11

Meta-analysis of Drug Safety Data

Gary G. Koch

University of North Carolina
Chapel Hill, North Carolina

Judith E. Schmid, Janet M. Begun, and William C. Maier

Pharmaceutical Product Development, Inc.
Morrisville, North Carolina

I. INTRODUCTION

In analyzing the safety data from phase I–IV clinical trials, a number of difficulties arise. Foremost among these is the matter of sample size. In many cases, an adverse drug experience of interest occurs rarely enough that it is not possible to test the association between the adverse event (AE) and the treatment drug within an individual study. One strategy to alleviate this is to use meta-analysis. In addition to the issue of sample size, the data structure may be complex, with time at risk and duration and frequency of adverse events varying among patients and across studies. Several methods of meta-analysis will be presented that deal with some of the complexities of safety data.

II. META-ANALYSIS OF SAFETY DATA

A. Issues and Advantages of Meta-analysis

Meta-analysis refers to a statistical analysis that combines the results of some collection of related studies to arrive at a single conclusion to the question at hand [1]. There are several advantages in using meta-analysis [2,3]. The increased sample size gained by including the data from several studies increases the power of the analysis to investigate questions that could not be adequately addressed in the individual studies. There may also be sufficient sample size for analysis of subpopulations. Qualitatively, by juxtaposing different studies, dif-

ferences in quality and design can be evaluated, and questions of generalizability of the result can be addressed.

There are several issues to be addressed before performing a meta-analysis. The most immediate of these is the decision as to which studies to include in the analysis [4-6]. Should studies be included that differ in scientific quality or in other factors such as design, population studied, type of data collected, or outcome measurement? The question of quality has to do with the believability of the results. Some authors will only include studies with certain characteristics such as randomized double-blind studies, while others include all studies regardless of quality. Differences in the other study factors raise questions regarding whether and how to combine studies that may measure slightly different outcomes. If only very similar studies are combined, the analysis is simpler and the result is clearer than if very diverse studies are combined [7]. When diverse studies are combined, the generalizability of the result is more evident; however, an effect that is present only under specific conditions may be obscured.

Most methods of meta-analysis assume that the individual studies are independent. This may not be the case, e.g., if investigators or patients are involved in more than one study [8]. The effect of variables such as investigator, center, and year of study should be examined and adjusted for as necessary.

In deciding how to combine the studies, it must first be decided how to weight the studies. In most cases, if quality of the studies is not an issue, studies are weighted in a way that reflects sample size. Larger studies are assumed to give a better estimate of the result under study and so are given more weight in the analysis. Another approach is to assume that each study gives an equally valid estimate and to weight all studies equally [7].

There has been much discussion of the merits of combining the results, rather than the data of the individual studies, and so adjusting for individual study differences [9]. For instance, one study may have a higher rate of occurrence of an adverse event but still show the same difference in occurrence rates across treatment groups as another study. In the analysis of rare events, however, the individual study results may be negligible, and analysis of the pooled data may be necessary.

B. Application to Drug Safety Data

Meta-analysis has been used in many fields, in each for slightly different reasons and with slightly different methods. In clinical trials, as analysis of safety data and efficacy data require different approaches, so meta-analysis differs in each case. Efficacy data can often be analyzed using a straightforward and traditional meta-analysis, while safety data often present additional difficulties. In both safety and efficacy meta-analyses, determining the presence of an association between a treatment and an outcome is generally the main focus of the analysis;

however, in the efficacy analysis, evaluation of the consistency of the individual studies in this association is also of major concern. In a safety analysis, the occurrence of the event in question may be infrequent enough that an evaluation of this type is not possible. If the event occurs more frequently, this analysis of homogeneity may be of interest [10-13].

A situation that is ideal for meta-analysis occurs when the studies to be combined are all very similar. If the studies all have similar design, treatment regimens, indications, populations, methods, and times for collecting data, duration, and early withdrawal rates, then the more traditional methods of meta-analysis, described in Section III.A, may be used. For many of these factors, the studies can be categorized into a few groups, or the factor can be considered a numerical variable (such as dose) to be determined for each patient or within each study. Then, either by using stratification or by adding factors to a model, the methods of Section III.A can still be used when the studies differ in one or more characteristics. The total sample size and the rate of occurrence of the adverse event may still limit the number of factors that the analyses can take into account.

For instance, if the studies vary in design, stratification or model factors can often be used to adjust for differences in outcome that are associated with the design. When some studies are multicenter studies, stratification by center or adding center as a predictor to the model will adjust for differences in outcome related to center. If it can be shown that center is not associated with outcome or if the sample size is too small to allow stratification by center, then the data may be pooled for the centers in each study. If some studies are unblinded and often blinded, then stratification for this factor should take place unless it can be shown that there is no association between blindedness of the study and the outcome.

If some studies are crossover studies, however, it is difficult to combine them with parallel studies. In any case, the data of crossover studies can be added to the analysis up to the point of the first crossover, but beyond this point it becomes more difficult to know which treatment the event should be classified under and how the time of exposure should be evaluated.

There are two situations that commonly occur in collections of safety studies that may require very different analysis strategies. First, the studies may vary in factors related to time [14]. If the studies have different treatment or follow-up times, or the patients of different studies have different withdrawal rates, or reports of safety problems are collected at different intervals, then an analysis that takes time into account is appropriate. Some methods which account for time are presented in Section III.B.

Second, the data structure may be complicated [15,16]. In many cases the occurrence of an adverse event or an abnormal laboratory finding can be thought of as dichotomous, i.e., for each patient it is either present or not. In some cases severity or duration of the event is of interest. The methods of Section III.A are

useful in analyzing dichotomous and ordinal outcomes (where an outcome, such as severity, can be classified into one of several discrete, ordered categories, such as absent, mild, moderate, or severe). Analysis of outcomes that are continuous numerical values are not discussed in this chapter, but methods can be found in the Refs. 9, 12, and 17. If a patient can experience an event more than once in a study, then the outcome can still be considered ordinal (the number of events per patient), and the methods of Section III.A can be used. If, however, the total number of events across all patients is the outcome of interest, then the events for each patient must be considered as correlated and the analysis adjusted accordingly. Severity or duration as well as multiple events for each patient may also be of interest. One analysis strategy for these two cases is discussed in Section III.C and other methods can be found in Refs. 18 and 19.

The methods of this chapter deal only with the univariate analysis of single adverse drug reactions. In reality, many adverse experiences are analyzed in each study and may be of interest in the meta-analysis. In many cases, an analysis is considered exploratory, and separate tests are performed for several different event outcomes, without adjusting for multiplicity of tests or considering possible correlations between events. Outcomes that show significant results are then subjected to further testing and examined in future studies. References that deal with multiple testing issues and multivariate analysis methods include Refs. 14–16, 20–22.

