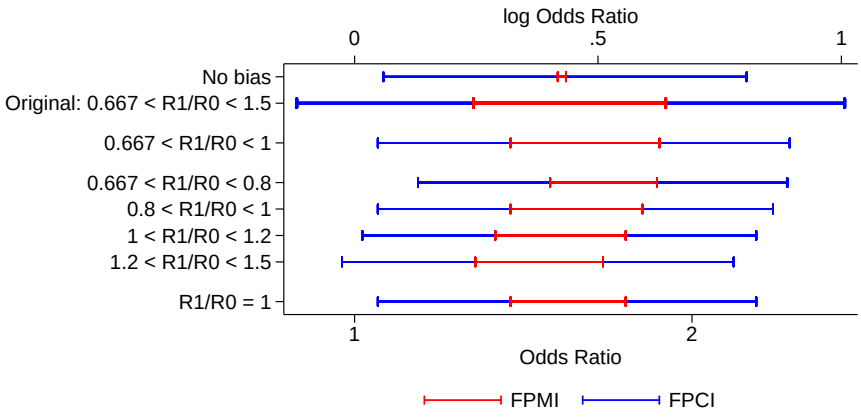


Figure 5.5: Sensitivity analysis of changes to the Feasible Posterior Intervals of θ in response to changes in the feasible range of R_1/R_0 in the FE meta-analysis

Summary:

- In this chapter, I have briefly outlined two commonly used models for adjustment of non-participation bias — the *selection* model and the *pattern mixture* model. In a meta-analysis setting, the *pattern mixture* model appears more straightforward to apply as estimates of some the parameters, namely the participation probabilities (Q_1 and Q_0), are often available in reports.

- The other parameters, R_1 and R_0, which gives the odds ratio between participation and exposure status, are harder to estimate, although often the exposed (if they know of their exposure) are more likely to participate, and that if controls and cases are invited in similar ways, R_1 and R_0 are likely to be similar. This means it may be reasonable to limit the ratio R_1/R_0.

- I applied the pattern-mixture model to the Greenland meta-analysis, assigning feasible ranges to the parameters Q_1, Q_0, R_1, and R_0. As in the previous chapter, I demonstrated the use of the algorithm of this thesis to search through the feasible region for the most extreme $\hat{\boldsymbol{\Theta}}$.

- As in the previous chapter, multimodality is a problem for optimization in the problems of this chapter. However, it appears at most two modes are found for each study.

- For sensitivity analyses, I demonstrated how the location of $\hat{\boldsymbol{\Theta}}$ is sensitive to the values of R_1/R_0, and also to the values of R_1 and R_0 themselves (if the range of allowable R_1/R_0 stays the same). Greater values of R_1, R_0 as well as smaller values of R_1/R_0 tend to lead to more positive $\hat{\boldsymbol{\Theta}}$. The effect of R_1/R_0 on the value of $\hat{\boldsymbol{\Theta}}$ might be predicted from theory, whereas the reason for the effect of the values of R_1, R_0 themselves on $\hat{\boldsymbol{\Theta}}$ are not entirely clear.

Chapter 6

Accounting for uncertainty due to incomplete control of confounding

In epidemiology, confounding refers to the true (causal) effect of an exposure on a disease being *confounded* by some other factors, leading to a spurious association (Greenland and Robins, 1986). In clinical trials, bias due to confounding can often be probabilistically controlled for by randomization. In observational studies, however, randomization is generally not possible, and stratification in either the sampling stage or the analysis stage over known confounders is believed to reduce the potential for bias due to confounding. Commonly, stratification is replaced by *adjustment* in a regression context, and is expected to give reasonable approximation to the stratification results (Greenland and Maldonado, 1994). However, generally we do not expect such strategies to completely eliminate confounding bias. This is because (1) the exact relationship (i.e. dose-response relationship) between the confounder and the disease may not be known or modeled correctly, (2) the confounder may be subject to measurement error/dichotomization, (3) some relevant confounder data may not be measured or collected, and (4) even when all relevant confounders are measured precisely and modeled correctly, there will likely be residual confounding (due to residual "randomness").

In the literature, a number of proposals have been raised to address bias due to such incomplete control of confounding, particularly due to (3) (e.g. Rosenbaum and Rubin, 1983a; Arah *et al.*, 2008; McCandless *et al.*, 2009). However, nearly all of these techniques assume we have access to measured confounder data. In a meta-analysis setting, this is usually not the case. Thus, we are even more limited than usual in our ability to deal with bias due to incomplete control of confounding. Among the three types of biases considered in this thesis, bias due to uncontrolled confounding might therefore be the one that is most difficult to quantify satisfactorily.

6.1 Making use of what is available to adjust for bias due to measured confounders in a meta-analysis

However, many studies report at least 2 estimates of the effect of a risk factor on a disease – one unadjusted for confounders and one adjusted for confounders. A common strategy for meta-analysis that takes into account of study-specific adjustment for measured confounders is to extract the "adjusted" estimate ($\hat{\theta}_s^{\text{adj}}$) together with its standard error[1] ($\hat{\sigma}_s^{\text{adj}}$) and form a weighted average of ($\hat{\theta}_s^{\text{adj}}$) with weights being inversely proportional to the variance of ($\hat{\theta}_s^{\text{adj}}$), estimated as $\hat{\sigma}_s^{\text{adj}^2} + \hat{\sigma}_\delta^2$, where $\hat{\sigma}_\delta^2$ is an estimate of the random-effects (between-study) variance (e.g. Deeks *et al.*, 2001; Petitti, 2000):

$$\text{Estimate of meta-analytic } \theta = \sum_s w_s \hat{\theta}_s^{\text{adj}} \tag{6.1}$$

$$w_s \propto \frac{1}{\hat{\sigma}_s^{\text{adj}^2} + \hat{\sigma}_\delta^2} \tag{6.2}$$

While the justification of the "inverse-variance" weights is usually given as it being the minimum variance unbiased estimator of θ in the fixed-effects (FE) meta-analysis (Deeks *et al.*, 2001), an alternative justification which applies to both the FE and the random-effects (RE) model is that they lead to the maximum likelihood estimate for θ under a model with Normally distributed random effects and errors and known σ_δ^2:

$$\hat{\theta}_s^{\text{adj}} \sim N(\theta_s, \hat{\sigma}_s^{\text{adj}^2}) \tag{6.3}$$

$$\theta_s = \theta + \delta_s \tag{6.4}$$

$$\delta_s \sim N(0, \sigma_\delta^2) \tag{6.5}$$

This model has also formed the basis for general Bayesian meta-analysis inference (Sutton and Abrams, 2001; Higgins *et al.*, 2009), presumably because in the situation where only the effect ($\hat{\theta}_s$) and its standard error ($\hat{\sigma}_s$) is available, the Normal distribution of $\hat{\theta}_s$ around its true effect θ_s can be justified through the Central Limit Theorem for large enough samples. A disadvantage of this model, however, is that we cannot easily extend it to account for bias due to selection and exposure misclassification using the techniques of the previous chapters, because it does not involve the exposure probability parameters π_{si}. To overcome this problem, I propose the following modification to the standard model.

[1]The standard error is often converted from the 95% confidence interval, assuming the estimate follows a Normal distribution.

6.1.1 A proposal

Supposing (6.3) to be an adequate approximation of the true model (i.e. the fully Bayesian model which properly accounts for measured confounders), we note that (6.3) can be equivalently written as (i.e. it leading to the same likelihood for θ_s):

$$\hat{\theta}_s^{\text{crude}} \sim N(\theta_s^{**}, \hat{\sigma}_s^{\text{crude}^2}) \tag{6.6}$$

$$\theta_s^{**} = \theta_s - (\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}) + \varepsilon_s \tag{6.7}$$

$$\varepsilon_s \sim N(0, \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) \tag{6.8}$$

given $\hat{\sigma}_s^{\text{adj}^2} \geq \hat{\sigma}_s^{\text{crude}^2}$. Thus, if we replace equation (6.6) with:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{6.9}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s^{**} \tag{6.10}$$

our inference for θ_s should not differ very much from if we had used model (6.6-6.8) since the justification for the use of (6.6) is that it approximates (6.9), (6.10). However, in the form of (6.9-6.10), we can easily extend it to incorporate adjustment for non-participation bias and bias due to exposure misclassification using the techniques of the previous chapters. Moreover, in the light of evidence from Appendix D, it appears likely that this model may give even closer approximation to the true model than (6.3) since some of the bias introduced through the Normal approximation of (6.6) may be reversed when we revert back to a Binomial model. This, however, awaits confirmation in a future study.

6.2 The Greenland meta-analysis example

In this section, the above techniques are demonstrated in the Greenland meta-analysis example. Table 6.1 presents the data that are extracted from the reports of the original studies. We note that not all studies have adjusted their estimates for confounders, and that among those that have adjusted for confounders, all except one (study 6) included a measure of socioeconomic status (SES), such as parental education or income. Moreover, let us note that the adjusted estimates were generally not available for the exposure contrast we have been considering (i.e. not for $> 0.3\mu$T vs. $< 0.3\mu$T), but for some other contrasts. This is because not all studies considered the contrast between $> 0.3\mu$T and $< 0.3\mu$T as their main analysis, as Greenland did. Table 6.1 also gives the estimates of the confounding bias due to measured confounders ($\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}$) and the difference in variance of the estimates ($\hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}$), as derived from the published estimates and confidence intervals.

Index	Study	Crude estimate (95% CI)	Adjusted estimate (95% CI)	$\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}$	$\hat{\sigma}_s^{\text{adj}\,2} - \hat{\sigma}_s^{\text{crude}\,2}$	Variables adjusted for
1	Coghill *et al.* (1996)	No adjusted analysis		NA	NA	NA
2	Dockerty *et al.* (1998)	4.9 (0.6, 42.7)	15.5 (1.1, 224)[a]	1.15	0.656	Mother's education, maternal smoking, residence on farm
3	Feychting and Ahlbom (1993)	2.7 (1.1, 6.48)	3.1 (1.1, 8.6)[b]	0.14	0.071	sex, age, country, type of building, year of diagnosis, SES, NO2 levels
4	Kabuto *et al.* (2006)	2.56 (0.76, 8.58)	2.77 (0.8, 9.57)[c]	0.08	0.018	Father's education, Mother's education
5	Linet *et al.* (1997)	1.27 (0.88, 1.81)	1.24 (0.86, 1.79)[d]	-0.02	0.001	Age, mother's education, family income
6	London *et al.* (1991)	1.48 (0.66, 3.29)	1.69 (0.71, 4)[e]	0.13	0.027	pesticide use, hair dryer use, B& W TV use, father's occupational exposure
7	McBride *et al.* (1999)	0.72 (0.42, 1.23)	0.59 (0.32, 1.07)[f]	-0.20	0.020	Maternal age, education, household income, ethnicity, no. of residences since birth
8	Michaelis *et al.* (1998)	2.65 (1, 6.97)	2.3 (0.8, 6.7)[g]	-0.14	0.049	SES, degree of urbanization
9	Olsen *et al.* (1993)	Results not reported[h]		NA	NA	Population density, SES, family mobility
10	Savitz *et al.* (1988)	Results not reported[i]		NA	NA	Unknown
11	Tomenius (1986)	No adjusted analysis		NA	NA	NA
12	Tynes and Haldorsen (1997)	Results not reported[j]		NA	NA	SES, no. of residences
13	UKCCS (1999)	1.62 (0.39, 6.77)	1.68 (0.4, 7.1)[k]	0.04	0.008	Deprivation index
14	Verkasalo *et al.* (1993)	No adjusted analysis		NA	NA	NA

[a] From Table 3. Exposure contrast $>0.2\mu$T vs $<0.1\mu$T. Conditional logistic regression analysis.
[b] From Table 6. Exposure contrast $>0.2\mu$T vs $<0.1\mu$T.
[c] From Table 1. Exposure contrast $>0.4\mu$T vs $<0.1\mu$T.
[d] From Table 2. Exposure contrast $>0.2\mu$T vs $<0.065\mu$T.
[e] From Table 9. Exposure contrast $>0.268\mu$T vs $<0.067\mu$T.
[f] From Table 4. Exposure contrast $>0.27\mu$T vs $<0.08\mu$T.
[g] From Table 2. Exposure contrast $>0.2\mu$T vs $<0.2\mu$T. Conditional logistic regression analysis for adjusted analysis, but not crude. Crude analysis derived from raw counts.
[h] Adjustment "did not affect results".
[i] Adjustment "did not affect results".
[j] Adjustment "did not affect results".
[k] From Table 4. Exposure contrast $>0.4\mu$T vs $<0.1\mu$T. Conditional logistic regression analysis.

Table 6.1: Crude and adjusted estimates of odds ratio between EMF exposure and childhood leukaemia

6.2.1 Adjustment for measured confounders only

As a first example, let us consider a meta-analysis with adjustment only for SES as measured by these studies. The studies that have included an estimate of θ adjusted for a measure of SES are: 2, 3, 4, 5, 7, 8, 13. (Although study 6 reported an estimate of θ adjusted for confounders, the confounder was not SES.) For these studies, the model I use is:[2]

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{6.11}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s^{**} \tag{6.12}$$

$$\left.\begin{array}{l} \theta_s^{**} = \theta_s - (\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}) + \varepsilon_s \\ \varepsilon_s \sim N(0, \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) \end{array}\right\} s = 2, 3, 4, 5, 7, 8, 13 \tag{6.13}$$

$$\left.\theta^{**} = \theta_s \qquad\right\} s = 1, 6, 9, 10, 11, 12, 14 \tag{6.14}$$

$$\theta_s = \theta + \delta_s \tag{6.15}$$

For δ_s and θ, I follow the previous chapters in having $\theta \sim N(0, 100)$ and

$$\delta_s = 0 \tag{6.16}$$

for a FE model and

$$\delta_s \sim N(0, 0.15) \tag{6.17}$$

for a RE model.

Note that we can rewrite the (6.11 - 6.17) as:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{6.18}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{6.19}$$

$$\theta_s = \theta + \delta_s \tag{6.20}$$

with

$$\begin{aligned} \delta_s \sim N(-(\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}), \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) \qquad & s = 2, 3, 4, 5, 7, 8, 13 \\ \delta_s = 0 \qquad & s = 1, 6, 9, 10, 11, 12, 14 \end{aligned} \tag{6.21}$$

[2]Normally, it may not be reasonable to include estimates from studies that adjust for confounders as well as studies that do not adjust in the same meta-analysis model, since they may be estimating different quantities. Here, however, they are included in order to facilitate comparison with previous and subsequent results.

for a FE model and:

$$\delta_s \sim N(-(\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}), 0.15 + \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) \qquad s = 2, 3, 4, 5, 7, 8, 13$$
$$\delta_s \sim N(0, 0.15) \qquad s = 1, 6, 9, 10, 11, 12, 14 \tag{6.22}$$

for a RE model, and we see that these models are equivalent to the standard meta-analysis model given in chapter 2 (2.14, 2.15, 2.16) except for the prior given to δ_s. Given these specifications, we find that our posterior median and 95% interval of θ is 0.41 (0.13, 0.69) for the FE model and 0.44 (0.05, 0.82) for the RE model. This compares to 0.42 (0.14, 0.69) for the FE model and 0.43 (0.06, 0.81) for the RE model without adjusting for measured confounding. Thus, we see that the 95% posterior interval is only very slightly widened. Overall, the posterior distribution is not much different. This is likely because the bias due to the measured confounders were small for most studies.

6.2.2 Accounting for uncertainty in unmeasured confounders or incomplete confounding

In the above, we estimated the bias due to confounding by SES by subtracting the unadjusted estimate ($\hat{\theta}^{\text{crude}}$) from the adjusted estimate ($\hat{\theta}^{\text{adj}}$) for studies that did report these estimates. It should be noted that this estimate of bias, like the estimates $\hat{\theta}^{\text{crude}}$ and $\hat{\theta}^{\text{adj}}$ themselves, is subject to sampling/random error. This error has not been ignored in the above model, but has rather been absorbed into the random error of the Binomial model (6.11). Nonetheless, there are still many reasons why this estimate may still depart from the true confounding bias due to SES, such as:

1. As noted above, the exposure contrasts with which the studies derived their adjusted estimates were not generally the one we are interested in, namely $> 0.3\mu$T vs $< 0.3\mu$T. We have assumed that the bias is the same for all exposure contrasts, which is probably not a very realistic assumption.

2. We assumed that the studies adjusted *only* for SES, used the correct model for adjustment, and that SES was measured without error. Evidently, many studies adjusted for other factors in addition to SES, and the inclusion of these additional factors may bias our estimates. Likewise, measurement error in the assessment of SES is also likely to contribute to bias (Greenland and Robins, 1985).

Apart from the above reasons, it was also found that the validity of the Normal approximation of model (6.3) was inadequate in the current case, as many studies have very few numbers of exposed cases and controls. We saw in Chapter 3 that without adjustment for measured confounders, the median and 95% posterior interval of $\exp\theta$ was 1.52 (1.16, 1.99) for the FE model and 1.54 (1.06, 2.24) for the RE model (Table 3.1), which translate to a median and 95% posterior interval of 0.42 (0.14, 0.69) and 0.43

(0.06, 0.81) for θ. However, when we replace

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{6.23}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{6.24}$$

with

$$\hat{\theta}_s^{\text{crude}} \sim N(\theta_s, \hat{\sigma}_s^{\text{crude}^2}) \tag{6.25}$$

the posterior median and 95% posterior interval of θ becomes 0.56 (0.28, 0.84) in the FE model and 0.60 (0.22, 0.97) in the RE model, such that the posterior distribution is shifted to the right by about 0.15. I demonstrate in Appendix D that this is most likely the result of the assumptions of the Normal approximation not being met.

Given these reasons, together with the fact that not all studies provided estimates of θ_s adjusted for SES, we cannot be considered to have sufficiently accounted for the uncertainty of the bias due to possible confounding with SES in the above model. We can, however, extend the model in a simple way to allow for this uncertainty, e.g., by adding an extra bias parameter, as in chapter 3, i.e. replacing

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{6.26}$$

with:

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s^* \tag{6.27}$$

$$\theta_s^* = \theta_s + \eta_s^{\text{conf}} \tag{6.28}$$

As in chapter 3, we can give η_s^{conf} uncertainty ranges, say

$$l_s \leq \eta_s^{\text{conf}} \leq u_s \tag{6.29}$$

These ranges can be elicited directly, or using the method described below.

