ADJUSTED AND SYNTHETIC LIKELIHOODS
FOR COMBINING EMPIRICAL EVIDENCE

Robert Wolpert
ISDS, Duke University

DP# 91-A14
Adjusted and Synthetic Likelihoods
For Combining Empirical Evidence

by Robert L. Wolpert

Several methods are introduced and illustrated for combining evidence from clinical series, controlled trials, expert opinion and other sources within the Bayesian paradigm. One of these methods, the adjustment of likelihood functions, is described in detail with examples and special cases.

1. INTRODUCTION

The hazards of naively combining evidence from multiple sources are well known. In the case of medical trials, variations in patient attributes and treatment regimens will threaten the validity of methods which simply pool the data and will threaten the power of methods which are insensitive to the probability distributions governing the observed outcomes in the individual trials. A number of authors have used methods (often referred to collectively as meta-analysis) intended to permit evidence to be combined, usually in order to construct estimates and confidence intervals for the overall “effect size” of some treatment.

More recently Bayesian methods have been brought to bear on the problem of combining evidence by DuMouchel and Harris (1983), Eddy (1985, 1986), Wolpert (1986), and others. In developing his Confidence Profile approach to Bayesian meta-analysis Eddy introduced the idea of “adjusting” likelihood functions to accommodate certain variations in patient attributes, treatment protocol, and provider skill and experience (which he terms biases). In Eddy’s approach the statistician determines “what result would have been delivered” by an experiment in the absence of these variations and then replaces the observed likelihood function with one adjusted to reflect what the statistician believes an unbiased experiment would have yielded. The present approach differs from Eddy’s by introducing any needed adjustments through the choice of a joint prior probability distribution for the unknown quantities of interest and for the parameters governing the experimental observables’ probability distributions.

I would like to thank David Eddy, Vic Hasselblad, Greg Critchfield, Jim Berger, and Raphie Adar for many fruitful conversations and helpful suggestions.
1.1 A Need For Adjustment

When evidence about some parameter of interest is available from two or more observational studies, and when none of the studies is fully compelling on its own (or when some are in apparent disagreement), there is a need to combine evidence from the trials. If we can calculate likelihood functions $L_i(\theta \mid X_i)$ reflecting the evidence bearing on the unknown parameter $\theta$ obtained by observing the random variable $X_i$ in the $i^{th}$ of some set $I$ of independent experiments then the synthesis of all the evidence can be captured in the joint likelihood function

$$
L_I(\theta \mid X_I) = \prod_{i \in I} L_i(\theta \mid X_i). 
$$

(1)

A Bayesian statistician might view this as a reflection of Bayes' formula expressing the posterior probability distribution for $\theta$ after observing the $i^{th}$ trial as the normalized product

$$
\pi_I(d\theta \mid X_i) = \frac{L_i(\theta \mid X_i)\pi(d\theta)}{\int L_i(\theta \mid X_i)\pi(d\theta')} 
$$

(2)

of the $i^{th}$ likelihood function and the prior probability distribution $\pi(d\theta)$; if $\pi_i$ is used as the prior distribution for a subsequent observation of $X_j$ the resulting posterior distribution will be $\pi_{i,j}(d\theta \mid X_i, X_j) \propto L_i(\theta \mid X_i)L_j(\theta \mid X_j)\pi(d\theta)$. The result of applying (2) successively for each $i \in I$ is

$$
\pi_I(d\theta \mid X_i: i \in I) = c\pi(d\theta)\prod_{i \in I} L_i(\theta \mid X_i), 
$$

(3)

where $c$ is defined implicitly by the requirement that $1 = \int \pi_I(d\theta \mid X_i: i \in I)$. This posterior distribution depends on the data only through the joint likelihood function $L_I$ given by (1).

1.1.1 The Binomial Example: Clinical Series

For example, if $\theta = p$ represents the success probability in some experimental medical procedure, and if we have evidence available from $i$ independent clinical series (the $i^{th}$ of which consists of the pair $X_i = (s_i, f_i)$ of success and failure counts among $n_i = s_i + f_i$ experimental subjects), then we can write the individual likelihood functions as

$$
L_i(p \mid X_i) = c_i 1_{[0,1]}(p)p^{s_i}(1-p)^{f_i}. 
$$

(4)

Here $c_i$ represents the combinatorial factor $\binom{n_i}{s_i} = \frac{n_i!}{s_i!f_i!}$ and $1_A(x)$ denotes the indicator function of the set $A$, equal to one if $X \in A$ and zero otherwise. Since $c_i$ does not depend upon $p$, we can renormalize the likelihood function by setting $c_i = 1$ (and omit $c_i$ from the notation) without affecting the evidence about $p$.

