Departament d'Estadística i I.O., Universitat de València.
Facultat de Matemàtiques, 46100–Burjassot, València, Spain.
Tel. 34.6.363.6048, Fax 34.6.363.6048 (direct), 34.6.386.4735 (office)
Internet: bernardo@uv.es, Web: http://www.uv.es/˜bernardo/

Printed on December 9, 1996
Invited paper for the *Bhattacharya Memorial Volume*, to be published in the *Far East J. Math. Sci.*

The Concept of Exchangeability and its Applications

JOSÉ M. BERNARDO
Universitat de València, Spain

SUMMARY

The general concept of *exchangeability* allows the more flexible modelling of most experimental setups. The representation theorems for exchangeable sequences of random variables establish that any coherent analysis of the information thus modelled requires the specification of a joint probability distribution on all the parameters involved, hence forcing a *Bayesian* approach. The concept of partial exchangeability provides a further refinement, by permitting appropriate modelling of related experimental setups, leading to coherent information integration by means of so-called *hierarchical models*. Recent applications of hierarchical models for combining information from similar experiments in education, medicine and psychology have been produced under the name of *meta-analysis*.

Keywords: BAYESIAN INFERENCE; EXCHANGEABILITY; HIERARCHICAL MODELS; META-ANALYSIS; REFERENCE DISTRIBUTIONS; REPRESENTATION THEOREMS.

1. INTRODUCTION

It is generally agreed that the uncertainty relative to the possible values of the *observable* outcome $\boldsymbol{x}{=}\{x_1,\ldots,x_n\}$ of an experiment of size n is appropriately described by its joint probability distribution with, say, density $p(\boldsymbol{x}) = p(x_1,\ldots,x_n)$, so that the probability that $\boldsymbol{x}$ belongs to a region A is

$$P(\boldsymbol{x} \in A) = \int_A p(\boldsymbol{x})\, d\boldsymbol{x}$$

and, hence, by standard arguments of probability theory, the (predictive) probability that a future 'similar' observation x_{n+1} belongs to an interval I given the information provided by $\boldsymbol{x}$ is

$$P(x_{n+1} \in I \,|\, \boldsymbol{x}) = \int_I p(x_{n+1} \,|\, x_1,\ldots,x_n)\, dx_{n+1},$$

$$p(x_{n+1} \,|\, \boldsymbol{x}) = \frac{p(x_1,\ldots,x_{n+1})}{p(x_1,\ldots,x_n)}\ .$$

José M. Bernardo is Professor of Statistics at the University of Valencia. Research partially funded with grant PB93-1204 of the DGICYT, Madrid, Spain.

It follows that, in order to predict a future *observable* quantity given a sequence of 'similar' observations —which is one of the basic problems in the analysis of scientific data— it is necessary and sufficient to assess, for any n, the form of the joint probability density $p(x_1, \ldots, x_n)$.

In Section 2, we describe the concept of *exchangeability*, which makes precise the sense in which the observations must be 'similar'. In Section 3 we discuss the radical consequences of the exchangeability assumptions which are implied by the so-called representation theorems. In Section 4, we describe a further elaboration, introducing hierarchical models as an appropriate tool for information integration. Finally, Section 5 contains additional remarks, and some references to recent applied work on the combination of information, which makes use of the concepts reviewed here.

2. EXCHANGEABILITY

The joint density $p(x_1, \ldots, x_n)$, which we have to specify, must encapsulate the type of *dependence* assumed among the individual random quantities x_i. In general, there is a vast number of possible assumptions about the form such dependencies might take, but there are some particularly simple forms which accurately describe a large class of experimental setups.

Suppose that in considering $p(x_1, \ldots, x_n)$ the scientist makes the judgement that the subscripts, the 'labels' identifying the individual random quantities are 'uninformative', in the sense the information that the x_i's provide is independent of the order in which they are collected. This judgement of 'similarity' or 'symmetry' is captured by requiring that

$$p(x_1, \ldots, x_n) = p(x_{\pi(1)}, \ldots, x_{\pi(n)}),$$

for all permutations π defined on the set $\{1, \ldots, n\}$. A sequence of random quantities is said to be *exchangeable* if this property holds for every finite subset of them.

