7

Observational Studies

7.1 INTRODUCTION

The RCT is generally considered the 'gold-standard' methodology in evaluating health-care interventions, but there are circumstances in which randomisation is either impossible or unethical (*e.g.* evaluating the health effects of smoking) or where there is substantial valuable information available in non-randomised or 'observational' data (Concato *et al.*, 2000). In many circumstances such observational data would form part of an evidence synthesis, which is dealt with in Chapter 8.

It is important to understand that the probability models used in Bayesian analysis are expressions of personal or group uncertainty and so do not need to be based on randomisation. Therefore in principle non-randomised studies can be analysed in exactly the same manner as randomised comparisons. In Section 7.2 we describe how both case–control and cohort designs provide a likelihood which can be combined with prior information using standard Bayesian methods, perhaps with extra attention to adjusting for covariates in an attempt to control for possible baseline differences in the treatment groups with respect to uncontrolled risk factors or exposures.

Of course, the dangers associated with the use of observational studies in evaluating health-care interventions have been well described in the medical literature (Byar *et al.*, 1976). For example, Dunn *et al.* (2002) compare randomised and non-randomised evidence collected according to a common protocol, and find a potentially misleading treatment comparison based on the observational data. Essentially, randomised studies should provide an unbiased likelihood for the parameter of interest, while observational studies may have a degree of systematic bias. In this book we do not argue the case for or against the use of non-randomised studies, but suggest that *if* observational studies are to be used, then their analysis falls naturally into a Bayesian framework. Specifically, the possibility of bias leads inevitably to a degree of subjective judgement about the comparability of studies, and this fits well into the

Bayesian Approaches to Clinical Trials and Health-Care Evaluation D. J. Spiegelhalter, K. R. Abrams and J. P. Myles
© 2004 John Wiley & Sons, Ltd ISBN: 0-471-49975-7

acknowledged judgement underlying all Bayesian reasoning. Hence, in Section 7.3 we consider the explicit modelling of potential biases, building on the structure developed in the context of evidence-based priors (Section 5.4) and using historical controls (Section 6.9), in each of which a range of methods are possible for 'downweighting' studies to allow for doubts about their degree of relevance.

Finally, in Section 7.4 we consider the specific issue of making institutional comparisons, also known as 'profiling'. This fits naturally into a hierarchical modelling framework, and we also show how a Bayesian approach allows direct probability statements about the rank of an institution.

7.2 ALTERNATIVE STUDY DESIGNS

Case–control studies involve retrospective investigation of risk factors for a sample of cases and controls, possibly matched for known risk factors. Inference is generally on the odds ratio, which is directly estimable from this design. Bayesian approaches have generally relied on analytic approximations in order to obtain reasonably simple analyses (Zelen and Parker, 1986; Marshall, 1988; Nurminen and Mutanen, 1989; Zelen, 1990); for example, Ashby *et al.* (1993) examine two case–control studies studying leukaemia following chemotherapy treatment for Hodgkin's disease, and consider the consequences of various prior distributions based on a cohort study. However, all the techniques for analysing clinical trials can be adopted, with the additional complication in relation to judgements on the potential for bias and appropriateness of the prior. Example 7.1 describes the analysis of Lilford and Braunholtz (1996) concerning potential side-effects of oral contraceptives using a likelihood arising from case–control studies.

A large *cohort* study or *registry* database may provide observational evidence on the 'natural history' of a disease, which might be used to model the consequences of an intervention; for example, Craig *et al.* (1999) describe an analysis of a population-based cohort of patients with diabetic retinopathy in order to evaluate different screening policies. It is, of course, possible to directly estimate apparent effects of different interventions from registry data, although again the potential for bias should be acknowledged: Example 9.3 illustrates one technique for downweighting registry and single cohort data in an evidence synthesis.

There is also a substantial literature on Bayesian methods for complex epidemiological modelling, particularly spatial correlation (Heisterkamp *et al.*, 1993; Bernardinelli *et al.*, 1995; Richardson *et al.*, 1995; Ashby and Hutton, 1996), measurement error (Richardson and Gilks, 1993) and missing covariate data (Raghunathan and Siscovick, 1996).

7.3 EXPLICIT MODELLING OF BIASES

Bayesian techniques for explicitly modelling potential bias, both within studies and in the attempt to generalise studies outside their target population, were pioneered by Eddy *et al.* (1992) under their general title of the 'confidence profile method' (Section 8.1).