The result of using (1) and (4) to compute the likelihood function for $p$ is

$$
L_I(p \mid X_I) = \prod_{i \in I} L_i(p \mid X_i) = 1_{[0,1]}(p)p^{\sum s_i}(1-p)^{\sum f_i},
$$

(5)
exactly the same as if \( S = \sum_{i \in T} s_i \) successes and \( F = \sum_{i \in T} f_i \) failures had been observed in one grand experiment.

In many cases, this is just not reasonable—the evidence from the separate trials is not identical with the evidence from one grand trial with the same total number of successes and failures. The several trials almost always differ in important respects, some recognized and some not, some reported and some not, which affect the distribution of observable quantities \( X_i \) from that trial and hence affect the evidence about \( p \) represented by the observation of \( X_i \).

For example, often the details of an experimental medical procedure vary from one independent clinical trial to another. The skill and experience of the practitioners will vary, as will the demographic and medical characteristics of the experimental subjects in the trials. Adjuvant therapy may vary from one study to another, as may the length of follow-up observation. All of these will affect the probability of recorded “successes” within the trials. The effect of ignoring this variation is systematically to underrepresent uncertainty in the posterior distribution.

### 1.1.2 Randomized Controlled Trials

A related example arises when each experiment yields evidence about a pair of binomial probabilities \( \theta = (p^T, p^C) \), possibly the success probabilities in the treated and control groups of a randomized controlled trial with reported success and failure counts \( X_i = (s_i^T, f_i^T, s_i^C, f_i^C) \). Here the “quantity of interest” is usually not the vector \( \theta = (p^T, p^C) \) but rather some measure of the treatment effect \( \epsilon \) indicating how \( p^T \) differs from \( p^C \); we will return to such measures (which we call binomial contrasts) below.

For any joint prior probability distribution \( \pi(dp^T, dp^C) \) we can calculate the posterior distribution for \( p^T \) and \( p^C \) from (1) and (3) as

\[
\pi_I(dp^T, dp^C|X_1, \ldots, X_I) = c \prod_{i \in T} L_i(p^T, p^C|X_i) \pi(dp^T, dp^C)
\]

\[
= c (p^T)^{s_i^T} (1 - p^T)^{f_i^T} (p^C)^{s_i^C} (1 - p^C)^{f_i^C} \pi(dp^T, dp^C) \tag{6}
\]

where \( S^T = \sum s_i^T, F^T = \sum f_i^T, S^C = \sum s_i^C, \) and \( F^C = \sum f_i^C \). Again it is the same as if all the data had come from one grand randomized controlled trial.

This is even more unreasonable for controlled trials than for clinical series. Again the populations and protocols will be expected to differ in significant ways, and both the treated groups’ and untreated groups’ success probabilities must be expected to vary among the several trials.
1.2 Compound Models and Adjustment of Likelihood Functions

One criticism of the argument leading to equations (1) and, in the examples, to (5) and (6) is that it is unreasonable to expect all the pieces of evidence to bear directly on the same parameter \( \theta \). If we denote by \( \theta \) any quantity of interest to the experimenter, it is unusual to have even one of the several experiments give direct evidence bearing exactly on \( \theta \). For example, if \( \theta \) represents the success probability in a hypothetical binomial experiment assessing a carefully prescribed medical procedure, performed under precisely controlled conditions on a population with specified patient attributes by providers with specified levels of experience and skill, then it is hard to justify taking at face value the evidence \( L_i(\theta | X_i) \) from any one trial, and even harder to justify the assumption which underlies equation (1): that every one of the experiments gives independent evidence about exactly the same parameter \( \theta \).

It is more plausible (but still bears checking) that the different experiments offer independent evidence about different parameters \( \{\theta_i\}_{i \in I} \), each related in some way to \( \theta \). We call a model parameterized by a parameter of interest \( \theta \) and also a family \( \{\theta_i\}_{i \in I} \) of parameters, one for each experiment, a compound model, and use the notation \( \theta_I \) to represent the composite \( \{\theta_i\}_{i \in I} \).