III. STATISTICAL METHODS FOR META-ANALYSIS

A. Basic Methods

The methods in this section are appropriate when the studies are similar in all aspects of design that can affect the outcome. If the studies differ in a factor that has only a few categories, the factor may be used as a stratification variable in addition to study. A categorical or numerical factor may be added as a predictor in a model.

1. Analysis of Data with a Dichotomous Outcome

In the simplest case, there are two treatment groups (test and reference) and the outcome of occurrence or nonoccurrence of the adverse event (AE) can be considered to be dichotomous. The data for each study can then be arranged in a 2×2 table (Table 1). In the table, $a + b + c + d = N$, the total number of patients in the study. The row and column totals of the 2×2 table ($a + b$, $c + d$, $a + c$, $b + d$) are considered to be fixed, the row totals by assignment to treatment group, and the column totals by occurrence of the event in the study population, under the hypothesis of no association with treatment.

If the occurrence of the adverse drug experience is rare, so that stratification by study is not practical, then the data from the individual studies may be

Table 1 2 × 2 Table

Drug	AE occurs	AE does not occur
Test	*a*	*b*
Reference	*c*	*d*

combined into one 2 × 2 table. (It may be of interest to combine the data in any case to obtain a first approximation to a result.) Fisher's exact test can be used to test for an association between the treatment and the occurrence of the adverse event. In Fisher's test, the exact probability is calculated for each table with the same row and column totals as exhibited by the data. The p value is calculated by summing the probabilities of all tables as likely or less likely to occur, under the null hypothesis of no association between treatment and AE occurrence. The probability of each table under the null hypothesis is:

$$\text{pr}\ (a|H_0) = [(a + b)\,!\,(c + d)\,!\,(a + c)\,!\,(b + d)\,!\,] / [N\,!\,a!b!c!d!]$$

Fisher's exact test is easily available from software products such as the SAS procedure FREQ [23] and StatXact [24].

If the occurrence rate and sample size are sufficiently large that within-study estimates of the association between the AE and the treatment groups can be calculated, the individual study results can be combined as in classical meta-analysis. The sign test or the signed rank test may be used to determine if the AE occurs more frequently in one treatment group, while weighting all studies equally [7].

The Mantel–Haenszel procedure also tests the AE treatment association and weights larger studies more heavily than smaller studies [3,25]. In the Mantel–Haenszel (M-H) test, the data from each study is arranged in a 2 × 2 table (as above). For each study the expected value of a, $E(a) = (a + b)(a + c)/N$, and the variance of a, $V(a) = (a + b)(c + d)(a + c)(b + d)/N^2(N - 1)$, are calculated. The M-H statistic, using the subscript h to indicate the hth study, is:

$$Q = \frac{\{\Sigma_h[a_h - E\ (a_h)]\}^2}{\Sigma_h V(a_h)}$$

which has a χ^2 distribution, with 1 df. This statistic may be calculated using the SAS procedure FREQ. The M-H test is more powerful when the direction of association between treatment and outcome is the same for most of the studies. For adverse events with moderately large occurrence rates, the homogeneity of the association of treatment with outcome should be examined.

A logistic regression model can also be used to determine if there is an association between the occurrence of the AE and the test drug group, while

taking into account study and any other variables that may be important. The model is specified by:

$$\pi_{hik} = \frac{1}{1 + e^{-x_{hik}\beta}} \quad \text{or} \quad \log \frac{\pi_{hik}}{1 - \pi_{hik}} = x'_{hik}\beta$$

where π_{hik} is the probability of occurrence of the AE for the kth patient in the ith treatment group in the hth study or group within a study, x_{hik} is the vector of predictor values for the patient, and β is the vector of parameters to be estimated for the model. Many software packages are available for logistic regression, including the SAS procedures CATMOD and LOGISTIC.

2. Analysis of Data with an Ordinal Outcome

If the outcome is not dichotomous but ordinal, e.g., if the AE is classified according to severity, then extensions of the M-H test and the logistic regression model may be used [26]. The data from each study can be arranged in a $2 \times r$ table as shown in Table 2. In the table, if a set of scores (a_1, a_2, a_3, a_4) is assigned to the outcome possibilities [e.g., the integer scores (1, 2, 3, 4); for other types of scoring, see Ref. 26], then the extended M-H test would look at the difference in mean scores between treatment groups. Using the subscript g to indicate the gth level of severity, the mean score for the individual study for the test treatment group is:

$$m_t = \frac{\Sigma_g a_g n_{tg}}{n_{t+}}$$

and for the reference group is:

$$m_r = \frac{\Sigma_g a_g n_{rg}}{n_{r+}}$$

Under the hypothesis of no association between treatment and AE severity,

$$E(m_t) = \frac{\Sigma_g a_g n_{+g}}{N}$$

Table 2 $2 \times r$ Table: Adverse Event Severity

Drug	Severe	Moderate	Mild	None	Totals
Test	n_{t1}	n_{t2}	n_{t3}	n_{t4}	n_{t+}
Reference	n_{r1}	n_{r2}	n_{r3}	n_{r4}	n_{r+}
Totals	n_{+1}	n_{+2}	n_{+3}	n_{+4}	N

and

$$V(m_t) = \frac{n_{r+}}{n_{t+}\,(N-1)} \sum_s \frac{n_{g+}}{n} [a_g - E(m_t)]^2$$

$$= \frac{n_{r+}}{n_{t+}\,(N-1)} V_a$$

If the data are to be pooled across centers without adjustment, then

$$Q = \frac{[m_t - E(m_t)]^2}{V(m_t)}$$

has a χ^2 distribution, with 1 df. The extended M-H χ^2,

$$Q = \frac{\{\sum_h [m_{ht} - E(m_{ht})]\}^2}{\sum_h [n_{ht} n_{hr+} / (N_h - 1)]\, V_{ha}}$$

Also with 1 df, tests the association between treatment and AE severity while adjusting for individual study differences.

Similar tests can be performed in the case where there are several ordinal treatment categories, such as levels of dose, as well as an ordinal outcome variable. Scores are assigned to the levels of treatment and to the levels of outcome and trends in the mean scores across treatment levels are examined. If the data are pooled into one contingency table, a correlation χ^2 statistic can be calculated. There is also an extended M-H test that incorporates stratification by study into the analysis.

An extension of logistic regression that can take into account an ordinal outcome variable is the proportional odds model [16,26]. This looks at the successive logits of more vs. less severe outcomes; in this case, within one study or group within a study, the logit for severe vs. less than severe AE $\{\log[\pi_1/(\pi_2 + \pi_3 + \pi_4)]\}$, the logit for moderate or severe vs. less than moderate AE $\{\log[(\pi_1 + \pi_2)/(\pi_3 + \pi_4)]\}$, and the logit for any AE vs. no AE $\{\log[(\pi_1 + \pi_2 + \pi_3)/\pi_4]\}$. The model assumes that each successive logit has a different intercept but the same vector of regression parameters, i.e., $\text{logit}(\psi_{ik}) = \alpha_k + x'_h\beta$, for the kth of the successive logits within the hth study or group within study. With β_t as the coefficient estimated by the model for test drug, $\exp(\beta_t)$ is the ratio of test drug to reference drug for the odds of each more severe to less severe AE occurrence category, as predicted by the model. The proportional odds model can be applied with the SAS procedure LOGISTIC.