Example. ***(Physiological responses)***. Suppose $(x_1, \ldots, x_n)$ are real-valued measurements of a specific physiological response in human subjects when a particular drug is administered. If the drug is administered at more than one dose level, and if they are male and female subjects from different ethnic groups, one would be reluctant to make a judgement of exchangeability for the entire sequence of results. However, within each combination of dose-level, sex, and ethnic group, an assumption of exchangeability would often be reasonable.

We shall now review the important consequences of an exchangeability assumption implied by the general representation theorem.

3. THE GENERAL REPRESENTATION THEOREM

The 'similarity' assumption of exchangeability has strong mathematical implications. Formally, the general representation theorem provides an integral representation of the joint density $p(x_1, \ldots, x_n)$ of any subset of exchangeable random quantities. More specifically, if $\{x_1, x_2, \ldots\}$ is an exchangeable sequence of real-valued random quantities, then *there exists* a parametric *model*, $p(x \mid \boldsymbol{\theta})$, labeled by some parameter $\boldsymbol{\theta} \in \Theta$ which is the limit (as $n \to \infty$) of some function of the x_i's, and *there exists* a probability distribution for $\boldsymbol{\theta}$, with density $p(\boldsymbol{\theta})$, such that

$$p(x_1, \ldots, x_n) = \int_\Theta \prod_{i=1}^n p(x_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})\, d(\boldsymbol{\theta}).$$

In more conventional terminology, this means that, if a sequence of observations is judged to be exchangeable, *then*, any finite subset of them *is a random sample of some model* $p(x_i \mid \boldsymbol{\theta})$, and there *exists a prior* distribution $p(\boldsymbol{\theta})$ which has to describe the initially available information about the parameter which labels the model.

It then follows from standard probability arguments involving Bayes' theorem —hence the adjective 'Bayesian'— that the available information about the value of $\boldsymbol{\theta}$ after the outcome $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ of the experiment has been observed is described by its *posterior* density

$$p(\boldsymbol{\theta} \,|\, x_1, \ldots, x_n) = \frac{\prod_{i=1}^{n} p(x_i \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(x_1, \ldots, x_n)} \;;$$

Similarly, the available information about the value of a future observation x after $\boldsymbol{x}$ has been observed is described by

$$p(x \,|\, x_1, \ldots, x_n) = \int_{\Theta} p(x \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, x_1, \ldots, x_n) \, d\boldsymbol{\theta}.$$

It is important to realise that if the observations are conditionally independent, —as it is implicitly assumed when they are considered to be a random sample from some model—, then they are necessarily exchangeable. The representation theorem, —a pure probability theory result— proves that if observations are judged to be *exchangeable*, then they *must* indeed be a random sample from some model *and* there *must exist* a prior probability distribution over the parameter of the model, hence requiring a *Bayesian* approach.

Note however that the representation theorem is an *existence* theorem: it generally does not specify the model, and it never specifies the required prior distribution. The additional assumptions which are usually necessary to specify a particular model are described in particular representation theorems. An additional effort is necessary to assess a prior distribution for the parameter of the model.

Example. ***(Continued)***. If the measurements $\{x_1, x_2, \ldots\}$ of the response to the administration of a certain dose-level of some drug to a group of females of the same ethnic group are judged to be exchangeable *and* it is considered that

$$\overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i. \qquad s_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2,$$

the sample mean and variance, are *sufficient* statistics, in the sense that they are assumed to capture *all* the relevant information about the structure of the x_i's contained in $\{x_1, \ldots, x_n\}$, *then* the x_i's *must* necessarily be regarded as a random sample from a *normal* distribution $\mathrm{N}(x \,|\, \mu, \sigma)$ where

$$\mu = \lim_{n \to \infty} \overline{x}_n, \qquad \sigma = \lim_{n \to \infty} s_n,$$

and there *must exist* a prior distribution $p(\mu, \sigma)$ describing the available initial information about μ and σ. If, furthermore, it is assumed that no relevant information about either μ or σ is initially available, then the *reference* prior (Bernardo, 1979) $\pi(\mu, \sigma) \propto \sigma^{-1}$ may be used , and one finds that the available information about the population mean μ, and about a future observation x, after the outcome $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ of the experiment has been collected is respectively described by the Student t distributions

$$\pi(\mu \,|\, x_1, \ldots, x_n) = \mathrm{St}(\mu \,|\, \overline{x}, \frac{s}{\sqrt{n-1}}, n-1),$$

$$\pi(x \,|\, x_1, \ldots, x_n) = \mathrm{St}(x \,|\, \overline{x}, s\sqrt{\frac{n+1}{n-1}}, n-1).$$

In the next Section we extend the concept of exchangeability to be able to *integrate* the information available from different, related sources.