Biases to internal validity mean that the effect of interest is not being appropriately estimated within the circumstances of the study. For example, suppose we suspect that a proportion p of patients in a study did not comply with the intended treatment, although we do not know who these patients are. If we are interested in estimating the treatment effect θ_t in those who actually received the treatment, then the overall underlying treatment effect in the trial will be $\theta = (1-p)\theta_t + p\theta_0$, where θ_0 is the effect in non-compliers. A likelihood for θ can thus be transformed into a likelihood for θ_t, provided there is other evidence or prior opinion concerning p and θ_0. The likelihood therefore provides information on a *function* of the parameters of interest, and a fairly complex example is provided in Example 8.7.

Eddy *et al.* (1992) identify a range of potential biases that can be modelled in this manner: these include dilution and contamination due to those who are offered a treatment not receiving it, errors in measurement of outcomes, errors in ascertainment of exposure to an intervention, loss to follow-up, and patient selection and confounding in which the groups differ with respect to measurable features. These biases may occur singly or in combination.

Biases to external validity concern the ability of a study to generalise to defined populations or to be combined with studies carried out on different groups, and may be relevant even if a study has been meticulously carried out and has obtained an unbiased assessment of the treatment effect within its own study population. These include 'population bias' in which the study and general population differ with respect to known characteristics, 'intensity bias' in which the 'dose' of the intervention is varied when generalised, and differences in lengths of follow-up.

We have previously discussed the use of historical data as a basis for prior opinion (Section 5.4) or as historical controls in clinical trials (Section 6.9), and in each case examined ways of 'discounting' the data from their face-value interpretation. In each of these contexts it has been assumed that the current observed data, for example in a randomised trial, directly depend on the parameter of interest. The potential biases, whether internal or external, in observational studies can be modelled using similar techniques, but in this context the current likelihood may be adjusted.

As a simple example, we assume a normal likelihood

$$y_m \sim \mathrm{N}[\theta_{\mathrm{Int}}, \sigma^2/m],$$

where θ_{Int} represents an 'internal' parameter that is being estimated in the current study. Following the development in Sections 5.4 and 6.9, we might assume a bias δ so that $\theta_{\text{Int}} = \theta + \delta$, where θ is the parameter of real interest. Options then include the following:

1. Assuming δ is known.
2. Assuming δ has a known distribution with mean 0, indicating a non-systematic bias. If we assume $\delta \sim \text{N}[0,\ \sigma^2/n_\delta]$, from (2.25) we obtain a likelihood for the parameter of interest θ,

$$y_m \sim \text{N}\left[\theta,\ \sigma^2\left(\frac{1}{m}+\frac{1}{n_\delta}\right)\right],$$

   i.e. the sample variance is inflated to allow for the potential bias.
3. If we suspect systematic bias in one direction, we might take δ to have a known distribution with non-zero mean, say $\delta \sim \text{N}[\mu_\delta,\ \sigma^2/n_\delta]$. We then obtain a likelihood

$$y_m \sim \text{N}\left[\theta+\mu_\delta,\ \sigma^2\left(\frac{1}{m}+\frac{1}{n_\delta}\right)\right],$$

   or equivalently

$$y_m - \mu_\delta \sim \text{N}\left[\theta,\ \sigma^2\left(\frac{1}{m}+\frac{1}{n_\delta}\right)\right]. \tag{7.1}$$

   Hence, after subtracting the assumed mean bias μ_δ from the observation y_m, (7.1) provides a likelihood for the parameter of interest that can be combined with an appropriate prior distribution for θ.

Each of these approaches is illustrated in Example 7.1.

In practice, analytic solutions will rarely be possible and MCMC techniques will be necessary. More serious are the assumptions required concerning the extent of the biases, since although data may be available on which to base accurate estimates, there is likely to be considerable judgemental input. Any unknown quantity can, of course, be given a prior distribution, and Eddy *et al.* (1992) claim this obviates the need for sensitivity analysis. They also argue strongly against simple downweighting using the 'power prior' model (Section 5.4) in which the effective sample size is reduced: they claim this is an arbitrary technique and that potential biases should be explicitly modelled. In fact, as we showed in Section 5.4, the models are effectively equivalent when handling a single study. We also note the increasing pace of research concerning the quantitative bias of observational studies: see, for example, Kunz and Oxman (1998), Britton *et al.* (1998), Benson and Hartz (2000), Ioannidis *et al.* (2001), Reeves *et al.* (2001) and Sanderson *et al.* (2001).

