Adjusted Likelihoods For Synthesizing Empirical Evidence from Studies That Differ in Quality and Design: Effects of Environmental Tobacco Smoke

Robert L. Wolpert* Kerrie L. Mengersen†

April 19, 2001

Abstract

Methods are introduced and illustrated for synthesizing evidence from case-control and cohort studies and controlled trials, accounting for differences among the studies in their design, length of follow-up, and quality. The methods, based on hierarchical but nonexchangeable Bayesian models, are illustrated in a synthesis of disparate information about the health effects of passive exposure to environmental tobacco smoke.

Key words: Adjustment, Meta-analysis, nonexchangeable Bayesian hierarchical model, quality of studies, environmental tobacco smoke.

*Robert L. Wolpert is Associate Professor, Institute of Statistics and Decision Sciences, and Associate Professor, School of the Environment, at Duke University, Durham, NC, 27708-0251.

†Kerrie L. Mengersen is Senior Lecturer in the School of Mathematical Sciences, Queensland University of Technology, Gardens Point, Brisbane 4001, Australia.

1 Introduction

The hazards of combining evidence naïvely from multiple sources are well known. Even among sources intended to be homogeneous, such as multi-center controlled medical trials, there will be some variation in patient attributes and treatment regimens. This variation is strongly exacerbated in epidemiological studies because of differing study designs and methods and population characteristics: designs may vary according to their retrospective or prospective nature, their case-control, cohort, or cross-sectional structure, length of followup, admission and classification criteria, control of biases, consideration of confounders, and data collection procedures. Study populations will vary in their age structure, diet and other lifestyle factors, genetic makeup, and environmental exposure to competing and contributing risks. Investigators also vary with respect to experience, motivation and skill. All of these issues will threaten the validity of methods which simply pool the data and will reduce the power of methods which are insensitive to the probability distributions governing the observed outcomes in the individual studies.

Beginning with Fisher's (1934) early attempt to avoid the problem of heterogeneity in field trials by combining individual p-values to assess the overall significance level, a number of authors have introduced novel statistical methods intended to quantify an overall "effect size" and assess its variation. These methods, referred to collectively as *meta-analysis*, are described and compared by many authors including Hedges and Olkin (1985) and the National Research Council review (1992).

More recently Bayesian and Bayesian-inspired methods have been brought to bear on the meta-analysis problem. The landmark DuMouchel and Harris (1983) introduced the idea of constructing hierarchical Bayesian models to synthesise information from five types of environmental studies of the effect on human and animal subjects of exposure to nine related environmental agents; Dempster et al. (1983) synthesised historical and contemporary clinical evidence; and Berry (1990a), Carlin (1992),DuMouchel (1990, 1996), Morris and Normand (1992), Smith et al. (1996), and Spiegelhalter et al. (1995, in the BUGS software examples), offered broad guides to the use of Bayesian hierarchical models. All of these approaches make at least some accommodation for study heterogeneity, commonly through inclusion of a single extra variance term in the hierarchical model structure, but each treats all the studies as exchangeable (at least within groups).

The idea of "adjusting" likelihood functions to accommodate study-specific biases arising from variations in patient attributes, treatment protocol, and provider skill and experience was introduced by Eddy (1989) through replacement of the actual likelihood function with subjective estimates of "what would have been observed" had the biases not been present. More recently explicit sources of bias, misclassification and heterogeneity have been reflected in non-exchangeable Bayesian hierarchical models (Wolpert and Warren-Hicks, 1992), conditional independence models (Richardson and Gilks, 1993), and augmented likelihoods (Givens et al., 1995). In the present paper we build on these ideas by constructing a hierarchical Bayesian model with submodels for each study or other source of evidence, that reflect their study-specific differences and hence are not exchangeable. Under some simplifying conditional-independence assumptions the model's composite likelihood function reduces to the simple product of "adjusted" likelihoods from the individual studies. The method differs from earlier efforts by introducing the necessary adjustments through the choice of a suitable joint prior probability distribution for the unknown quantitities of interest and for the parameters governing the experimental observables' probability distributions, rather than through alteration of the data. This is consistent with the Likelihood Principle (Berger and Wolpert, 1988) and remains within the Bayesian paradigm.

The method is illustrated through application to the problem of sythesizing evidence from twenty-nine case-control and cohort studies of the relative risk of lung cancer associated with exposure to environmental tobacco smoke (ETS), as detailed in the U.S. Environmental Protection Agency report (EPA Review, 1992). A number of meta-analyses of these data have been conducted using frequentist fixed and random effects models (Lee, 1993, 1994; LeVois and Layard, 1994; Mengersen et al., 1995; NRC Committee on Passive Smoking, 1986; EPA Review, 1992; Tweedie and Mengersen, 1992; Wald et al., 1986) and through Bayesian hierarchical models (Besag et al., 1997; Biggerstaff et al., 1994; Tweedie et al., 1995). The issue still under active debate, however, is the impact of study quality: while it is widely acknowledged that these epidemiological studies differ substantially in their efforts to control the problems of bias and misclassification, only broadbrush (if any) account has been taken of this issue in previous meta-analyses.

1.1 Evidence from Individual Studies

Epidemiological studies are usually intended to ascertain if exposed and unexposed subjects from a specified population differ with respect to an outcome of interest (that is, become cases or noncases), or if cases and non-cases within this population differ with respect to their exposure status, and to quantify the possible differences. In general the exposure may be intentional or unintentional, active or passive, controlled or uncontrolled by the experimenter. In the ETS example of Section 3 the study population is taken to be women who are not current or former smokers ("never-smokers"); "exposure" is being married to a cigarette smoker; and "cases" are subjects who contract or die from lung cancer.

Epidemiological studies offer direct evidence about the conditional probabilities of outcome status (for cohort studies) or exposure (for case-control studies) for the study populations and, through these, indirect evidence about the quantity of interest: usually, some measure of the association between exposure and outcome. Evidence is often available about the size and impact of varying age distributions, misclassification rates, and other confounding features and possible biases.

1.1.1 Cohort Studies

In a cohort study (CHS) subjects are drawn from a specified population and classified as exposed or unexposed (Breslow and Day, 1987). They are then followed for a specified length of time and classified with respect to their disease status. Evidence bears directly on the conditional probabilities of becoming a case, $p_{c|e}$ and $p_{c|\bar{e}}$, in the exposed and unexposed study arms respectively, through the likelihood function (Berger and Wolpert, 1988); in the simple case where age and other covariates are not considered this is

$$L_{\mathsf{CHS}}(p_{c|e}, p_{c|\bar{e}}) = (p_{c|e})^{n_{ce}}\,(1-p_{c|e})^{n_{\bar{c}e}}\,(p_{c|\bar{e}})^{n_{c\bar{e}}}\,(1-p_{c|\bar{e}})^{n_{\bar{c}\bar{e}}}. \qquad (1)$$

Occasionally these probabilities may themselves be useful but more often interest lies in some measure of their difference quantifying the association between outcome and exposure, such as:

$$\begin{array}{ll} \text{Simple Difference} & \varepsilon_{\mathsf{SD}} \;= p_{c|e} - p_{c|\bar{e}} \\ \text{Log Relative Risk} & \varepsilon_{\mathsf{LRR}} = \log(p_{c|e}/p_{c|\bar{e}}) \\ \text{Log Odds Ratio} & \varepsilon_{\mathsf{LOR}} = \log(p_{c|e}\,p_{\bar{c}|\bar{e}}/p_{\bar{c}|e}\,p_{c|\bar{e}}) \end{array}$$

Each of these measures will be positive if greater exposure is associated with higher levels of disease, zero if they are unrelated, and each can be estimated in a cohort study by its maximum likelihood estimator (MLE) such as $\hat{\varepsilon}_{\mathsf{LOR}} = \log(n_{ce}\,n_{\bar{c}\bar{e}}/n_{c\bar{e}}\,n_{\bar{c}e})$. They and other measures have been discussed, illustrated and compared by many authors including Cox (1970, *p.* 20) and Freeman (1987, *pp.* 66,95).