In the analysis of safety data, it is usually assumed that there is a true rate of occurrence for an adverse event and that each study (or each subgroup within

each study) provides an estimate of this rate. Analyses with this assumption use a fixed effects approach to analysis. Alternatively, a random effects approach assumes that different populations may have different rates of occurrence and that there is not one true rate. In this case a mean rate is estimated. When the occurrence rate is fairly large, this may be of interest. If the studies to be combined demonstrate similar trends in the association between the treatments and the occurrence of the AE, then a fixed effects approach is adequate. If, however, the studies are not homogeneous in this way, a random effects approach may be more appropriate. One type of analysis that incorporates a random effects strategy is the use of survey regression methods that consider each study as a cluster of patients and adjust the estimate and its variance accordingly. Sources for the random effects approach include Refs. 10, 12, and 17.

B. Methods that Account for Time at Risk

If the studies have different durations, or if there is considerable variation between patients in the length of time on treatment or time of follow-up, then several strategies can be considered. The strategies discussed earlier did not take time into account, although time could be included as a stratification variable or a model factor. One approach that does account for time at risk is consideration of the incidence density as the rate of interest, rather than the proportion of patients experiencing the adverse event.

1. Incidence Density

The incidence density is the ratio of the total number of events to the total number of time units of risk (both sums across all patients) [27]. If the patients are considered in the analysis only to the time where the first occurrence of the AE takes place, then the observations can be considered independent. If the number of events occurring during each time interval is considered under the null hypothesis of no association with treatment to have a Poisson distribution with a rate of λ per unit of time, then within each study the expected number of AEs in the test drug treatment group is $E(n_t) = \lambda N_t = nN_t/N$, where n is the number of AEs for all treatments in the study, N the sum of time at risk for all patients in the study, and n_t and N_t are the number of AEs and amount of time at risk for the test drug group. The variance of $(n_t - E(n_t))$ is $nN_t N_r/N^2$, where N_r is the amount of time at risk for the reference group. The M-H statistic adjusting for study, with h indexing the hth study, is:

$$\frac{\{\Sigma_h[n_{ht} - E(n_{ht})]\}^2}{\Sigma_h V[n_{ht} - E(n_{ht})]}$$

and has a χ^2 distribution with 1 df.

Use of the incidence density assumes that the risk of an AE occurring remains constant over time. This may not be true [14]. Adverse events associated with

drug use occur in many different patterns, which may not be accurately accounted for by an incidence density. For instance, the incidence density underestimates the risk of occurrence of an AE that appears only after a long exposure to a drug. Life table rates can be used to calculate a cumulative incidence rate, while accounting for the changing number of patients at risk over time and estimating the risk of AE occurrence separately in each time interval.

2. Life Table Analysis

For each treatment group in each study, the data can be arranged in a life table (Table 3). The censored patients, those lost to follow-up or withdrawn from the study during the interval, are assumed to have left the study uniformly throughout the interval, and so are assigned a time at risk of $t/2$. It is also assumed that censoring is not associated with the effects of the drug and that the rate of AE occurrence for the censored patients for the remainder of the study would not have been different from the occurrence rate of those remaining in the study. Patients who experience an AE during the interval are sometimes assigned a time at risk of $t/2$, which assumes that the events occurred uniformly throughout the interval. Often, however, they are assigned a time of t, since the outcome for the interval is known.

The cumulative probability of completing the study throughout the sth time interval without experiencing the AE is the product of the probabilities of completing each of the intervals without the AE: $(1 - P) = \Pi_s (1 - p_s)$, where p_s is the probability of experiencing the event in the sth interval. The cumulative probability of experiencing the event is $P = 1 - \Pi_s (1 - p_s)$. If the number of patients is large throughout the study, Greenwood's formula for estimating the standard error can be applied [28]:

$$SE = (1 - P) \left(\Sigma_s \frac{n_{sa}}{T_s (T_s - n_{sa})} \right)^{1/2}$$

If the data from the individual studies have been collected at similar time intervals, one strategy would be to combine the data for each treatment within

Table 3 Life Table

Interval	Time in interval	No. patients at beginning of interval	No. with AE	No. lost or withdrawn	Person-time at risk	Probability of event
1	t_1	N_1	n_{1a}	n_{1b}	$T_1 = (N_1 - n_{1b}/2)t_1$	$p_1 = n_{1a}/T_1$
2	t_2	N_2	n_{2a}	n_{2b}	$T_2 = (N_2 - n_{2b}/2)t_2$	$p_2 = n_{2a}/T_2$
3	—	—	—	—	—	—
4	—	—	—	—	—	—

each time interval and compare the treatment-specific cumulative rates directly. A weighted average of the within-study difference scores could also be evaluated.

When the actual time of each adverse event is known the cumulative rate can be estimated by the Kaplan–Meier product limit method. In this case, a new time interval begins with the occurrence of each AE, so that each interval usually contains one occurrence. Intervals containing more than one event are possible when two or more AEs have the same actual time of occurrence.

The Mantel–Cox test (also called the Mantel–Haenszel test and the log-rank test) can be used to compare the cumulative AE rates of the treatments across all studies [16,28]. The data are rearranged as shown in Table 4. Each row can be viewed as a 2 × 2 table, and the M-H χ^2 statistic,

$$Q = \frac{\{\Sigma_s[n_{st1} - E(n_{st1})]\}^2}{\Sigma_s V(n_{st1})} \qquad \text{(with 1 df)}$$

can be used to test the difference between treatment groups in AE occurrence. Where the data from the individual studies are combined within each time interval, the sum in the M-H statistic is over all intervals. When the data are stratified by study, the sum is over all study by interval combinations. The Mantel–Cox test is most effective where there is a consistent trend in the difference between treatment groups throughout the time intervals. The SAS procedure LIFETEST will calculate the cumulative incidence rate and its standard error and the log-rank test.

3. Piecewise Exponential Model

A model that can be used which takes time at risk into account in the piecewise exponential model [16,26]. The model assumes a separate independent exponential distribution of AE occurrences within each time interval. Within each time interval, the incidence density within each subpopulation for AE occurrence is described with the structure $\lambda_{his} = \exp(x'_{his}\beta)$, where x'_{his} is a vector of predictor variables encompassing study, treatment, and time interval and β is a

Table 4 Arrangement of Data for Mantel–Cox Test

	Test		Reference	
Interval	No. with AE	No. without AE	No. with AE	No. without AE
1	n_{t1}	n_{t2}	n_{r1}	n_{r2}
2	n_{2t1}	n_{2t2}	n_{2r1}	n_{2r2}
3	—	—	—	—
4	—	—	—	—

vector of parameters to be estimated for the model. Since the likelihood function for the piecewise exponential model is proportional to the likelihood function from a Poisson regression, Poisson regression software such as procedure NLIN of SAS can be used to estimate the regression parameters. (The SAS procedure LOGISTIC can also be used.)