For example, if the patient selection criteria, experimental protocol, etc. of each of several clinical series allow the outcomes for individual subjects within each to be regarded as independent Bernoulli events, then the \( i \)th trial does give evidence about a binomial parameter \( \theta_i \)—but the \( \{\theta_i\}_{i \in I} \) are not all equal to \( \theta \).

In general the evidence from each trial about its parameter \( \theta_i \) can again be captured in a likelihood function \( L_i(\theta_i | X_i) \), but in order make inference about the parameter of interest \( \theta \) we must formalize the notion that \( \theta_i \) is "related in some way to \( \theta \)." The Bayesian paradigm suggests that we do this by specifying a joint prior probability distribution \( \pi(d\theta, d\theta_I) \) for \( \theta \) and all the \( \{\theta_i\}_{i \in I} \). With such a joint prior probability distribution the evidence from the several trials can be combined to give evidence about \( \theta \) by calculating the joint posterior distribution

\[
\pi(d\theta, d\theta_I | X_i: i \in I) = c \left[ \prod_{i \in I} L_i(\theta_i | X_i) \right] \pi(d\theta, d\theta_I) \tag{7a}
\]

and from it the posterior marginal distribution for \( \theta \):

\[
\pi(d\theta | X_i: i \in I) = c \int_{\Theta_I} \left[ \prod_{i \in I} L_i(\theta_i | X_i) \right] \pi(d\theta, d\theta_I). \tag{7b}
\]

The joint prior probability distribution plays a crucial role in compound models as the link through which evidence about \( \theta_i \) in the \( i \)th trial lends evidence about \( \theta \); the methods of adjustment discussed in the remainder of this paper all center on different ways of generating appropriate joint prior probability distributions.

If \( \theta \) reflects all unknown features common to the sampling distributions of the trials, then the \( \{\theta_i\}_{i \in I} \) may be taken to be conditionally independent given \( \theta \) and hence the
joint prior distribution can be factored into a marginal distribution for $\theta$ times a product of conditional distributions for each $\theta_i$:

$$
\pi(d\theta, d\theta_I) = \pi(d\theta) \prod_{i \in I} \pi_i(d\theta_i | \theta).
$$

This allows the multiple integral in equation (7b) to be factored into a product of lower-dimensional integrals (a considerable numerical advantage), representing the marginal posterior distribution for $\theta$ as:

$$
\pi(d\theta \mid X_i: \ i \in I) = c \pi(d\theta) \prod_{i \in I} \left[ \int L_i(\theta_i \mid X_i) \pi(d\theta_i | \theta) \right]. \tag{8}
$$

Because of the similarity between equations (8) and (3), we call the term in square brackets the \underline{adjusted likelihood function}

$$
L_i^{\text{adj}}(\theta \mid X_i) = \int L_i(\theta_i \mid X_i) \pi_i(d\theta_i | \theta). \tag{9}
$$

The modeling problem for the Bayesian statistician is to

(i) Choose a quantity of interest $\theta$ (say, the parameter that would govern a paradigm trial with specified patient attributes, practitioner skills, etc.);

(ii) Specify the marginal prior distribution $\pi(d\theta)$ for $\theta$;

(iii) For each $i$ specify a conditional distribution $\pi_i(d\theta_i | \theta)$ reflecting any departures from the paradigm conditions known or suspected in the $i^{th}$ trial;

(iv) Compute the marginal posterior distribution for $\theta$ via equation (8).

Notice that one factor comprising the posterior distribution (8) is the prior probability distribution $\pi(d\theta)$; this suggests that we interpret the remainder as a marginal likelihood function for $\theta$,

$$
L_I(\theta \mid X_I) = \prod_{i \in I} \int L_i(\theta_i \mid X_i) \pi_i(d\theta_i | \theta) \tag{10}
= \prod_{i \in I} L_i^{\text{adj}}(\theta \mid X_i),
$$

from which the posterior distribution for $\theta$ can be recovered for any marginal prior for $\theta$ by the relation

$$
\pi(d\theta \mid X_i: i \in I) = c \pi(d\theta) L_I(\theta \mid X_I). \tag{11}
$$

Because the task (i) of choosing a parameter of interest depends so much on the circumstances of the analysis and the needs of the experimenter it will not be considered further here, except in the special context of selecting binomial contrasts in marginal models below. The introduction of an adjusted likelihood function in equation (9) above allows us to postpone the selection (ii) of a marginal prior probability distribution for $\theta$ until the last step of an analysis, and the task (iv) of calculating a posterior distribution
is solved by (8) or (11). The only remaining task is (iii), that of specifying the conditional distribution $\pi_i(d\theta_i|\theta)$ for $\theta_i$ given $\theta$. We now consider several ways of generating such conditional distributions.