Example 7.1 *OC: interpreting case–control studies in pharmacoepidemiology*

Reference: Lilford and Braunholtz (1996).

Intervention: Third-generation oral contraceptives (OCs).

Aim of study: Suspicions had been raised as to whether 'third-generation' OCs increased the risk of venous thromboembolism compared to second-generation OCs. The aim of Lilford and Braunholtz (1996) was to assess the evidence from a Bayesian perspective.

Study design: Interpretation of a meta-analysis of four case–control studies.

Outcome measure: Odds ratio for venous thromboembolism, OR < 1 being in favour of 3rd-generation OCs.

Planned sample size: Not applicable.

Statistical model: Normal likelihood for pooled estimate of log(OR) derived from the meta-analysis of case–control studies, discounted for potential biases according to the methods described in Section 7.3. Lilford and Braunholtz (1996) consider a potential bias δ in the meta-analysis with a normal distrtribution: in the notation of Section 7.3, $\delta \sim \mathrm{N}[\mu_\delta,\ \sigma^2/n_\delta]$. They examine the effect of both a non-systematic and a systematic bias, as detailed below under 'Sensitivity analysis'.

Prospective analysis?: No.

Prior distribution: Prior beliefs were elicited from two gynaecologists with an interest in family planning. Expert 1 thought that a 20% risk reduction in venous thromboembolism would be associated with third-generation compared to second-generation OCs, i.e. OR $= 0.8$, but that the OR could be between 0.4 and 1.6. Assuming this corresponds to a 95% interval of a normal distribution, the true log(odds ratio), θ, can be assumed to have mean $\mu = \log(0.8) = -0.22$ and standard deviation $(\log(1.6) - \log(1.4))/(2 \times 1.96) = 0.35$. Equivalently, if we take $\sigma = 2$, we obtain a prior 'number of events' $n_0 = (\sigma/0.35)^2 = 31.9$.

Expert 2 thought that there was an equal chance of third-generation OCs reducing the OR of venous thromboembolism or increasing it, *i.e.* OR $= 1.0$, but was suitably uncertain as to think that the true OR was likely to be between 0.5 and 2.0. Using the same argument as for Expert 1, we assume an $\mathrm{N}[0,\ \sigma^2/31.9]$ prior for Expert 2.

Loss function or demands: No.

Computation/software: Conjugate normal model.

Evidence from study: The meta-analysis of case–control studies produced a pooled odds ratio of 2.0 with a 95% CI from 1.4 to 2.7. On a log(OR)

scale, this provides a likelihood with mean $\log(2.0) = 0.69$ and standard deviation $(\log(2.7) - \log(1.4))/(2 \times 1.96) = 0.17$. Equivalently, taking $\sigma = 2$, we obtain a sample 'number of events' $m = (\sigma/0.17)^2 = 142.5$.

Bayesian interpretation: Combining the evidence from the meta-analysis with each expert's prior beliefs produced the posterior distributions seen in Figure 7.1(a). Given that both gynaecologists were a priori quite uncertain as to the true odds ratio, their corresponding posterior distributions are influenced considerably by the data, so that the posterior distributions for both experts indicate less than 0.02% probability that third-generation OCs reduce the OR of venous thromboembolism.