4. Analysis of Data with Multiple Occurrences of the Outcome

When each patient can experience the AE more than once, a ratio can be formed that is analogous to an incidence density. Let y_{hik} be the number of events for the kth patient in the ith treatment group in the hth study, N_{hik} the time of exposure of the patient, and n_{hi} the number of patients in the hth study and ith group. The weight given to the hth study is $w_h = [(n_{ht}\, n_{hr}\, /n_h)/\Sigma_h(n_{ht}\, n_{hr}\, /n_h)]$.

For each treatment group, the ratio can be formed:

$$R_i = \frac{f_i}{g_i} = \frac{\Sigma_h \Sigma_k\, (w_h\, y_{hik}\, /n_{hi})}{\Sigma_h \Sigma_k\, (w_h\, N_{hik}\, /n_{hi})}$$

where f_i is a weighted sum of adverse event occurrences and g_i is a weighted sum of time at risk. The variance of R_i is

$$V(R_i) = R_i^2 \left[\frac{V(f_i)}{f_i^2} - \frac{2\ \mathrm{Cov}\ (f_i,\, g_i)}{f_i g_i} + \frac{V(g_i)}{g_i^2} \right]$$

where $V(f_i) = \Sigma_h w_h^2\, (1/n_{hi}(n_{hi} - 1))\, \Sigma_k\, (y_{hik} - \bar{y}_{hi})^2$ with $\bar{y}_{hi} = \Sigma_k\, (y_{hik}\, /n_{hi})$;

$V(g_i) = \Sigma_h w_h^2 (1/n_{hi}(n_{hi} - 1))\, \Sigma_k (N_{hik} - \bar{N}_{hi})^2$ with $\bar{N}_{hi} = \Sigma_k\, (N_{hik}\, /n_{hi})$;

and $\mathrm{Cov}\ (f_i,\, g_i) = \Sigma_h w_h^2\, (1/n_{hi}\, (n_{hi} - 1))\, \Sigma_k\, (y_{hik} - \bar{y}_{hi})\, (N_{hik} - \bar{N}_{hi})$.

In a similar way, the ratio can be formed where y is a measure that incorporates number of times the event occurred, duration, and/or severity. For instance, y could be the number of events multiplied by their average severity or the sums of durations over which the events took place.

IV. EXAMPLES

The methods of this chapter will be illustrated with artificial data from five studies, each with approximately 500 patients. A summary of the data is shown in Tables 5–8. The studies had durations between 4 and 16 weeks, with information on the adverse event collected at between 2 and 4 time points. A total of 129 of the 2417 patients experienced the AE at least once, and 31 of these had two or more occurrences of the event. The patients were also categorized by sex and age category (<40 and ≥40 years).

Table 5 Number of Adverse Events Reported by Time of Exposure

Study	Treatment	Weeks on study drug						Total patients
		2	4	6	8	12	16	
A	Reference	4	2	0	—	—	—	249
A	Test drug	2	1	0	—	—	—	237
B	Reference	4	4	0	—	—	—	256
B	Test drug	4	4	1	—	—	—	235
C	Reference	7	1	—	—	—	—	217
C	Test drug	10	1	—	—	—	—	228
D	Reference	—	6	—	6	14	6	232
D	Test drug	—	8	—	6	7	4	261
E	Reference	—	16	—	10	3	3	269
E	Test drug	—	10	—	12	5	2	233

Table 6 Frequencies for Patients Experiencing Adverse Events (AE) By Sex

Study	Treatment	Sex	Patients with AE	Total patients
A	Reference	F	2	122
A	Reference	M	2	127
A	Test drug	F	2	126
A	Test drug	M	0	111
B	Reference	F	4	135
B	Reference	M	1	121
B	Test drug	F	3	119
B	Test drug	M	2	116
C	Reference	F	8	111
C	Reference	M	0	106
C	Test drug	F	6	109
C	Test drug	M	5	119
D	Reference	F	13	117
D	Reference	M	13	115
D	Test drug	F	11	126
D	Test drug	M	10	135
E	Reference	F	13	131
E	Reference	M	14	138
E	Test drug	F	11	116
E	Test drug	M	9	117

Table 7 Frequencies for Patients Experiencing Adverse Events by Age Category

Study	Treatment	Age category	Patients with AE	Total patients
A	Reference	<40	3	165
A	Reference	≥40	1	84
A	Test drug	<40	2	152
A	Test drug	≥40	0	85
B	Reference	<40	3	159
B	Reference	≥40	2	97
B	Test drug	<40	4	161
B	Test drug	≥40	1	74
C	Reference	<40	1	140
C	Reference	≥40	7	77
C	Test drug	<40	4	153
C	Test drug	≥40	7	75
D	Reference	<40	11	146
D	Reference	≥40	15	86
D	Test drug	<40	9	169
D	Test drug	≥40	12	92
E	Reference	<40	13	173
E	Reference	≥40	14	96
E	Test drug	<40	11	148
E	Test drug	≥40	9	85

Table 8 Frequencies for Patients Experiencing One or More Adverse Events

Study	Treatment	Patients with AE			Total patients
		0 AE	1 AE	≥2 AE	
A	Reference	245	2	2	249
A	Test drug	235	1	1	237
B	Reference	251	2	3	256
B	Test drug	230	1	4	235
C	Reference	209	8	0	217
C	Test drug	217	11	0	228
D	Reference	206	21	5	232
D	Test drug	240	17	4	261
E	Reference	242	22	5	269
E	Test drug	213	13	7	233

A. Example 1: Illustration of Basic Methods

1. Dichotomous Adverse Event Outcome with Dichotomous Treatment Categories

For the first illustrative analysis, the outcome was considered as dichotomous (occurrence or nonoccurrence of the event), and the duration of the studies was not taken into account. To examine the effect of treatment on the outcome, the data were arranged in a 2 × 2 table, and Fisher's exact test was applied (Table 9). For Fisher's exact test, the p value of .416 was compatible with no association between the treatment and the occurrence of the AE. Similarly, Fisher's exact test was performed to examine the association of sex and age category with the outcome when AE frequencies were pooled across studies. AE frequencies relative to age category are shown in Table 7; frequencies relative to sex are shown in Table 6. For sex, the p value was .148, and for age category $p < .001$. Although more females than males reported the adverse event, there was not a significant association between sex and the outcome, whereas older patients did experience the event significantly more often than younger patients.