A variety of methods (reviewed and extended by Berger et al., 1997) have been proposed for extracting information about a quantity of interest like

ε amid nuisance parameters; the Bayesian approach is to select a (possibly noninformative) conditional prior distribution $\pi(d\theta|\varepsilon)$ and summarize the evidence about ε alone by $L_{\mathsf{CHS}}(\varepsilon) = \int L_{\mathsf{CHS}}(\theta)\, \pi(d\theta|\varepsilon)$. A change of variables to the log odds ratio $\varepsilon = \varepsilon_{\mathsf{LOR}}$ and independent Jeffreys' beta Be(1/2, 1/2) reference priors for both case probabilities $p_{c|e}$ and $p_{c|\bar{e}}$ leads to likelihood

$$\begin{aligned} L_{\mathsf{CHS}}(\varepsilon) &\propto \varepsilon^{-1}\,(e^{\varepsilon}-1)\,e^{n_{ce}\,\varepsilon} \int_0^1 \frac{p^{n_c}\,(1-p)^{n_{\bar{c}}}}{[1-p(1-e^{\varepsilon})]^{n_e+1}}\,dp \\ &\propto \varepsilon^{-1}\,(e^{\varepsilon}-1)\,e^{n_{ce}\,\varepsilon}\, {}_2F_1(n_e+1,\, n_c+1;\, n+2;\, 1-e^{\varepsilon}), \end{aligned} \tag{2}$$

where ${}_2F_1(a,b;c;z)$ is the confluent hypergeometric function (Abramowitz and Stegun, 1964, $p.\,558$). With the implied prior $\pi(\varepsilon) = \varepsilon/2\pi^2 \sinh(\varepsilon/2)$, the log odds ratio $\varepsilon_{\mathsf{LOR}}$ has posterior mean and variance

$$\begin{aligned} \mu &= \psi(n_{ce}+1/2) - \psi(n_{\bar{c}e}+1/2) - \psi(n_{c\bar{e}}+1/2) + \psi(n_{\bar{c}\bar{e}}+1/2) \\ &\approx \log(n_{ce}\, n_{\bar{c}\bar{e}} \,/\, n_{\bar{c}e}\, n_{c\bar{e}}) \qquad (3) \\ \sigma^2 &= \psi'(n_{ce}+1/2) + \psi'(n_{\bar{c}e}+1/2) + \psi'(n_{c\bar{e}}+1/2) + \psi'(n_{\bar{c}\bar{e}}+1/2) \\ &\approx 1/n_{ce} + 1/n_{\bar{c}e} + 1/n_{c\bar{e}} + 1/n_{\bar{c}\bar{e}} \qquad (4) \end{aligned}$$

where $\psi(z+1/2) = \log z + \mathcal{O}(z^{-2})$ and $\psi'(z+1/2) = 1/z + \mathcal{O}(z^{-3})$ are the digamma and trigamma functions (Abramowitz and Stegun, 1964, $p.\,158$–260).

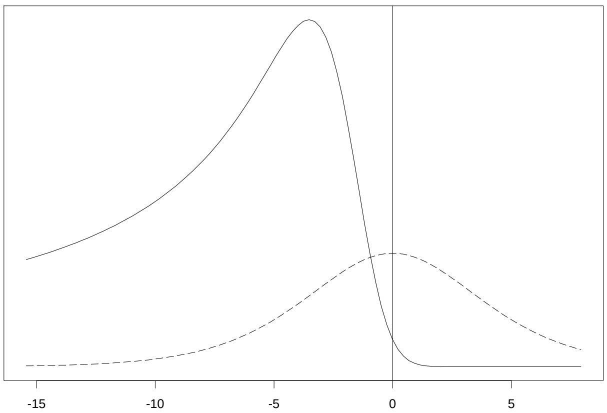

Figure 1: Marginal likelihood (solid) and prior density (dashed) for log odds ratio of survival in 1985 ECMO trial.

For example, Figure 1 shows $L_{\mathsf{CHS}}(\varepsilon)$ for the famous 1985 extracorporeal membrane oxygenation (ECMO) trial (see Ware, 1989), in which 6 of 10 subjects survived in the control group and 9 of 9 survived who were "exposed" to ECMO. The solid line is $L_{\mathsf{CHS}}(\varepsilon)$ from (2), the dashed line is the marginal prior density function $\pi(\varepsilon)$, and the vertical bar at $\varepsilon_{\mathsf{LOR}} = 0$ marks the null effect; the posterior probability of greater risk for ECMO is $\mathrm{P}[\varepsilon_{\mathsf{LOR}} > 0] \approx 0.0107$ with this prior (see Kass and Greenhouse (1989) and Lavine et al. (1991) for discussions of prior distributions for this problem).

For less extreme contingency tables, including all of those in the ETS study considered below, the distribution is very well approximated by a normal with the same mean and variance or with the approximate means and variances given in (3–4) above. Similar expressions are available for the other measures of exposure effect in cohort studies.

1.1.2 Case Control Studies

In a case control study (CCS) some number n_c of cases are matched with $n_{\bar{c}}$ non-cases (controls) on the basis of demographic variables and other covariates (Breslow and Day, 1980). Within the case and control groups subjects are further classified into those who are exposed (n_{ce}, $n_{\bar{c}e}$) and unexposed ($n_{c\bar{e}}$, $n_{\bar{c}\bar{e}}$).

This evidence bears directly on the conditional probabilities of exposure within case and non-case groups, $p_{e|c}$ and $p_{e|\bar{c}}$ through the likelihood function

$$L_{\mathsf{CCS}}(p_{e|c}, p_{e|\bar{c}}) = (p_{e|c})^{n_{ce}}\,(1 - p_{e|c})^{n_{ce}}\,(p_{e|\bar{c}})^{n_{\bar{c}e}}\,(1 - p_{e|\bar{c}})^{n_{\bar{c}\bar{e}}}, \tag{5}$$

but only indirectly on the exposure-specific conditional outcome probabilities $p_{c|e}$ and $p_{c|\bar{e}}$: if the outcome rate p_c and exposure rate p_e in the target population are known independently, for example, then $p_{c|e} = p_{e|c}\,p_c/p_e$, $p_{c|\bar{e}} = p_{\bar{e}|c}\,p_c/(1 - p_e)$, and so

$$\begin{array}{ll} \text{Simple Difference} & \varepsilon_{\mathsf{SD}} = p_{e|c}\,p_c/p_e - p_{\bar{e}|c}\,p_c/(1 - p_e) \\ \text{Log Relative Risk} & \varepsilon_{\mathsf{LRR}} = \log\left(p_{e|c}\,p_c\,(1 - p_e)/p_{\bar{e}|c}\,p_c\,p_e\right) \\ \text{Log Odds Ratio} & \varepsilon_{\mathsf{LOR}} = \log\left(p_{e|c}\,p_{\bar{e}|\bar{c}}/p_{\bar{e}|c}\,p_{e|\bar{c}}\right). \end{array}$$

Only one of p_c and p_e is actually required, since each can be computed from the other ($p_e = p_{e|c}\,p_c + p_{e|\bar{c}}\,(1 - p_c)$, $p_c = (p_e - p_{e|\bar{c}})/(p_{e|c} - p_{e|\bar{c}})$). Unfortunately it is rare for either p_c or p_e to be available at the study-specific level required. The fact that $\varepsilon_{\mathsf{LOR}}$ does not depend on these collateral quantities (and in particular that its MLE $\hat{\varepsilon}_{\mathsf{LOR}} = \log(n_{ce}\,n_{\bar{c}\bar{e}}/n_{c\bar{e}}\,n_{\bar{c}e})$ exists) contributes to its popularity in case-control studies, and its appeal in the present analysis. Inference can be made about any of the exposure effect measures with a suitable conditional prior distribution through the relation $L_{\mathsf{CCS}}(\varepsilon) = \int L_{\mathsf{CCS}}(\theta)\, \pi(d\theta|\varepsilon)$. Independent Jeffreys' Be(1/2, 1/2) distributions for the exposure probabilities, for example, lead again to

$$L_{\mathsf{CCS}}(\varepsilon) \;\; \propto \;\; \varepsilon^{-1}\, (e^{\varepsilon} - 1)\, e^{n_{ce}\,\varepsilon}\, {}_2F_1(n_c + 1,\, n_e + 1;\, n + 2;\, 1 - e^{\varepsilon}), \qquad (6)$$

exactly the same as (2) and again approximately normal with mean $\mu \approx \log(n_{ce}\, n_{\bar{c}\bar{e}} \,/\, n_{\bar{c}e}\, n_{c\bar{e}})$ and variance $\sigma^2 \approx 1/n_{ce}+1/n_{c\bar{e}}+1/n_{\bar{c}e}+1/n_{\bar{c}\bar{e}}$. Similar expressions are not available for the other measures of exposure effect in case control studies without information about p_c or p_e in the study population.