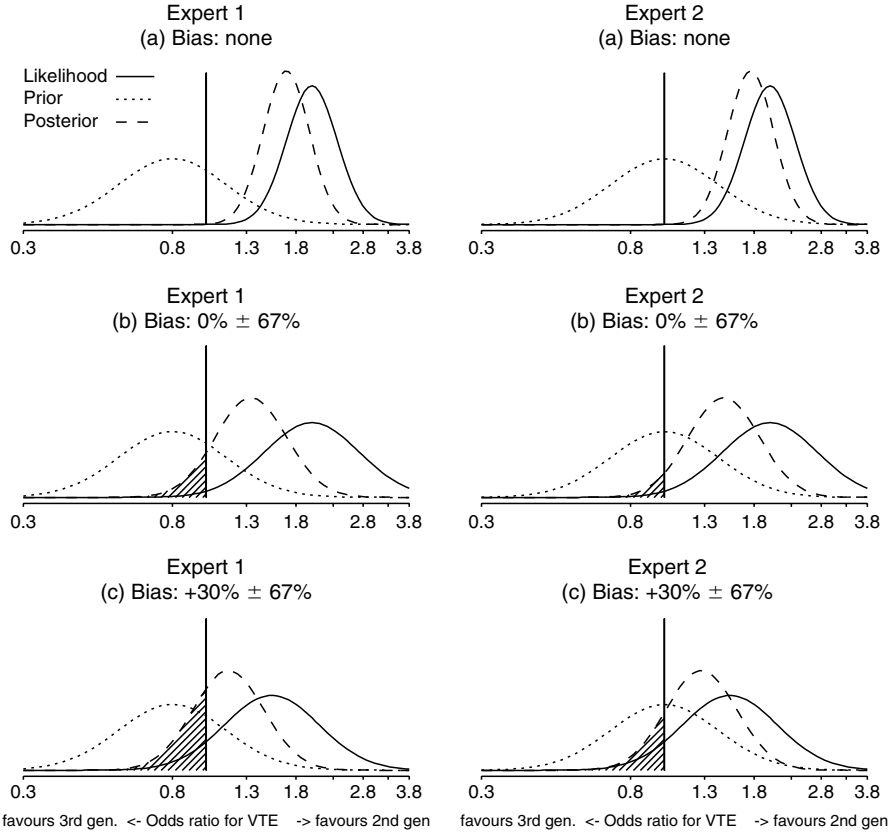

Figure 7.1 Likelihood, prior and posterior distributions for the oral contraceptive meta-analysis, showing the prior distributions for two experts and the results of (a) taking the meta-analysis at face value, (b) discounting the evidence by assuming the possibility of a random bias with standard deviation 30% on the HR scale, and (c) assuming an additional systematic bias of 30% on the HR scale.

Sensitivity analysis: It may be appropriate not to consider the evidence from such a meta-analysis at 'face value' since such retrospective epidemiological studies are known to be prone to various biases. Figure 7.1(b) shows an analysis in which the evidence from the meta-analysis is discounted using the non-systematic bias model described in Section 7.3.

Figure 7.1(b) shows the influence of a non-systematic $(\mu_\delta = 0)$ bias such that the odds ratio θ_{lnt} being estimated may be between 60% and 167% of the true odds ratio θ, *i.e*. up to a 67% bias in either direction. This corresponds, on a log(OR) scale, to a bias with standard deviation $\log(1.67)/\ 1.96 = 0.26$, equivalent, if we take $\sigma = 2$, to $n_\delta = (\sigma/0.26)^2 = 58.7$. The resulting posterior distributions for the two experts now give 11% and 5% probability to the notion that third-generation OCs may reduce the relative risk.

Figure 7.1(c) shows a further series of analyses in which the evidence from the meta-analysis is not only discounted, but also adjusted for the belief that case–control studies may have a systematic bias in which odds ratios are *overestimated* by a median of 30%: this is modelled by assuming $\mu_\delta = \log(1.3) = 0.26$, so that $\delta \sim \text{N}[0.26,\ 0.26^2]$. In this case the resulting posterior distributions show 27% and 15% probability that third-generation OCs may reduce the relative risk. Thus reasonable assumptions about the potential bias in the epidemiological studies, combined with a reasonably sceptical prior distribution, lead to substantial uncertainty as to the true effect of third-generation OCs.

Comments: There was great publicity surrounding the publication of this meta-analysis in 1995. Notification of family doctors in the UK was carried out in a 'panic' atmosphere, leading to a sudden drop in use of third-generation OCs, and reports of subsequent excess abortions. This Bayesian analysis suggests that such consternation may have been unfounded. A court case against the makers of third-generation OCs brought by 99 women who suffered strokes, deep vein thromboses and pulmonary embolisms was settled in July 2002 in the English courts, when the judge ruled that there was 'not, as a matter of probability, any increased relative risk' associated with the pills. It is notable that both sides in the case agreed that a doubling of risk had to be shown, in order that it was 'as likely as not' that any side-effect was caused by the third-generation OC. In view of this demand, it is hardly surprising the case against the companies failed.