6.2.3 Eliciting the range of η_s^{conf} through the consideration of a categorical unmeasured confounder

One possible method for eliciting the possible degree of confounding bias is by considering the various bias formulae described by Arah *et al.* (2008), which relate a crude, unadjusted odds ratio (OR^{crude}) to one standardized for an unmeasured confounder (OR^{std}) through the odds ratio between the confounder and the disease and the exposure. For example, assuming a confounder (U) with 2 levels, and with

standardization with respect to the exposed population, the authors showed that

$$OR^{\text{std}} = OR^{\text{crude}} \frac{(\Gamma\Lambda + 1)(\Omega\Lambda + 1)}{(\Gamma\Omega\Lambda + 1)(\Lambda + 1)} \tag{6.30}$$

$$\Gamma = \text{Odds ratio of disease between level 1 and level 0 of } U \text{ among the unexposed} \tag{6.31}$$

$$\Lambda = \text{Odds of the prevalence of } U \text{ among the unexposed} \tag{6.32}$$

$$\Omega = \text{Odds ratio of exposure between level 1 and level 0 of } U \tag{6.33}$$

This is also the formula Greenland (2005a) and Greenland and Kheifets (2006) made use of for adjustment for unmeasured confounding. Defining Ψ as:

$$\Psi = OR^{\text{crude}} / OR^{\text{std}} \tag{6.34}$$

it can be shown that:

$$\frac{\partial \Psi}{\partial \Gamma} \begin{cases} > 0 & \text{if } \Omega > 1 \\ < 0 & \text{if } \Omega < 1 \end{cases} \tag{6.35}$$

$$\frac{\partial \Psi}{\partial \Omega} \begin{cases} > 0 & \text{if } \Gamma > 1 \\ < 0 & \text{if } \Gamma < 1 \end{cases} \tag{6.36}$$

Thus, given $a_s \leq \Gamma_s \leq b_s$ and $c_s \leq \Omega_s \leq d_s$, we have:

$$\min(\Psi_s) = \min \left(\begin{array}{c} \frac{(a_s d_s \Lambda_s + 1)(\Lambda_s + 1)}{(a_s \Lambda_s + 1)(d_s \Lambda_s + 1)} \\ \frac{(b_s c_s \Lambda_s + 1)(\Lambda_s + 1)}{(b_s \Lambda_s + 1)(c_s \Lambda_s + 1)} \end{array} \right) \tag{6.37}$$

$$\max(\Psi_s) = \max \left(\begin{array}{c} \frac{(a_s c_s \Lambda_s + 1)(\Lambda_s + 1)}{(a_s \Lambda_s + 1)(c_s \Lambda_s + 1)} \\ \frac{(b_s d_s \Lambda_s + 1)(\Lambda_s + 1)}{(b_s \Lambda_s + 1)(d_s \Lambda_s + 1)} \end{array} \right) \tag{6.38}$$

Thus, if we elicit feasible ranges for Γ_s and Ω_s, and fix Λ_s, we imply a feasible range for Ψ_s. For example, let us assume that among studies that have adjusted for SES, U_s represents a variable that summarizes all incomplete control of confounding due to SES, with $U_s = 1$ indicating having a higher SES and $U_s = 0$ a lower SES. Further, because the definition of high and low SES is arbitrary, we can set the prevalence of U_s to 0.5, i.e. $\Lambda_s = 1$. Now, if high SES confers a 0.5 to 2 times increase in risk (odds) of childhood leukaemia, we have:

$$0.5 \leq \Gamma_s \leq 2 \tag{6.39}$$

If high SES are also 0.5 to 2 times more likely to be exposed, then

$$0.5 \leq \Omega_s \leq 2 \tag{6.40}$$

Given these ranges, we have:

$$8/9 \leq \Psi_s \leq 10/9 \qquad s = 2, 3, 4, 5, 7, 8, 13 \tag{6.41}$$

through equations (6.37) and (6.38). For studies that have not adjusted for SES, we may assume:

$$1/3 \leq \Gamma_s \leq 3 \tag{6.42}$$

$$1/3 \leq \Omega_s \leq 3 \tag{6.43}$$

i.e. we give it a wider range to represent our greater uncertainty. This then implies:

$$3/4 \leq \Psi_s \leq 5/4 \qquad s = 1, 6, 9, 10, 11, 12, 14 \tag{6.44}$$

Note that Ψ_s denote the ratio between the standardized and the unstandardized Odds Ratio when there is no adjustment for additional confounders. If we want to use $\log \Psi_s$ as estimates for η_s^{conf}, we have to make the additional assumption that this bias due to the missing confounder U is the same in the presence of the other confounders as in their absence. If this was the case, we have:

$$-0.12 \leq \eta_s^{\text{conf}} \leq 0.11 \qquad s = 2, 3, 4, 5, 7, 8, 13 \tag{6.45}$$

$$-0.29 \leq \eta_s^{\text{conf}} \leq 0.22 \qquad s = 1, 6, 9, 10, 11, 12, 14 \tag{6.46}$$

Using the model:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{6.47}$$

$$\operatorname{logit} \pi_{s1} = \operatorname{logit} \pi_{s0} + \theta_s^* \tag{6.48}$$

$$\theta_s^* = \theta_s + \eta_s^{\text{conf}} \tag{6.49}$$

$$\theta_s = \theta + \delta_s \tag{6.50}$$

with

$$\delta_s \sim N(-(\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}), \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) \qquad s = 2, 3, 4, 5, 7, 8, 13$$

$$\delta_s = 0 \qquad s = 1, 6, 9, 10, 11, 12, 14 \tag{6.51}$$

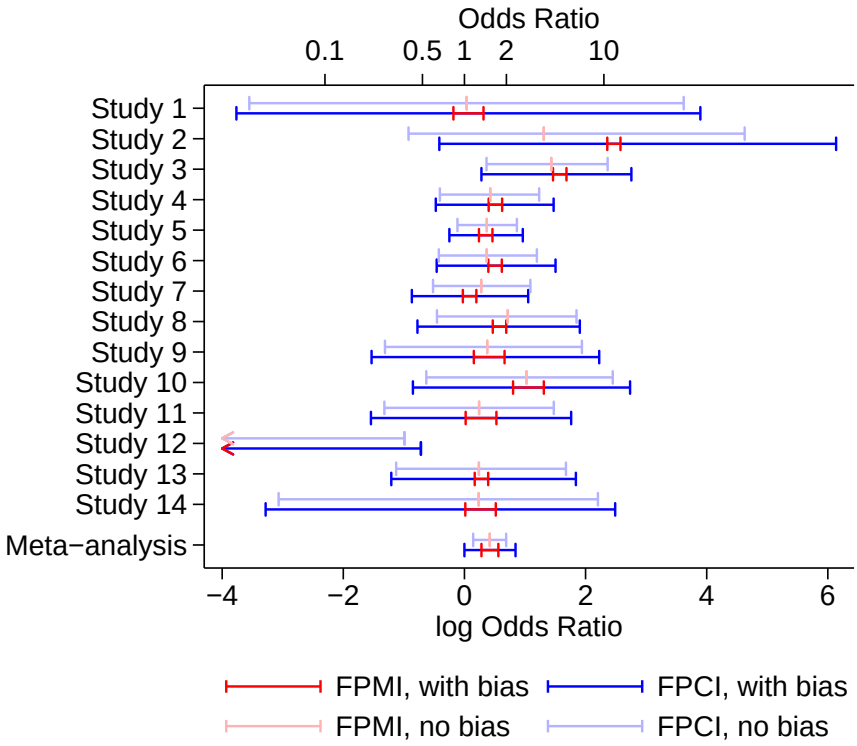


Figure 6.1: FPIs of θ with simple adjustment for bias due to incomplete control of confounding in a FE model

for the FE model and:

$$\begin{aligned} \delta_s &\sim N(-(\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}), 0.15 + \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) \qquad s = 2,3,4,5,7,8,13 \\ \delta_s &\sim N(0, 0.15) \qquad s = 1,6,9,10,11,12,14 \end{aligned} \tag{6.52}$$

for the RE model, I optimized over η_s^{conf} for max/min $\hat{\boldsymbol{\Theta}}$. The FPIs for the meta-analysis of this situation are given in Figure 6.1 for the FE model and Figure 6.2 for the RE model, compared to the no-bias scenario. Because the implementation of bias adjustment in this model is the same as that in the simple bias model of chapter 3, I had similar experience optimizing for max/min $\hat{\boldsymbol{\Theta}}$ in these models as I had for models in chapter 3. Multi-modality was not found in any of the problems.

Finally, let us note that in certain cases, we may know the direction of confounding bias due to an unmeasured confounder. If $1 \leq \Gamma \leq b$ and $1 \leq \Omega \leq d$ (or alternatively $a \leq \Gamma \leq 1$ and $c \leq \Omega \leq 1$), then we have $\Psi \geq 1$. If $1 \leq \Gamma \leq b$ and $c \leq \Omega \leq 1$ (or alternatively $1 \leq \Gamma \leq b$ and $c \leq \Omega \leq 1$), then $\Psi \leq 1$. Thus, if we have a clearer idea of how SES might be related to childhood leukaemia and EMF exposure, we may know the direction of possible confounding bias due to SES, and this can help us narrow the range of $\hat{\boldsymbol{\Theta}}$ considerably. This has been noted by Flanders and Khoury (1990), and expressed more generally by VanderWeele (2008), among others. It is also of note that simply knowing the direction of

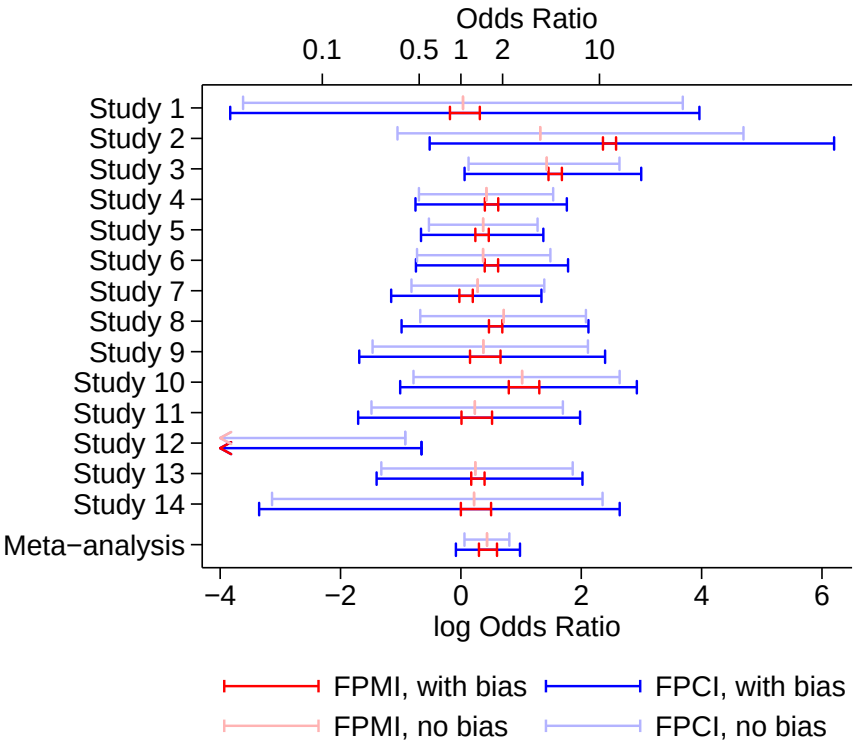


Figure 6.2: FPIs of θ with simple adjustment for bias due to incomplete control of confounding in a RE model

one of Γ and Ω is insufficient in determining the direction of Ψ.

Summary:

- In a meta-analysis setting, we do not generally have access to the raw data. As a result, a full model involving both the exposure and confounders is not possible. In a meta-analysis, one solution is to consider the reported adjusted estimate (an estimate of the effect adjusted for confounders) and model the estimate using a Normal distribution with zero mean and standard deviation estimated by the standard error.

- A disadvantage of this approach is that the model cannot be extended to account for non-participation bias and exposure misclassification bias in case-control studies using the techniques of the previous chapters. In this chapter I propose a modified model that allows for easy extension.

- However, for many reasons, we are unlikely to have completely controlled for confounding by adopting the proposed model. Resulting uncertainty due to incomplete control of confounding can be summarized by an additional parameter (η^{conf}). As in previous chapters, we can assign a feasible range for this parameter and examine the range of possible posterior inference given this range.

- There is, however, typically very little information by which we may choose suitable limits for η^{conf}. One possible method is to suppose the existence of a single unmeasured binary confounder, and elicit the range of η^{conf} by giving ranges to the odds ratio between the confounder and the exposure, and between the confounder and the disease.

- The model used for accounting for incomplete control of confounding is thus the same as the simple bias model used in chapter 3.

Chapter 7

Meta-analysis with multiple bias adjustment

More frequently than not, observational studies are subject to more than one type of bias, and authors such as Greenland (2005a) and Lash *et al.* (2009) have called for multiple bias modelling to examine the impact of all biases together. The models used in this thesis for exposure misclassification, non-participation bias, and incomplete confounding bias are all formulated in a way to make multiple bias modelling a straightforward extension from single bias models. As a reminder, for exposure misclassification, we have the model:

$$Y_{si} \sim Bin(N_{si}, p_{si}) \tag{7.1}$$

$$p_{si} = \pi_{si} sens_{si} + (1 - \pi_{si})(1 - spec_{si}) \tag{7.2}$$

$$\operatorname{logit} \pi_{s1} = \operatorname{logit} \pi_{s0} + \theta_s \tag{7.3}$$

$$\theta_s = \theta + \delta_s \tag{7.4}$$

For the non-participation bias model, we have:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \tag{7.5}$$

$$\pi_{si} = f(\pi^*_{si}, Q_{si}, R_{si}) \tag{7.6}$$

$$\operatorname{logit} \pi^*_{s1} = \operatorname{logit} \pi^*_{s0} + \theta_s \tag{7.7}$$

$$\theta_s = \theta + \delta_s \tag{7.8}$$

For incomplete confounding bias, we have:[1]

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \tag{7.9}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s^* \tag{7.10}$$

$$\theta_s^* = \theta_s + \eta_s^{\text{conf}} + \beta_s \tag{7.11}$$

$$\theta_s = \theta + \delta_s \tag{7.12}$$

$$\beta_s \begin{cases} \sim N(-(\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}), \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) & \text{if adjusted estimates were available} \\ = 0 & \text{otherwise} \end{cases} \tag{7.13}$$

To put all three models together, we have:

$$Y_{si} \sim Bin(N_{si}, p_{si}) \tag{7.14}$$

$$p_{si} = \pi_{si} sens_{si} + (1 - \pi_{si})(1 - spec_{si}) \tag{7.15}$$

$$\pi_{si} = f(\pi_{si}^*, Q_{si}, R_{si}) \tag{7.16}$$

$$\text{logit}\, \pi_{s1}^* = \text{logit}\, \pi_{s0}^* + \theta_s^* \tag{7.17}$$

$$\theta_s^* = \theta_s + \eta_s^{\text{conf}} + \beta_s \tag{7.18}$$

$$\theta_s = \theta + \delta_s \tag{7.19}$$

$$\beta_s \begin{cases} \sim N(-(\hat{\theta}_s^{\text{adj}} - \hat{\theta}_s^{\text{crude}}), \hat{\sigma}_s^{\text{adj}^2} - \hat{\sigma}_s^{\text{crude}^2}) & \text{if adjusted estimates were available} \\ = 0 & \text{otherwise} \end{cases} \tag{7.20}$$

Using the methods of this thesis, we provide uncertainty ranges for $sens_{si}, spec_{si}, Q_{si}, R_{si}, \eta_s^{\text{conf}}$ and examine the FPI of θ. Because exposure misclassification is involved, it will be necessary to provide informative prior distributions for π_{s0} or π_{s0}^*, as well as for θ_s. In the Greenland example, I believe it is better to assign prior distributions to π_{s0}, the prevalence of exposure in the *biased* population, rather than π_{s0}^*, the prevalence of exposure in the *target* population. This is because we have information on p_{s0}, the prevalence of misclassified exposure in the *biased* population, but not the target population. In chapter 5, I also gave π_{s0} rather than π_{s0}^* the prior distribution of $U(0, 1)$. As for $\theta_s = \theta + \delta_s$, I use:

$$\theta \sim N(0, 0.5) \tag{7.21}$$

$$\delta_s = 0 \qquad s = 1 \ldots 14 \tag{7.22}$$

[1] I have represented the model slightly differently than in chapter 6, by introducing the parameter β_s, representing the bias due to measured confounders. This enables us to treat δ_s in the same way as in the other chapters.

for the FE model and

$$\theta \sim N(0, 0.35) \tag{7.23}$$

$$\delta_s \sim N(0, 0.15) \qquad s = 1 \ldots 14 \tag{7.24}$$

for the RE model, as in chapter 4.

If we use the ranges given to $sens_{si}$, $spec_{si}$, Q_{si}, and R_{si} in previous chapters of this thesis (c.f. Tables 4.1 and 5.2), and the prior distribution given to π_{s0} in Table 4.1 together with the constraints:

$$1.05 \leq sens_{si} + spec_{si} \leq 1.7 \qquad i = 0, 1 \tag{7.25}$$

$$|sens_{s1} - sens_{s0}| \leq 0.05 \tag{7.26}$$

$$|spec_{s1} - spec_{s0}| \leq 0.02 \tag{7.27}$$

$$s = 1 \ldots 14 \tag{7.28}$$

and the bias due to measured confounders as given in Table 6.1, we have the following Feasible Posterior Intervals for θ in the meta-analyses:

$$[\min \hat{\theta}_M, \max \hat{\theta}_M] = [-1.38, 4.97] \tag{7.29}$$

$$[\min \hat{\theta}_L, \max \hat{\theta}_U] = [-2.38, 5.67] \tag{7.30}$$

for the FE model, and

$$[\min \hat{\theta}_M, \max \hat{\theta}_M] = [-1.14, 4.37] \tag{7.31}$$

$$[\min \hat{\theta}_L, \max \hat{\theta}_U] = [-2.02, 5.09] \tag{7.32}$$

for the RE model. In Figure 7.1 and 7.2, I compare these Feasible Posterior Intervals (FPIs) with the FPIs derived in single-bias meta-analyses. It is apparent that the majority of the uncertainty is due to exposure misclassification bias. Because there are more bias parameters in these multiple bias models than the single bias models of previous chapters, there are also more modes when carrying out optimization, with 75% of the meta-analysis problems having more than one mode.