1.2 Exchangeable Combination of Evidence

Both cohort studies and case control studies give evidence about the same vector $\theta = (p_{ce}, p_{c\bar{e}}, p_{\bar{c}e}, p_{\bar{c}\bar{e}})$ of probabilities that a subject drawn randomly from the target population will be an exposed or unexposed case, exposed or unexposed non-case, respectively, through their respective likelihood functions $L_{\mathsf{CHS}}(\theta)$ and $L_{\mathsf{CCS}}(\theta)$, so it is convenient to express all studies' evidence in terms of θ before trying to combine them across study types.

1.2.1 Simple Pooling

The evidence about a common parameter vector θ obtained by observing a set I of *independent* studies is captured by the joint likelihood function $L^I(\theta) = \prod_{i\in I} L_i(\theta)$; for case control studies with exactly the same conditional exposure probabilities, for example, this is given by

$$L^I_{\mathsf{CCS}}(\theta) \;\; = \;\; \prod_{i\in I} (p_{e|c})^{n^i_{ce}}\, (1 - p_{e|c})^{n^i_{c\bar{e}}}\, (p_{e|\bar{c}})^{n^i_{\bar{c}e}}\, (1 - p_{e|\bar{c}})^{n^i_{\bar{c}\bar{e}}}, \qquad (7)$$

exactly the same as if $n^+_{ce} \equiv \sum_{i\in I} n^i_{ce}$ exposed and $n^+_{c\bar{e}} \equiv \sum_{i\in I} n^i_{c\bar{e}}$ unexposed cases, $n^+_{\bar{c}e} \equiv \sum_{i\in I} n^i_{\bar{c}e}$ exposed and $n^+_{\bar{c}\bar{e}} \equiv \sum_{i\in I} n^i_{\bar{c}\bar{e}}$ unexposed non-cases had been observed in a single pooled study, leading (for Jeffreys' priors) to

$$L^I_{\mathsf{P}}(\varepsilon) \;\; \propto \;\; \left[\frac{\varepsilon\, e^{n^+_{ce}\,\varepsilon}}{e^{\varepsilon} - 1}\right] {}_2F_1(n^+_c + 1,\, n^+_e + 1;\, n^+ + 2;\, 1 - e^{\varepsilon}), \qquad (8)$$

approximately normal with mean and variance

$$\mu_{\mathsf{P}} \approx \log(n^+_{ce}\, n^+_{\bar{c}\bar{e}} \,/\, n^+_{\bar{c}e}\, n^+_{c\bar{e}}) \qquad\qquad \sigma^2_{\mathsf{P}} \approx 1/n^+_{ce} + 1/n^+_{c\bar{e}} + 1/n^+_{\bar{c}e} + 1/n^+_{\bar{c}\bar{e}}.$$

In most cases this is just not reasonable—the evidence from the separate studies is *not* identical with the evidence from one pooled trial with the same total number of subjects. The several studies almost always differ in important respects that affect their probabilities of exposure and outcome and hence their evidence about θ. The effect of ignoring this variation is systematically to underrepresent uncertainty in the posterior distribution, and sometimes to distort the location of the distribution as well.

1.2.2 Fixed Effects

A first step toward accommodating study variation is to allow the parameter vector $\theta^i = (p^i_{ce}, p^i_{c\bar{e}}, p^i_{\bar{c}e}, p^i_{\bar{c}\bar{e}})$ to vary across studies, while still treating the measure of interest ε as constant across all studies. Changing variables to $\varepsilon = \varepsilon_{\mathsf{LOR}}$ and $p^i_{e|\bar{c}} = p^i_{\bar{c}e}/p^i_{\bar{c}}$ for cohort studies and $p^i_{c|\bar{e}} = p^i_{c\bar{e}}/p^i_{\bar{e}}$ for case control studies leads to

$$L^I_{\mathsf{CHS}}(\varepsilon, p^I_{c|\bar{e}}) = e^{\varepsilon n^+_{ce}} \prod_{i\in I} \left[(p^i_{c|\bar{e}})^{n^i_c} (1 - p^i_{c|\bar{e}})^{n^i_{\bar{c}}} [1 - p^i_{c|\bar{e}}(1 - e^{\varepsilon})]^{-n^i_e} \right] \tag{9}$$

$$L^I_{\mathsf{CCS}}(\varepsilon, p^I_{e|\bar{c}}) = e^{\varepsilon n^+_{ce}} \prod_{i\in I} \left[(p^i_{e|\bar{c}})^{n^i_e} (1 - p^i_{e|\bar{c}})^{n^i_{\bar{e}}} [1 - p^i_{e|\bar{c}}(1 - e^{\varepsilon})]^{-n^i_c} \right] \tag{10}$$

and in each case to $L^I(\varepsilon) = \int L(\varepsilon, p^I)\, \pi(d\,p^I|\varepsilon)$ for the evidence about ε—for independent Jeffreys' prior distributions, for example, to

$$L^I_{\mathsf{FE}}(\varepsilon) \propto \frac{\varepsilon^{|I|}\, e^{n^+_{ce}\,\varepsilon}}{(e^{\varepsilon} - 1)^{|I|}} \prod_{i\in I} {}_2F_1(n^i_c + 1,\, n^i_e + 1;\, n^i + 2;\, 1 - e^{\varepsilon}), \tag{11}$$

approximately normal again but now with mean and variance

$$\mu_{\mathsf{FE}} \approx \frac{\sum_{i\in I} \mu_i/\sigma^2_i}{\sum_{i\in I} 1/\sigma^2_i} \qquad\qquad \sigma^2_{\mathsf{FE}} \approx \frac{1}{\sum_{i\in I} 1/\sigma^2_i},$$

the precision-weighted mean of the sample means $\mu_i \approx \log(n^i_{ce}\, n^i_{\bar{c}\bar{e}} \,/\, n^i_{\bar{c}e}\, n^i_{c\bar{e}})$ using variances $\sigma^2_i \approx [1/n^i_{ce} + 1/n^i_{\bar{c}e} + 1/n^i_{c\bar{e}} + 1/n^i_{\bar{c}\bar{e}}]$. The best choice of $\pi(d\,p|\varepsilon)$ will depend on the application; in the ETS example of Section 3 both reference and informed prior distributions are illustrated.

1.2.3 Random Effects

While some invariance may be offered by using a measure of exposure effect that is thought to be relatively insensitive to study variation (for example, using $\varepsilon_{\mathsf{LOR}}$ if the variations are expected to scale incidence odds by equal factors; see (Wolpert, 1986)), in general the evidence cannot meaningfully be combined without some sort of explicit accommodation of effect variability. A second step toward accommodating study variation is to model the variation across studies in the effect measures ε^i as well.

If we take $\varepsilon^i \sim \mathrm{N}(\varepsilon, \sigma^2)$, for example, independent normal variates with common unknown mean ε and known variance σ^2, and ascribe a diffuse conjugate distribution $\varepsilon \sim \mathrm{N}(0, \tau^2)$, the earlier normal approximations to $L_{\mathsf{CHS}}(\varepsilon_{\mathsf{LOR}})$ and $L_{\mathsf{CCS}}(\varepsilon_{\mathsf{LOR}})$ lead to the familiar random effects model with approximately normal posterior distribution with mean and variance

$$\mu_{\mathsf{RE}} \approx \frac{\sum_{i\in I} \mu_i/(\sigma^2+\sigma_i^2)}{1/\tau^2 + \sum_{i\in I} 1/(\sigma^2+\sigma_i^2)} \qquad \sigma^2_{\mathsf{RE}} \approx \frac{1}{1/\tau^2 + \sum_{i\in I} 1/(\sigma^2+\sigma_i^2)}.$$

Uncertainty about hyperparameters τ^2 and σ^2 is reflected through prior distributions in the ETS example of Section 3.