Whilst this, and many other analyses have concentrated on the potentially negative effects of third-generation OCs, there has been evidence published that their use has been associated with a reduced relative risk of myocardial infarction compared to second-generation oral contraceptives. However, this example serves to illustrate the fact that in many situations in which there are numerous outcomes, both positive and

negative, consideration of one in isolation is fraught with danger. It is also notable that policy decisions should depend on differences in expected utilities whch in turn depend on risk differences rather than odds ratio (Section 3.14), and hence this analysis, strictly speaking, is not in a suitable form for decision-making.

7.4 INSTITUTIONAL COMPARISONS

If we consider an individual clinician, a medical team or a hospital as representing a class of 'intervention', then the use of performance indicators to compare outcomes could be considered as a form of evaluation. There are many complex issues surrounding such 'profiling' of institutions, including risk adjustment, choice of indicator, frequency of analysis, public reporting and so on, but these are beyond the scope of this book. Bayesian approaches to institutional comparisons have been suggested by Goldstein and Spiegelhalter (1996), Normand *et al.* (1997) and Christiansen and Morris (1997a), while fully Bayesian methods have also been used in the analysis of panel agreement data on the appropriateness of coronary angiography (Ayanian *et al.*, 1998).

A popular method when comparing institutions is to plot the observed performance (possibly risk-adjusted) and 95% confidence interval; see, for example, the New York cardiac surgery indicators (New York State Department of Health, 1998). If the interval does not overlap a benchmark then attention focuses on that centre. However, by chance alone one can expect 2.5% of centres to be identified as 'significantly' below standard, even if they are actually performing at the benchmark level. This indicates the need for caution in interpreting 'statistically significant' results, as this is essentially testing the hypothesis that each surgeon has exactly the same underlying patient mortality rate, which is neither plausible nor particularly interesting. We can deal with this 'multiplicity' problem (Section 3.17) in an analogous way to subset estimation (Section 6.8.1) and meta-analysis (Section 8.2), in using hierarchical models to make inferences based on estimating a common prior distribution, leading to 'shrunken' estimates for each centre. Furthermore, *regression to the mean* describes the tendency for institutions that have been identified as 'extreme' to become less extreme when monitored in the future – put simply, part of the reason for their extremity was a run of good or bad luck. This simple phenomenon could lead to spurious claims being made about the benefit of interventions to 'rescue' failing institutions. Shrinkage estimation is intended to counter this difficulty (Christiansen and Morris, 1997a).

An additional benefit of using Markov chain Monte Carlo methods (Section 3.19) is the ability to derive uncertainty intervals around the rank order of each

institution (Marshall and Spiegelhalter, 1998). Example 7.2 describes an analysis of success rates in *in vitro* clinics, in which Bayesian methods are used both to make inferences on the true rank of each clinic and to estimate the true underlying success rates with and without an exchangeability assumption.

Benefits of the Bayesian approach to institutional comparisons therefore include:

- methods for reporting probabilities that any specified centre's true rate exceeds any particular threshold of interest;
- a natural way of dealing with 'regression to the mean';
- explicit allowance for between-centre variability;
- an opportunity to incorporate covariates both at the patient and institutional level of the model;
- inferences on the true rank of the institution.

Example 7.2 *IVF: estimation and ranking of institutional performance*

Reference: Marshall and Spiegelhalter (1998).

Intervention: In vitro fertilisation (IVF).

Aim of study: The UK Human Fertilisation and Embryology Authority (HFEA) monitors clinics licensed to carry out donor insemination (DI) and IVF, and to help people who are considering fertility treatment to understand the services offered by licensed clinics and to decide which clinic is best for them (Human Fertilisation and Embryology Authority, 1996). They publish risk-adjusted live birth rates per treatment cycle started, and we are concerned with whether one can rank the institutions with any confidence.

Study design: Retrospective analysis of prospectively collected data on 52 clinics carrying out IVF treatment in the UK between April 1994 and March 1995.

Outcome measure: Estimated adjusted live birth rate $\hat{p}_k$, with 95% intervals, per treatment cycle started, where the case-mix adjustment is based on a pooled logistic regression of all IVF treatments.