4. HIERARCHICAL MODELS

One often has several related sequences of exchangeable random quantities, the distribution of which depends on separately sufficient statistics in the sense that, if one has m of such sequences,

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \prod_{i=1}^{m} p(\boldsymbol{x}_i \,|\, \boldsymbol{t}_i),$$

where $\boldsymbol{t}_i$ is a sufficient statistic for $\boldsymbol{x}_i$. The corresponding integral representation is then of the form

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \int_\Theta \prod_{i=1}^{m} \prod_{j=1}^{n_i} p_i(x_{ij} \,|\, \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)\, d\boldsymbol{\theta}_1, \ldots, d\boldsymbol{\theta}_m.$$

Most often, the fact that the m sequences are being considered together means that all random sequences relate to the same measurement procedure, so that one typically has $p_i(x \,|\, \boldsymbol{\theta}_i) = p(x \,|\, \boldsymbol{\theta}_i)$.

Example. ***(Continued)***. If $x_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, m$ denote the responses to a drug from females from m ethnic groups in the conditions described above, then one would typically have a representation of the form

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \int_\Theta \prod_{i=1}^{m} \prod_{j=1}^{n_i} \mathrm{N}(x_{ij} \,|\, \mu_i, \sigma) p(\mu_1, \ldots, \mu_m, \sigma)\, d\mu_1, \ldots, d\mu_m d\sigma.$$

Thus, $\{x_{i1}, \ldots, x_{in_i}\}$ must be considered as a random sample from a normal distribution $\mathrm{N}(x \,|\, \mu_i, \sigma)$ and there must exist a prior distribution $p(\mu_1 \ldots, \mu_m, \sigma)$ which has to describe our assumptions on the relationship among the means of the m groups.

In this context, judgements of exchangeability are not only appropriate *within* each of the m separate sequences of observations, but also *between* the m corresponding parameters, so that its joint distribution has an integral representation of the form

$$p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m) = \int_\Phi \prod_{i=1}^{m} p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\phi}) p(\boldsymbol{\phi})\, d\boldsymbol{\phi},$$

and, hence, the parameter values which correspond to each sequence may be seen as a random sample from some parameter population with density $p(\boldsymbol{\theta} \,|\, \boldsymbol{\phi})$, and there must exist a prior distribution $p(\boldsymbol{\phi})$ describing the initial information about the *hyperparameter* $\boldsymbol{\phi}$ which labels $p(\boldsymbol{\theta} \,|\, \boldsymbol{\phi})$.

Example. ***(Continued)***. If the mean responses $\{\mu_1, \ldots, \mu_m\}$ within each of the m ethnic groups considered are judged to be exchangeable, then

$$p(\mu_1, \ldots, \mu_m, \sigma) = \prod_{i=1}^{m} p(\mu_i \,|\, \sigma) p(\sigma).$$

If, furthermore, their mean and standard deviations are judged sufficient to capture all relevant information about the μ_i's, then

$$p(\mu_1, \ldots, \mu_m, \sigma) = p(\sigma) \int \prod_{i=1}^{m} \mathrm{N}(\mu_i \,|\, \mu_0, \sigma_0) p(\mu_0, \sigma_0)\, d\mu_0, d\sigma_0,$$

and, hence, the μ_i's, the means of the responses within each ethnic group, may be regarded as a random sample from a normal population of female mean responses with overall mean μ_0 and standard deviation σ_0.

If no prior information is assumed on μ_0, σ_0 and σ, one may use the corresponding reference prior and derive the appropriate posterior distributions for each of the μ_i's, $\pi(\mu_i \,|\, \boldsymbol{x}_1, \ldots \boldsymbol{x}_m)$. It is easily verified that, contidional on σ and σ_0, the posterior means of the μ_i's are given by

$$E[\mu_i \,|\, \boldsymbol{x}_1, \ldots \boldsymbol{x}_m, \sigma, \sigma_0] = \omega_i \overline{x}_i + (1 - \omega_i)\overline{x},$$

with

$$\overline{x} = \frac{\sum_{i=1}^m \omega_i \overline{x}_i}{\sum_{i=1}^m \omega_i}, \qquad \omega_i = \frac{n_i \sigma_0^2}{n_i \sigma_0^2 + \sigma^2}\,.$$

This shows that there is a *shrinkage* from the sample mean within the i-th group, $\overline{x}_i$, towards the overall weighted mean, $\overline{x}$, which reflects the fact that the exchangeability assumption about the μ_i's makes *all* data relevant to draw conclusions about each of them, thus providing a very simple example of information integration. For details, see Berger and Bernardo (1992b).