1.2.4 Exchangeable Hierarchical Models

It is a small generalization of the random effect model to model the variation in study-specific parameters θ_i explicitly. Modeling them as exchangeable is equivalent (by DeFinetti's theorem) to treating them as conditionally independent identically distributed random vectors, given an overall hyperparameter θ; if the exposure effect ε is conditionally independent of the $\{\theta_i\}$, given θ (for example, if it can be written as a function of θ) then the joint prior distribution for ε, θ, and all the $\{\theta_i\}$ can be factored as $\pi(d\varepsilon)\,\pi(d\theta|\varepsilon)\,\prod_{i\in I}\pi(d\theta_i|\theta)$, so the marginal likelihood for the exposure effect has the representation:

$$L^I_{\mathsf{EHM}}(\varepsilon) \;=\; \int \left[\prod_{i\in I} \int L_i(\theta_i)\,\pi(d\theta_i|\theta)\right] \pi(d\theta|\varepsilon). \tag{12}$$

This reduces to the Random Effects model for $\theta \equiv (\mu, \tau, \sigma)$ and normal logistic conditional distributions $\pi(d\theta_i|\theta)$; to the Fixed Effects model for $\theta \equiv \varepsilon$ and beta conditional distributions $\pi(d\,p^i_{c|\bar{e}}|\theta)$ and $\pi(d\,p^i_{e|\bar{c}}|\theta)$; and to Simple Pooling for $\theta \equiv (p_{ce}, p_{c\bar{e}}, p_{\bar{c}e}, p_{\bar{c}\bar{e}})$ with unit point masses $\pi(d\theta_i|\theta)$ at $\theta_i = \theta$.

2 Synthesizing Heterogeneous Evidence

Early exploratory studies usually show stronger effects than later confirmatory ones, as do small studies (more suseptible to publication bias) and studies from more susceptible population or those with greater exposure. Studies vary in the degree to which confounding effects and possible biases have been recognized and accomodated and in the levels of similarity between their subject populations and the study population. The four synthesis methods presented in Section 1 (simple pooling, fixed and random effect models, and exchangeable hierarchical models) allow for increasing degrees of heterogeneity among studies, but they do not make use of collateral evidence about the size and impact of varying age and exposure distributions, misclassification rates, and other differences that might be expected to affect the evidence.

Several methods have been proposed for discounting evidence thought to be less reliable or to apply less directly to the questions of interest in meta-analysis (some of which were first developed to address the similar problems that arise in synthesizing the opinions of several experts (Makridakis and Winkler, 1983; Genest and Zidek, 1986; Wolpert, 1989)), including:

1. Threshold exclusion, in which studies thought to be less applicable are simply excluded from the analysis;

2. Weighted likelihood functions, in which fractional powers of likelihood functions are used to discount less applicable studies;

3. Block mixtures, in which similarly-applicable studies are included in the analysis in blocks, with a monotonicity constraint;

4. Mixtures, in which individual studies are included in the analysis with probabilities based on their relative applicability;

5. Hierarchical models, in which treatment effects are estimated separately for homogeneous groups of studies, to reveal systematic variation of apparent effects across groups.

This paper investigates another alternative, intended to extract, reconcile, and synthesize information about a quantity of interest despite study variations and flaws:

6. Systematic adjustment of the different studies within the Bayesian paradigm to accomodate (and, where possible, correct for) their differences from one another and from the intended object of study.

2.1 Adjustment

Putting aside for the moment any apprehensions about overstating our certainty about the relationship between the i^{th} study and the paradigm, we have available an entire class of *adjustments* of the form $\theta_i = \phi_i(\theta)$, i.e.,

$$\pi_i(d\theta_i|\theta) = \delta(\theta_i - \phi_i(\theta))d\theta_i, \tag{13}$$

where the arbitrary function $\phi_i(\theta)$ represents the certain value of θ_i when the paradigm parameter value is θ. The specification of the function $\phi_i(\theta)$ by the experimenter now determines the conditional prior distribution completely, and through (15) the adjusted likelihood

$$L_i^{\text{Adj}}(\theta) = L_i(\phi_i(\theta)).$$

If the analyst studying a series of clinical trials believes that the success probability in the i^{th} trial should be only a fraction (say, half) of the success probability θ under paradigm conditions, perhaps because half of the subjects in that trial were noncompliant, then he might use $\phi_i(\theta) = \theta/2$ for that trial; of course, this choice implies that his joint prior probability distribution for the pair (θ, θ_i) is concentrated on the one-dimensional set $\{(\theta, \theta_i) : 0 \leq \theta_i \leq \frac{1}{2}, \theta = 2\theta_i\}$ and in particular that $\theta_i \leq \frac{1}{2}$. Similarly one analyzing a series of cohort studies might use $\phi_i(\theta) = 0.9\theta + 0.1\theta_0$ if 10% of the subjects were thought to be misclassified from a group with event probability θ_0 rather the paradigm probability θ.

More generally, for any adjustment function $\theta_i = \phi_i(\theta)$ and any prior probability distribution $\pi(d\theta)$ there is an *implicit* marginal prior probability distribution for θ_i

$$\begin{aligned} \pi_i(d\theta_i) &= \int \pi_i(d\theta_i|\theta)\pi(d\theta) \\ &= \left[\int \delta(\theta_i - \phi_i(\theta))\pi(d\theta)\right] d\theta_i; \end{aligned}$$

in the common case in which $\pi(d\theta) = \pi(\theta)d\theta$ has a density function and the relation $\theta_i = \phi_i(\theta)$ can be inverted by a differentiable function $\theta = \phi_i^{-1}(\theta_i)$, the integral yields a density function for $\pi_i(d\theta_i) = \pi_i(\theta_i)d\theta_i$ given by:

$$\pi_i(\theta_i) = \pi(\phi_i^{-1}(\theta_i)) \left| \frac{d}{d\theta_i}\phi_i^{-1}(\theta_i) \right|.$$

For example, the adjustment $\phi_i(\theta) = \theta/2$ for any prior probability density function $\pi(\theta)$ implies the prior probability density $\pi_i(\theta_i) = 2 \times \pi(2\theta_i)$ for θ_i, supported on the set $\{0 \leq \theta_i \leq \frac{1}{2}\}$. If an experimenter finds this implicit prior for θ_i untenable, the choice of the adjustment function $\phi_i(\theta)$ or of the marginal prior $\pi(d\theta)$ should be reconsidered.

2.1.1 Parametric Adjustments

A useful generalization of the conditionally certain models are the *parametric adjustment models* in which the adjustment function

$$\theta_i = \phi_i(\theta, \alpha_i)$$

depends explicitly on a parameter α_i, leading to a degenerate conditional prior probability distribution

$$\pi_i(d\theta_i|\theta) = \delta(\theta_i - \phi_i(\theta, \alpha_i))d\theta_i. \tag{14}$$

For example, we may allow a shift in the binomial probability parameter θ by setting

$$\phi_i(\theta, \alpha_i) = \theta + \alpha_i$$

or a logistic shift by specifying $\frac{\phi_i}{1-\phi_i} = e^{\alpha_i}\frac{\theta}{1-\theta}$, i.e.,

$$\phi_i(\theta, \alpha_i) = \frac{\theta e^{\alpha_i}}{\theta e^{\alpha_i} + 1 - \theta}.$$

In each case the adjusted likelihood function is calculated by a simple change of variables:

$$L_i^{\text{Adj}}(\theta) = L_i(\phi_i(\theta, \alpha_i)).$$

Although (13) offers no real increase in generality over the nonparametric adjustments made in (12), it is often possible to simplify the modeling and analysis by using the same function $\phi_i(\theta, \alpha_i) = \phi(\theta, \alpha_i)$ for several studies and allow only the parameter α_i to change from one to another, simplifying the problem of eliciting conditional prior distributions. Several examples will illustrate the parametric adjustment approach below.

2.1.2 Uncertain Adjustments

If the parameter α_i in (13) is regarded as uncertain, and therefore (in the Bayesian context) random with a prior probability distribution $\pi_i^\alpha(d\alpha_i|\theta)$, then we can form a conditional distribution for θ_i given θ by averaging (13) over the possible values of α_i:

$$\pi_i(d\theta_i|\theta) = \int \delta(\theta_i - \phi_i(\theta, \alpha_i))\pi_i^\alpha(d\alpha_i|\theta).$$

Combined with (15), this relation yields a simple expression for the adjusted likelihood function:

$$L_i^{\text{Adj}}(\theta) = \int L_i(\phi_i(\theta, \alpha_i))\pi_i^\alpha(d\alpha_i|\theta). \tag{15}$$

It is important *not* to use a noninformative prior for α_i, since this ordinarily results in a constant likelihood function $L_i^{\text{Adj}}(\theta)$ lending no evidence whatsoever about θ; this is the mathematical reflection of the fact that an instrument whose bias or scale is entirely unknown can give no evidence about a measured quantity.