Statistical model: If there are n_k treatments in the kth clinic, we calculate $r_k = \hat{p}_k n_k$ as the effective number of successful live births. The log-odds on success for each clinic are denoted y_k and estimated to be $y_k = \log[(r_k + 0.5)/(n_k - r_k + 0.5)]$, with estimated variance $s_k^2 = 1/(r_k + 0.5) + 1/(n_k - r_k + 0.5)$ (Section 2.4.1). Then we assume

$$y_k \sim \mathrm{N}[\theta_k, s_k^2],$$

where θ_k is the true log-odds on success in the kth clinic; an exact likelihood based on the binomial distribution is possible but makes negligible difference in this example due to the substantial number of treatments.

Two models for the θ_ks are considered. First, that they are *independent*. Second, the clinics are assumed to be fully *exchangeable* (Section 3.4), with the true rates (on a logit scale) being drawn from a common normal distribution: if, after adjusting for case-mix, we can find no other contextually meaningful way to differentiate between the institutions, then the assumption of their exchangeability seems justified. Hence we assume

$$\theta_k \sim \mathrm{N}[\mu, \tau^2].$$

Prior distributions:

Independence model. Originally assume the θ_k each have an independent uniform distribution: this is used for the ranking exercise.

Exchangeable model. Uniform priors are adopted for μ, τ.

Computation/software: MCMC techniques in the WinBUGS software are used to derive posterior distributions for the ranks of the institutions: this is done by calculating the current rank of each institution at each iteration of the simulation, and then summarising the distribution of these calculated ranks after many thousands of iterations.

Evidence from study: The raw data are shown in Figure 7.2.

Bayesian interpretation: It is clear from Figure 7.2 that there is substantial shrinkage towards the overall mean performance when assuming exchangeability, although there are still a number of clinics that would be considered 'significantly' above or below average. It can be argued that this adjustment is an appropriate means of dealing with the problem of multiple comparisons. In addition, this shrinkage should deal with 'regression to the mean', in which extreme institutions will tend back towards the overall average when they recover from their temporary run of good or bad luck.

Figure 7.3 shows that there is considerable uncertainty in the true rank of an institution, even when they show substantial differences in performance.