Hierarchical modelling provides a powerful and flexible approach to the representation of assumptions about observables in complex data structures. This section merely provides a brief sketch to the basic ideas. A comprehensive discussion of hierarchical models requires a dedicated monograph.

5. DISCUSSION

The representation theorems are mainly due to de Finetti (1930, 1970/1974), Hewitt and Savage (1955) and Diaconis and Freedman (1984, 1987); for a recent review at a textbook level see Bernardo and Smith (1994, Ch. 4). The detailed mathematics of the representation theorems are involved, but their main message is very clear: if a sequence of observations is judged to be exchangeable, then any subset of them must be regarded as a random sample from some model, *and* there exist a prior distribution on the parameter of such model, hence requiring a *Bayesian* approach.

A simple hierarchical model is of the form

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \,|\, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m) = \prod_{i=1}^m p(\boldsymbol{x}_i \,|\, \boldsymbol{\theta}_i),$$

$$p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m \,|\, \boldsymbol{\phi}) = \prod_{i=1}^m p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\phi}),$$

$$p(\boldsymbol{\phi}),$$

which is to be interpreted as follows. Sequences of observables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are available from m different, but related sources: for example m clinical trial centres involved in the same study. The first stage of the hierarchy specifies the parametric model of each of the m sequences. Since the sequences are 'similar', the parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ are themselves judged to be exchangeable; the second and third stages of the hierarchy thus provides a prior of the form

$$p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m) = \int_\Phi \prod_{i=1}^m p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\phi}) p(\boldsymbol{\phi}) \, d\boldsymbol{\phi},$$

where the hyperparameter $\boldsymbol{\phi}$ has typically an interpretation in terms of the characteristics — for example mean and variance— of the population (for example trial centers) from which the m data sequences are drawn. Very often, no reliable prior information is available on the hypermarameter ϕ, in which case, a *reference* prior $\pi(\phi)$ may be used; for details, see Bernardo (1979), Berger and Bernardo (1992a) or Bernardo and Smith (1994, Sec. 5.4).

The information about both the unit characteristics, $\boldsymbol{\theta}_i$ and the population characteristics $\boldsymbol{\phi}$ after the data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ have been collected is described by their corresponding posterior densities, which may be obtained, from standard probability arguments involving Bayes' theorem, as

$$p(\boldsymbol{\theta}_i \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \int_\Phi p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\phi}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) p(\boldsymbol{\phi} \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)\, d\boldsymbol{\phi},$$

$$p(\boldsymbol{\phi} \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \propto p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \,|\, \boldsymbol{\phi}) p(\boldsymbol{\phi}),$$

where

$$p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\phi}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \propto p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \,|\, \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\phi}),$$

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \,|\, \boldsymbol{\phi}) = \int_\Theta p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \,|\, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m) p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m \,|\, \boldsymbol{\phi})\, d\boldsymbol{\theta}_1, \ldots, d\boldsymbol{\theta}_m.$$

Naturally, actual implementation requires the evaluation of the appropriate integrals, and this is not necessarily trivial.

Some key references on hierarchical models are Good (1965, 1980), Lindley (1971), Lindley and Smith (1972), Smith (1973), Mouchart and Simar (1980), Goel and DeGroot (1981), Berger and Bernardo (1992), George *et al.* (1994) and Morris and Christiansen (1996).

The term 'meta-analysis' is often used in the education, medicine and psychology literature to describe the practice of combining results from similar independent experiments. Hierarchical models provide the appropriate tool for understanding and generalizing those analysis. Some key references within this specialized topic are Cochran (1954), Edwards *et al.* (1963), Hodges and Olkin (1985), DerSimonian and Laird (1986), Berlin *et al.* (1989), Goodman (1989), DuMouchel (1990), Wachter and Straf (1990), Cook *et al.* (1992), Morris and Normand (1992), Wolpert and Warren-Hicks (1992), Cooper and Hedges (1994) and Petitti (1994).

REFERENCES

Berger, J. O. and Bernardo, J. M. (1992a). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35–60 (with discussion).