2.2 Specific Examples of Adjustment

In this section we will examine three specific examples of adjustments, each of the parametric or uncertain type described above, in order to illustrate the concepts.

In both case control and cohort studies, each subject is classified three times: once when assessing eligibility, once when assessing exposure, and once when assessing outcome (case or non-case status). Thus three fundamentally different sorts of classification error may affect the analysis: eligibility violation, exposure misclassification, and case misclassification.

2.2.1 Eligibility Violation

Under both CCS and CHS designs, criteria which determine eligibility may be subjective or may be based on information which is potentially inaccurate. Moreover, in a CCS, there may be inaccuracies in the determination of case or noncase status of potential subjects, leading to the erroneous inclusion of some subjects in these arms of the study. These eligibility violations impact directly on the probabilities p_c and p_e, and also indirectly on the conditional probabilities of exposure or case status because these are usually different for erroneously included subjects (else there would be no need for the eligilibility criteria).

Denote by α_i^1 the probability that a "case" will be misclassified (in Experimenters wishing to exclude the evidence from misclassified subjects would use a noninformative prior probability distribution for the true probabilities of being eligible to enter the study, while those wishing to adjust the evidence and include it in the analysis would specify an informative conditional prior probability distribution between the true and purported probabilities.

If treatment or exposure status (in a CCS) or successes and failures (in a ChS) are misreported within the trials, then the parameter p_i governing the *reported* results X_i will differ from the parameter p governing the unobservable *true* results.

2.3 Other Adjustment Issues

2.3.1 Variable Follow-up

If p_i represents the probability of survival for some period of time in the i^{th} study, and if the period of time (say, T_i years) varies from one study to another, some sort of adjustment must be made before all the evidence can be brought to bear on any quantity of interest θ. If a constant hazard rate model seems appropriate then each p_i might be expected to be of the form $p_i = e^{-HT_i}$ for some unknown hazard H; if it is the hazard itself which is selected as the "quantity of interest" $\theta = H$, then the model function

$$p_i = \phi_i(\theta, T_i) = e^{-\theta T_i}$$

would be appropriate, while choosing five-year survival $\theta = e^{-5H}$ suggests

$$p_i = \phi_i(\theta, T_i) = \theta^{T_i/5}.$$

The adjusted likelihood function for $\theta = H$ upon observing s_i survivors and f_i nonsurvivors in T_i years is

$$L_i^{\text{Adj}}(\theta) = (e^{-\theta T_i})^{s_i}(1 - e^{-\theta T_i})^{f_i}.$$

It is even more important to adjust evidence from RCTs in which the length of follow-up differs between the control and treated groups. The simple constant hazard model suggests again that if $\theta_i = (p_e^i, p_{\bar{e}}^i)$ represents the vector of success probabilities in the i^{th} trial and if $\theta = (H^1, H^0)$ represents the vector of hazards in the treated and control groups, then the appropriate adjustment is

$$p_e^i = e^{-H^1 T_i^1} \qquad p_{\bar{e}}^i = e^{-H^0 T_i^0}$$

or, for five-year survival $\theta = (p^1, p^0)$,

$$p_e^i = (p^1)^{T_i^1/5} \qquad p_{\bar{e}}^i = (p^0)^{T_i^0/5}.$$

For simple binomial evidence consisting of counts s_i^1, f_i^1 of survivors and nonsurvivors in the treated group, and s_i^0, f_i^0 in the control group, the adjusted likelihood function for p_e^i and $p_{\bar{e}}^i$ in this latter case is

$$L_i^{\mathrm{Adj}}(p^1, p^0|X_i) = ((p^1)^{T_i^1/5})^{s_i^1}\,((1-(p^1)^{T_i^1/5})^{f_i^1}\,((p^0)^{T_i^1/5})^{s_i^0}\,((1-(p^0)^{T_i^0/5})^{f_i^0}.$$

2.3.2 Bias and Confounding

If subjects in a CCS are not well matched on all relevent criteria, or if those in a ChS differ in ways not causally related to the exposure, or if subjects in an RCT are assigned to treatments without randomization, it is possible that the subjects assigned to the case (resp. exposed, resp. treated) group will differ systematically from those assigned to the control group, lending misleading evidence to an effect. In the presence of such a *bias* it can be misleading to interpret the trial results at face value without making adjustments. Here is one way to make such adjustments.

Choose a measure of treatment effect $\varepsilon = \varepsilon(\theta)$ and suppose that we have available evidence about $\varepsilon_i = \varepsilon(\theta_i)$ from each of a set I of independent trials in the form of marginal likelihood functions (see Section xx)

$$L_i(\varepsilon_i|X_i);$$

recenter and rescale if necessary to arrange that $\varepsilon > 0$ if the treatment is helpful, $\varepsilon = 0$ if the treatment is ineffective, and $\varepsilon < 0$ if the treatment is harmful. In the presence of bias the data may offer evidence that $\varepsilon_i \approx \beta_i$ for some $\beta_i \neq 0$ even in the absence of a true treatment effect.

In the simplest adjustment model the differences $\beta_i = \varepsilon_i - \varepsilon$ would be taken as conditionally independent given ε, each with a conditional prior distribution $\pi_i^\beta(d\beta_i|\varepsilon)$. Now the adjusted likelihood function from the i^{th} trial would be

$$L_i^{\mathrm{Adj}}(\varepsilon|X_i) = \int L_i(\varepsilon + \beta_i|X_i)\pi_i^\beta(d\beta_i|\varepsilon).$$

2.3.3 Disease Severity, Treatment Intensity

For some CCSs, ChSs, or RCTs it makes sense to quantify the *intensity* or *severity* of the disease, exposure, or treatment, respectively, or at least the relative intensity or severity when two or more studies are compared. For a drug therapy the intensity might be depend on the delivered dose (the drug concentration at its site of action within the body) or the average dose over time, while the "intensity" of a screening program intended to help discover some disease in its early stages might depend on the screening frequency or on the skill and the technology of those performing the screen.

If several trials differ in severity or intensity from each other or from the paradigm, then some sort of adjustment is called for before combining the evidence or applying it. Denote by τ_i the severity or intensity in the i^{th} trial, with unity $\tau_i = 1$ for disease severity, treatment intensity, or exposure intensity identical to that of the paradigm.

To adjust the evidence from a trial with $\tau_i \neq 1$ we select a measure of treatment effect ε as above and postulate that the effectiveness in the observed trial will be $\varepsilon_i = \tau_i \times \varepsilon$. This leads to the adjusted likelihood function

$$L_i^{\text{Adj}}(\varepsilon|X_i) = \int L_i(\tau_i\varepsilon|X_i)\pi_i^{\tau}(d\tau_i|\varepsilon)$$

for a trial with intensity τ_i governed by the prior (conditional) probability distribution $\pi_i^{\tau}(d\tau_i|\varepsilon)$.

2.3.4 Simultaneous Uncertain Adjustments

Usually conditions within each study will differ in several different ways from those of the paradigm, and so multiple adjustments will be needed. When more than one of these is an uncertain adjustment (e.g., with parameters α_i^j for the j^{th} adjustment for the i^{th} study) then a *joint* adjustment distribution $\pi_i^{\alpha}(d\alpha_i^j|\theta)$ must be supplied; for example, simultaneous adjustments for bias and intensity (where we have used the letters β and τ in place of the generic α^j) can be combined in the form

$$L_i^{\text{Adj}}(\varepsilon|X_i) = \int L_i(\tau_i\varepsilon + \beta_i|X_i)\pi_i^{\tau,\beta}(d\tau_i, d\beta_i|\varepsilon).$$

If the uncertainty in the adjustments is small relative to other sources of uncertainty in the analysis, then the adjustment parameters α_i^j will have (nearly) degenerate distributions, simplifying the calculations; in other cases a subjective joint distribution must be elicited and fit, a challenging prospect.