1.2.1 Conditionally Certain Models: Adjusting the Evidence

In many interesting applications the conditional prior distributions for $\theta_i$ given $\theta$ are not absolutely continuous, i.e. do not have well-behaved distributions. For example, if the conditional distribution of each $\theta_i$ given $\theta$ is certain and, say, equal to $\theta$ (so that the conditional distribution would be the point mass (or Dirac delta function) $\pi_i(d\theta_i|\theta) = \delta(\theta_i - \theta) d\theta_i$), then the integrals appearing in (8) are trivial and (9) reduces to the unadjusted likelihood function $L_i(\theta|X_i)$ while (10) reduces to (1); heuristically this expresses a prior belief that each experiment gives evidence bearing exactly on $\theta_i = \theta$, and once again the effect of combining evidence from several trials is just like pooling the data. What made this conditional distribution degenerate is not specifically that $\theta_i$ was equal to $\theta$, but merely that it was conditionally certain given the value of $\theta$... i.e. that we express no doubt about how the circumstances of the $i^{th}$ study differ from those of the paradigm. This may be unrealistic in practice.

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.

$$\pi_i(d\theta_i|\theta) = \delta(\theta_i - \phi_i(\theta)) d\theta_i,$$

where the arbitrary function $\phi_i(\theta)$ represents the certain value of $\theta_i$ when the paradigm parameter value is $\theta$. The specification of the function $\phi_i(\theta)$ by the experimenter now determines the conditional prior distribution completely, and through (8) the adjusted likelihood

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

If the experimenter believes that the success probability in the $i^{th}$ trial should be only a fraction (say, half) of the success probability $\theta$ under paradigm conditions, perhaps because half of the subjects in the $i^{th}$ trial refused treatment, then he might use $\phi_i(\theta) = \frac{1}{2} \theta$ for that trial; of course, this choice implies that his joint prior probability distribution for the pair $(\theta, \theta_i)$ is concentrated on the one-dimensional set $\{((\theta, \theta_i) : 0 \leq \theta_i \leq 1/2, \theta = 2\theta_i\}$ and in particular that $\theta_i \leq 1/2$.

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 $\theta_i$

$$\pi_i(d\theta_i) = \int \pi_i(d\theta_i|\theta) \pi(d\theta)$$

$$= \left[ \int \delta(\theta_i - \phi_i(\theta)) \pi(d\theta) \right] d\theta_i,$$

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

### 1.2.2 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 \( \alpha_i \), leading to a degenerate conditional prior probability distribution

\[
\pi_i(d\theta_i|\theta) = \delta(\theta_i - \phi_i(\theta, \alpha_i))d\theta_i.
\] (13)

For example, we may allow a shift in the binomial probability parameter \( \theta \) by setting

\[
\phi_i(\theta, \alpha_i) = \theta + \alpha_i
\]
or a logistic shift by specifying \( \phi_i = e^{\alpha_i \theta} \), i.e.

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

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

\[
L_i^{\text{adj}}(\theta|X_i) = L_i(\phi_i(\theta, \alpha_i)|X_i).
\]

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

If the parameter $\alpha_i$ 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 $\theta_i$ given $\theta$ by averaging (13) over the possible values of $\alpha_i$:

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

Combined with (9), this relation yields a simple expression for the adjusted likelihood function:

$$
L_i^{\text{Adj}}(\theta|X_i) = \int L_i(\phi_i(\theta, \alpha_i)|X_i)\pi_i^\alpha(d\alpha_i|\theta). 
\quad (14)
$$

It is important not to use a noninformative prior for $\alpha_i$, since this ordinarily results in a constant likelihood function $L_i^{\text{Adj}}(\theta|X_i)$ lending no evidence whatsoever about $\theta$; 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.