An M-H test was performed to examine the association between treatment and the outcome while controlling for study (see Table 8). For instance, for study A the 2 × 2 table is presented in Table 10. Under the hypothesis of no association between treatment and AE occurrence, $E(a) = (249 \times 6)/486 = 3.07$ and $V(a) = (249 \times 237 \times 6 \times 480)/(486^2 \times 485) = 1.48$. The M-H statistic, $Q = \{\Sigma_h[a_h - E(a_h)]\}^2/\Sigma V(a_h)$ (where the sum is taken across all five studies) was 0.890, with a p value of .345, which supports the nonassociation between the treatment drug and occurrence of the AE. The Breslow–Day test for homogeneity of the odds ratios was also given by the SAS procedure FREQ ($Q = 1.797$, with 4 df and $p = .773$), indicating that the studies were similar in the association between the treatment and the outcome.

An M-H test was additionally performed to examine the association of sex with AE occurrence while controlling for study. Here, $Q = 2.776$, with 1 df and $p = .096$, suggesting some association between sex and the outcome. For the M-H test for association of age category with AE while controlling for study, Q

Table 9 2 × 2 Table for Treatment vs. Adverse Event Occurrence

	AE occurrence		
Treatment	Yes	No	Total
Reference	70	1153	1223
Test	59	1135	1194
Total	129	2288	2417

Table 10 2 × 2 Table for Study A for Treatment vs. Adverse Event Occurrence

	AE occurrence		
Treatment	Yes	No	Total
Reference	4	245	249
Test	2	235	237
Total	6	480	486

= 17.740 and $p < .001$. There is a strong association between age category and AE occurrence. For age, the Breslow–Day statistic was 10.004, with 4 df and $p = .040$, which indicates some heterogeneity in the age effect across the different studies.

A logistic regression model was applied to describe the relationship between AE occurrence and explanatory variables for study, treatment, sex, and age category. The parameter estimates given by the SAS procedure CATMOD is shown in Table 11. The difference in adverse event occurrence between older and younger patients is evident; the odds of occurrence to nonoccurrence is exp(0.7430) = 2.10 times higher for the patients over 40 years old. The nonsignificant likelihood ratio test is consistent with a reasonably good fit of the model to the data and thereby supports the assumption of homogeneity of the effects for each of the explanatory variables in the model relative to the others.

are

Table 11 Parameter Estimates from SAS Procedure CATMOD for Logistic Regression Model

Effect	Estimate	Standard error	χ^2	Prob.
Intercept	−5.5095	0.5136	115.09	.0000
Study B	0.5030	0.5211	0.93	.3345
Study C	1.2860	0.4738	7.37	.0066
Study D	2.1453	0.4394	23.84	.0000
Study E	2.1153	0.4393	23.19	.0000
Treatment	−0.1656	0.1853	0.80	.3714
Sex	0.2545	0.1862	1.87	.1718
Age group	0.7430	0.1849	16.14	.0001
Goodness-of-fit:	df	χ^2	p value	
Likelihood ratio statistic	32	35.30	.3151	

Chi-square

Chi-square

i.e. Greek chi 2x

Show X^2 as Chi-square in words if poss. b/c

move df to left

2. Ordinal Adverse Event Outcome with Dichotomous Treatment Categories

To illustrate methods for an ordinal outcome, the patients were categorized based on the number of occurrences of the adverse event: 0, 1,and 2 or more (see Table 12). The mean score for the reference group is m_r = [0(1153) + 1(55) + 2(15)]/1223 = 0.0695. Under the hypothesis of no association, $E(m_r)$ = [0(2288) + 1(98) + 2(31)]/2417 = 0.0662 and $V(m_r)$ = (1194/1223 × 2416)[(2288/2417)(0 − .0662)^2 + (98/2417)(1 − 0.0662)^2 + (31/2417)(2 − 0.0662)^2] = 3.53 × 10^{-5}. The χ^2 statistic for comparing the mean scores of the two treatment groups (from the SAS procedure FREQ)is Q = (m_r − $E(m_r)$)^2/$V(m_r)$ = 0.309, with 1 df and p = .578. The distribution of the number of AE occurrences within each study is shown in Table 8. The analogous M-H statistic for testing if the row mean scores differ for the treatments, while controlling for study, is Q = 0.356, with 1 df and a p value of .551.

The proportional odds model extension of logistic regression analysis was used to describe the relationship between the number of AE occurrences (0, 1, and ≥2) and the explanatory variables for study, sex, age category, and treatment. In this model, the odds of at least one AE vs. none and the odds of at least two AEs vs. at most one AE have parallel relationships to the explanatory variables through a single set of corresponding regression parameters, but they have distinct intercepts. The parameter estimates and χ^2 tests from the SAS procedure LOGISTIC are shown in Table 13.

As noted for other analyses, age category is significant, with older patients having exp(0.7436) = 2.10 times the odds of more prevalent occurrence of AE than younger patients. The individual studies are also seen to differ from each other in the odds of AE occurrence. One possible reason for the difference between studies is the variation in the length of study. The studies varied from 4 to 16 weeks in length, with patients in the longer studies having greater opportunity to experience and report AEs. For purposes of completeness, one can note that the LOGISTIC procedure provides a score statistic for evaluating the appropriateness of the proportional odds assumption. However, its result of Q = 45.87 with 7 df and p < .001 should be viewed skeptically because of the

Table 12 2 × 3 Table for Treatment vs. Adverse Event Occurrence

	AE occurrences			
Treatment	0	1	≥2	Total
Reference	1153	55	15	1223
Test drug	1135	43	16	1194
Total	2288	98	31	2417

Table 13 Parameter Estimates from SAS Procedure LOGISTIC for Proportional Odds Model

Variable	Parameter estimate	Standard error	Wald X^2	Pr > X^2	Standardized estimate
Intercept 1	−7.0118	0.5387	169.4152	0.0001	—
Intercept 2	−5.5129	0.5131	115.4574	0.0001	—
Study B	0.5091	0.5196	0.9600	0.3272	0.112951
Study C	1.2678	0.4739	7.1569	0.0075	0.270961
Study D	2.1385	0.4387	23.7640	0.0001	0.475180
Study E	2.1133	0.4385	23.2276	0.0001	0.472732
Sex	0.2619	0.1863	1.9768	0.1597	0.072214
Treatment	−0.1613	0.1852	0.7587	0.3837	−0.044479
Age group	0.7436	0.1849	16.1739	0.0001	0.195840

potentially spurious influence of 0 frequencies for ≥ 2 AE in study C and the small frequencies (≤4) of both 1 AE and ≥ 2 AEs in studies A and B. If the model is simplified by pooling study B with study C and study D with study E, the score statistic for the proportional odds assumption is $Q = 6.66$ with 5 df and $p = .247$. This result is interpreted as providing reasonable support for the use of the proportional odds model.