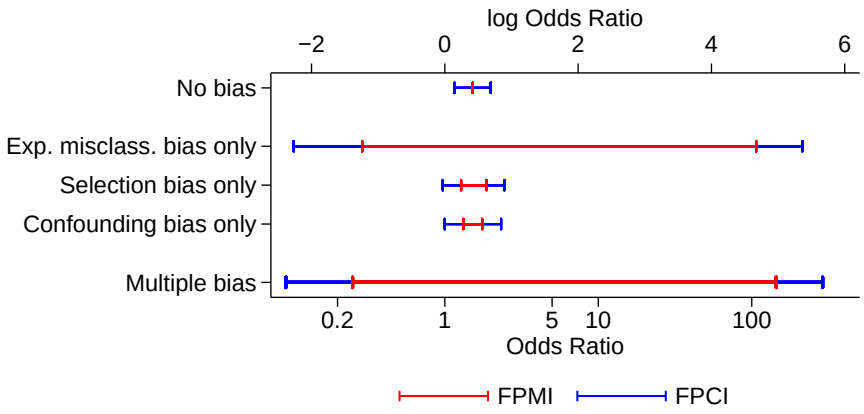


Figure 7.1: FPIs of θ in multiple bias FE meta-analysis as compared to single-bias meta-analysis

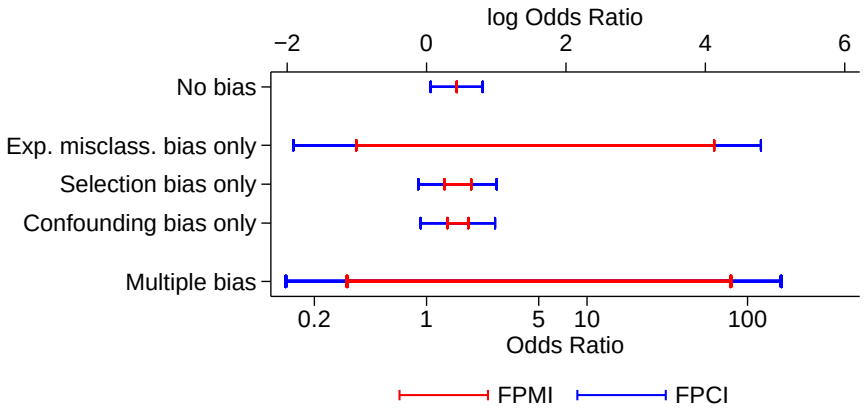


Figure 7.2: FPIs of θ in multiple bias RE meta-analysis as compared to single-bias meta-analysis

Chapter 8

Discussion

As discussed at the beginning of this thesis, although randomization or random sampling are often not employed in observational studies, standard practice in quantifying uncertainty has relied heavily on methodology developed for situations involving some form of randomization or random sampling (Greenland, 1990). As a result, commonly used measures of uncertainty of parameter estimates such as confidence intervals and standard error generally do not capture the "true" degree of uncertainty reliably. To overcome such limitations, there has been considerable interest recently to employ methods which extend traditional techniques, discussed in the biostatistics literature under topics such as "bias", or "causal inference".

In statistical parlance, "bias" often refers to systematic (long-run) departures of estimated quantities from their true values due to assumptions not being met. Uncertainty due to bias is not quantified in standard measures such as confidence intervals. The recent proposals, however, aimed to rectify this, and both probabilistic and non-probabilistic approaches have been advanced to quantify uncertainty due to bias. This thesis represents a valuable addition to the literature on non-probabilistic approaches to quantifying uncertainty due to bias. Previously, such approaches have been limited to the use of sensitivity analyses (e.g. Groenwold *et al.*, 2010; Arah *et al.*, 2008; Lash *et al.*, 2009), or the estimation of bounds of estimates in simple scenarios (e.g. Birnbaum and Sirken, 1950; Balke and Pearl, 1997; Vansteelandt *et al.*, 2006; VanderWeele, 2008; Kuroki *et al.*, 2010). In this thesis, I proposed a method for the derivation of bounds to estimates that can be applied to problems with many more parameters, through the use of an optimization algorithm.

The thesis also represents a significant addition to the quantitative evidence synthesis literature. Meta-analysis is perhaps the most widely practiced form of quantitative evidence synthesis. However, the lack of any randomization mechanisms in observational studies renders the meta-analyses of observational studies a less convincing method even than other forms of meta-analyses (Stroup and Thacker, 2005), which themselves are not without criticisms (Eysenck, 1978; Iyengar and Greenhouse, 1988; Shapiro, 1994; Egger *et al.*, 1998). Again, the presence of biases in study estimates, which are unac-

counted for in traditional forms of uncertainty summaries, has been at the heart of the problem (Turner *et al.*, 2009; Thompson *et al.*, 2011). Proposals have been made to quantify uncertainty in study-specific biases probabilistically (Greenland, 2005a; Turner *et al.*, 2009; Thompson *et al.*, 2011). These authors suggested assigning subjectively-elicited probability distributions to quantities/parameters that cannot be estimated from the data, in order to "adjust" standard models or estimates to provide better estimates and uncertainty measures for the target parameter of interest. As discussed in the first chapter of this thesis, probability distributions for this purpose can be very difficult to elicit accurately, since the number of parameters involved can be so many. Furthermore, if more than one elicitation is carried out (i.e. more than one "expert" is consulted), there is also the additional issue of finding a suitable way to summarize the different experts' probability distributions. The non-probabilistic approach of this thesis offers a solution to this problem. Since experts provide ranges rather than probability distributions, a natural way to combine different ranges specified by different experts is to seek the *union* of the ranges. Following the theory given in chapter 2 of this thesis, we can be sure that the Feasible Posterior Intervals (FPIs) derived in this way for the parameter of interest encompasses the FPIs derived based on any individual expert's information. This approach therefore has considerable appeal in a policy/decision making setting where the consultation of more than one expert is involved. More generally, when communicating the uncertainty in the results of bias-adjusted meta-analysis to the public, this method allows the reader to compare the results based on his/her own "prior information" to that given in the paper more readily. The reader can be invited to examine whether his/her ranges for the unknown parameters fall within those specified in the paper. If they do, then he/she can be certain that his/her FPIs also fall within that given in the paper.

For this reason, it may be recommended that when applying the method of this thesis in practice, one reports the FPIs based on at least several scenarios of differing "conservativeness", since we can expect our readers to have different degree of belief over the possible extent of bias. This enables readers of a wide range of different opinions to form their own conclusions based on the analyses of the study.

8.1 Some issues that are not addressed in the thesis

Because the main innovation of this project concerns the use of a non-probabilistic means to quantifying uncertainty due to biases, this thesis focuses on the challenges this presents, and does not address issues associated with the use of probability distributions in the quantification of the other parameters, including the target parameter θ, the between-study variation parameter δ_s, and the study-specific (true) prevalence of exposure among the controls π_{s0}. In the misclassification bias chapter, for example, I fixed the distribution of θ and δ_s at $\theta \sim N(0, 0.5), \delta_s = 0$ for a fixed-effects (FE) model and $\theta \sim N(0, 0.35), \delta_s \sim N(0, 0.15)$ for a random-effects (RE) model. The choice of the Normal distribution is by convention, and the choice of 0.35 and 0.15 as variances of these parameters is based on considerations

in the Greenland and Kheifets (2006) paper (although their considerations were not based on estimates from research findings, but rather subjective opinion). It is natural to expect the FPI of θ to be sensitive to its own prior distribution, and likely more so than to the prior distributions of other parameters. In theory, a possible improvement to this is to allow the user to specify a *set* of feasible prior distributions. A simple case would be to assume a Normal distribution for these parameters, and assign *feasible ranges* to the mean and variance parameters and use the algorithm of this thesis to seek the maximum and minimum posterior inference for θ within this range, and thus quantify uncertainty in the variance non-probabilistically. In this thesis, however, I avoided this complication in order to reduce computational burden and simplify discussion. The program that was written, however, does allow the user to provide ranges to the variance parameters, assuming Normal distributions.

Another issue not addressed in this thesis is the model that is used for the meta-analysis. As given in chapter 2, the "basic" meta-analysis model is:

$$Y_{si} \sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \tag{8.1}$$

$$\text{logit}\, \pi_{s1} = \text{logit}\, \pi_{s0} + \theta_s \tag{8.2}$$

This was based on convention, although by the arguments of chapter 1, one might question the appropriateness of the Binomial distribution for the number of exposed cases and controls, since they are frequently not a random sample from a large, well-defined population. If the recruited cases constitute a large proportion of the underlying population, for example, a more appropriate model might be the Hypergeometric distribution, which assumes sampling without replacement. In theory, this can be implemented within the current framework, although the evaluation of the log likelihood for the Hypergeometric distribution may prove computationally much more costly than for the Binomial.

A further issue which has not received attention in this thesis is the appropriateness of the RE model in the presence of heterogeneity. It has been argued that if study-specific effects of interest (θ_s) differ from one another, a summary estimate of the mean of the study-specific effect is irrelevant (Al khalaf *et al.*, 2011). Moreover, some have suggested if adjustment for biases is adequate, then there should be no remaining heterogeneity (Turner *et al.*, 2009). This argument may have validity to the extent that the definition of *bias* encompasses all departures from a particular target effect of interest. However, in this thesis, only three types of biases have been considered. It would not be surprising to find that even if all studies involved no biases due to exposure misclassification, non-participation bias, or unmeasured confounding, and were sufficiently large to ensure the extent of random error is small, some heterogeneity in effect estimates would remain due to differences in the underlying population being sampled, the period during which it is sampled, the background incidence of childhood leukaemia in the region/country specific to the study, and so on. Nonetheless, even when it is accepted that there are real between-study differences in the odds ratio, it remains a question as to whether the *mean* of the

study-specific effects is of interest. For example, if we were interested in the odds ratio of childhood leukaemia due to EMF in the UK, the odds ratio from a particular study in the UK might be more relevant than the global mean. In such a situation, it may be argued that a weighted mean, with more weight given to UK-based studies, is more appropriate.

Although the above issues are important and have been considered during the course of the development of this project, they were set aside in the main investigation of this thesis, which aims to focus on the practical aspects of implementing the proposed combined non-probabilistic/probabilistic approach to quantifying uncertainty. Because the proposed method is novel, a more important aim of the thesis was to explore and identify potential issues in the application of the technique in practice, particularly in the meta-analysis situation. These issues have been discussed to various extent in chapters 3 to 7. In the following section, I present a summary and some concluding remarks.

8.2 Issues concerning the application of the proposed method of this thesis

8.2.1 Wide FPIs

Throughout this thesis, we see that one important feature of the proposed method in a meta-analysis setting is that it allows the pooling of information across studies to reduce uncertainty that is considered random (i.e. aleatory uncertainty) but not uncertainty due to bias, which is not considered random (i.e. epistemic uncertainty). While such behaviour may be desirable philosophically, in practice it may allow too little uncertainty to be reduced by the pooling of studies in meta-analyses. This happens in situations where epistemic uncertainty dominates, and in the example of this thesis, we see this in the consideration of bias due to exposure misclassification, where sometimes the meta-analytic Feasible Posterior Credible Interval (FPCI) is even wider than the study-specific FPCI. When an FPCI is wide and straddles over the null, the analysis gives virtually no evidence for or against an association between the exposure and disease and is therefore not immediately useful in that particular situation. One feature of the methodology is that whenever the feasible region for the bias parameters $\boldsymbol{\eta}$ is tightened, the FPCI will also be narrowed, and therefore uncertainty due to bias will diminish if we know more about our bias parameters. Thus, the more information we have about the bias parameters, the more useful and applicable this method becomes. On the other hand, however, if the amount of data or the number of studies increases without concomitant increases in knowledge about the biases, then uncertainty due to residual randomness will decrease, and uncertainty due to biases dominate, and the method of this thesis becomes less and less useful. This remains a paradox in the application of this method in meta-analyses. In this thesis, I proposed we can also use the method as a tool for sensitivity analysis, which is less vulnerable to this problem, and is further discussed in the next subsection. I provide some other

possible developments to overcome this problem in the Future directions section (section 8.3).

8.2.2 The proposed method as a method for sensitivity analyses

One of the most pertinent issues with traditional sensitivity analyses when applied to models with many parameters (such as a meta-analysis model with bias adjustment) is that the number of potential scenarios we want to investigate may often be too many to carry out or to report. Using the methods of this thesis, instead of giving bias parameters fixed values for the sensitivity analyses, we can assign them feasible regions, and thus greatly reduce the number of potential scenarios that need to be carried out. However, as we see in chapter 4, the "wide FPI" problem we saw earlier for the standard analyses can remain a problem in sensitivity analyses. One solution to this problem would be to further subdivide our feasible regions into smaller and smaller sub-regions, since this guarantees narrower and narrower FPI. But this eventually leads us to the same problems that affect traditional sensitivity analyses. Nonetheless, in this thesis, it appears that the use of this method as sensitivity analyses has given us some useful insight into the effect of biases in modifying the apparent odds ratio of EMF on childhood leukaemia not easily obtained otherwise. In chapter 4, for example, we see that both *sensitivity* and *specificity* need to be high for all studies in order that an overall positive effect can be seen. We also see that posterior inference is fairly sensitive to the non-differential misclassification assumption, such that even a small departure from the assumption can have significant impact on the posterior inference for θ. In chapter 5 we see that a positive relationship between EMF and childhood leukaemia is favoured when exposed individuals are more likely to consent to study participation (i.e. when R_1 and R_0 are positive).

8.2.3 The assumption of independence between the bias parameters $\boldsymbol{\eta}$ and other parameters

Throughout this thesis, it is assumed that beliefs concerning the range of the bias parameters $\boldsymbol{\eta}$ and the other parameters that are treated probabilistically, such as θ, δ_s, π_{s0}, are independent. Sometimes, it may be more reasonable to assume that they are dependent. For example, in chapter 4, feasible regions are given to the parameters $sens_{s0}$, $sens_{s1}$, $spec_{s0}$, and $spec_{s1}$, but not to the parameters p_{s0} (the observed prevalence of exposure among the controls) or π_{s0} (the true prevalence of exposure among the controls) or θ_s (the study s-specific odds ratio). It may be desirable to give feasible regions to these parameters also, if only to avoid the conflict between prior distribution and likelihood, as we see in section 4.2.3. Giving a feasible region to a parameter already treated probabilistically such as θ is the same as restricting its domain, and is not difficult to implement in practice. However, to ensure that the prior distribution does not conflict with the likelihood severely, it may be necessary to impose dependence between the domain/distribution of θ and $\boldsymbol{\eta}$. For example, in the exposure misclassification

model, in order that the prior distribution with respect to p_{si} does not conflict with the likelihood, we may have to impose dependence between π_{si} and $sens_{si}$ and $spec_{si}$, since p_{si}, π_{si}, $sens_{si}$, and $spec_{si}$ are related together through the identity:

$$p_{si} = \pi_{si} sens_{si} + (1 - \pi_{si})(1 - spec_{si}) \quad i = 0, 1 \tag{8.3}$$

While such dependence between the prior distribution of π_{si} and bias parameters $sens_{si}$, $spec_{si}$ does not present too many additional problems computationally, it results in the Feasible Posterior Intervals losing their interpretation as Robust Bayes inference, since a key assumption in this interpretation is that the distribution of the non-bias parameters must be independent of the bias parameters (c.f. section 2.1). In an earlier stage of the development of this thesis, I considered the option of treating π_{s0} non-probabilistically together with $sens_{si}$ and $spec_{si}$ and thus avoid conflict between p_{s0} and $sens_{s0}$ and $spec_{s0}$. This, however, was dropped in favour of the treatment as presented in this thesis because of its tendency to produce FPIs that are even wider than those given in this thesis, and also because of difficulties in optimization. However, further research would be useful to determine what is the best solution to this problem.

8.2.4 Optimization issues

In this thesis, I proposed the use of the cyclic coordinate method for seeking the maximum and minimum of $\hat{\boldsymbol{\Theta}}$. One reason for using this method is that it takes advantage of the factorization of the likelihood function in a meta-analysis, such that not all parts of the likelihood need be evaluated at every iteration. Nonetheless, it is generally an inefficient algorithm. In this thesis, it was also found that multi-modality appears a fairly frequent phenomena and thus renders global optimization difficult. In this thesis, optimizations were repeated a number of times until new modes are no longer discovered. Of course, this still does not guarantee finding the global optima. Future research should focus on algorithms that focus on certain "hot-spots" where optima are likely to be found. Theoretical work may also be needed to characterize these "hot-spots".

8.3 Future research directions

In view of the above limitations of the method of this thesis, I believe useful research may be conducted in the following areas:

8.3.1 Use of additional constraints

To overcome the problem of FPIs being too wide, in addition to the proposals of this thesis, another option is to introduce additional constraints. For example, one type of constraint that can be introduced may require the *average* of a number of parameters to be within a certain region. An example might be a constraint given to the mean of the *sensitivity* parameter across studies:

$$l \leq \sum_{s} sens_{si}/n \leq u \tag{8.4}$$

This then disallows parameters to take their extreme values all at the same time in a meta-analysis. The above constraint is still linear, and in theory, can still be accommodated using the technique of this thesis. However, as we see in chapter 4, constraints that involve more than one parameter requires us to introduce additional search directions. In fact, this increase is exponential. Thus, to implement constraints such as the above, we are likely to require a different optimization strategy to the one used in this thesis.