1.3 Specific Examples of Adjustment

In this section we will examine certain specific examples of adjustments, each of the parametric or uncertain type described above, in order to illustrate the concepts. In several cases the parameter of interest will be the vector $\theta = (p^T, p^C)$ of success probabilities in a hypothetical randomized controlled trial conducted under paradigm conditions, and the evidence will consist of success and failure counts $X_i = (s_i^T, f_i^T, s_i^C, f_i^C)$ from a trial which differed in some respects from the paradigm. Each trial’s success and failure probabilities $\theta_i = (p_i^T, p_i^C)$ are to be related to $\theta$ through adjustment within a compound model.

1.3.1 Measurement Errors

If successes or failures are occasionally 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. If only the treated group is subject to this misreporting, with a probability $\alpha_i^{T+}$ of false positives (i.e. of a true failure being misreported as a success) and $\alpha_i^{T-}$ of false negatives, then the reported successes may include both true successes and true failures with total probability

$$
p_i^T = p^T(1 - \alpha_i^{T-}) + (1 - p^T)(\alpha_i^{T+}).
$$

If the control group is subject to misreporting as well, then

$$
p_i^C = p^C(1 - \alpha_i^{C-}) + (1 - p^C)(\alpha_i^{C+}).
$$

If the four error probabilities $\alpha_i = (\alpha_i^{T+}, \alpha_i^{T-}, \alpha_i^{C+}, \alpha_i^{C-})$ are known exactly (perhaps some are zero, or perhaps $\alpha_i^{T+} = \alpha_i^{C+}$ and $\alpha_i^{T-} = \alpha_i^{C-}$) then the parametric adjustment model with these known parameters would be appropriate; often they will
have to be estimated, perhaps subjectively by experts, leading to a prior distribution $\pi_i^\alpha(d\alpha_1|\theta)$ to be used in equation (14) above and to the adjusted likelihood function:

$$L_i^{\pm}(p^T, p^C|X_i) =$$

$$\int \int \int \int [p^T(1 - \alpha_i^{T-}) + (1-p^T)(\alpha_i^{T+})]^s_i^T [1 - p^T(1 - \alpha_i^{T-}) - (1-p^T)(\alpha_i^{T+})]^i_i^T$$

$$\times [p^C(1 - \alpha_i^{C-}) + (1-p^C)(\alpha_i^{C+})]^s_i^C [1 - p^C(1 - \alpha_i^{C-}) - (1-p^C)(\alpha_i^{C+})]^i_i^C \pi_i^\alpha(d\alpha_i|\theta).$$

1.3.2 Protocol Departures

In randomized controlled trials of readily-available drugs or treatments it sometimes happens that one or more subjects in the control group will seek and find treatment outside the study, possibly without the knowledge of the experimenters. This contamination of the control group must be accounted for in interpreting reported experimental results. A similar and more common problem arises when one or more subjects designated as treated in fact refuses treatment, diluting the treated group.

Several alternatives are available for adjusting experimental evidence to diminish the effects of contamination and dilution. The experimenter may wish to exclude the reported successes and failures of control group subjects who find treatment, or may prefer to include those successes and failures (perhaps following adjustment of some sort) with those of the treated group. Similarly the successes and failures of those who are offered treatment but refuse it may be excluded or adjusted or simply included with those of the controls.

If the protocol departures are observed, i.e. if it is known which subjects received treatment outside the experimental protocol and which ones refused treatment, the experimenter can correct the success and failure counts to eliminate the effects of contamination and dilution. If the protocol departures are unobserved, then their frequency must be estimated in order to make an adjustment.

Denote by $\alpha_i^T$ the probability that a subject offered treatment will refuse it (causing dilution), and by $q_i^T$ the probability of success for such a subject. Denote by $\alpha_i^C$ the probability that a subject in the control group will nevertheless find treatment outside the study (introducing contamination), and by $q_i^C$ the probability of success for such a subject. Then the probability of a reported success in the ostensibly treated group is

$$p_i^T = (1 - \alpha_i^T)p^T + (\alpha_i^T)q_i^T$$

while that in the designated control group is

$$p_i^C = (1 - \alpha_i^C)p^C + (\alpha_i^C)q_i^C.$$
conditional prior probability distribution \( \pi(dq^T_i, dq^C_i|p^T_i, p^C_i) \). In the simplest case, the experimenter might use a conditionally certain distribution for \( q^T_i \) and \( q^C_i \), concentrated on the points \( q^C_i = p^T \) (reflecting a belief that those who find treatment outside the experimental protocol will respond to it just as those in the treated group do) and \( q^T_i = p^C \) (treating those who refuse treatment as if they had been assigned to the control group), leading to the adjusted likelihood function