3. Ordinal Adverse Event Outcome with Ordinal Treatment Categories

In Tables 14 and 15, frequencies of 0, 1, and ≥2 adverse events are shown for a modification of the studies under consideration to compare four treatment groups: a placebo and three doses of a test drug. The frequencies within each study are shown in Table 14. After pooling the data across studies, the numbers of patients experiencing 0, 1, or ≥ 2 adverse events are shown in Table 15. The correlation χ^2 statistic using integer scores is 3.29, with 1 df and $p = .070$; there is a nearly significant tendency for AE occurrences to be more prevalent at higher doses. The extended M-H statistic, testing for correlation between dosage and AE occurrence with adjustment for study, is $Q = 3.359$, with 1 df and $p = .067$.

B. Example 2: Illustration of Methods Accounting for Time at Risk

A restructuring of the data from Example 1 that considers the first occurrence of the AE and takes time into account is provided in Table 16. The previous discussion considered only patients who had either completed the study without an AE or had experienced at least one AE. Patients who withdrew from the study before its completion without having an AE are now also included in the analysis.

Table 14 Adverse Event Occurrences for Patients from a Four-Dose Study

Study	Treatment	Patients with AE			Total patients
		0 AE's	1 AE	≥2 AE	
A	Placebo	121	1	1	123
A	Dose 1	122	1	0	123
A	Dose 2	132	0	0	132
A	Dose 3	105	1	2	108
B	Placebo	128	0	2	130
B	Dose 1	113	2	3	118
B	Dose 2	129	1	0	130
B	Dose 3	111	0	2	113
C	Placebo	112	3	0	115
C	Dose 1	111	4	0	115
C	Dose 2	108	2	0	110
C	Dose 3	95	10	0	105
D	Placebo	118	10	1	129
D	Dose 1	112	7	4	123
D	Dose 2	106	14	2	122
D	Dose 3	110	7	2	119
E	Placebo	128	6	3	137
E	Dose 1	106	8	2	116
E	Dose 2	123	10	2	135
E	Dose 3	98	11	5	114

1. Incidence Density

Since the exact times of withdrawal from the study or occurrence of the adverse event were not known, those events were regarded as happening (on average) halfway through the appropriate time interval. Time at risk was then calculated over all patients by summing the weeks of the completed intervals and half the weeks from intervals where an AE occurred or a patient withdrew from the

Table 15 4 × 3 Table of Treatment vs. Adverse Event Occurrence

Treatment	AE occurrences			Total
	0	1	≥2	
Placebo	607	20	7	634
Low dose	564	22	9	595
Middle dose	598	27	4	629
High dose	519	29	11	559
Total	2288	98	31	2417

Table 16 Patient Outcome and Incidence Density Calculations

Study	Treatment outcome	Interval ending at week 2	4	6	8	12	16	Completed study	Weeks of study	Time at risk	Incidence density
A	Reference										
	AE	4	0	0	—	—	—				
	Withdrawn	29	6	1	—	—	—	245	6	1526	.0026
	Test drug										
	AE	2	0	0	—	—	—				
	Withdrawn	33	0	3	—	—	—	235	6	1460	.0014
B	Reference										
	AE	4	1	0	—	—	—				
	Withdrawn	32	1	3	—	—	—	251	6	1563	.0032
	Test drug										
	AE	4	0	1	—	—	—				
	Withdrawn	24	4	3	—	—	—	230	6	1440	.0035
C	Reference										
	AE	7	1	—	—	—	—				
	Withdrawn	50	3	—	—	—	—	209	4	905	.0088
	Test drug										
	AE	10	1	—	—	—	—				
	Withdrawn	52	0	—	—	—	—	217	4	933	.0118
D	Reference										
	AE	—	6	—	6	11	3				
	Withdrawn	—	25	—	3	5	0	206	16	3614	.0066
	Test drug										
	AE	—	8	—	6	7	0				
	Withdrawn	—	23	—	0	7	0	240	16	4078	.0049
E	Reference										
	AE	—	16	—	6	2	3				
	Withdrawn	—	19	—	4	4	0	242	16	4104	.0065
	Test drug										
	AE	—	10	—	5	3	2				
	Withdrawn	—	24	—	1	2	0	213	16	3590	.0056

study. The incidence density was calculated for each of the two treatment groups for each study as (number of events/time at risk). Relative to the hypothesis of no association between treatment and event occurrence, and using the Poisson distribution for AE frequency, for study A, $E(n_i) = ((4 + 2) \times 1526)/(1526 + 1460) = 2.93$, and $V(n_i) = ((4 + 2) \times 1526 \times 1460)/(1526 + 1460)^2 = 1.50$. The M-H statistic was calculated using this method for the Poisson distribution, as discussed earlier (Section IV.B). M-H χ^2 (with 1 df) was 0.679, with p value of .410.

III. B.)

2. Life Table Methods

The SAS procedure LIFETEST was used to calculate life table statistics. In determining time at risk, LIFETEST uses an estimated number of patients at risk for each of the respective time intervals. This quantity is the sum of all patients completing the interval, patients experiencing the event in the interval, and one half the patients withdrawing during the interval. Patients who completed a study were viewed as having withdrawn from the study during the day after the end of the study (within 0.14 week). To standardize the end points of the intervals for all studies, the events occurring between weeks 4 and 8 in studies D and E were

Table 17 Life Table Survival Estimates from SAS Procedure LIFETEST

Study	Interval (weeks)	Number failed (with AE)	Number withdrawn	Effective sample size	Conditional probability of AE	No AE to beginning of interval
A	0–4	6	68	524.0	.0115	1.0000
	4–4.14	0	0	484.0		0.9885
	4.14–6	0	4	482.0		0.9885
	6–6.14	0	480	240.0		0.9885
B	0–4	9	61	527.5	.0171	1.0000
	4–4.14	0	0	488.0		0.9829
	4.14–6	1	6	485.0	.00206	0.9829
	6–6.14	0	481	240.5		0.9809
C	0–4	19	105	497.5	.0382	1.0000
	4–4.14	0	426	213.0		0.9618
D	0–4	14	48	532.0	.0263	1.0000
	4–4.14	0	0	494.0		0.9737
	4.14–6	6	2	493.0	.0122	0.9737
	6–6.14	0	0	486.0		0.9618
	6.14–8	6	1	485.5	.0124	0.9618
	8–8.14	0	0	479.0		0.9499
	8.14–12	18	12	473.0	.0381	0.9499
	12–16	3	0	449.0	.00668	0.9138
	16–16.14	0	446	223.0		0.9077
E	0–4	26	43	534.5	.0486	1.0000
	4–4.14	0	0	487.0		0.9514
	4.14–6	5	2	486.0	.0103	0.9514
	6–6.14	0	0	480.0		0.9416
	6.14–8	6	3	478.5	.0125	0.9416
	8–8.14	0	0	471.0		0.9298
	8.14–12	5	6	468.0	.0107	0.9298
	12–16	5	0	460.0	.0109	0.9198
	16–16.14	0	455	227.5		0.9098

equally divided between weeks 4–6 and weeks 6–8, and 4, 6, 8, 12, and 16 weeks were used as the applicable end points of the interval. The lift table results for the pooled treatment groups are shown in Table 17. The results for the log-rank (Mantel–Haenszel) tests from the procedure LIFETEST are shown in Table 18.