8.3.2 Treating some bias parameters probabilistically

In this thesis, I have divided parameters into those that are "bias" parameters and those that are "non-bias" parameters, where the uncertainty of the former is quantified non-probabilistically and the latter probabilistically. This approach may be seen as a compromise between a purely probabilistic and a purely non-probabilistic approach. It overcomes some of the problems associated with the purely probabilistic approach, where specifying suitable probability distributions for a large number of parameters is difficult, and the purely non-probabilistic approach for specifying feasible regions, where the lack of a mechanism for downweighting extreme scenarios may lead to uncertainty intervals that are too wide to be useful. The division of parameters into "bias" and "non-bias" parameters in this way also means that the number of parameters whose uncertainty is quantified probabilistically is limited, and it is thus feasible to use exact numerical integration to evaluate the posterior medians and 95% intervals.

However, sometimes, it may be preferable to consider some of the "bias" parameters probabilistically as well. For example, uncertainty in the parameters Q_0 and Q_1, the proportion of people in the population who would not participate in the study, is estimated in this thesis by the confidence interval limits of the proportion of participation in the sample. In other words the extent of uncertainty is informed by reference to a random sampling mechanism. If this is reasonable, then a probabilistic quantification of uncertainty for these parameters will also be reasonable. In other situations, we may want to revert to using a probabilistic quantification of uncertainty simply because the non-probabilistic approach leads to uncertainty intervals that are too wide to be useful. The challenge at present in adopting an alternative strategy to divide parameters into "probabilistic" and "non-probabilistic" is mainly computational —

as the number of parameters that are quantified probabilistically increases, so does the computational burden in evaluating the posterior percentiles exactly. This problem may be overcome if we use an approximation method for the estimate, however (section 8.3.3).

I believe the potential for future development in this regards is enormous. As discussed in the introduction of this thesis, observational studies often do not have any kind of random sampling mechanism, and so the separation of parameters into "bias" and "non-bias" may not always correspond to the division between *epistemic* and *aleatory* uncertainty, and hence the treatment of one type probabilistically and another type non-probabilistically is still arbitrary to an extent. It remains to be seen what is the most informative or useful way of dividing parameters into the two categories in order to facilitate understanding of the data in different applications in epidemiology.

8.3.3 Use of an approximation method for evaluating $\hat{\boldsymbol{\Theta}}$

In this thesis, I have relied on the use of numerical integration for evaluating $\hat{\boldsymbol{\Theta}}$, i.e. the target percentile of the posterior distribution given the bias parameters. While the advantage of this is that calculation is exact, it is also slow and limits the flexibility of the bias model that can be used for the meta-analysis. Moreover, if the integration scheme of this thesis were not used, there would also be less reason for using the inefficient cyclic coordinate method for optimization, and would open us to a wider variety of options for optimization. I mentioned in chapter 2 that we could alternatively use approximation methods available for evaluating $p(\theta|X,\boldsymbol{\eta})$ and from there we can calculate $\hat{\boldsymbol{\Theta}} = (\hat{\theta}_L, \hat{\theta}_M, \hat{\theta}_U)$. Perhaps the most well known is Laplace Approximation where we approximate $p(\theta|X,\boldsymbol{\eta})$ by $\tilde{p}(\theta|X,\boldsymbol{\eta})$, where:

$$\tilde{p}(\theta|X,\boldsymbol{\eta}) = \frac{p(\theta,\boldsymbol{\theta'},X,\boldsymbol{\eta})|\boldsymbol{R}|^{-\frac{1}{2}}}{\int_\theta p(\theta,\boldsymbol{\theta'},X,\boldsymbol{\eta})|\boldsymbol{R}|^{-\frac{1}{2}}\,\mathrm{d}\,\theta} \tag{8.5}$$

$$\boldsymbol{R} = -\left[\frac{\partial^2 \log p(\theta,\boldsymbol{\theta'}|X,\boldsymbol{\eta})}{\partial(\boldsymbol{\theta'}\boldsymbol{\theta'}^T)}\right]_{\boldsymbol{\theta'}=\hat{\boldsymbol{\theta}}'(\theta)} \tag{8.6}$$

and $\hat{\boldsymbol{\theta}}'(\theta)$ is the value of $\boldsymbol{\theta'}$ that maximizes $\log p(\theta,\boldsymbol{\theta'}|X,\boldsymbol{\eta})$ given θ (Tierney and Kadane, 1986; Leonard and Hsu, 1999). The use of Laplace Approximation avoids evaluating $p(\theta|X,\boldsymbol{\eta})$ as a multidimensional integral:

$$p(\theta|X,\boldsymbol{\eta}) = \int_{\boldsymbol{\theta'}} p(\theta,\boldsymbol{\theta'}|X,\boldsymbol{\eta})\,\mathrm{d}\,\boldsymbol{\theta'} \tag{8.7}$$

and therefore saves time. It may also be viewed as an extension of the 'profile likelihood' technique for avoiding nuisance parameters (Leonard and Hsu, 1999, p.191). However, as is clear from (8.5), integration over one dimension is still needed. Furthermore, at every abscissa of θ, we need to find $\hat{\boldsymbol{\theta}}'(\theta)$ through a search algorithm. In preliminary work for this thesis, the computational burden for this appears comparable to full integration. Nonetheless, there is still an important advantage of using this

approximation, as described below.

In this thesis, I have relied on reusing parts of the integrands in reducing the computation burden in integration over 3 dimensions (section 2.2). This in turn necessitates the use of Newton-Cotes formula for integration (Appendix A.4), for which precise control over error tolerances is difficult. Even if the use of Laplace Approximation does not decrease computational burden, it allows us to avoid using Newton-Cotes formula for integration, such that tolerances for computational error can be more readily controlled (i.e. it allows me to avoid having to check for integration errors as I did in Appendix A.2).

8.3.4 Using non-Bayesian estimates of θ

It may be argued that a non-Bayesian approach to estimating θ is more consistent with the philosophy of not treating epistemic uncertainty probabilistically. In chapter 2, however, I mentioned that a Bayesian estimate is used for θ because this allows us to incorporate prior information in estimating θ through the use of subjectively-elicited prior distributions. However, as can be seen in chapter 4, the specification of a suitable prior distribution can be quite challenging and moreover, it introduces some counterintuitive results (section 3.2.1) which can be difficult to explain, particularly to a non-statistician audience. In principle, non-Bayesian methods can be employed also in the estimation of θ, though these can be expected to produce much wider uncertainty intervals for θ, since we do not have a prior. An advantage in using non-Bayesian methods, however, is that we can often avoid making distributional assumptions for parameters. This may be particularly useful for the misclassification model, for which the specification of priors for θ and logit π_{s0} is difficult in the Bayesian approach (section 4.2.1). Another potential benefit of the non-Bayesian approach is in the estimation of θ in the RE meta-analysis model. Unlike the Bayesian approach, the standard non-Bayesian method of estimation makes no distribution assumption on the random effects variance (Dersimonian and Laird, 1986). A disadvantage with non-Bayesian procedures, however, is that the construction of confidence intervals can be a difficult issue. Standard approaches for the construction of confidence intervals in complicated estimation problems in general rely on asymptotic theories. Their validity in meta-analyses with bias modelling needs to be explored further.

8.3.5 Alternative optimization strategies

The literature on optimization of smooth functions is huge, and it is likely that a more efficient algorithm can be found than the one proposed in this thesis. A promising option might be the gradient projection method of Rosen (1960) for optimization of problems with linear inequality constraints, as we have in this thesis. This method focuses attention on optimization within the space defined by the active constraints, and may be more effective than the cyclic coordinate method of this thesis.

8.3.6 Overcoming problems with multi-modality

The existence of multiple local optima in problems of this kind makes it difficult to find the global optimum. However, throughout this thesis, I have given some simple mathematical results to suggest where the extreme is likely to lie. Often, extremes are found at the expected locations, and in this thesis, this is uniformly the case for the simple bias model. It should be noted that by the Central Limit Theorem, likelihoods are approximately Normal for large samples. And if the likelihood is Normal, then analytical results are often available to allow us to determine the maximum/minimum directly without recourse to a search algorithm. However, we note in Appendix D that the Normal approximation can be poor, especially in small samples. It would be interesting to see if a compromise can be achieved, such that an approximation model can be found which gives reliable results even with small study sizes.

Appendix A

Integration details

The four integrals that we need to evaluate in this thesis are given on page 36. To evaluate these integrals, I used Newton-Cotes formulas to evaluate the integrand at a number of abscissas within a finite lower and upper limit a and b. a and b need to be chosen to encompass most of the mass of the probability density function. An *ad hoc* algorithm which allows us to adapt the limits of integration to the probability function at hand is given in the A.1. In section A.2, I discuss choosing the number of abscissas within $[a, b]$. In section A.3, I discuss deriving approximate estimates for the posterior variance of λ_s and γ_s, which is used to inform the choice of abscissas in sections A.1 and A.2. In section A.4, I discuss the Newton-Cotes formula used to compute the integrals from a fixed number of equally-spaced abscissas.

A.1 Choosing the integration limits

1. First, start with any suitable chosen range $[a, b]$. In my program, this range is $\bar{x} \pm k\bar{\sigma}$, where I have used x to denote a general parameter over which integration is needed. In integral A1 and A2, this is θ. In integral B, this is λ_s, and in integral C, this is γ_s. $\bar{x}$ is the mode from the previous integration (or else if it is the first ever integration, 0). $\bar{\sigma}$ is a rough estimate of the standard deviation of the distribution. For γ_s and λ_s, this is the square root of a preliminary rough estimate of the posterior variance of γ_s (see subsection A.3). For θ, this gets adapted based on the empirical posterior standard deviation of the previous evaluation. (In fact, in order to retain the same integration points, it is necessary that this does not change very frequently, since every time $\bar{\sigma}_\theta$ is changed, all previous calculations that involve θ have to be abandoned. $\bar{\sigma}_\theta$ is therefore readjusted only in "exceptional" circumstances, when the posterior standard deviation appears to be vastly different from $\bar{\sigma}$ that is being used.) k is a constant chosen to be 7 for γ_s, 7 for λ_s, and 6 for θ. These parameters were chosen by trial and error to optimize speed. They do not affect the accuracy of the algorithm.

2. Carry out numerical integration by summing up the integrand at n equally spaced points within the interval. The number to use for n will be discussed in the next section. Denote this integral by I, i.e.:

$$I = \sum_{1}^{n} f(x_i) \tag{A.1}$$

Note that $x_1 = a$ and $x_n = b$.

3. Check whether the range $[a, b]$ is sufficient by the following:

   (a) Evaluate the following

   $$I_1 = \sum_{a}^{a+\kappa} f(x_i) \tag{A.2}$$

   and

   $$I_2 = \sum_{b-\kappa}^{b} f(x_i) \tag{A.3}$$

   where

   $$\kappa = 0.05(b - a) \tag{A.4}$$

   represents the extreme 5% distance from both ends of the range $[a, b]$.

   (b) If $I_1/I > tol$, then change the range of integration to $[a', b]$, where $a' = a - 0.3(b - a)$ and repeat steps 3 and 4. (0.3 was chosen as a scaling factor to optimize speed.) If $I_2/I > tol$, then change the range of integration to $[a, b']$ where $b' = b + 0.3(b - a)$. (i.e. extending the range of integration either on the left or the right hand side if $[a, b]$ is deemed insufficient.) For the problems in this thesis, I have chosen tol to be 5×10^{-7}. This ensures that the extreme 5 percent of the range being considered covers at most 0.0000005 of the mass of the entire probability distribution, if the distribution is unimodal.

4. Repeat until both I_1/I and $I_2/I < tol$.

It appears that even with a distribution with as heavy tails as a Cauchy distribution, using the method to derive suitable limits for integration results in error of $< 1 \times 10^{-5}$ compared to the exact integral (see Box A.1).

In theory, it is possible for the method to fail in situations where the probability distribution is seriously "pathological", e.g. a multimodal distribution with modes separated by more than 5 standard deviations. Such scenarios, however, seem highly unlikely, especially when we use Normal prior distributions for most parameters in the thesis. Evaluation of the integral will and should fail if the integral is improper. This is one of the reasons why proper prior distributions are used in this thesis unless it is clear the posterior distribution cannot be improper.

The Cumulative distribution function of a standard Cauchy distribution is:

$$\frac{1}{\pi}\arctan(z) + 0.5$$

Therefore, given integration limits $[-a, a]$, with $a > 0$, the integral is:

$$\frac{1}{\pi}(\arctan(a) - \arctan(-a))$$

The integral of the extreme 5% of $[-a, a]$, i.e. $[-a, -0.9a]$, is:

$$\frac{1}{\pi}(\arctan(-0.9a) - \arctan(-a))$$

The following table shows the various values of these integrals as a function of a, assuming $f(x)$ is a standard Cauchy probability distribution function:

a	$I = \int_{-a}^{a} f(x)\,\mathrm{d}\,x$	$I_1 = \int_{-a}^{-0.9a} f(x)\,\mathrm{d}\,x$	Proportion (I_1/I)
0.1	0.0635	0.0032	0.0497
1	0.5	0.0167	0.0335
10	0.937	0.00350	0.00373
100	0.994	0.000354	0.000356
1000	0.99936	3.537×10^{-5}	3.539×10^{-5}
10000	0.999936	3.537×10^{-6}	3.537×10^{-6}
100000	0.9999936	3.537×10^{-7}	3.537×10^{-7}
70736	0.9999910	5×10^{-7}	5×10^{-7}

We see that in order that the extreme 5% of the range $[-a, a]$ has less than 5×10^{-7} coverage of the integral, a needs to be at least 70736. When a is 70736, the integral covers over 0.99999 of the mass of the distribution, and so error due to the use of this definite integral instead of the improper integral is less than 1×10^{-5}.

Box A.1: Demonstration that the algorithm for adapting the domain of integration results in error of less than 1×10^{-5} even when computing the integral of a Cauchy distribution

A.2 Choosing the number of integration points

In this section, I discuss choosing the number of abscissas between the integration limits $[a, b]$ for integration. Obviously, the greater the number of abscissas n we use for integration, the greater the resulting accuracy of the integral (generally). However, this also increases computational burden. It is therefore a matter of trial and error to find the optimal compromise.

In the application of this thesis, $F^{-1}_{\theta|X,\boldsymbol{\eta}}(p)$ does have to be evaluated to a high degree of accuracy, since if the error is large, then the surface over which $\hat{\boldsymbol{\Theta}}$ varies with $\boldsymbol{\eta}$ may become rough and the search algorithm (section 2.3) would fail. (Reminder: Our goal is to search for $\hat{\boldsymbol{\Theta}} = F^{-1}_{\theta|X,\boldsymbol{\eta}}(p)$ over values of $\boldsymbol{\eta}$ within $\mathcal{E}$.) However, it is difficult to determine a minimum threshold for accuracy that must be achieved. In this thesis, I simply adopt a precision level that appears to be acceptable, and does not cause problems in the optimization algorithm that is discussed in section 2.3.

Through trial and error, I have determined that for integral C, I use roughly $n = (a-b)/\bar{\sigma}$. For integral B, I use roughly $n = 3(a-b)/\bar{\sigma}$, where $\bar{\sigma}$ is defined in the previous section. For integrals A1 and A2, I use roughly $n = 7(a-b)/\bar{\sigma}$. These results in error less than 1×10^{-3} generally, although it is likely that for the examples used in this thesis the error is much less, in the order of 1×10^{-6} or so. These estimates of error are determined by comparing the results obtained to that obtained when n is doubled. In the following two subsections, I report the investigation I did to examine the precision of the integrals (1) when there is no bias, and (2) when there is misclassification bias.

A.2.1 No bias

Using the data given in Table 3.1, and the model of (2.14), (2.15), and (2.16), I calculated the 2.5%-ile, 50%-ile, and 97.5%-ile of θ, when the prior distributions of θ and δ_s are:

$$\theta \sim N(0, 0.35), \delta_s = 0 \tag{1}$$

$$\theta \sim N(0, 0.35), \delta_s \sim N(0, 0.15) \tag{2}$$

$$\theta \sim N(0, 100), \delta_s = 0 \tag{3}$$

$$\theta \sim N(0, 100), \delta_s \sim N(0, 0.15) \tag{4}$$

Table A.1 shows the percentiles for the four situations, calculated with different values of $n_\theta, n_\lambda, n_\gamma$. It can be seen that $n_\theta = 7(a-b)/\bar{\sigma}_\theta, n_\lambda = 3(a-b)/\bar{\sigma}_\lambda, n_\gamma = (a-b)/\bar{\sigma}_\gamma$ gives the same estimate as $n_\theta = 14(a-b)/\bar{\sigma}_\theta, n_\lambda = 6(a-b)/\bar{\sigma}_\lambda, n_\gamma = 2(a-b)/\bar{\sigma}_\gamma$ to at least 5 decimal places and so it suffices to use $n_\theta = 7(a-b)/\bar{\sigma}_\theta, n_\lambda = 3(a-b)/\bar{\sigma}_\lambda, n_\gamma = (a-b)/\bar{\sigma}_\gamma$ in this situation.

	Prior 1			Prior 2			Prior 3			Prior 4		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = 1/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = l/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = l/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = l/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	0.13008	0.39571	0.66031	0.03636	0.39358	0.74765	0.14471	0.41740	0.68910	0.05927	0.43411	0.80529

Table A.1: The 2.5%, 50%, and 97.5%-ile of the meta-analytic posterior distribution of θ in the Greenland meta-analysis under four different prior distributions, evaluated with different precisions. $l = (a - b)$.