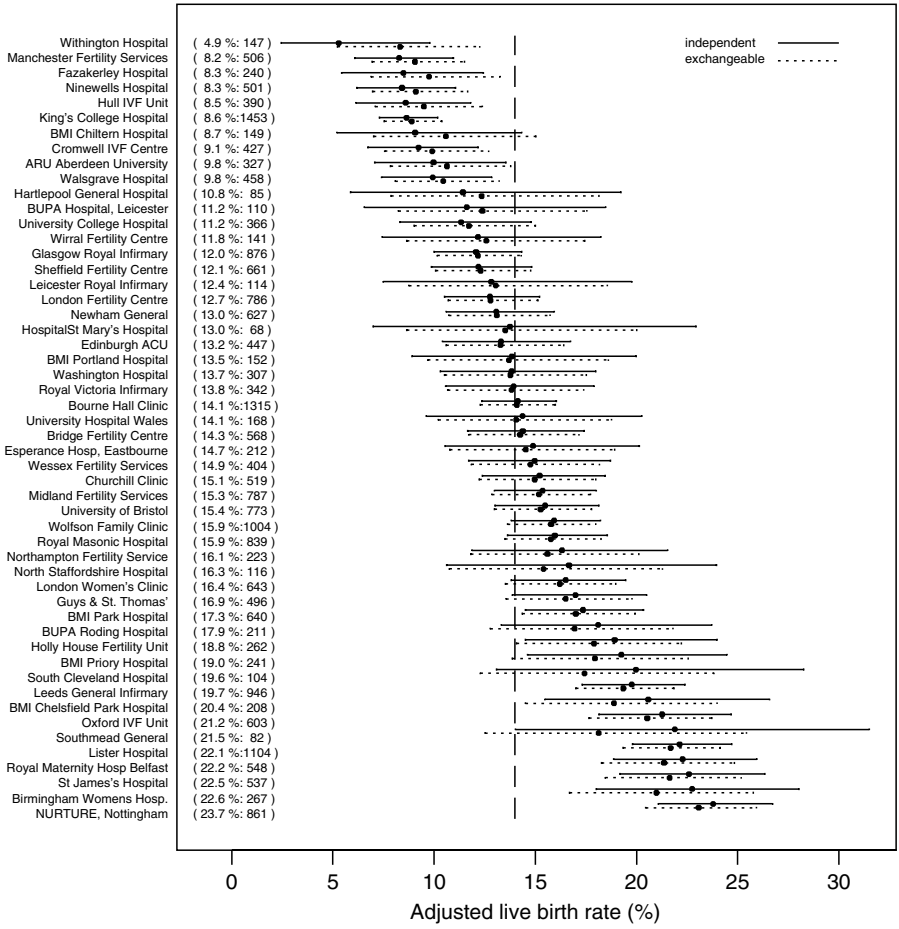

Figure 7.2 Estimates and 95% intervals for the adjusted live birth rate in each clinic, assuming both independent and exchangeable rates. The vertical lines represent the national average of 14%. The estimated adjusted live birth rate for each clinic is given in brackets, together with the number of treatment cycles started.

The consequence of assuming exchangeability is to reduce the differences between clinics and hence to make their ranks even more uncertain. Figure 7.3 shows this is the case to a limited extent, although since many of the extreme clinics are also fairly large, their rank is not unduly effected.

Sensitivity analysis: The results are extremely insensitive to the prior on τ and the use of a full binomial likelihood.

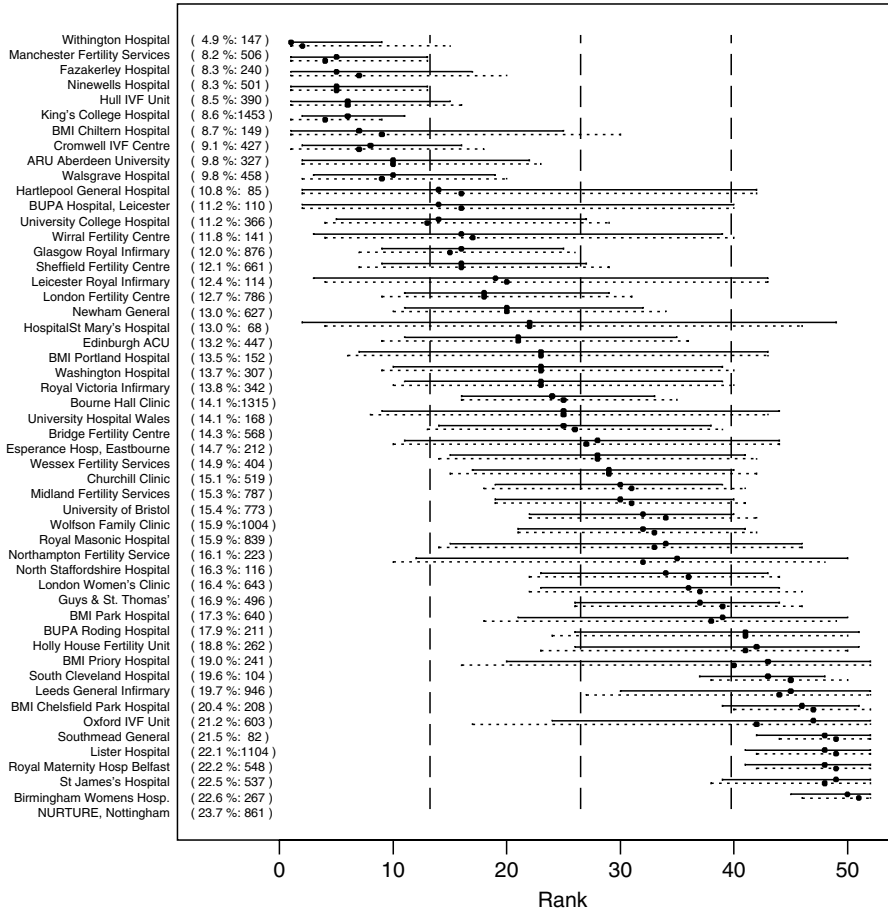

Figure 7.3 Median and 95% intervals for the rank of each clinic, assuming both independent and exchangeable rates. The dashed vertical lines divide the clinics into quarters according to their rank.

7.5 KEY POINTS

1. Data from observational studies may, in principle, be analysed in exactly the same framework as for randomised trials.
2. Imperfections in the design and conduct, and generalisation to other populations, may be approached by adopting a more complex model.
3. There are likely to be increased demands for Bayesian analysis, particularly in areas such as institutional comparisons and gene–environment interactions.