3. Piecewise Exponential Model

A Poisson regression analysis was used to estimate parameters in a piecewise exponential model for incidence densities like those in Table 16. For each patient, the time at risk was defined as the length (in weeks) of each completed interval or one half the length of an interval in which the patient withdrew from the study or experienced the adverse event for the first time. For each interval, the number of events and sums of time at risk were then determined for the patients in each subpopulation with respect to the study by treatment sex age category cross-classification. After multiplication of the times at risk by 100 to stimulate Poisson distributions (see Ref. 29), these quantities were then analyzed in a (numerator/denominator) model with the SAS procedure LOGISTIC. The model included explanatory variables for study, treatment, sex, age category, and time interval. The results are shown in Table 19. (The estimate for the intercept was obtained by adding $\log_e(100)$ to the value of -11.8663 obtained from the LOGISTIC procedure in order to remove the influence of the multiplication of the times at risk by 100.)

As in previously described analyses, age category is significant, with the rate of AE occurrence for older patients being exp(0.7082) = 2.03 times higher than that for younger patients. There were significantly lower rates of occurrence in the later time intervals than in the first; for instance, in the 4- to 8-week interval, the rate of AE occurrence is 0.47 time lower than in the 0- to 4-week interval.

C. Example 3: Illustration Using Multiple Occurrence Data

Finally, to take into account multiple occurrences of the event experienced by individual patients, a ratio of a weighted mean of AE occurrences to a weighted mean of time at risk was formed, where the mean was calculated over all patients within each study, and then over all studies. The time at risk was calculated as

Table 18 Univariate X^2 for the Log-Rank Tests from SAS Procedure LIFETEST

Variable	Test statistic	Standard deviation	X^2	$Pr > X^2$
Treatment	5.1354	5.6704	0.8202	0.3651
Sex	−9.0996	5.6777	2.5686	0.1090
Age category	−23.3704	5.4027	18.7119	0.0001

Table 19 Analysis of Maximum Likelihood Estimates for Piecewise Exponential Model from SAS Procedure LOGISTIC

Variable	Parameter estimate	Standard error	Wald X^2	Pr > X^2	Standardized estimate
Intercept	−7.2611	0.5083	544.8830	0.0001	
Study B	0.5909	0.5076	1.3552	0.2444	0.118881
Study C	1.6646	0.4746	12.3032	0.0005	0.235131
Study D	1.6470	0.4395	14.0442	0.0002	0.418740
Study E	1.6398	0.4397	13.9110	0.0002	0.417227
Treatment	−0.1383	0.1762	0.6163	0.4324	−0.038122
Sex	0.2639	0.1776	2.2074	0.1374	0.072746
Age category	0.7082	0.1761	16.1767	0.0001	0.185779
4–8 weeks	−0.7479	0.2412	9.6117	0.0019	−0.191023
8–12 weeks	−0.5426	0.2550	4.5302	0.0333	−0.106618
12–16 weeks	−1.5817	0.3828	17.0750	0.0001	−0.308605

the sum of completed intervals and one half the sum of intervals in which a patient withdrew from the study. (No patients with AEs withdrew before the completion of the study; had they withdrawn, their time at risk would have been determined in the same way as for those who withdrew and had no adverse events.) Table 20 shows the calculations for total events and total time at risk in each study.

The weight for study A was (285 × 273)/(285 + 273) divided by (285 × 273)/(285 + 273) + (292 × 266)/(292 + 266) + (270 × 280)/(270 + 280) + (265 × 291)/(265 + 291) + (296 × 260)/(296 + 260) for 139.435/693.200 = 0.20015. The weighted values for the portion of f and g (ratio numerator and denominator) due to study A for the reference group were 0.00423 and 1.09114, respectively, and the ratio across all studies for the reference group was R_r = 0.061359/8.72658 = 0.0070. For the patients on the test drug, R_t = 0.056255/8.73963 = 0.0064. The weights and numerator and denominator values for each study are displayed in Table 21.

For the reference group, the variance of the numerator was $V(f_r) = 5.739 \times 10^{-5}$, the variance of the denominator was $V(g_r) = 5.378 \times 10^{-3}$, and the covariance between them was $\text{Cov}(f_r, g_r) = 5.096 \times 10^{-5}$. For the test drug group, $V(f_t) = 6.049 \times 10^{-5}$, $V(g_t) = 5.550 \times 10^{-3}$, and $\text{Cov}(f_t, g_t) = 4.4787 \times 10^{-5}$. The variances of the two ratios, then, were $V(R_r) = 7.477 \times 10^{-7}$ and $V(R_t) = 7.874 \times 10^{-7}$. Because of the large sample size, $Z = (R_t - R_r)/[V(R_t) + V(R_r)]^{1/2}$ is approximately normally distributed and is equal to −0.489, with a two-sided p value of .631. Thus, for the combined studies, the weighted rates of

Table 20 Total Event and Time at Risk Calculations for Ratio Estimate

Study	Treatment	No. patients	Outcome	Total events	Time (weeks) End of interval	Time at risk	Total time
A	Reference	29	Withdrawn	0	2	1	29
		6	Withdrawn	0	4	3	18
		1	Withdrawn	0	6	5	5
		245	No AE	0	6	6	1470
		2	1 AE	2	6	6	12
		2	2 AE	4	6	6	12
Totals		285		6			1546
A	Test drug	33	Withdrawn	0	2	1	33
		3	Withdrawn	0	6	5	15
		235	No AE	0	6	6	1410
		1	1 AE	1	6	6	6
		1	2 AE	2	6	6	6
Totals		273		3			1470
B	Reference	32	Withdrawn	0	2	1	32
		1	Withdrawn	0	4	3	3
		3	Withdrawn	0	6	5	15
		251	No AE	0	6	6	1506
		2	1 AE	2	6	6	12
		3	2 AE	6	6	6	18
Totals		292		8			1586
B	Test drug	24	Withdrawn	0	2	1	24
		4	Withdrawn	0	4	3	12
		3	Withdrawn	0	6	5	15
		230	No AE	0	6	6	1380
		1	1 AE	1	6	6	6
		4	2 AE	8	6	6	24
Totals		266		9			1461
C	Reference	50	Withdrawn	0	2	1	50
		3	Withdrawn	0	4	3	9
		209	No AE	0	4	4	836
		8	1 AE	8	4	4	32
Totals		270		8			927
C	Test drug	52	Withdrawn	0	2	1	52
		217	No AE	0	4	4	868
		11	1 AE	11	4	4	44
Totals		280		11			964