A.2.2 Meta-analysis model with exposure misclassification adjustment

Here, I investigate how the evaluation of the $\hat{\boldsymbol{\Theta}}$ varies with n, the number of abscissas when the meta-analysis model includes adjustment for exposure misclassification. Within each of the four scenarios above, i.e.:

$$\theta \sim N(0, 0.35), \delta_s = 0 \tag{1}$$
$$\theta \sim N(0, 0.35), \delta_s \sim N(0, 0.15) \tag{2}$$
$$\theta \sim N(0, 100), \delta_s = 0 \tag{3}$$
$$\theta \sim N(0, 100), \delta_s \sim N(0, 0.15) \tag{4}$$

I have the following sub-scenarios:

$$sens_0 = sens_1 = 0.3; spec_0, spec_1 : \text{ low}; \pi_{s0} \sim U(0, 1) \tag{x.1}$$
$$sens_0 = sens_1 = 0.3; spec_0, spec_1 : \text{ low}; \pi_{s0} \sim \text{as in Table 4.1} \tag{x.2}$$
$$sens_0 = sens_1 = 0.3; spec_0 = spec_1 = 1; \pi_{s0} \sim U(0, 1) \tag{x.3}$$
$$sens_0 = sens_1 = 0.3; spec_0 = spec_1 = 1; \pi_{s0} \sim \text{as in Table 4.1} \tag{x.4}$$
$$sens_0 = sens_1 = 0.6; spec_0, spec_1 : \text{ low}; \pi_{s0} \sim U(0, 1) \tag{x.5}$$
$$sens_0 = sens_1 = 0.6; spec_0, spec_1 : \text{ low}; \pi_{s0} \sim \text{as in Table 4.1} \tag{x.6}$$
$$sens_0 = sens_1 = 0.6; spec_0 = spec_1 = 1; \pi_{s0} \sim U(0, 1) \tag{x.7}$$
$$sens_0 = sens_1 = 0.6; spec_0 = spec_1 = 1; \pi_{s0} \sim \text{as in Table 4.1} \tag{x.8}$$
$$sens_0 = sens_1 = 0.9; spec_0, spec_1 : \text{ low}; \pi_{s0} \sim U(0, 1) \tag{x.9}$$
$$sens_0 = sens_1 = 0.9; spec_0, spec_1 : \text{ low}; \pi_{s0} \sim \text{as in Table 4.1} \tag{x.10}$$
$$sens_0 = sens_1 = 0.9; spec_0 = spec_1 = 1; \pi_{s0} \sim U(0, 1) \tag{x.11}$$
$$sens_0 = sens_1 = 0.9; spec_0 = spec_1 = 1; \pi_{s0} \sim \text{as in Table 4.1} \tag{x.12}$$

In the above, "$spec_0, spec_1 :$ low" means they take the value of the lower bound *low* specificity given in Table 4.2.

There are therefore $4 \times 12 = 48$ scenarios. Taking the condition with the most number of abscissas, i.e. $n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$, as the gold standard, Table A.2 compares the maximum deviation in the evaluation of $\hat{\theta}_L, \hat{\theta}_M, \hat{\theta}_U$ with lesser n_θ, n_λ, or n_γ as compared to the gold standard over the 12 sub-scenarios.

Here, it can be seen under Prior 1 and 2, the maximum error from using $n =$ half the number of the gold standard is less than 4.3×10^{-5}. When using Priors 3 and 4, this can be considerably more, with error of up to 0.0013. We see that doubling the number of abscissas for λ and γ makes very little difference,

	Prior 1			Prior 2			Prior 3			Prior 4		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = l/\bar{\sigma}_\gamma$	1.7e-05	4.3e-05	3.9e-05	1.7e-05	3.4e-05	3.3e-05	1.9e-05	1.7e-04	1.3e-03	2.2e-05	1.7e-04	1.0e-03
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	1.7e-05	4.3e-05	3.9e-05	1.7e-05	3.4e-05	3.3e-05	1.9e-05	1.7e-04	1.3e-03	2.2e-05	1.7e-04	1.0e-03
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = l/\bar{\sigma}_\gamma$	1.9e-05	3.7e-05	3.7e-05	1.7e-05	3.4e-05	3.3e-05	2.0e-05	1.7e-04	1.3e-03	2.2e-05	1.7e-04	1.0e-03
$n_\theta = 7l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	1.9e-05	3.7e-05	3.7e-05	1.7e-05	3.4e-05	3.3e-05	2.0e-05	1.7e-04	1.3e-03	2.2e-05	1.7e-04	1.0e-03
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = l/\bar{\sigma}_\gamma$	3.4e-06	6.0e-06	5.7e-06	5.4e-08	1.7e-07	2.3e-06	4.5e-06	8.8e-06	1.4e-05	6.3e-08	1.2e-07	1.6e-07
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 3l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	3.4e-06	6.0e-06	5.7e-06	4.4e-08	2.1e-07	2.3e-06	4.5e-06	8.8e-06	1.4e-05	1.8e-08	3.6e-08	3.7e-08
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = l/\bar{\sigma}_\gamma$	0	0	0	2.5e-07	3.6e-08	1.5e-07	0	0	0	2.4e-07	3.8e-08	1.8e-07
$n_\theta = 14l/\bar{\sigma}_\theta, n_\lambda = 6l/\bar{\sigma}_\lambda, n_\gamma = 2l/\bar{\sigma}_\gamma$	0	0	0	0	0	0	0	0	0	0	0	0

Table A.2: The maximum absolute deviation from the gold standard in the evaluation of $\hat{\theta}_L, \hat{\theta}_M, \hat{\theta}_U$ in the Greenland meta-analysis under 48 different scenarios with 4 prior distributions. $l = (a - b)$.

so the error is due to insufficient n_θ. Increasing n_θ, however, significantly increases computational time, especially if we need to get the error level down to around 1×10^{-5} or so. Nonetheless, in this thesis, because we do not use flat or nearly flat prior distributions for θ when we consider the exposure misclassification models, the error should in general be not more than 4.3×10^{-5}.

I did not investigate the precision of integration for models involving non-participation and incomplete confounding bias, since these models do not affect the shape of the likelihood significantly, except in altering its central location, and therefore precision of integration using these models should be similar to the model without bias considered in the previous subsection.

A.3 Deriving approximate posterior variances for γ_s and λ_s

In this section, I describe the methods for deriving approximate posterior variances for γ_s given λ_s and θ, and for λ_s given θ. These approximate posterior variances are used to inform σ_γ and σ_λ. First, I describe the method used in the case where no biases are involved, and then in the case where exposure misclassification biases are involved.

A.3.1 Assuming no biases

A.3.1.1 Posterior variance of γ_s given λ_s and θ

Throughout this section, I make use of the following property: If the likelihood of a parameter is proportional to a Normal distribution, and a Normal prior distribution is used, then the posterior distribution is also Normal with variance

$$\mathrm{Var}^{\mathrm{post}} = 1/(1/\mathrm{Var}^{\mathrm{prior}} + 1/\mathrm{Var}^{\mathrm{Lik}}) \tag{A.5}$$

where $\mathrm{Var}^{\mathrm{post}}$ denotes the posterior variance, $\mathrm{Var}^{\mathrm{prior}}$ the prior variance and $\mathrm{Var}^{\mathrm{Lik}}$ the variance of the Normal distribution to which the likelihood is proportional (Gelman *et al.*, 2004, p.47). In other words, the posterior variance is a *harmonic mean* of the prior variance and the "likelihood variance" divided by 2.

The conditional likelihood of γ_s given λ_s, and θ is proportional to:

$$l(Y_{s0}, N_{s0}, Y_{s1}, N_{s1}|\gamma_s; \lambda_s, \theta) \propto (\mathrm{expit}\,\gamma_s)^{Y_{s0}}(1-\mathrm{expit}\,\gamma_s)^{N_{s0}-Y_{s0}}(\mathrm{expit}(\lambda_s+\theta))^{Y_{s1}}(1-\mathrm{expit}(\lambda_s+\theta))^{N_{s1}-Y_{s1}} \tag{A.6}$$

$$\propto (\mathrm{expit}\,\gamma_s)^{Y_{s0}}(1-\mathrm{expit}\,\gamma_s)^{N_{s0}-Y_{s0}} \tag{A.7}$$

which is the typical Binomial likelihood.[1] The Fisher's Information for γ_s is thus

$$I(\gamma_s) = (Y_{s0}(N_{s0} - Y_{s0}))/N_{s0} \tag{A.11}$$

which implies that in large samples, the likelihood is proportional to a Normal distribution with variance

$$1/I(\gamma_s) = 1/Y_{s0} + 1/(N_{s0} - Y_{s0}) \tag{A.12}$$

Thus, we may use equation (A.12) to give us an approximate estimate of the "likelihood variance" Var^{Lik}. One problem is that if Y_{s0} or $N_{s0} - Y_{s0}$ is 0, in which case (A.12) is undefined. To overcome this problem, I replaced equation (A.12) with:

$$\hat{\text{Var}}^{\text{Lik}} = 1/(Y_{s0} + 0.5) + 1/(N_{s0} - Y_{s0} + 0.5) \tag{A.13}$$

One reason for adding 0.5 to the denominators of (A.12) is that it leads to posterior inferences based on Jeffreys' prior for γ_s (Gelman *et al.*, 2004, p.63).

As for the prior distribution of γ_s given λ_s and θ, we observe that:

$$p(\gamma_s|\lambda_s, \theta) = p(\gamma_s|\lambda_s) \qquad \text{(by assumption of independence of } \theta \text{ and } \gamma_s) \tag{A.14}$$

$$= \frac{p(\lambda_s|\gamma_s)p(\gamma_s)}{p(\lambda_s)} \tag{A.15}$$

$$\propto p(\gamma_s)p_{\delta_s}(\lambda_s - \gamma_s) \qquad (\text{since } \delta_s = \lambda_s - \gamma_s) \tag{A.16}$$

where $p(\gamma_s)$ is the prior distribution of γ_s and $p_{\delta_s}(\lambda_s - \gamma_s)$ the prior distribution of δ_s. In this thesis, these two distributions are both Normal, and hence the prior variance of γ_s given λ_s and θ is:

$$\text{Var}^{\text{prior}} = 1/(1/\text{Var}(\gamma_s) + 1/\text{Var}(\delta_s)) \tag{A.17}$$

Combining the prior (A.17) and the likelihood (A.13) variance using equation (A.5), we have an estimate of the posterior variance of γ_s given λ_s and θ. Note that if $\delta_s = 0$, then the posterior variance is also 0. This is because if $\delta_s = 0$, then $\gamma_s = \lambda_s$, and hence γ_s is known (given λ_s).

[1] Recall that:

$$\gamma_s = \text{logit}\, \pi_{s0} \tag{A.8}$$

$$\lambda_s = \gamma_s + \delta_s \tag{A.9}$$

$$\text{logit}\, \pi_{s1} = \lambda_s + \theta \tag{A.10}$$

A.3.1.2 Posterior variance of λ_s given θ

Here, I again derive a formula for this posterior variance based on (A.5). In the following, I first derive a formula for $\text{Var}^{\text{prior}}$ and then a formula for Var^{Lik} and then combine the two to derive an estimate for Var^{post}.

$\text{Var}^{\textbf{prior}}$

By assumption, θ, δ_s, and γ_s are independent *a priori*, and hence the prior variance of λ_s given θ is:

$$\text{Var}^{\text{prior}} = \text{Var}(\lambda_s|\theta) = \text{Var}(\lambda_s) = \text{Var}(\gamma_s) + \text{Var}(\delta_s) \tag{A.18}$$

$\text{Var}^{\textbf{Lik}}$

The conditional likelihood of λ_s given θ, $Lik(X_s|\lambda_s;\theta)$ can be split up into the *case* component $Lik(X_{s1}|\lambda_s;\theta)$ and the *control* component $Lik(X_{s0}|\lambda_s;\theta)$

$$Lik(X_s|\lambda_s;\theta) = Lik(X_{s1}|\lambda_s;\theta)Lik(X_{s0}|\lambda_s;\theta) \tag{A.19}$$

where $X_{si} = \{Y_{si}, N_{si}\}, i = 1, 2$ and $X_s = \{X_{s1}, X_{s0}\}$. By the same logic as (A.5), I derive Var^{Lik} as:

$$\text{Var}^{\text{Lik}} = 1/(1/\text{Var}^{\text{Lik(X1)}} + 1/\text{Var}^{\text{Lik(X0)}}) \tag{A.20}$$

$\text{Var}^{\textbf{Lik(X1)}}$

$$Lik(X_{s1}|\lambda_s;\theta) \propto (\text{expit}(\lambda_s + \theta))^{Y_{s1}}(1 - \text{expit}(\lambda_s + \theta))^{N_{s1}-Y_{s1}} \tag{A.21}$$

Because θ is fixed, the likelihood with respect to λ_s is simply the likelihood with respect to $\text{logit}\,\pi_{s1}$ shifted to the left by θ (since $\text{logit}\,\pi_{s1} = \lambda_s + \theta$). The likelihood with respect to $\text{logit}\,\pi_{s1}$ is a standard logit-Binomial likelihood, and hence has Fisher's Information.

$$I(\text{logit}\,\pi_{s1}) = (Y_{s1}(N_{s1} - Y_{s1}))/N_{s1} \tag{A.22}$$

In large samples, the variance of the Normal distribution which approximates the shape of the likelihood (both with respect to to $\text{logit}\,\pi_{s1}$ and λ_s) would therefore have variance

$$1/I(\text{logit}\,\pi_{s1}) = 1/Y_{s1} + 1/(N_{s1} - Y_{s1}) \tag{A.23}$$

Again, to cope with possible zeros for Y_{s1} or $N_{s1} - Y_{s1}$, I used:

$$\hat{\text{Var}}^{\text{Lik(X1)}} = 1/(Y_{s1} + 0.5) + 1/(N_{s1} - Y_{s1} + 0.5) \tag{A.24}$$

$\mathrm{Var}^{\mathbf{Lik(X0)}}$

As for $Lik(X_{s0}|\lambda_s;\theta)$, let us note that:

$$Lik(X_{s0}|\lambda_s;\theta) \propto \frac{Lik(X_{s0}|\lambda_s,\theta)\tilde{p}(\lambda_s|\theta)}{p(X_{s0}|\theta)} \qquad (\text{if } \tilde{p}(\lambda_s|\theta)=1) \tag{A.25}$$

$$= \tilde{p}(\lambda_s|X_{s0},\theta) \tag{A.26}$$

where $\tilde{p}(\lambda_s|\theta)=1$ is an improper flat prior and $\tilde{p}(\lambda_s|X_{s0},\theta)$ denotes the posterior density of λ_s given this flat prior. Therefore, the likelihood is proportional to the posterior distribution of λ_s given θ, if a flat prior distribution were used for λ_s given θ. Now,

$$\mathrm{Var}_{\tilde{p}(\lambda_s|X_{s0},\theta)}(\lambda_s|X_{s0}) = \mathrm{Var}_{\tilde{p}(\lambda_s|X_{s0},\theta)}(\gamma_s+\delta_s|X_{s0}) \tag{A.27}$$

$$= \mathrm{Var}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\gamma_s|X_{s0}) + \mathrm{Var}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\delta_s|X_{s0}) + 2\,\mathrm{Cov}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\delta_s,\gamma_s|X_{s0}) \tag{A.28}$$

Therefore, we seek to estimate $\mathrm{Var}^{\mathrm{Lik(X0)}}$ as a sum of estimates of $\mathrm{Var}_{\tilde{p}(\gamma_s|\theta)}(\gamma_s|X_{s0})$ and $\mathrm{Var}_{\tilde{p}(\delta_s|\theta)}(\delta_s|X_{s0})$ and $2\,\mathrm{Cov}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\delta_s,\gamma_s|X_{s0})$. However, while we defined $\tilde{p}(\lambda_s|X_{s0},\theta)$, $\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)$ remains undefined. To be consistent, $\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)$ must be the posterior distribution of (δ_s,γ_s) given certain prior distribution such that the prior distribution of $\gamma_s+\delta_s=\lambda_s$ is flat. One possible option is if γ_s and δ_s were independent and γ_s were flat and δ_s is Normal with given prior variance $\sigma^2_{\delta_s}$. If this were the case, then in large samples,

$$\mathrm{Var}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\gamma_s|X_{s0}) \approx 1/(Y_{s0}) + 1/(N_{s0}-Y_{s0}) \tag{A.29}$$

since with a flat prior, the posterior is proportional to the likelihood. Here, the variance of the likelihood is approximated by the inverse Fisher's Information.

Also,

$$\mathrm{Var}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\delta_s|X_{s0}) = \sigma^2_{\delta_s} \tag{A.30}$$

$$\mathrm{Cov}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\delta_s,\gamma_s|X_{s0}) = 0 \tag{A.31}$$

since X_{s0} doe not inform δ_s, and δ_s and γ_s are independent *a priori*.

Again, to cope with possible $Y_{s0}=0$ or $N_{s0}-Y_{s0}=0$, I replaced equation (A.29) with:

$$\hat{\mathrm{Var}}_{\tilde{p}(\delta_s,\gamma_s|X_{s0},\theta)}(\gamma_s|X_{s0}) = 1/(Y_{s0}+0.5) + 1/(N_{s0}-Y_{s0}+0.5) \tag{A.32}$$

Altogether, we have:

$$\hat{\mathrm{Var}}^{\mathrm{Lik(X0)}} = 1/(Y_{s0}+0.5) + 1/(N_{s0}-Y_{s0}+0.5) + \sigma^2_{\delta_s} \tag{A.33}$$

Applying equation (A.5) to the estimates of $\text{Var}^{\text{Lik(X0)}}$, $\text{Var}^{\text{Lik(X0)}}$, and $\text{Var}^{\text{Lik(X0)}}$ derived in equations (A.18), (A.24) and (A.33), we derive an estimate of the posterior variance of λ_s given θ as:

$$\hat{\text{Var}}^{\text{post}} = 1/(1/\hat{\text{Var}}^{\text{Lik(X0)}} + 1/\hat{\text{Var}}^{\text{Lik(X1)}} + 1/\text{Var}^{\text{prior}}) \tag{A.34}$$

A.3.2 In the presence of exposure misclassification bias

In the previous subsection, we assumed there are no biases. In the presence of non-participation bias and bias due to incomplete control of confounding, because the data are not "diluted" as in bias due to exposure misclassification (see chapters 5 and 6), the posterior variance of γ_s and λ_s should not be very different from the case without bias (although the posterior mean might be considerably different). Therefore, I continue to use the above as estimates of the posterior variance. In the presence of exposure misclassification, the data are in effect "diluted", and hence the posterior variance would generally be larger. Using the above estimates for the posterior variance would generally lead to an underestimate. This, however, only increases the accuracy of the integration, though at the expense of computation time. To avoid such underestimation, I adjusted the posterior variances calculated above by the following factor:

$$\text{Var}^{\text{bias adjusted}} = \text{Var}^{\text{unadjusted}} \left(\frac{\hat{p}(1-\hat{p})}{\hat{\pi}(1-\hat{\pi})}\right)^2 \frac{1}{(sens + spec - 1)^2} \tag{A.35}$$

where $\hat{p}$ is the observed prevalence and $\hat{\pi}$ the true prevalence, derived using equation (4.5) in chapter 4. This is because:

$$1/I(\text{logit}\,\pi) = 1/I(\text{logit}\,p) \left(\frac{p(1-p)}{\pi(1-\pi)}\right)^2 \frac{1}{(sens + spec - 1)^2} \tag{A.36}$$

where $I(\text{logit}\,\pi)$ and $I(\text{logit}\,p)$ are Fisher's information with respect to $\text{logit}\,\pi$ and $\text{logit}\,p$, respectively, assuming fixed *sens* and *spec* and $p = \pi sens + (1-\pi)(1-spec)$ (c.f. equation 4.5).