\[
L_{i}^{\text{Adj}}(p^T_i, p^C_i|X_i) = \int \int [(1 - \alpha_i^T)p^T + (\alpha_i^T)p^C]^s_i [(1 - (1 - \alpha_i^T)p^T - (\alpha_i^T)p^C]^f_i \times \\
[(1 - \alpha_i^C)p^C + (\alpha_i^C)p^T]^s_i [(1 - (1 - \alpha_i^C)p^C - (\alpha_i^C)p^T]^f_i \times \\
\pi^a(dx_i^T, dx_i^C).
\]

The appropriate prior probability distribution \( \pi^a(dx_i^T, dx_i^C) \) for the probabilities \( (\alpha_i^T, \alpha_i^C) \) of protocol departure must usually be selected subjectively.

### 1.3.3 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 \( \theta \). 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 \( f_i \) non-survivors 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 randomized controlled trials in which the length of follow-up differs between the control and treated groups. The simple constant hazard model suggests again that if \( \theta_i = (p_i^T, p_i^C) \) represents the vector of success probabilities in the \( i \)th trial and if \( \theta = (H^T, H^C) \) represents the vector of hazards in the treated and control groups, then the appropriate adjustment is

\[
p_i^T = e^{-H^T T_i^T} \quad p_i^C = e^{-H^C T_i^C}
\]

or, for five-year survival \( \theta = (p^T, p^C) \),

\[
p_i^T = (p^T)^{T_i^T/5} \quad p_i^C = (p^C)^{T_i^C/5}.
\]
For simple binomial evidence consisting of counts \( s_i^T, f_i^T \) of survivors and nonsurvivors in the treated group, and \( s_i^C, f_i^C \) in the control group, the adjusted likelihood function for \( p_i^T \) and \( p_i^C \) in this latter case is

\[
L_{i,\text{Adj}}^\text{Adj}(p^T, p^C|X_i) = \left( \frac{(p^T)^{T_i^T/5}}{s_i^T} \right)^{s_i^T} \left( \frac{(1 - (p^T)^{T_i^T/5})}{f_i^T} \right)^{f_i^T} \left( \frac{(p^C)^{C_i^C/5}}{s_i^C} \right)^{s_i^C} \left( \frac{(1 - (p^C)^{T_i^C/5})}{f_i^C} \right)^{f_i^C}.
\]

### 1.3.4 Selection Bias

Often controlled trials are undertaken in order to discover whether or not some treatment is effective, and to estimate some “measure of treatment effect” quantifying how the experiences of those in a treated group differed from those of control group members. Examples of treatment effect measures include the difference in systolic blood pressure before and after administration of a drug, the difference in life expectancy (or probability of five-year survival) between treated and control groups in a medical trial, or the increase in the number of drink-free days in the first year following a treatment for alcoholism over that for a control group.

In each case it is possible to construct some measure of treatment effect \( \varepsilon \) with the property that \( \varepsilon > 0 \) if the treatment is helpful, \( \varepsilon = 0 \) if the treatment is ineffective, and \( \varepsilon < 0 \) if the treatment is harmful. Choose such an effect measure \( \varepsilon \) and suppose that we have available evidence about the treatment effect \( \varepsilon_i \) from each of \( I \) independent trials in the form of likelihood functions

\[
L_i(\varepsilon_i|X_i).
\]

If subjects in a controlled trial are assigned to treatments without randomization it is possible that the subjects assigned to the control group will differ systematically from those assigned to the treated group, lending evidence that \( \varepsilon_i \approx \beta_i \) for some \( \beta_i \neq 0 \) even in the absence of a true treatment effect. In the presence of such a selection bias it can be misleading to interpret the trial results at face value without making adjustments.

In the simplest adjustment model the differences \( \beta_i = \varepsilon_i - \varepsilon \) would be taken as conditionally independent given \( \varepsilon \), 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

\[
L_{i,\text{Adj}}^\text{Adj}(\varepsilon|X_i) = \int L_i(\varepsilon + \beta_i|X_i)\pi_i^\beta(d\beta_i|\varepsilon).
\]
1.3.5 Treatment Intensity

For some sorts of treatments it makes sense to quantify the intensity of treatment, or at least the relative intensity when two or more versions of the treatment 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 treatment 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 $\tau_i$ the treatment intensity in the $i^{th}$ trial, with unit intensity $\tau_i = 1$ for a treatment identical in intensity to that of the paradigm.