3 Is exposure to ETS associated with lung cancer in female nonsmokers?

Whether or not exposure to other people's tobacco smoke, or *passive smoking*, is harmful to health is an issue which has been widely debated over the past decade and which has broad implications in current tort and public policy decisions. While many meta-analyses of the various epidemiological studies have been published in recent decades (Wald et al., 1986; NRC Committee on Passive Smoking, 1986; Tweedie and Mengersen, 1992; EPA Review, 1992; Lee, 1992; OSHA, 1994), there remains an open question about the impact of *study quality* on the individual and overall results. In particular, because of the small magnitude of the possible effect, possible biases and misclassification have not been entirely ruled out as explanations for any observed increase in relative risk of deleterious health effects for those thought to be exposed to environmental tobacco smoke (ETS).

Three main quality issues are commonly identified as influential in the analysis of the association between ETS exposure and lung cancer in adults:

1. **Misclassification of active smokers as nonsmokers**. If active smokers are included among the subjects in a study of nonsmokers, any observed excess risk may be misattributed to exposure instead of to active smoking; since active smokers' relative risks are well-known to be elevated, spuriously raised relative risks maybe observed with relatively small levels of such misclassification. Considerable attention has been paid to this source of bias, but efforts to correct it predominantly comprise a single "adjustment" to the combined relative risk estimate or adjustments to each study-specific relative risk estimate prior to meta-analysis.

2. **Misclassification of disease**. Classification is not verified histologically in many studies and inaccuracy in diagnosis of lung cancer deaths based on death certificates or clinical diagnoses is widely recognised (Lee, 1992, *pp.* 128–129; Table 23.3, *p.* 87). McFarlane et al. (1986) report that such misclassification is differential between active smokers and nonsmokers, but (Lee, 1992, *p.* 129) counters that these differentials have not been established for those exposed or unexposed to ETS. Supporting analyses have simply compared the overall relative risk for those studies which are almost entirely histologically confirmed with that for the remaining studies.

3. **Misclassification of exposure**. The measure of exposure to ETS is

most often indirect and inadequate, leading to speculation of under- or over-reporting and possible misclassification as 'exposed' or 'unexposed'. Most studies in our dataset do not use any objective measure of exposure (such as cotinine analysis, which itself is prone to criticism since it measures only recent exposure). Although this problem is not sufficient, on its own, to explain the observed association between lung cancer and ETS exposure, its effect in combination with other biases is unknown.

The Environmental Protection Agency Review (1992) and Lee (1992) provide comprehensive discussions and literature reviews of these problems for thirty-one studies available at that time, of which relevant data are available for twenty-nine. As a result, we focus on this dataset and ignore more recent studies (Besag et al., 1997; Givens et al., 1995, see) of the ETS/lung cancer association.

Table 1 lists and summarizes the twenty-nine studies we have used to shed light on this association, based on Tables 5-2 and 5-5 from EPA Review (1992) and Table 3.13 Lee (1992), where detailed references appear. The 'Quality Tiers' (1='best'–4='worst') are scores assigned by the EPA (EPA Review, 1992) to reflect the ability of each study to control for various quality issues. The score for each study was based on summation of "penalty points" (in the range -5 (bonus) to 1.0) awarded in each of 19 categories, comprising the study's account of never-smoker status, ETS-exposure criteria, lung cancer indication, interview type, proxy respondents, followup, design issues, and analysis issues (control of age, control of other confounders, statistical methods). Studies in the fourth tier were regarded as being unsuitable for inclusion in a meta-analysis.

That heterogeneity exists between studies in this dataset is indicated by the fact that the EPA Review undertakes meta-analyses within but not across country groups. These analyses are based on a fixed-effects model, with the stated implicit assumption that studies within country groups are homogenous with respect to all other confounders (p. 5–31).

The impact of study quality is also observed in the EPA Review, in which substantial difference is found between the relative risk estimates obtained from cumulative meta-analyses of studies in the various quality tiers (Tier 1, Tiers 1-2, Tiers 1-3), again within country groups. For interest, these results are reproduced in Table 2 and form a baseline against which to compare the various modifications considered in this paper. In further analyses, the EPA pays specific attention to the quality issue of misclassification of active smokers as neversmokers. Unlike previous adjustments by Wald (1986), Lee

	Yrs	Cases Unex	Cases Exp	Non-cases Unex	Non-cases Exp	Qual Tier	Ctry Grp
Case Control Studies							
1 Akiba et al. (1986)	71–80	21	73	82	188	2	3
2 Brownson et al. (1987)	79–82	15	4	41	6	3	4
3 Buffler et al. (1984)	76–80	8	33	32	164	3	4
4 Chan and Fung (1982)	76–77	50	34	73	66	4	2
5 Correa et al. (1983)	79–82	8	14	72	61	2	4
6 Fontham et al. (1991)	85–88	126	294	288	492	1	4
7 Gao et al. (1987)	84–86	57	189	99	276	3	6
8 Garfinkel et al. (1985)	71–81	43	91	148	254	2	4
9 Geng et al. (1988)	1983	20	34	52	41	4	6
10 Humble et al. (1987)	80–82	5	15	71	91	2	4
11 Inoue and Hirayama (1988)	73–83	4	18	17	30	4	3
12 Kabat and Wynder (1984)	71–80	11	13	10	15	2	4
13 Kalandidi et al. (1990)	87–89	26	65	46	74	1	1
14 Koo et al. (1987)	81–83	35	51	70	66	1	2
15 Lam T et al. (1987)	83–86	84	115	183	152	2	2
16 Lam W et al. (1985)	81–84	23	37	80	64	3	2
17 Lee et al. (1986)	77–82	10	22	21	45	2	5
18 Liu et al. (1991)	85–86	9	45	26	176	4	6
19 Pershagen et al. (1987)	60–80	33	37	141	153	1	5
20 Shimizu et al. (1988)	82–85	38	52	72	91	2	3
21 Sobue et al. (1990)	86–88	64	80	336	395	2	3
22 Svensson et al. (1989)	83–85	10	24	60	114	2	5
23 Trichopoulos et al. (1983)	78–82	24	53	109	116	3	1
24 Wu et al. (1985)	80–81	10	19	24	38	2	4
25 Wu-Williams et al. (1990)	85–87	212	205	271	331	4	6
Cohort Studies							
26 Butler (1988)	76–82	5	3	43052	10579	2	2
27 Garfinkel (1981)	60–72	65	88	81794	94792	3	4
28 Hirayama (1984)	66–81	37	163	21858	69482	2	3
29 Hole et al. (1989)	72–85	1	5	488	1290	1	5

Table 1: Twenty-nine studies of the association between lung cancer and exposure to spousal smoking among neversmoking females.

(1986) and the NRC (1986), in which the overall relative risk was adjusted, the EPA Report computes "corrected" relative risk estimates for each study and the consequent bias expressed as a ratio of the corrected and uncorrected estimates. The methodology on which this is based is due to Wells and Stewart (Wells, 1990), and is founded on misclassification rates found in a small number of cotinine studies and studies of discordant answers. Different

Table 2. Overall Estimates of RR, reproduced from EPA Report (1992, Table 5-7).

Country Group	All studies RR (90% CI)	Tiers 1-2 RR (90% CI)
Greece	2.01 (1.42, 2.84)	1.92 (1.13, 3.23)
Hong Kong	1.48 (1.21, 1.81)	1.61 (1.25, 2.07)
Japan	1.41 (1.18, 1.69)	1.39 (1.16, 1.66)
USA	1.19 (1.04, 1.35)	1.23 (1.04, 1.42)
W. Europe	1.17 (0.84, 1.62)	1.17 (0.85, 1.64)
China	0.95 (0.81, 1.12)	-

(and larger) biases have been found by Lee (1992) using essentially the same methodology but different baseline estimates.

These observations confirm the necessity of accounting for both inhomogeneity between studies and study quality in a meta-analysis of these studies. In this paper we investigate alternative methods for doing this, based not on crude overall or individual study adjustment, but on integrating information about these quality issues into the likelihood itself.

While one may not agree with the scoring method used by the EPA or the categories chosen, this Review and the associated evidence provided by Lee (1992) provide a rich foundation from which to conduct a more sensitive meta-analysis through adjusted likelihoods.

3.1 Naïve Meta-analysis of the ETS Data

3.1.1 One Combined Study

As discussed in Section 1, the most naïve method of combining the twenty-nine studies would be to consider them as a single large study. However, because the two study designs give rise to different conditional probabilities, they must be considered separately. Hence $(n_{ec}, n_{e\bar{c}}, n_{\bar{e}c}, n\bar{e}\bar{c})$ become $(1617, 3560, 946, 2424)$ for the case control studies and $(259, 176143, 108, 147192)$ for the cohort group.