Proof. Denote by $s = \text{logit}\,p$ and $t = \text{logit}\,\pi$, where $p = \pi sens + (1-\pi)(1-spec)$, and $Y \sim Bin(N, p)$.

$$I(t) = I(\text{logit}\,\pi) = -E_Y\left(\frac{\partial^2 l}{\partial t^2}\right) \tag{A.37}$$

where l is the log-likelihood. Now,

$$\frac{\partial l}{\partial t} = \frac{\partial l}{\partial s}\frac{\partial s}{\partial p}\frac{\partial p}{\partial \pi}\frac{\partial \pi}{\partial t} \tag{A.38}$$

Note that because

$$\frac{\partial p}{\partial \pi} = sens + spec - 1 \tag{A.39}$$

is a constant,

$$\frac{\partial^2 l}{\partial t^2} = \frac{\partial p}{\partial \pi}\left[\frac{\partial l}{\partial s}\left(\frac{\partial s}{\partial p}\frac{\partial^2 \pi}{\partial t^2} + \frac{\partial \pi}{\partial t}\frac{\partial^2 s}{\partial p^2}\right) + \left(\frac{\partial s}{\partial p}\right)^2\left(\frac{\partial \pi}{\partial t}\right)^2 \frac{\partial^2 l}{\partial s^2}\frac{\partial p}{\partial \pi}\right] \tag{A.40}$$

(by differentiating (A.38) with respect to t.) Note that because

$$E_Y\left(\frac{\partial l}{\partial s}\right) = \frac{\partial p}{\partial s} E_Y\left(\frac{\partial l}{\partial p}\right) \tag{A.41}$$

since $\frac{\partial p}{\partial s} = p(1-p)$ and is independent of Y, and that

$$E_Y\left(\frac{\partial l}{\partial p}\right) = E_Y\left(\frac{Y}{p} - \frac{N-Y}{1-p}\right) = 0 \tag{A.42}$$

after taking expectation (over Y), the first term on the right side of (A.40) $= 0$, and hence combining (A.37) and (A.40) gives:

$$I(t) = -E_Y\left(\frac{\partial^2 l}{\partial t^2}\right) \tag{A.43}$$

$$= -E_Y\left(\left(\frac{\partial p}{\partial \pi}\right)^2 \left(\frac{\partial s}{\partial p}\right)^2 \left(\frac{\partial \pi}{\partial t}\right)^2 \frac{\partial^2 l}{\partial s^2}\right) \tag{A.44}$$

It turns out that $\frac{\partial p}{\partial \pi}$, $\frac{\partial s}{\partial p}$, $\frac{\partial \pi}{\partial t}$ are all independent of Y, and hence,

$$I(t) = -\left(\frac{\partial p}{\partial \pi}\right)^2 \left(\frac{\partial s}{\partial p}\right)^2 \left(\frac{\partial \pi}{\partial t}\right)^2 E_Y\left(\frac{\partial^2 l}{\partial s^2}\right) \tag{A.45}$$

$$= \left(\frac{\partial p}{\partial \pi}\right)^2 \left(\frac{\partial s}{\partial p}\right)^2 \left(\frac{\partial \pi}{\partial t}\right)^2 I(s) \tag{A.46}$$

$$= \left(\frac{\partial p}{\partial \pi}\right)^2 \left(\frac{\partial s}{\partial p}\right)^2 \left(\frac{\partial \pi}{\partial t}\right)^2 I(\operatorname{logit} p) \tag{A.47}$$

Verifying that:

$$\frac{\partial p}{\partial \pi} = sens + spec - 1 \tag{A.48}$$

$$\frac{\partial s}{\partial p} = \frac{1}{p(1-p)} \tag{A.49}$$

$$\frac{\partial \pi}{\partial t} = \pi(1-\pi) \tag{A.50}$$

completes the proof. □

In the problems of this thesis *sens* and *spec* often take on a range of values rather than being fixed at particular values. From (A.35) we may expect $\text{Var}^{\text{bias adjusted}}$ to be smaller as $sens+spec$ becomes further away from 1. Because it is worse to overestimate than to underestimate the variance (overestimation of variance leads to loss of accuracy; underestimation leads to increase in computation time), I choose the value of *sens* and *spec* within their feasible regions that result in $sens + spec$ furthest away from 1 to

use in (A.35).

It is important to note that $\hat{\pi}$ is not always defined, i.e. not always between 0 and 1, as calculated from equation (4.5). Moreover, using equation (A.35) can often lead to variance estimates that are unrealistically large. Thus, I also impose a "cap" on $\text{Var}^{\text{bias adjusted}}$: $\text{Var}^{\text{bias adjusted}}$ must not exceed its prior variance. If it exceeds its prior variance, it is replaced by the prior variance. This agrees with intuition that in general the posterior variance is not larger than the prior variance.

A.4 Newton-Cotes formula for evaluating an integral

Suppose we have evaluated the integrand $f(x)$ at $n+2$ equally-spaced abscissas from a to b, i.e. $f(x_0), f(x_1), \ldots, f(x_n), f(x_{n+1})$, where $x_0 = a$ and $x_{n+1} = b$. The integral can be written as a sum of sub-integrals:

$$\int_{x=x_0}^{x_{n+1}} f(x)\,\mathrm{d}\,x = \sum_{i=0}^{n} \int_{x=x_i}^{x_{i+1}} f(x)\,\mathrm{d}\,x \tag{A.51}$$

For any integral of the function $f(x)$ over the space $[x_1, x_2]$, Newton-Cotes formulae approximates the integral as a weighted sum of $f(x_{(1)}), f(x_{(2)}), f(x_{(3)}), \ldots$, where $x_{(1)}, x_{(2)}, x_{(3)}, \ldots$ are generally chosen to be in the neighbourhood of $[x_1, x_2]$. An $n-$point Newton-Cotes formula is a weighted sum of n different evaluations, and is exactly correct if $f(x)$ over the intervals of the n points (not just over $[x_1, x_2]$) can be exactly represented by a polynomial of order $n-1$. For example, if $f(x)$ over the interval of $[x_1, x_2]$ can be exactly represented by a cubic equation, then the integral can be exactly evaluated as $\Delta(-\frac{1}{24}f(x_1-\Delta) + \frac{13}{24}f(x_1) + \frac{13}{24}f(x_2) - \frac{1}{24}f(x_2+\Delta))$ where $\Delta = x_2 - x_1$. When evaluating the integral between x_1 and x_n, with equally spaced abscissas, we can simply add up the individual $[x_1, x_2]$, $[x_2, x_3]$, $[x_3, x_4]$, …, $[x_{n-1}, x_n]$ integrals, resulting in the formula:

$$\begin{aligned}\int_{x_1}^{x_n} f(x)\,\mathrm{d}\,x \approx \Delta\Bigg(&-\frac{1}{24}f(x_0) + \frac{12}{24}f(x_1) + \frac{25}{24}f(x_2) + f(x_3) + f(x_4) + f(x_5) + \cdots \\ &+ f(x_{n-3}) + f(x_{n-2}) + \frac{25}{24}f(x_{n-1}) + \frac{12}{24}f(x_n) - \frac{1}{24}f(x_{n+1})\Bigg)\end{aligned} \tag{A.52}$$

It has been shown that when the integrand is not a polynomial, integration using this scheme results in error of the order $1/\Delta^5$, i.e. when Δ is halved, the accuracy increases roughly 32 times (Press *et al.*, 2007).

Appendix B

A simple algorithm for optimization in one dimension

In Box 2.1 of chapter 2, step 3 involves optimizing the objective function over 1 dimension within given limits. In many mathematical programming environments, this is straightforward, as there will be a pre-packaged function for this. In the programming environment that I used for this thesis (The Mata programming language in Stata 10.1), however, only unconstrained optimizers are available. Although it is possible to transform the problem into an unconstrained optimization problem by adding in barrier penalties (e.g. Nocedal and Wright, 1999), this was deemed too cumbersome. Moreover, when I first set about programming for this thesis, I had hoped to solve a very general problem (more general than the ones that are discussed in this thesis, including problems with discontinuous constraints), and wanted a very general algorithm that copes with discontinuity as well as multi-modality and missing values. I therefore developed the following simple line search algorithm for this purpose, based on the golden section search algorithm (Press *et al.*, 2007, p.492).

A simple line search algorithm

Consider finding **min** $f(\eta)$ subject to $l \leq \eta \leq u$:

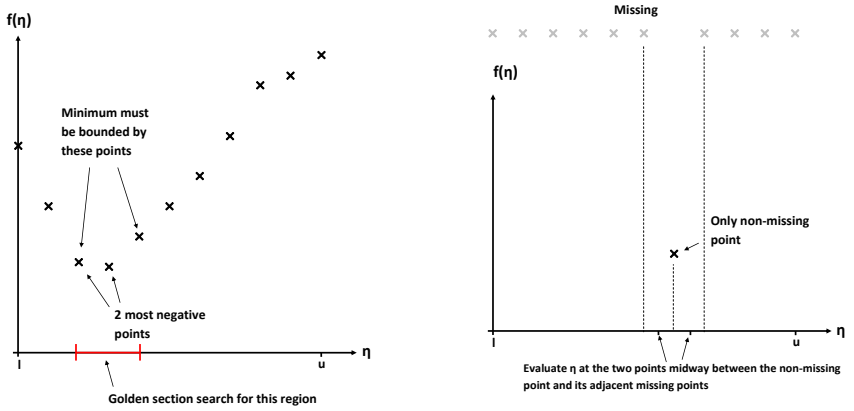
1. Evaluate $f(\eta)$ at n equally spaced intervals between l and u. For the problems in this thesis, I chose n to be 11.

2. Identify the two most negative points. If the most negative point is at the boundary and the second most negative point is next to it, then carry out a golden section search (Press *et al.*, 2007, p.492) based on these two points. If the two most negative points are next to one another, but the most negative point is not at the boundary, then carry out a golden section search, based on these two points and the other point that is next to the minimum point (see Figure B.1a.). The golden section search should finally converge to a point $\dot{\eta}$ when neighbouring η's do not differ by

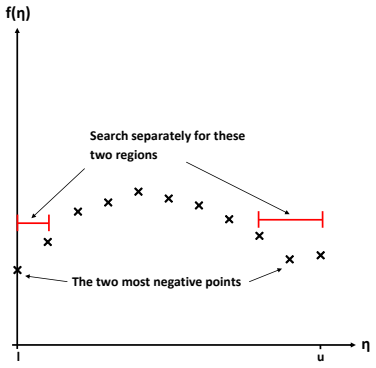
more than a certain tolerance level. I used 1×10^5 for this tolerance. Otherwise, a flat region is encountered, when $f(\eta)$ is essentially the same across a wide range of η's. In this case, declare a flat region is encountered and set $\dot{\eta}$ as a randomly chosen η among the η's with the same $f(\eta)$.

3. If only one point is non-missing, then evaluate $f(\eta)$ for two more points which are halfway between the adjacent points of the only non-missing point. Keep repeating until at least one other non-missing point is found, in which case follow step 2. If no other non-missing point is found, then the only non-missing point is naturally also the minimum of the function. (see Figure B.1b)

4. If the initial two most negative points were not adjacent to one another, then repeat step 3 for the two most negative points separately, and compare the final optimized result of these two "modes", and select the one with the more negative $f(\dot{\eta})$. (See Figure B.1c).



(a) (b)



(c)

Figure B.1: (a) When the two minimum points are adjacent, we perform a Golden section search for the region bounded by the two points adjacent to the lowest point. (b) When only one point is non-missing, we continue to search around the point until we find another non-missing point. (c) When the two minimum points are not adjacent, we search through the two regions separately and find the minimum of the two.

Appendix C

Estimating misclassification probabilities of EMF exposure assessment

In my Master's thesis (Mak, 2008), I estimated the correlation between the true EMF measurement and various imperfect measurement of EMF that has been used in the literature. This was based on various studies in the literature that examined the correlation between various different types of exposure measurement, and between the same measurements taken at different time points. I proposed that the correlation between a particular EMF measurement and a child's true exposure levels, defined as the average EMF level in the three months prior to diagnosis, be estimated as a product of three components — the *stability*, the *non-personal-monitor penalty*, and the *incomplete coverage penalty*.

The *stability* is defined as the correlation between the study's measure with an average of the measure made in exactly the same way except that it is made during the target exposure period. This depends on the mode of estimation (e.g. measurements based on personal monitor tend to be less stable, and 24-hr measurements are assumed to be more stable than "spot" measurements). The non-personal-monitor penalty is defined as the correlation between the average exposure during the target exposure period and the average of *personal monitor* measurements made during this period. The *coverage penalty* is an arbitrary penalty given to studies depending on the time lapse between their measurement period (typically post diagnosis), and the target exposure period.

Some estimates for each of these quantities were available in the literature for various types of measurements, and I used the most relevant estimate for each of the studies in the Greenland (2005a) meta-analysis to derive an overall correlation estimate, as given in Table C.1.

As can be seen in Table C.1, studies with direct measurement of EMF (such as room measurement) tend to have lower stability but smaller (i.e. more positive) non-personal-monitor penalty. In contrast, studies with calculated EMF estimates (based on distance to nearest power lines) such as Tynes and Haldorsen (1997) have perfect stability but a greater non-personal-monitor penalty. Nonetheless, it was clear that huge scope for error exists in these estimates.

Study	Stability	Non-personal monitor penalty	Coverage penalty	Overall Correlation
Coghill *et al.* (1996)	0.837	0.75	1	0.63
Dockerty *et al.* (1998)	0.837	0.75	1	0.63
Feychting and Ahlbom (1993)	1	0.4	1	0.4
Kabuto *et al.* (2006)	0.837	0.75	1	0.63
Linet *et al.* (1997)	0.837	0.71	0.8	0.50
London *et al.* (1991)	0.837	0.75	0.9	0.56
McBride *et al.* (1999)	0.548 0.837	1 0.64	0.9	0.49*
Michaelis *et al.* (1998)	0.837	0.75	1	0.63
Olsen *et al.* (1993)	1	0.4	0.7	0.28
Savitz *et al.* (1988)	0.837	0.75	1	0.63
Tomenius (1986)	0.837	0.64	1	0.54
Tynes and Haldorsen (1997)	1	0.4	0.7	0.28
UKCCS (1999)	0.812	0.8	1	0.65
Verkasalo *et al.* (1993)	1	0.4	0.7	0.28

* Calculated as the average of the two rows, because some children are measured in one way and some in another

Table C.1: Estimates of overall correlation between measurements of EMF and target "true" exposure

If we, however, believe that the rationale behind the estimation was basically sound, i.e. that the true correlation can be estimated as the product of the three components, it does appear unlikely that the true correlation between measured EMF exposure and true exposure can be high. In my Master's thesis, I also explored how these correlation coefficients may be converted to the misclassification probabilities *sens* and *spec*, through the fitting of a bivariate Normal distribution. This is taken further below.