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

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

for a trial with intensity $\tau_i$ governed by the prior (conditional) probability distribution $\pi_i^{\tau_i}(d\tau_i | \epsilon)$. Adjustments for intensity and bias can be combined in the form

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

to yield a generalization of the “intensity and additive bias adjustment” of Eddy (1986).

1.4 Alternatives to Adjustment

1.4.1 Marginal Models

When the parameter governing each trial $\theta_i$ is a vector with components $\{\theta_i^j\}_{1 \leq j \leq J}$ it is often not the whole parameter $\theta = \{\theta_i^j\}_{1 \leq j \leq J}$ governing a hypothetical trial under paradigm conditions which is of interest but only some lower-dimensional (often one-dimensional) quantity $\epsilon = \psi(\theta)$. Even though the quantity $\theta_i$ may be expected to vary from one study to the next, it may be possible to choose an interesting quantity $\epsilon_i = \psi(\theta_i)$ which does not vary appreciably from one study to the next. If we can “marginalize” the evidence from each study to find its bearing on $\epsilon_i = \epsilon$, and represent that evidence in the form of a likelihood function $L_i(\epsilon|X_i)$, then under certain conditions the evidence can be combined just as in (1).

For example, in the case of randomized controlled trials with success probabilities $\theta_i = (p_{i}^{T}, p_{i}^{C})$, we may be able to choose as a measure of effect a binomial contrast $\epsilon = \psi(p_{i}^{T}, p_{i}^{C})$ with the property that, although the treated and control group success probabilities $p_{i}^{T}$ and $p_{i}^{C}$ might be expected to vary individually from trial to trial, their contrast $\epsilon$ would not. In some cases the appropriate measure of treatment effect is simply the difference $\epsilon = p^{T} - p^{C}$ between the two success probabilities, while in others it is better to calculate and report the log relative-risk $\epsilon = \log(\frac{1-p^{C}}{1-p^{T}})$ or the log odds-ratio $\epsilon = \log(\frac{1-p^{T}(1-p^{C})}{1-p^{C}(1-p^{T})})$. These and other choices are considered and compared in Wolpert
(1986), where conditions are given under which marginal likelihood functions \( L_i(\epsilon_i | X_i) \) can be calculated giving independent evidence about the treatment effect in separate trials.

We can often calculate for each \( i \) a posterior distribution \( \pi_i(d \theta_i^T, p_i^C | X_i) \) and from that a marginal posterior distribution for \( \epsilon, \pi_i^*(d \epsilon | X_i) \). From this marginal posterior distribution a marginal likelihood function for \( \epsilon \) can be constructed, and the resulting likelihood functions \( L_i(\epsilon | X_i) \) can be combined as in equation (1). See Wolpert and Berger (1991) for details and for conditions under which the combination is justified.

### 1.4.2 Hierarchical Models

In the compound models considered above the evidence from each trial about its parameter \( \theta_i \) could be converted directly (or "adjusted") to give evidence about \( \theta \) via equation (10). The evidence from the several trials remained independent; in (10), just as in equation (1), the likelihood function for \( \theta \) is expressed as a product of \( I \) terms, one from each experiment.

In many problems the conditional distributions of \( \theta_i \) and \( \theta_j \) are not independent, even when conditioned on \( X_i, X_j \), and \( \theta \); for example, the presence of nuisance parameters that are common across experiments or of other uncertain features shared by more than one trial will lead to this dependence. In some cases we can introduce a hyperparameter \( \gamma \) \( \in \Gamma \) such that \( \{ \theta, \theta_i \} \) are conditionally independent given \( \gamma \), and write the joint distribution for \( \{ \theta, \theta_i \} \) in the form \( \pi(\theta, \theta_i) = \int_{\Gamma} \pi(d \theta | \gamma) \prod_{i \in I} \pi(d \theta_i | \gamma) \Pi(d \gamma) \) or, after applying Bayes’ formula to \( \gamma \) and \( \theta \) to write \( \pi(d \theta | \gamma) \Pi(d \gamma) = \pi(d \theta) \Pi(d \gamma | \theta) \), in the form \( \pi(\theta, \theta_i) = \pi(d \theta) \int_{\Gamma} \prod_{i \in I} \pi(d \theta_i | \gamma) \Pi(d \gamma | \theta) \). This leads to the following expression for the synthetic likelihood for \( \theta \):

\[
L_I(\theta | X_I) = c \int_{\Theta_i} \prod_{i \in I} \left[ \int_{\Theta_i} L_i(\theta_i | X_i) \pi(d \theta_i | \gamma) \right] \Pi(d \gamma | \theta).
\]

The product is just the adjusted likelihood for \( \gamma \), and \( L_I(\theta | X_I) \) is its (prior) conditional expectation, given \( \theta \).