In Section 1 the three models for the single large trial were discussed. Under model **Mp**, the $p_{e|c}, p_{e|\bar{c}}, p_{c|e}$ and $p_{c|\bar{e}}$ were modeled directly; under model **Mle**, the LOR was modelled directly and exactly; and under model **Mla**, the posterior distribution of the LOR was approximated as gaussian. Assuming no other information about $p_{e|c}, p_{e|\bar{c}}, p_{c|e}$ and $p_{c|\bar{e}}$, we might impose $\text{Beta}(1/2, 1/2)$ priors on each of these parameters. Some summary posterior estimates for the ETS data are given in Table 3.

In this example, however, information is available to construct quite informative priors. Thus....The results are in Table 3. It is clear that the utilization of this additional information through the priors has ...

3.1.2 Exchangeable Studies

A first accommodation of study heterogeneity is through exchangeable models, assuming either fixed or random effects. Using the information described in the previous section, we can construct priors corresponding to the four formulations described in Section 1 (FEp: fixed effects Mp model; FEla: fixed effects Mla model; REp and REla for the analogous random effects models). For the *ith* case control study these become

$$\mathbf{FEp} : p^i_{e|\bar{c}} \sim \text{Be } (1/2, 1/2); \quad p^i_{e|c} = e^{\varepsilon} p^i_{e|\bar{c}}/(1 + p^i_{e|\bar{c}}(1 + e^{\varepsilon})); \quad \varepsilon \sim N(?, ?)$$

$$\mathbf{FEla} : \text{ logit } (p^i_{e|\bar{c}}) = \mu^i; \text{ logit } (p^i_{e|c}) = \mu^i + \varepsilon$$

$$\varepsilon \sim N(?, ?); \quad \mu^i \sim N(?, ?)$$

$$\mathbf{REp} : p^i_{e|c} = e^{\varepsilon i} p^i_{e|\bar{c}}/(1 + p^i_{e|\bar{c}}(1 + e^{\varepsilon i}))$$

$$\varepsilon^i \sim N(\varepsilon, \tau); \quad \varepsilon \sim N(?, ?); \quad \tau \sim \text{Ga } (?, ?)$$

$$\mathbf{REla} : \text{logit } (p^i_{e|c}) = \mu^i + \varepsilon^i$$

$$\varepsilon^i \sim N(\varepsilon, \tau); \quad \varepsilon \sim N(?, ?); \quad \tau \sim \text{Ga } (?, ?)$$

Analogous formulations are derived for the cohort studies, with [describe changes in above values].

Using MCMC to compute the resultant posterior distribution, the mean and 90% credible interval for the overall LOR for the ETS datset are given in Table 3. [Decide how much of the mcmc method needs to be put in here or somewhere.]

3.2 Nonexchangeable Models

As discussed earlier in this section, the assumption of exchangeability for the ETS studies is not tenable. We consider below three specific examples of nonexchangeability, recognising that there are many others within this dataset, but illustrating through these the implementation and impact of adjusted likelihoods described in Section 2.

We consider in detail the adjustments required for cohort studies, omitting corresponding details for the case control studies.

Table 3. Posterior estimates under naïve meta-analyses

	$p_{e\|c}$	$p_{e\|\bar{c}}$	$p_{c\|e}$	$p_{c\|\bar{e}}$	$\varepsilon_{\mathsf{LOR}}$
One Trial					
Case Controls					
Mean	$6.31E-1$	$5.95E-1$			
±Std Dev	$\pm 6.35E-3$	$\pm 9.13E-5$			
Cohorts					
Mean			$1.47E-3$	$7.37E-4$	
±Std Dev			$\pm 7.07E-5$	$\pm 7.07E-5$	
Fixed Effects					
Model FEp					
Mean					
±Std Dev					
Model FEla					
Mean					
±Std Dev					
Random Effects					
Model REp					
Mean					
±Std Dev					
Model REla					
Mean					
±Std Dev					

3.2.1 Eligibility Violation

Two sources of misclassification in the initial intake of subjects are inclusion of current and former smokers as nonsmokers and misdiagnosis of cases, so that for study i where Similarly where

3.2.2 Misclassification of smoking status

From (Lee, 1992, Tables 3.33, 3.35 and 3.36 and the discussion on *p.* 157) and in accord with estimates used elsewhere (Lee (1992); EPA Review (1992)), we construct the following estimates, based on cotinine tests and requestioning: α_{1i1} is between 0.5 and 2.0% for current smokers and 10-15% for former smokers, giving an overall estimate of 2% for eversmokers; α_{1i0} is between 1.0 and 3.0% for current smokers and 13-18% for former smokers, giving an overall estimate of 1.5% for eversmokers. Because of the debate about differences in rates between countries (Lee (1992, 1993), (EPA Review, 1992, *pp.* 2-24, 2-25, B-13), we impose an error of 0.005 for US and European studies and 0.01 for other country groups.

Estimates of q_{1i}^1 and q_{1i}^0 may be derived from (EPA Review, 1992, Table B-11), which provides for each study relative risk estimates for eversmokers, and from the individual 2×2 data. We can show that $\varepsilon_{\mathsf{LRR}}(\mathsf{s}) \approx q_{1i}^1$, and a corresponding range for q_{1i}^0 might be constructed by considering additive and multiplicative models. Under the former, we might take $q_{1i}^0 = q_{1i}^1$. Under the latter, the RR is the same among eversmokers as among neversmokers, so that $q_{1i}^0 = q_{1i}^1/\varepsilon_{\mathsf{LRR}}^i$. For each study, we might express our uncertainty about ${}_1q_i^0$ via a Gaussian distribution with mean at the midpoint of the estimates under the additive and multiplicative model and standard deviation reflecting the corresponding range.

3.2.3 Misclassification of exposure

Tweedie et al. (1995) provide support for the following estimates: 47% of currently nonsmoking wives of smokers have less than one hour of exposure at home; 40–50% of women with nonsmoking spouses have some ETS exposure.

From this we might infer, for neversmoking subjects, $\mathrm{P}(\bar{E} \mid `E') = 0.53 \pm 0.05$ and $\mathrm{P}(E \mid `\bar{E}') = 0.55 \pm 0.05$. Furthermore, we adopt for $\mathrm{P}(`E' \mid NS)$ the range $0.20 \pm 0.03$, reflecting the estimates of 0.17% and 0.22% reported from large studies by Venters (Lee, 1992, *p.* 161) and Lee (1988), respectively. After manipulating the corresponding $2 \times 2$ table of $`\bar{E}'/`E'$ and

$\bar{E}/E$ subjects, $\begin{bmatrix} E & \bar{E} \\ `E' & `\bar{E}' \end{bmatrix}$, we obtain the following estimates: $\alpha_{2i1}, \mathrm{P}(\bar{E} \mid `E', NS) = 0.11 \pm 0.02; \alpha_{2i0}, \mathrm{P}(E \mid `\bar{E}', NS) = 0.44 \pm 0.05$.

As indicated above, EPA Review (1992) attempts to categorise studies with respect to their ability to control this source of bias, in particular through the amount of histological verification undertaken. We employ these classifications in allocating to the individual studies estimates within the above ranges. Acknowledging the potential simplification, we now allow that $q^1_{2i} = p^0 \qquad q^0_{2i} = p^1$.

3.2.4 Misclassification of lung cancer

A typical error to which cohort studies are prone after initial classification is that of misclassification of lung cancer. Here we introduce the notation

Unlike classification based on active smoking, there are neither sufficient data nor rationale to differentiate between misdiagnosis of lung cancer according to ETS exposure, so we let $\alpha_{-i} = \alpha_{-i1} = \alpha_{-i0}$ and $\alpha_{+i} = \alpha_{+i1} = \alpha_{+i0}$. From (Lee, 1992, $p.\,129$) we might infer that α_{-i} ranges from 0.0 to 0.15 and α_{+i} ranges from 0.0 to 0.30. Using the EPA classifications for each study's control of this particular source of bias, we make the corresponding allocations, with an error of 0.005 attached to each estimate.

4 Discussion

Probably we should add a discussion.

Acknowledgements

The authors would like to thank Professors Jim Berger, Richard Tweedie, and Bill DuMouchel for helpful conversations. This work was supported by United States Environmental Protection Agency grant CR-822047-01-0, United States National Science Foundation grant DMS-9626829, Australian Research Council grant ARC-000, and the Queensland University of Technology CISSaIM (School of Mathematical Sciences).