C.1 Converting correlations to misclassification probabilities

Let us denote by X the measured EMF levels and by T the true EMF levels that the child is exposed to and assume:

$$\begin{pmatrix} X \\ T \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_T \\ \rho\sigma_X\sigma_T & \sigma_T^2 \end{pmatrix}\right) \tag{C.1}$$

where μ_X and μ_T denote the mean, and σ_X and σ_T denote the standard deviation, and ρ denote the correlation of X and T. Whether a child is classified as exposed or not depends on whether X is greater than a certain threshold τ, such that:

$$p = Pr(X > \tau) \tag{C.2}$$

$$\pi = Pr(T > \tau) \tag{C.3}$$

$$sens = Pr(X > \tau | T > \tau) \tag{C.4}$$

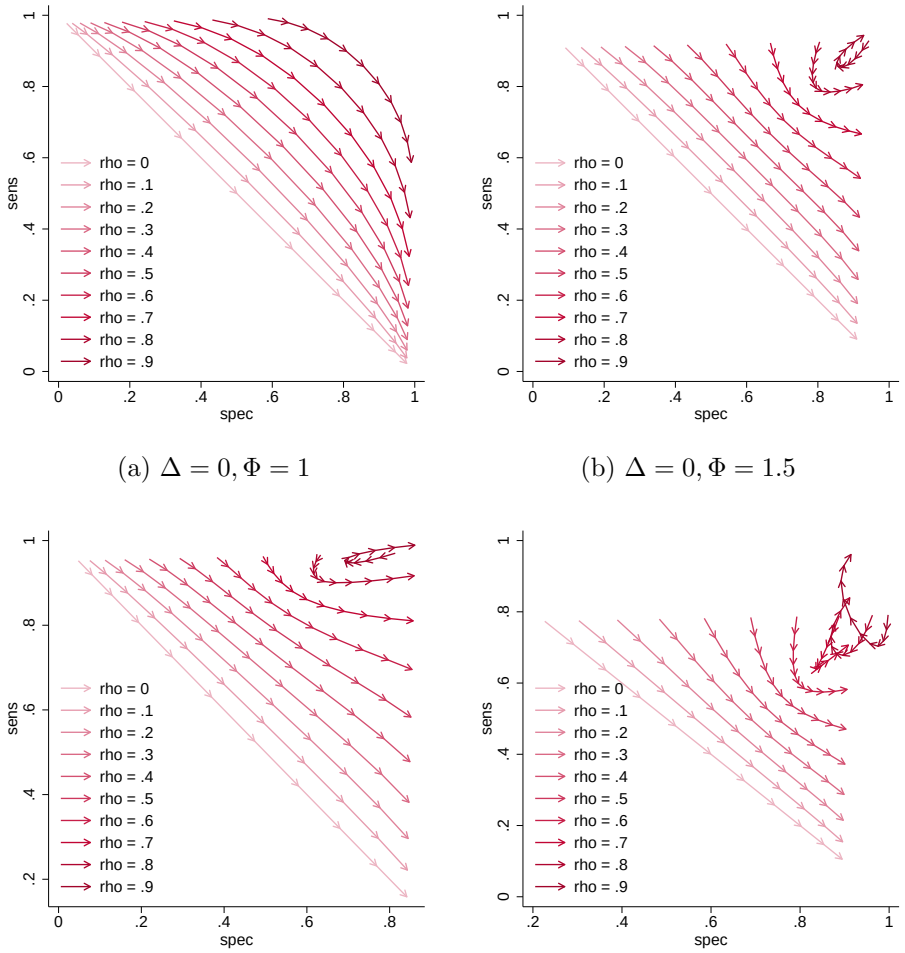
$$spec = Pr(X < \tau | T < \tau) \tag{C.5}$$

Given the form of the bivariate Normal distribution, it is clear that *sens* and *spec* are in fact functions of the parameters τ, ρ, Δ, and Φ, where $\Delta = \mu_X - \mu_T$ and $\Phi = \sigma_X^2/\sigma_T^2$.

In Figure C.1, I examine how *sens* and *spec* varies with different values of Δ, Φ, ρ, and $\tau' = (\tau - \mu_T)/\sigma_T$. In Figure C.1a, we see that when $\Delta = 0$, and $\Phi = 1$, i.e. when the observed exposure and the true exposure have the same mean and variance, the Resceiver Operation Characteristic (ROC) curve moves from the diagonal and become closer to the upper and right boundary with increasing ρ. The Area Under the Curve (AUC) naturally also increases. It is of note that the maximum $sens + spec$ is achieved when $\tau' = 0$. Because in this thesis, we use a high level of EMF as a cut-off point, we expect τ' to be positive. Note that when τ' approaches ∞ or $-\infty$, however, either $sens \to 1$ or $spec \to 1$. With a high threshold, we tend to have high *spec* and low *sens*. With a low threshold, we get the reverse.

However, it may not be the case that the observed exposure and the true exposure have the same mean and variance. In fact, under the classical measurement error model, i.e.:

$$X = T + \epsilon$$



(a) $\Delta = 0, \Phi = 1$ (b) $\Delta = 0, \Phi = 1.5$

(c) $\Delta = 0.5, \Phi = 1.5$ (d) $\Delta = 0, \Phi = 2$

Figure C.1: Diagrams showing how *sens* and *spec* vary with Δ, Φ, ρ, and τ'. The direction of the arrows is from low values of τ' to high values. The minimum τ' is -2 and the maximum is 2. The lines that interpolate the $(sens, spec)$ are called Receiver Operating Characteristics (ROC) curves. These are in darker shade of red for higher ρ lighter shade for smaller ρ

where T and ϵ are independent, the observed exposure (X) is expected to have a greater variance than the variance for the true exposure (T), i.e. $\Phi > 1$. Figure C.1b shows the case for $\Phi = 1.5$. Here, it is interesting that for $\rho < 0.7$ or so, the ROC curves are much more linear, and almost parallel to lines defined by $sens + spec = k$. When $\rho = 0.9$, *sens* no longer tends to 0 as $\tau' \to \infty$. Instead, *sens* tends to 1, both when $\tau' \to \infty$ and when $\tau' \to -\infty$.

Figures C.1c and C.1d show how the ROC curves vary with other values of Δ and Φ. In general, the ROC curve for high ρ is more unpredictable but for lower values of ρ, as long as Φ is around 1.5 - 2, it is roughly parallel to lines defined by $sens + spec = k$.

Therefore, based on arguments that the correlation between X and T are unlikely to be high (most likely less than 0.7), and that *sens* appears to have a nearly linear relationship with *spec* for $\rho < 0.7$ or so, a reasonable upper bound for the parameters $sens + spec$ might be:

$$sens + spec \leq 1.7 \tag{C.6}$$

On the other hand $sens + spec > 1$ corresponds to:

$$Pr(\text{Truly exposed}|\text{Observed to be exposed}) > Pr(\text{Truly exposed}|\text{Observed to be unexposed}) \tag{C.7}$$

i.e.

$$sens + spec > 1 \iff Pr(T > \tau | X > \tau) > Pr(T > \tau | X < \tau) \tag{C.8}$$

(I give the proof of (C.7) in the next section.) Since it seems reasonable that those classified as exposed are more likely to be truly exposed than those otherwise classified, I believe a reasonable lower bound for $sens + spec$ for any EMF exposure assessment, for both cases and controls, would be:

$$1.05 \leq sens + spec \tag{C.9}$$

C.2 Proof that $sens + spec > 1$ is equivalent to (C.7)

Let a, b, c, d denote the probabilities $Pr(T > \tau, X > \tau)$, $Pr(T > \tau, X < \tau)$, $Pr(T < \tau, X > \tau)$, and $Pr(T < \tau, X < \tau)$ respectively. The sensitivity (*sens*) and specificity (*spec*) are therefore defined as:

$$\begin{aligned} sens &= Pr(X > \tau | T > \tau) = \frac{a}{a+b} \\ spec &= Pr(X < \tau | T < \tau) = \frac{d}{c+d} \end{aligned} \tag{C.10}$$

Likewise, the positive predictive value (PPV) and negative predictive value (NPV) are defined as:

$$\begin{aligned} PPV &= Pr(T > \tau | X > \tau) = \frac{a}{a+c} \\ NPV &= Pr(T < \tau | X < \tau) = \frac{d}{b+d} \end{aligned} \tag{C.11}$$

We want to show that: as long as *sens*, *spec*, PPV, and NPV are all defined, i.e.:

$$a + b > 0,\ c + d > 0,\ a + c > 0,\ b + d > 0$$

then:

$$sens + spec > 1 \iff PPV + NPV > 1 \tag{C.12}$$

$$sens + spec < 1 \iff PPV + NPV < 1 \tag{C.13}$$

$$sens + spec = 1 \iff PPV + NPV = 1 \tag{C.14}$$

Proof of C.12.

$$\begin{aligned} & sens + spec > 1 \\ \iff \quad & \frac{a}{a+b} + \frac{d}{c+d} > 1 \\ \iff \quad & a(c+d) + d(a+b) > (a+b)(c+d) \\ \iff \quad & ac + ad + ad + bd > ac + bc + ad + bd \\ \iff \quad & ad > bc \\ \iff \quad & ab + ad + ad + cd > ab + bc + ad + cd \\ \iff \quad & a(b+d) + d(a+c) > (a+c)(b+d) \\ \iff \quad & \frac{a}{a+c} + \frac{d}{b+d} > 1 \\ \iff \quad & PPV + NPV > 1 \end{aligned}$$

□

The proof for C.13 and C.14 can be obtained analogously.

Appendix D

Comparison of meta-analysis based on the Binomial and the Normal model

Throughout the thesis, the meta-analysis model that has been employed is based on the distribution of exposed cases and controls being Binomial in case-control studies. Thus, denoting the number of exposed cases and controls in study s by Y_{s1} and Y_{s0}, and the total number of cases and controls by N_{s1} and N_{s0}, respectively, the model used is:

$$\begin{aligned} Y_{si} &\sim Bin(N_{si}, \pi_{si}) \qquad i = 0, 1 \\ \operatorname{logit} \pi_{s1} &= \operatorname{logit} \pi_{s0} + \theta_s \end{aligned} \tag{D.1}$$

where π_{s1} and π_{s0} are the probabilities of exposure in the case and the control group in study s, and θ_s is the log odds ratio. For the rest of this Appendix, (D.1) is referred to as the Binomial model.

In Chapter 6, I discuss extending this model to account for bias due to measured confounders. Many studies give an estimate of the effect θ_s adjusted for a number of confounders. Nonetheless, the entire dataset with the confounders is generally not available. Without these data, we cannot easily extend model (D.1) to account for bias due to measured confounders. An alternative model that does account for these measured confounders make use of the Normal approximation to the distribution of $\hat{\theta}_s^{\text{adj}}$, the estimate of θ_s adjusted for measured confounders:

$$\hat{\theta}_s^{\text{adj}} \sim N(\theta_s, \hat{\sigma}_s^{\text{adj}^2}) \qquad i = 0, 1 \tag{D.2}$$

Here, $\hat{\sigma}_s^{\text{adj}}$ is the standard error of $\hat{\theta}_s^{\text{adj}}$, which is often available by back-calculating from confidence intervals, and in this Appendix, model (D.2) is referred to as the Normal Approximation model. Because the likelihood for θ_s is approximately Normal in large samples, we may expect this to lead to very similar inference for θ in large samples. In the example data of this thesis, however, many studies do not have

Index	Study	Exposed case (Y_{s1})	Total case (N_{s1})	Exposed controls (Y_{s0})	Total controls (N_{s0})	$\hat{\theta}_s$ ($\hat{\sigma}_s$)
1	Coghill *et al.* (1996)	1	56	0.1	56	2.319 (3.322)
2	Dockerty *et al.* (1998)	3	87	0.1	82	3.376 (3.218)
3	Feychting and Ahlbom (1993)	6	38	22	554	1.512 (0.495)
4	Kabuto *et al.* (2006)	11	312	13	603	0.506 (0.416)
5	Linet *et al.* (1997)	42	638	28	620	0.399 (0.251)
6	London *et al.* (1991)	17	162	10	143	0.444 (0.416)
7	McBride *et al.* (1999)	14	297	11	329	0.358 (0.411)
8	Michaelis *et al.* (1998)	6	176	6	414	0.875 (0.585)
9	Olsen *et al.* (1993)	3	833	3	1666	0.695 (0.818)
10	Savitz *et al.* (1988)	3	36	5	198	1.255 (0.754)
11	Tomenius (1986)	3	153	9	698	0.426 (0.673)
12	Tynes and Haldorsen (1997)	0.1	148	31	2004	-3.146 (3.168)
13	UKCCS (1999)	5	1057	3	1053	0.509 (0.732)
14	Verkasalo *et al.* (1993)	1	32	5	320	0.709 (1.112)

Table D.1: The Greenland (2005a) childhood leukaemia-EMF data with estimated $\hat{\theta}$ and $\hat{\sigma}^2$

large samples. Therefore, my original intention was to investigate whether this causes significant bias to the inference for θ. A surprising result was that bias can result even in large samples. I report this investigation in the section below.

D.1 A small investigation on the bias of using the Normal approximation model

For this investigation, I employ a Bayesian FE model, with θ_s given as:

$$\theta_s = \theta \tag{D.3}$$

$$\theta \sim N(0, 100) \tag{D.4}$$

The data I am using for comparison is given in Table D.1. Table D.1 is equivalent to Table 3.1 except for Y_{s1} in the Coghill *et al.*, Dockerty *et al.*, and the Tynes and Haldorsen studies. This is because in these studies, either Y_{s1} or Y_{s0} is zero, and $\hat{\theta}_s$ is infinity and $\hat{\sigma}_s$ is not calculable from conventional statistics. To overcome this limitation, I replaced 0 with 0.1 for these studies and calculated $\hat{\theta}_s$ and $\hat{\sigma}_s$ from standard large sample maximum likelihood theory.

Here, we found that given the Binomial model (D.1), the posterior median and 95% interval for θ

Model	Posterior Median (95% interval)
Binomial	0.52 (0.46, 0.58)
Normal Approximation	0.59 (0.52, 0.65)

Table D.2: Posterior median and 95% credible intervals for θ in the Binomial model and the Normal Approximation model

is 0.42 (0.14, 0.69), which is almost identical to the results in chapter 3, which does not involve the correction of zero counts. When we use the Normal approximation of (D.2), however, the posterior median and 95% was 0.59 (0.31, 0.87). Thus, it appears that the posterior distribution was shifted to the right by around 0.17.

I suspected that this was because of the inadequacy of the Normal approximation, in turn due to small sample size. To see if this is the case, I increased all numbers in Table D.1 by 20 times. Results are given in Table D.2. Now the two models give more similar results. However, some discrepancy remains (around 0.07), which does not go away even when I increased the amount data by many more times, suggesting the difference in estimation is not entirely due to insufficient sample size.

To further investigate the problem, I identified the part of the data that appears instrumental in causing this discrepancy. For example, let us consider the last three studies of the dataset: The Tynes and Haldorsen (1997) study, the UKCCS (1999) study, and the Verkasalo *et al.* (1993) study. If we carry out a FE meta-analysis of these three studies alone using the Binomial model and the prior distributions of (D.4), the posterior median and 95% credible interval of θ is: -0.31 (-1.37, 0.58). However, when using the Normal Approximation model, the posterior median and 95% credible interval is 0.44 (-0.74, 1.61), considerably different to the results from the Binomial model.

If we increase the amount of data by 1000 times, the posterior median and 95% credible interval of θ in the Binomial model is -0.14 (-0.17, -0.11), whereas in the Normal Approximation model, it is 0.44 (0.40, 0.47). Therefore, although the credible interval is narrowed in both cases, the estimates remain far apart. In particular, the influence of the Tynes and Haldorsen (1997) study is considerably greater in the Binomial model.

It can be shown that this is not due to a Bayesian formulation of the model. Fitting a logistic generalized linear model to these data (1000 times the original), our maximum likelihood estimate of the common θ is -0.14 (-0.17, -0.11), almost identical to the Bayesian results. When the three studies' study-specific θ_s are estimated independently (using maximum likelihood in the generalized linear model), they are:

Study	Y_{s1}	N_{s1}	Y_{s0}	N_{s0}	$\hat{\theta}_s$	$\hat{\sigma}_s^2$
13	100	148000	31000	2004000	-3.15	0.0100
14	5000	1057000	3000	1053000	0.51	0.00053
15	1000	32000	5000	320000	0.71	0.00124

The inverse-variance weighted average $\hat{\theta}_s$ is 0.44, which equals the Normal Approximation model results. However, a full likelihood-based calculation does not employ inverse-variance weights. Therefore, we see that the discrepancy between the Binomial model and the Normal Approximation model results is not due to the Bayesian formulation, and remains even in large samples.

An intuitive explanation of the above results is that it is *not* sufficient for large samples to ensure that the likelihood of θ is accurately approximated by the Normal approximation model (contrast, e.g. comments from Deeks *et al.*, 2001, p.303). The accurate approximation instead requires the study-specific likelihood of θ be approximately Normal *in the region of combination of the study-specific likelihoods*, and not just in the region of the maximum likelihood estimate $\hat{\theta}_s$, as has been pointed out by O'Rourke (2007, p.9). This may not necessarily be guaranteed by large samples.

In conclusion, it is demonstrated in this Appendix that the use of the Normal Approximation model instead of the "true" Binomial model can introduce bias, in meta-analyses with small study sizes as well as large sizes. It may be the case that the proposal as given in chapter 6 of this thesis may to some extent reverse the bias, although this awaits further investigation.