Berger, J. O. and Bernardo, J. M. (1992b). Reference priors in a variance components problem. *Bayesian Analysis in Statistics and Econometrics* (P. K. Goel and N. S. Iyengar, eds.). Berlin: Springer, 323–340.

Berlin, J. A., Laird, N. M., Sacks, H. S. and Chalmers, T. C. (1989). A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine* **8**, 141–151.

Bernardo, J. M. (1979b). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion).

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.

Berry, D. A., Thor, A., Cirrincione, C., Edgerton, S., Muss, H., Marks, J., Lui, E., Wood, W., Budman, D., Perloff, M., Peters, W. and Henderson, I. C. (1996). Scientific inference and predictions; multiplicities and convincing stories: a case study in breast cancer therapy. *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 45–67, (with discussion).

Cochran, W, G. (1954). The combination of estimates from different experiments. *Biometrics* **10**, 101–129.

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V. Light, R. J., Jouis, T. A. and Mosteller, F. (1992). *Meta-Analysis for Explanation: A Casebook*. New York: Russel Sage Foundation.

Cooper, H. and Hedges, L. V. (eds.) (1994). *The Handbook of Research Synthesis*. New York: Russel Sage Foundation.

de Finetti, B. (1930). Funzione caratteristica di un fenomeno aleatorio. *Mem. Acad. Naz. Lincei* **4**, 86–133.

de Finetti, B. (1970/1974). *Teoria delle Probabilità* **1**. Turin: Einaudi. English translation as *Theory of Probability* **1** in 1974, Chichester: Wiley.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.

Diaconis, P. and Freedman, D. (1984). Partial exchangeability and sufficiency. *Statistics: Applications and New Directions* (J. K. Ghosh and J. Roy, eds.). Calcutta: Indian Statist. Institute, 205–236.

Diaconis, P. and Freedman, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré* **23**, 397–423.

DuMouchel, W. H. (1990). Bayesian Meta-Analysis. *Statistical Methodology in the Pharmaceutical Sciences*. (D. A. Berry, ed.). New York: Marcel Dekker, 509–529.

Edwards, W., Lindman, H. and Sacage, L. J. (1963). Bayesian statistical inference for psychological research.*Psychological Rev*. **70**, 193–242.

George, E. I., Makov, U. E. and Smith, A. F. M. (1994). Bayesian hierarchical analysis for exponential families via Markov chain Monte Carlo. *Aspects of Uncertainty: a Tribute to D. V. Lindley* (P. R. Freeman and A. F. M. Smith, eds.). Chichester: Wiley, 181–199.

Goodman, S. N. (1989). Meta-analysis and evidence. *Controlled Clinical Trials* **10**, 188–204.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470–501.

Hodges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-analysis*. New York: Academic Press.

Goel, P. K. and DeGroot, M. H. (1981). Information about hyperparameters in hierarchical models. *J. Amer. Statist. Assoc.* **76**, 140–147.

Good, I. J. (1965). *The Estimation of Probabilities. An Essay on Modern Bayesian Methods*. Cambridge, Mass: The MIT Press.

Good, I. J. (1980). Some history of the hierarchical Bayesian methodology. *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Valencia: University Press, 489–519.

Lindley, D. V. (1971). The estimation of many parameters. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Toronto: Holt, Rinehart and Winston, 435–453 (with discussion).

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc. B* **34**, 1–41 (with discussion).

Morris, C. N. and Christiansen, C. L. (1996). Hierarchical models for ranking and for identifying extremes, with applications. *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 277–298, (with discussion).

Morris, C. N. and Normand, S. L. (1992). Hierarchical models for combining information and for meta-analysis. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 321–344, (with discussion).

Mouchart, M. and Simar, L. (1980). Least squares approximation in Bayesian analysis. *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Valencia: University Press, 207–222 and 237–245 (with discussion).

Petitti, D. B. (1994). *Meta-Analysis, Decision analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*. Oxford: University Press

Smith, A. F. M. (1973). A general Bayesian linear model. *J. Roy. Statist. Soc. B* **35**, 67–75.

Wachter, K. W. and Straf, M. L. (eds.) (1990) *The Future of Meta-Analysis*. New York: Russel Sage Foundation.

Wolpert, R. L. and Warren-Hicks, W. J. (1992). Bayesian hierarchical logistic models for combining field and laboratory survival data. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 522–546, (with discussion).