References

Abramowitz, M. and Stegun, I. A., editors. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, Washington, D.C., 1964.

Berger, J. O., Liseo, B., and Wolpert, R. L. Integrated likelihood methods for eliminating nuisance parameters. Discussion Paper 97-01, Duke University ISDS, USA, January 1997. Submitted to *Statistical Science*.

Berger, J. O. and Wolpert, R. L. *The Likelihood Principle: A Review, Generalizations, and Statistical Implications*, volume 6 of *Institute of Mathematical Statistics Lecture Notes-Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, USA, second edition, 1988. (with Discussion).

Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors. *Bayesian Statistics 4*, Oxford, 1992. Oxford University Press.

Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors. *Bayesian Statistics 5*. Oxford University Press, Oxford, 1996.

Berry, D. A. A bayesian approach to multicenter trials and meta-analysis. In *ASA Proceedings of Biopharmaceutical Section*, pages 1–10. American Statistical Association, 1990a.

Berry, D. A., editor. *Statistical Methodology in the Pharmaceutical Sciences*, volume 104 of *Statistics: Textbooks and Monographs*. Marcel Dekker, New York, NY, USA, 1990b.

Berry, D. A. and Stangl, D. K., editors. *Bayesian Biostatistics*. Marcel Dekker, New York, NY, USA, 1996.

Besag, J., Higdon, D., and Mengersen, K. L. Logistic regression and meta-analysis with unobserved covariates. In preparation, 1997.

Biggerstaff, B. J., Tweedie, R. L., and Mengersen, K. L. Passive smoking in the workplace: Classical and Bayesian meta-analyses. *International Archives of Occupational and Environmental Health*, 66:269–277, 1994.

Breslow, N. E. and Day, N. E. *Statistical methods in cancer research (Vol. I): The analysis of case-control studies*, volume 32 of *IARC Scientific Publication Series*. Oxford University Press, Oxford, 1980.

Breslow, N. E. and Day, N. E. *Statistical methods in cancer research (Vol. II): The design and analysis of cohort studies*, volume 82 of *IARC Scientific Publication Series*. Oxford University Press, Oxford, 1987.

Carlin, J. B. Meta-analysis for 2×2 tables: A Bayesian approach. *Statistics in Medicine*, 11:141–158, 1992.

Cox, D. R. *Analysis of binary data*. Monographs on Statistics and Applied Probability. Chapman & Hall, New York, NY, USA, 1970.

Dempster, A. P., Selwyn, M. R., and Weeks, B. J. Combining historical and randomized controls for assessing trends in proportions. *Journal of the American Statistical Association*, 78(382):221–227, June 1983.

DuMouchel, W. H. Bayesian metaanalysis. In Berry (1990b), pages 509–529.

DuMouchel, W. H. Predictive cross-validation of Bayesian meta-analyses. In Bernardo et al. (1996), pages 107–127. (With discussion).

DuMouchel, W. H. and Harris, J. E. Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association*, 78(382):293–315, June 1983.

Eddy, D. M. The Confidence Profile method: A Bayesian method for assessing health technologies. *Journal of Operations Research*, 37:210–228, 1989.

EPA Review. Health effects of passive smoking: Assessment of lung cancer in adults and respiratory disorders in children. Document EPA/600/6-90/006F, U.S. Environmental Protection Agency, 1992.

Fisher, R. A. The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6:13–25, 1934.

Freeman, D. H. *Applied Categorical Data Analysis*, volume 79 of *Statistics: textbooks and monographs*. Marcel Dekker, New York, NY, USA, 1987.

Genest, C. and Zidek, J. V. Combining probability distributions. *Statistical Science*, 1(1):114–148, February 1986.

Givens, G. H., Smith, D. D., and Tweedie, R. L. Esimating and adjusting for publication bias using data augmentation in Bayesian meta-analysis. Submitted to *Statistical Science*, November 1995.

Hedges, L. V. and Olkin, I. *Statistical Methods for Meta-Analysis*. Academic Press, New York, NY, USA, 1985.

Kass, R. E. and Greenhouse, J. B. Comment: A Bayesian perspective. *Statistical Science*, 4(4):310–317, November 1989. (Discussion of Ware (1989)).

Kasuga, H., editor. *Environmental Tobacco Smoke*. Springer-Verlag, New York, NY, USA, 1993.

Larntz, K., editor. *Statistical Issues in Combining Environmental Studies*, Washington, D.C., 1986. EPA Office of Policy, Planning, and Evaluation.

Lavine, M. L., Wasserman, L., and Wolpert, R. L. Bayesian inference with specified prior marginals. *Journal of the American Statistical Association*, 86(416):964–971, December 1991.

Lee, P. N. *Misclassification of smoking habits and passive smoking: A review of the evidence*. Springer-Verlag, New York, NY, USA, 1988.

Lee, P. N. *Environmental Tobacco Smoke and Mortality*. Karger Press, Basel, CH, 1992.

Lee, P. N. An assessment of the epidemiological evidence relating lung cancer risk in never smokers to environmental tobacco smoke exposure. In Kasuga (1993), pages 28–70.

Lee, P. N. Lung cancer and ETS: Is the epidemiological evidence conclusive? Presented at Invited Session 271, American Statistical Association Annual Meeting, Toronto, August 1994.

LeVois, M. E. and Layard, M. W. Inconsistency between workplace and spousal studies of environmental tobacco smoke and lung cancer. *Regulatory Toxicology and Pharmacology*, 19:309–316, 1994.

Makridakis, S. and Winkler, R. L. Averages of forecasts: Some empirical results. *Management Science*, 29(9):987–996, September 1983.

McFarlane, M. J., Feinstein, A. R., and Wells, C. K. Necropsy evidence of detection bias in the diagnosis of lung cancer. *Archives of Internal Medicine*, 146:1695–1698, 1986.

Mengersen, K. L., Tweedie, R. L., and Biggerstaff, B. J. The impact of method choice in meta-analysis. To appear in *Australian Journal of Statistics*, 1995.

Morris, C. N. and Normand, S.-L. Hierarchical models for combining information and for meta-analysis. In Bernardo et al. (1992), pages 321–344.

NRC Committee on Applied and Theoretical Statistics. *Combining Information: Statistical Issues and Opportunities for Research.* National Academy Press, Washington, D.C., 1992.

NRC Committee on Passive Smoking. *Environmental Tobacco Smoke: Measuring Exposures and Assessing Health Effects.* National Academy Press, Washington, D.C., 1986.

OSHA. Proposed rule on indoor air quality. *Federal Register*, 59(65):15968–16039, 1994.

Richardson, S. and Gilks, W. R. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology*, 138:430–442, 1993.

Smith, T. C., Spiegelhalter, D. J., and Parmar, M. K. B. Bayesian meta-analysis of randomized trials using graphical models and BUGS. In Berry and Stangl (1996), pages 411–427.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. *BUGS: Bayesian Inference Using Gibbs Sampling Version 0.50.* MRC Biostatistics Unit, Cambridge, 1995. (software and examples manuals).

Tweedie, R. L. and Mengersen, K. L. Lung cancer and passive smoking: Reconciling the biochemical and epidemiological approaches. *British Journal of Cancer*, 66:700–705, 1992.

Tweedie, R. L., Mengersen, K. L., and Eccleston, J. A. Garbage in, garbage out: Can statisticians quantify the effects of poor data? *Chance*, 7(2):20–27, 1995.

Wald, N. J., Nanchahal, K., Thompson, S. G., and Cuckle, H. S. Does breathing other people's tobacco smoke cause lung cancer? *British Medical Journal*, 293:1217–1222, 1986.

Ware, J. H. Investigating therapies of potentially great benefit: ECMO. *Statistical Science*, 4(4):298–340, November 1989. (With discussion).

Wolpert, R. L. Choosing a measure of treatment effect. In Larntz (1986), pages 19–25.

Wolpert, R. L. Eliciting and combining expert judgments about uncertainty. *International Journal of Technology Assessment in Health Care*, 5 (4):537–557, 1989.

Wolpert, R. L. and Warren-Hicks, W. J. Bayesian hierarchical logistic models for combining field and laboratory survival data (with discussion). In Bernardo et al. (1992), pages 525–546.