The simplest examples are those from conjugate prior families. If \( \theta = p \) and \( \theta_i = p_i \) represent binomial success probabilities under paradigm conditions and those prevailing in the \( i^{th} \) trial, respectively, and if \( \{ p, p_i : i \in I \} \) are taken to be drawn independently from a common Beta probability distribution with uncertain parameters \( \alpha \) and \( \beta \), and if \( \Pi(d \alpha, d \beta) \) represents prior uncertainty about \( \alpha \) and \( \beta \), then the resulting likelihood function for \( p \) becomes

\[
L_I(p | X_I) = c^{-1}_{\alpha, \beta} \int [p^{\alpha-1} (1-p)^{\beta-1} \Gamma(\alpha + \beta)] \prod_{i \in I} \left[ \int \frac{p_i^{\alpha + \alpha_i - 1} (1-p_i)^{\beta + \beta_i - 1} dp_i} \Pi(d \alpha, d \beta) \right] \Pi(d \alpha, d \beta).
\]
\[ \prod_{i \in I} \left[ \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + s_i)\Gamma(\beta + f_i)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + n_i)} \right] \Pi(d\alpha, d\beta), \]

where

\[ c_{\alpha, \beta} = \int \int p^{\alpha-1}(1-p)^{\beta-1} \left( \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right) \Pi(d\alpha, d\beta). \]

Any hyperprior probability distribution \( \Pi(\alpha, \beta) \) may be chosen, including a non-informative one such as the Jeffreys prior

\[ \Pi(\alpha, \beta) = \sqrt{\psi'(\alpha)\psi'(\beta) - \psi'(\alpha + \beta)[\psi'(\alpha) + \psi'(\beta)]}; \]

\( \psi'(x) \) denotes the trigamma function \( \frac{d^2}{dx^2} \log \Gamma(x) \) (see Jeffreys (1961) for a discussion of noninformative priors, and Abramowitz and Stegun (1964) for the definition and properties of the trigamma function).

### 1.4.3 Regression Models

In the adjustment models the functions \( \theta_i = \phi_i(\theta) \) (or the parameters \( \alpha_i \) governing the adjustments \( \theta_i = \phi_i(\theta, \alpha_i) \)) must be supplied separately by the experimenter for each trial, i.e. for each \( i \). Often in practice the values of measured covariates are available from the several trials, quantities \( \{X_{ij}\}_{1 \leq j \leq i} \) which offer evidence about \( \theta_i \) and therefore \( \theta \). Although the experimenter can use these quantities informally as cues in selecting adjustment functions \( \phi_i(\theta) \) or parameters \( \alpha_i \) above, an appealing alternative is to try to reparameterize the distributions in such a way that a formal regression model can describe the conditional distribution of \( \theta_i \) given the observable quantities: for some unknown regression coefficients \( \beta_j \),

\[ \theta_i = \sum_{j=1}^{J} X_{ij}\beta_j + \zeta_i \tag{16} \]

where the \( \zeta_i \) are independent mean-zero random variables representing unexplainable variations whose distribution is the same for all \( i \). A typical reparametrization would be to set \( \theta_i \) equal to the logistic \( \log(p_i/1-p_i) \) for a binomial probability parameter \( p_i \), making (16) a multiple logistic regression model. Now information about the unknown parameters \( \theta_i \) (and in particular about any measure of treatment effect \( \epsilon \)) is represented in the vector \( \beta = \{\beta_j\} \), and each study offers additional evidence about \( \beta \). Bayesian linear regression methods can be used to complete the analysis by constructing a posterior probability distribution for \( \beta \) from empirical evidence and subjective prior information, and, from this, a marginal posterior probability distribution for the parameter of interest \( \epsilon \). For further development, examples and details, see Wolpert (1987).
2. REFERENCES