Table 20 *(Continued)*

Study	Treatment	No. patients	Outcome	Total events	Time (weeks) End of interval	Time at risk	Total time
D	Reference	25	Withdrawn	0	4	2	50
		3	Withdrawn	0	8	6	18
		5	Withdrawn	0	12	10	50
		206	No AE	0	16	16	3296
		21	1 AE	21	16	16	336
		4	2 AE	8	16	16	64
		1	3 AE	3	16	16	16
Totals		265		32			3830
D	Test drug	23	Withdrawn	0	4	2	46
		7	Withdrawn	0	12	10	70
		240	No AE	0	16	16	3840
		17	1 AE	17	16	16	272
		4	2 AE	8	16	16	64
Totals		291		25			4292
E	Reference	19	Withdrawn	0	4	2	38
		4	Withdrawn	0	8	6	24
		4	Withdrawn	0	12	10	40
		242	No AE	0	16	16	3872
		22	1 AE	22	16	16	352
		5	2 AE	10	16	16	80
Totals		296		32			4406
E	Test drug	24	Withdrawn	0	4	2	48
		1	Withdrawn	0	8	6	6
		2	Withdrawn	0	12	10	20
		213	No AE	0	16	16	3408
		13	1 AE	13	16	16	208
		5	2 AE	10	16	16	80
		2	3 AE	6	16	16	32
Totals		260		29			3802

events per week at risk are similar for the two treatments. Additionally, one can note that $\log_e(R_t/R_r) = 0.0883$ and $\mathrm{Var}\{\log_e(R_t/R_r)\} = \mathrm{Var}(R_t)/R_t^2 + \mathrm{Var}(R_r)/R_r^2 = 0.0341$; and so a .95 confidence interval for the ratio of weighted rates of events per week at risk for the two treatments is $\exp\{(-0.088) \pm 1.96/0.0341\}$ or (0.64, 1.31). The advantage of this confidence interval is that it quantifies the degree of similarity of the weighted rates of events per week at risk for the two treatments.

Should be $1.96\sqrt{0.0341}$ }

Alternatively,

Table 21 Weights and Contributions to the Ratio Numerator and Denominator for each Study

Study	Weight	Reference group contribution to		Test group contribution to	
		Numerator	Denominator	Numerator	Denominator
A	0.20115	0.00423	1.091	0.00221	1.083
B	0.20080	0.00550	1.091	0.00679	1.103
C	0.19829	0.00588	0.681	0.00780	0.683
D	0.20008	0.02416	2.892	0.01719	2.951
E	0.19968	0.02159	2.972	0.02227	0.920

V. DISCUSSION

Meta-analysis is a useful strategy in the analysis of drug safety data, particularly when the sample size in individual studies is insufficient to allow a meaningful interpretation. There are several issues to keep in mind while performing a meta-analysis, such as which studies to include and how much weight to give to each one, but the statistical methods for the most part are very straightforward. In many cases, methods that are used in multicenter clinical trials can be adapted for use in meta-analysis. The methods presented in this chapter take into account some of the difficulties common in drug safety data such as varying time at risk and multiple occurrences of an adverse event in individual patients.

REFERENCES

1. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res.* 1976;5: 3–8.
2. Peto R. Why do we need systematic overviews of randomized trials? *Stat Med.* 1987;6:233–240.
3. Chalmers TC, Buyse ME. Meta-analysis. In: Chalmers T. ed. *Data Analysis for Clinical Medicine: The Quantitative Approach to Patient Care in Gastoenterology.* Rome: International University Press; 1988:75–84.
4. Light RJ. Accumulating evidence from independent studies: What we can win and what we can lose. *Stat Med.* 1987;6:221–228.
5. Bulpitt CJ. Medical statistics: meta-analysis. *Lancet.* 1988;2:93–94.
6. Einarson TR. Meta-analysis of the pharmacotherapy literature. In: Hartzema AB, Porta M, Tilson HH. eds. *Pharmacoepidemiology: An Introduction.* 2nd ed. Cincinnati: Harvey Whitney; 1991:236–268.
7. Elashoff JD, Koch GG. Statistical methods in trials of anti-ulcer drugs. In: Swabb E, Szabo S. eds. *Ulcer Disease: Investigation and Basis for Therapy.* New York: Marcel Dekker; 1990:375–406.
8. Wachter KW. Disturbed by meta-analysis? *Science.* 1988;241:1407–1408.

9. Hedges L, Olkin I. *Statistical Method for Meta-analysis*. New York: Academic Press; 1985.
10. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials*. 1986;7:177–188.
11. Bailey KR. Interstudy differences: how should they influence the interpretation and analysis of results? *Stat Med*. 1987;6:351–358.
12. Schmid JE, Koch GG, LaVange LM. An overview of statistical issues and methods of meta-analysis. *J Biopharm Stat*. 1991;1(1):103–120.
13. Gail M, Simon R. Testing for qualitative interaction between treatment effects and patient subsets. *Biometrics*. 1985;6:341–348.
14. O'Neill RT. Assessment of safety. In: Peace KE. ed. *Biopharmaceutical Statistics for Drug Development*. New York: Marcel Dekker; 1988:543–604.
15. Koch GG. Discussion: statistical perspective. *Drug Inf J*. 1991;25(3):461–464.
16. Edwards S, Koch GG, Sollecito WA, Peace KE. Summarization, analysis, and monitoring of adverse events. In: Peace KE. ed. *Statistical Issues in Drug Research and Development*. New York: Marcel Dekker; 1990:19–170.
17. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Stat Med*. 1991;10:1665–1677.
18. Chevart B. A nonparametric model for multiple recurrences. *Appl Stat*. 1988;37(2):157–168.
19. Thall PF, Lachin JM. Analysis of recurrent events: Non-parametric methods for random-interval count data. *J Am Stat Assoc*. 1988;83(402):339–347.
20. Enas GG. Making decisions about safety in clinical trials; the case for inferential statistics. *Drug Inf J*. 1991;25(3):439–446.
21. Huster WJ. Clinical trial adverse events: the case for descriptive statistics. *Drug Inf J*. 1991;25(3):447–456.
22. D'Agostino RB, Heeren TC. Multiple comparisons in over-the-counter drug clinical trials with both positive and placebo controls. *Stat Med*. 1991;10(1):1–6.
23. SAS/STAT Users Guide. version 6. 4th ed. vols 1 and 2. Cary, NC: SAS Institute; 1990.
24. StatXact. Statistical software for exact nonparametric inference. User Manual. version 2. Cambridge, MA: Cytel Software Corporation; 1991.
25. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York: Wiley; 1986.
26. Koch GG, Edwards SE. Clinical efficacy trials with categorical data. In: Peace KE. ed. *Biopharmaceutical Statistics for Drug Development*. New York: Marcel Dekker; 1988:403–457.
27. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. New York: Van Nostrand Reinhold; 1982:98–115.
28. Harris EK, Albert A. *Survivorship Analysis and Clinical Studies*. New York: Marcel Dekker; 1991:1–54.
29. Vine MF, Schoenbach VJ, Hulka BS, Koch GG, Samsa G. A typical metaplasia and incidence of bronchogenic carcinoma. *Am J Epidemiol*. 1990;131:781–793.