4. The explicit modelling of potential biases in observational data may be widely applicable but needs some evidence base in order to be convincing.
5. Analysis of sensitivity to modelling and prior assumptions is even more important than in RCTs.

EXERCISES

7.1. Ashby *et al.* (1993) consider the association between treatment for Hodgkin's disease and the subsequent risk of leukaemia. An international case–control study reported data on 149 cases who had Hodgkin's disease followed by leukaemia and 411 matched controls who had Hodgkin's disease but no subsequent leukaemia. Table 7.1 displays cases and controls stratified according to treatment received.
   (a) Estimate the probability that cases with leukaemia had been treated with chemotherapy, *i.e.* $p(C|L)$, and compare this with the probability that controls without leukaemia had been treated with chemotherapy, *i.e.* $p(C|\overline{L})$.
   (b) Prove that from these quantities you can estimate the odds ratio associating leukaemia with treatment with chemotherapy, *i.e.* $[p(L|C)/p(\overline{L}|C)]/[p(L|\overline{C})/p(\overline{L}|\overline{C})]$.
   (c) Hence estimate the log(odds ratio) and its variance from the table.
   (d) Assuming a sceptical prior that doubts whether odds ratios as large as 10 are reasonable, how does this influence the conclusions?

7.2. Suppose that $r = 20$ people responded out of $n = 50$ given a particular drug. We then hear that $p = 20\%$ of individuals did not in fact take the drug. (a) Express the overall response rate θ in the experiment in terms of the true response rate θ_t of those who did take the drug, the proportion p of compliers, and the response rate θ_0 of those who did not take the drug. Assuming a uniform prior for θ_t, what inference would you make on θ_t, assuming (b) $\theta_0 = 0$, (c) a Beta[2,10] prior distribution for θ_0?

7.3. In Example 7.1, justify the statement that the bias is equivalent to a 'standard deviation of 30% on the HR scale'. How might you interrogate an expert concerning the potential size of a bias?

Table 7.1 Results from an international case–control study of leukaemia following treatment for Hodgkin's disease.

Treatment	Cases	Controls
No chemotherapy	11	160
Chemotherapy	138	251
Total	149	411

Table 7.2 Odds ratios and 95% CIs for venous thromboembolism in users of third-generation oral contraceptives compared to second-generation OCs.

Study	Odds ratio	95% CI
Farley *et al.*	2.6	1.4 to 4.8
Jick *et al.*	2.2	1.1 to 4.4
Bloemenkamp *et al.*	2.5	1.2 to 5.2
Spitzer *et al.*	1.5	1.1 to 2.2

7.4. Table 7.2 presents the results of the four case–control studies reported by Lilford and Braunholtz (1996) in Example 7.1. Estimate the log(odds ratio) assuming (a) a pooled-effects model and (b) a random-effects model, using the empirical Bayes methodology of Section 3.17. The analysis in Example 7.1 considers a conjugate normal analysis, using the results of a meta-analysis of the four studies to produce an approximate normal likelihood. (c) Examine the sensitivity of the conclusions to the assumptions underlying the meta-analysis.

7.5. In Example 7.2, investigate the claim that the findings are robust to the prior on τ and the use of a full binomial likelihood.

7.6. Goldstein and Spiegelhalter (1996) report the teenage conception rates shown in Table 7.3.

Table 7.3 Teenage conception rates (13–15-year-olds) in 1990–1992 for 15 health boards in Scotland.

Health Board	No. conceptions	Relevant population
Western Isles	6	1 935
Orkney	5	1 220
Highland	76	11 515
Borders	36	5 294
Lanark	230	31 944
Argyle	172	23 243
Forth	121	14 938
Glasgow	388	45 647
Shetland	13	1 512
Lothian	303	35 233
Dumfries	67	7 614
Grampian	267	27 526
Ayr	204	20 606
Fife	188	18 614
Tayside	208	20 000

(a) Calculate the observed conception rates per 10 000 population, and rank the health boards according to their rates.
(b) Assuming either Poisson or binomial responses, estimate the ranks of each health board in a 'league table', assuming both independent and exchangeable rates.
(c) What is the probability that Tayside truly has the highest rates?