References

Ades A (2003). A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Statistics in Medicine*, 22(19), 2995–3016

Ades A and Sutton A (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 169, 5–35

Ahlbom A, Day N, Feychting M, Roman E, Skinner J, Dockerty J, Linet M, McBride M, Michaelis J, Olsen J, Tynes T, and Verkasalo P (2000). A pooled analysis of magnetic fields and childhood leukaemia. *Br J Cancer*, 83(5), 692–698

Al khalaf M, Thalib L, and Doi S (2011). Combining heterogenous studies using the random-effects model is a mistake and leads to inconclusive meta-analyses. *Journal of Clinical Epidemiology*, 64(2), 119–123

Angrist J, Imbens G, and Rubin D (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444–455

Arah O, Chiba Y, and Greenland S (2008). Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Annals of Epidemiology*, 18(8), 637–646

Armstrong B, Whittemore A, and Howe G (1989). Analysis of case-control data with covariate measurement error – application to diet and colon cancer. *Statistics in Medicine*, 8(9), 1151–1163

Ayyub B and Klir G (2006). *Uncertainty Modeling and Analysis in Engineering and the Sciences*. Chapman and Hall/CRC, Boca Raton, FL

Balke A and Pearl J (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439), 1171–1176

Bayarri M and Berger J (2004). The interplay of frequentist and bayesian statistics. *Statistical Science*, 19(1), 58–80

Berger J (1984). The robust Bayesian viewpoint (with discussion). In: J Kadane (ed.) *Robustness of Bayesian Analysis*. North-Holland, Amsterdam, pp. 63–144

Berger J (1985). *Statistical Decision Theory and Bayesian Analysis*. Second edition. Springer-Verlag, New York

Berger J (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25, 303–328

Berger J (2006). The case for objective bayesian analysis. *Bayesian Analysis*, 1(3), 385–402

Birnbaum Z and Sirken M (1950). Bias due to non-availability in sampling surveys. *Journal of the American Statistical Association*, 45(249), 98–111

Borenstein M, Hedges L, Higgins J, and Rothstein H (2009). *Introduction to meta-analysis*. Wiley, Chichester

Bross I (1954). Misclassification in 2 by 2 tables. *Biometrics*, 10, 478–486

Carroll R, Ruppert D, and Stefanski L (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, London

Chiba Y, Sato T, and Greenland S (2007). Bounds on potential risks and causal risk differences under assumptions about confounding parameters. *Statistics in Medicine*, 26(28), 5125–5135

Chu H, Wang Z, Cole S, and Greenland S (2006). Sensitivity analysis of misclassification: a graphical and a bayesian approach. *Ann Epidemiol*, 16(11), 834–841

Chu R, Gustafson P, and Le N (2010). Bayesian adjustment for exposure misclassification in case-control studies. *Stat Med*, 29(9), 994–1003

Cochran W and Chambers S (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society Series A (General)*, 128(2), 234–266

Coghill R, Steward J, and Philips A (1996). Extra low frequency electric and magnetic fields in the bedplace of children diagnosed with leukaemia: a case-control study. *Eur J Cancer Prev*, 5(3), 153–158

Congdon P (2006). *Bayesian Statistical Modelling*. Second edition. John Wiley and Sons, Chichester

Copas J and Li H (1997). Inference for non-random samples. *Journal of the Royal Statistical Society Series B-Methodological*, 59(1), 55–77

Cordray D and Morphy P (2009). *Research Synthesis and Public Policy*. Second edition. Russell Sage Foundation, New York

Cornfield J, Haenszel W, Hammond E, Lilienfeld A, Shimkin M, and Wynder E (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst*, 22(1), 173–203

De Finetti B (1974). *Theory of Probability*. Wiley, New York

Deeks J, Altman D, and Bradburn M (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: M Egger, G Davey Smith, and D Altman (eds.) *Systematic Reviews in Health Care: Meta-analysis in context*. BMJ Publishing Group, London, pp. 285–312

Dersimonian R and Laird N (1986). Metaanalysis in clinical-trials. *Controlled Clinical Trials*, 7(3), 177–188

Diamond E and Lilienfeld A (1962). Effects of errors in classification and diagnosis in various types of epidemiological studies. *Am J Public Health Nations Health*, 52:1137-44., 1137–1144

Dockerty J, Elwood J, Skegg D, and Herbison G (1998). Electromagnetic field exposures and childhood cancers in New Zealand. *Cancer Causes Control*, 9(3), 299–309

Drews C and Greenland S (1990). The impact of differential recall on the results of case-control studies. *International Journal of Epidemiology*, 19(4), 1107–1112

DuMouchel W (1996). Predictive cross-validation of Bayesian meta-analyses. In: J Bernardo, J Berger, A Dawid, and A Smith (eds.) *Bayesian Statistics 5*. Oxford University Press, New York, pp. 63–144

Egger M, Schneider M, and Smith G (1998). Meta-analysis – spurious precision? Meta-analysis of observational studies. *British Medical Journal*, 316(7125), 140–144

Espeland M and Hui S (1987). A general-approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics*, 43(4), 1001–1012

Eysenck H (1978). Exercise in mega-silliness. *American Psychologist*, 33(5), 517–517

Ferson S, Joslyn C, Helton J, Oberkampf W, and Sentz K (2004). Summary from the epistemic uncertainty workshop: consensus amid diversity. *Reliability Engineering & System Safety*, 85(1-3), 355–369

Feychting M and Ahlbom A (1993). Magnetic fields and cancer in children residing near Swedish high-voltage power lines. *Am J Epidemiol*, 138(7), 467–481

Fishburn P (1986). The axioms of subjective probability. *Statistical Science*, 1(3), 335–345

Flanders W and Khoury M (1990). Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology*, 1(3), 239–246

Fox M, Lash T, and Greenland S (2005). A method to automate probabilistic sensitivity analyses of misclassified binary variables. *International Journal of Epidemiology*, 34(6), 1370–1376

Garthwaite P, Kadane J, and O'Hagan A (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470), 680–700

Gastwirth J (1987). The statistical precision of medical screening procedures: Application to polygraph and aids antibodies test data. *Statistical Science*, 2(3), 213–238

Gelman A, Carlin J, Stern H, and Rubin D (2004). *Bayesian Data Analysis*. 2nd edition. Chapman and Hall/CRC, Boca Raton, FL

Geneletti S, Richardson S, and Best N (2009). Adjusting for selection bias in retrospective, case-control studies. *Biostatistics*, 10(1), 17–31

Good I (1950). *Probability and the Weighing of Evidence*. Charles Griffin, London

Good I (1962). Subjective probability as the measure of a non-measurable set. In: *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford University Press, Stanford, pp. 319–329. Reprinted in I.J. Good (1983) Good Thinking: The Foundations of Probability and its Applicaionts. Minneapolis: University of Minnesota Press

Good I (1980). Some history of the hierarchical Bayesian methodology. In: J Bernardo, M DeGroot, D Lindley, and A Smith (eds.) *Bayesian Statistics 1*. University of Valencia Press, Valencia, Spain, pp. 489–519

Good I (1988). The interface between statistics and philosophy of science. *Statistical Science*, 3(4), 386–397

Greenhalgh T (1997). How to read a paper: Getting your bearings (deciding what the paper is about). *British Medical Journal*, 315(7102), 243–246

Greenland S (1989). On correcting for misclassification in twin studies and other matched-pair studies. *Statistics in Medicine*, 8(7), 825–829

Greenland S (1990). Randomization, statistics, and causal inference. *Epidemiology*, 1(6), 421–429

Greenland S (1994). Invited commentary: a critical look at some popular meta-analytic methods. *American Journal of Epidemiology*, 140(3), 290–296

Greenland S (1998). Probability logic and probabilistic induction. *Epidemiology*, 9(3), 322–332

Greenland S (2001). Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Analysis*, 21(4), 579–583

Greenland S (2005a). Multiple-bias modeling for analysis of observational data. *Journal of the Royal Statistical Society Series A*, 168, 267–306

Greenland S (2005b). Multiple-bias modelling for analysis of observational data – rejoinder. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 168, 302–306

Greenland S (2008). Maximum-likelihood and closed-form estimators of epidemiologic measures under misclassification. *Journal of Statistical Planning and Inference*, 138(2), 528–538

Greenland S (2009a). Bayesian perspectives for epidemiologic research III. Bias analysis via missing-data methods. *International Journal of Epidemiology*, 38(6), 1662–1673

Greenland S (2009b). Relaxation penalties and priors for plausible modeling of nonidentified bias sources. *Statistical Science*, 24(2), 195–210

Greenland S and Kheifets L (2006). Leukemia attributable to residential magnetic fields: Results from analyses allowing for study biases. *Risk Analysis*, 26(2), 471–482

Greenland S and Lash T (2008). Bias analysis. In: K Rothman, S Greenland, and T Lash (eds.) *Modern Epidemiology*, third edition. Lippincott Williams & Wilkins, Philadelphia, PA, pp. 345–380

Greenland S and Maldonado G (1994). The interpretation of multiplicative-model parameters as standardized parameters. *Stat Med*, 13(10), 989–999

Greenland S and Robins J (1985). Confounding and misclassification. *American Journal of Epidemiology*, 122(3), 495–506

Greenland S and Robins J (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, 15(3), 413–419

Greenland S, Sheppard A, Kaune W, Poole C, and Kelsh M (2000). A pooled analysis of magnetic fields, wire codes, and childhood leukemia. *Epidemiology*, 11(6), 624–634

Groenwold R, Nelson D, Nichol K, Hoes A, and Hak E (2010). Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *International Journal of Epidemiology*, 39(1), 107–117

Gustafson P (2005). On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Statistical Science*, 20(2), 111–129

Gustafson P, Hossain S, and MacNab Y (2006). Conservative prior distributions for variance parameters in hierarchical models. *Canadian Journal of Statistics-Revue Canadienne de Statistique*, 34(3), 377–390

Gustafson P, Le N, and Saskin R (2001). Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics*, 57(2), 598–609

Gustafson P, McCandless L, Levy A, and Richardson S (2010). Simplified Bayesian sensitivity analysis for mismeasured and unobserved confounders. *Biometrics*, 66(4), 1129–1137

Hansen M and Hurwitz W (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41(236), 517–529

Higgins J, Thompson S, and Spiegelhalter D (2009). A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*, 172(1), 137–159

Hill A (1965). The environment and disease: Association or causation? *Proc R Soc Med*, 58, 295–300

Hui S and Walter S (1980). Estimating the error rates of diagnostic-tests. *Biometrics*, 36(1), 167–171

International Agency for Research on Cancer (IARC) (2002). *Non-ionizing radiation, Part 1: Static and extremely-low-frequency (ELF) electric and magnetic fields*. Monographs of the Evaluation of Carcinogenic Risks to Humans, Volume 80, International Agency for Research on Cancer, Lyon, France

International Commission on Non-Ionizing Radiation Protection (ICNIRP) (2001). Review of the epidemiologic literature on EMF and health. *Environ Health Perspect*, 109 Suppl 6, 911–933

Iyengar S and Greenhouse J (1988). Selection models and the file drawer problem. *Statistical Science*, 3(1), 109–117

Kabuto M, Nitta H, Yamamoto S, Yamaguchi N, Akiba S, Honda Y, Hagihara J, Isaka K, Saito T, Ojima T, Nakamura Y, Mizoue T, Ito S, Eboshida A, Yamazaki S, Sokejima S, Kurokawa Y, and Kubo O (2006). Childhood leukemia and magnetic fields in Japan: a case-control study of childhood leukemia and residential power-frequency magnetic fields in Japan. *Int J Cancer*, 119(3), 643–650

Koopman B (1940). The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41, 269–292

Krommer A and Ueberhuber C (1998). *Computational Integration*. Society for Industrial and Applied Mathematics, Philadelphia, PA

Kuchenhoff H, Mwalili S, and Lesaffre E (2006). A general method for dealing with misclassification in regression: The misclassification simex. *Biometrics*, 62(1), 85–96

Kuroki M, Cai Z, and Geng Z (2010). Sharp bounds on causal effects in case-control and cohort studies. *Biometrika*, 97(1), 123–132

Lash T, Fox M, and Fink A (2009). *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer, New York, NY

Leonard T and Hsu J (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge University Press, Cambridge

Lin D, Psaty B, and Kronmal R (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, 54(3), 948–963

Linet M, Hatch E, Kleinerman R, Robison L, Kaune W, Friedman D, Severson R, Haines C, Hartsock C, Niwa S, Wacholder S, and Tarone R (1997). Residential exposure to magnetic fields and acute lymphoblastic leukemia in children. *N Engl J Med*, 337(1), 1–7

Little R (1982). Models for nonresponse in sample-surveys. *Journal of the American Statistical Association*, 77(378), 237–250

Little R (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421), 125–134

London S, Thomas D, Bowman J, Sobel E, Cheng T, and Peters J (1991). Exposure to residential electric and magnetic fields and risk of childhood leukemia. *Am J Epidemiol*, 134(9), 923–937

Lunn D, Thomas A, Best N, and Spiegelhalter D (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337

MacLehose R, Kaufman S, Kaufman J, and Poole C (2005). Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology*, 16(4), 548–555

Mak T (2008). *Estimating misclassification probabilities in a meta-analysis of studies looking at the relationship between childhood leukemia and magnetic fields*. Master's thesis, Imperial College London

Mardia K (1970). *Families of Bivariate Distributions*. Griffin, London

McBride M, Gallagher R, Theriault G, Armstrong B, Tamaro S, Spinelli J, Deadman J, Fincham S, Robson D, and Choi W (1999). Power-frequency electric and magnetic fields and risk of childhood leukemia in Canada. *Am J Epidemiol*, 149(9), 831–842

McCandless L, Gustafson P, and Austin P (2009). Bayesian propensity score analysis for observational data. *Statistics in Medicine*, 28(1), 94–112

McCandless L, Gustafson P, and Levy A (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine*, 26(11), 2331–2347

Michaelis J, Schuz J, Meinert R, Zemann E, Grigat J, Kaatsch P, Kaletsch U, Miesner A, Brinkmann K, Kalkner W, and Karner H (1998). Combined risk estimates for two German population-based case-control studies on residential magnetic fields and childhood acute leukemia. *Epidemiology*, 9(1), 92–94

Molitor N, Best N, Jackson C, and Richardson S (2009). Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth weight and water disinfection by-products. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 172, 615–637

Morrissey M and Spiegelman D (1999). Matrix methods for estimating odds ratios with misclassified exposure data: Extensions and comparisons. *Biometrics*, 55(2), 338–344

Nocedal J and Wright S (1999). *Numerical Optimization*. Springer-Verlag, New York

O'Hagan A and Oakley J (2004). Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering & System Safety*, 85(1-3), 239–248

Olsen J, Nielsen A, and Schulgen G (1993). Residence near high voltage facilities and risk of cancer in children. *BMJ*, 307(6909), 891–895

O'Rourke K (2007). *The combining of information: Investigating and synthesizing what is possibly common in clinical observations or studies via likelihood*. Ph.D. thesis, University of Oxford

Paulino C, Soares P, and Neuhaus J (2003). Binomial regression with misclassification. *Biometrics*, 59(3), 670–675

Petitti D (2000). *Meta-analysis, decision analysis, and cost-effectiveness analysis : methods for quantitative synthesis in medicine*. 2nd edition. Oxford University Press, New York

Plackett R (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60(310), 516–522

Politz A and Simmons W (1949). An attempt to get the 'not at homes' into the sample without callbacks. *Journal of the American Statistical Association*, 44(245), 9–16

Prekopa A (1995). *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands

Press W, Teukolsky S, Vetterling W, and Flannery B (2007). *Numerical Recipes – the art of scientific computing*. 3rd edition. Cambridge University Press, New York, NY

Rice K (2005). Multiple-bias modelling for analysis of observational data – discussion. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 168, 294

Robert C (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Second edition. Springer, New York, NY

Robins J (1988). Confidence-intervals for causal parameters. *Statistics in Medicine*, 7(7), 773–785

Rosen J (1960). The gradient projection method for nonlinear programming. Part I. Linear constraints. *Journal of the Society for Industrial and Applied Mathematics*, 8(1), 181–217

Rosenbaum P (2005). Sensitivity analysis in observational studies. In: B Everitt and D Howell (eds.) *Encyclopedia of Statistics in Behavioural Science*. John Wiley and Sons, Chichester, pp. 1809–1814

Rosenbaum P and Rubin D (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society Series B-Methodological*, 45(2), 212–218

Rosenbaum P and Rubin D (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55

Rosner B, Willett W, and Spiegelman D (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med*, 8(9), 1051–1069

Rubin D (1976). Inference and missing data. *Biometrika*, 63(3), 581–590

Rubin D (1977). Formalizing subjective notions about effect of non-respondents in sample-surveys. *Journal of the American Statistical Association*, 72(359), 538–543

Rue H, Martino S, and Chopin N (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 71(Part 2), 319–392

Sahinidis N (2004). Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering*, 28(6-7), 971–983

Savitz D, Wachtel H, Barnes F, John E, and Tvrdik J (1988). Case-control study of childhood cancer and exposure to 60-Hz magnetic fields. *Am J Epidemiol*, 128(1), 21–38

Selen J (1986). Adjusting for errors in classification and measurement in the analysis of partly and purely categorical-data. *Journal of the American Statistical Association*, 81(393), 75–81

Shapiro S (1994). Meta-analysis/shmeta-analysis. *American Journal of Epidemiology*, 140(9), 771–778

Smith T, Spiegelhalter D, and Thomas A (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine*, 14(24), 2685–2699

Stamey J, Young D, and Seaman J (2008). A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. *Stat Med*, 27(13), 2440–2452

StataCorp (2007). Stata statistical software: Release 10

Stijnen T, Hamza T, and Ozdemir P (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*, 29(29), 3046–3067

Stroup D and Thacker S (2005). Meta-analysis in epidemiology. In: P Armitage and T Colton (eds.) *Encyclopedia of Biostatistics*, 2nd edition. John Wiley and sons, Chichester, pp. 3448–3462

Sturmer T, Schneeweiss S, Avorn J, and Glynn R (2005). Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *American Journal of Epidemiology*, 162(3), 279–289

Sutton A and Abrams K (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, 10(4), 277–303

Tenenbein A (1970). A double sampling scheme for estimating from binomial data with misclassification. *Journal of the American Statistical Association*, 65(331), 1350–1361

Thompson S, Ekelund U, Jebb S, Lindroos A, Mander A, Sharp S, Turner R, and Wilks D (2011). A proposed method of bias adjustment for meta-analyses of published observational studies. *International Journal of Epidemiology*, 40(3), 765–777

Tian J and Pearl J (2000). Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4), 287–313

Tierney L and Kadane J (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393), 82–86

Tomenius L (1986). 50-Hz electromagnetic environment and the incidence of childhood tumors in stockholm county. *Bioelectromagnetics*, 7(2), 191–207

Turner R, Spiegelhalter D, Smith G, and Thompson S (2009). Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 172, 21–47

Tynes T and Haldorsen T (1997). Electromagnetic fields and cancer in children residing near Norwegian high-voltage power lines. *Am J Epidemiol*, 145(3), 219–226

UK Childhood Cancer Study Investigators (UKCCS) (1999). Exposure to power-frequency magnetic fields and the risk of childhood cancer. *Lancet*, 354, 1925–1931

VanderWeele T (2008). The sign of the bias of unmeasured confounding. *Biometrics*, 64(3), 702–706

Vansteelandt S, Goetghebeur E, Kenward M, and Molenberghs G (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16(3), 953–979

Vassiliadis V and Conejeros R (2009). Cyclic coordinate method. In: C Floudas and P Pardalos (eds.) *Encyclopedia of Optimization*, second edition. Springer, pp. 595–596

Verkasalo P, Pukkala E, Hongisto M, Valjus J, Jarvinen P, Heikkila K, and Koskenvuo M (1993). Risk of cancer in Finnish children living close to power lines. *BMJ*, 307(6909), 895–899

Welton N, Ades A, Carlin J, Altman D, and Sterne J (2009). Models for potentially biased evidence in meta-analysis using empirically based priors. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 172, 119–136

Wolpert R and Mengersen K (2004). Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: Effects of environmental tobacco smoke. *Statistical Science*, 19(3), 450–471

World Health Organization (WHO) (2007). *Extremely Low Frequency Fields*. Environmental Health Criteria Monograph No.238, WHO Press, Geneva, Switzerland

Zadeh L (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3–28