J Clin Epidemiol Vol. 45, No. 7, pp. 769–773, 1992
Printed in Great Britain

0895-4356/92 $5.00 + 0.00
Pergamon Press Ltd

VARIANCE IMPUTATION FOR OVERVIEWS OF CLINICAL TRIALS WITH CONTINUOUS RESPONSE

DEAN FOLLMANN,[1] PAUL ELLIOTT,[3] IL SUH[2] and JEFFREY CUTLER[1]

[1]Division of Epidemiology and Clinical Applications, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, U.S.A., [2]Department of Preventive Medicine and Public Health, Yonsei University College of Medicine, Seoul, Korea and [3]Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, England

(Received in revised form 2 January 1992)

Abstract—Overviews of clinical trials are an efficient and important means of summarizing information about a particular scientific area. When the outcome is a continuous variable, both treatment effect and variance estimates are required to construct a confidence interval for the overall treatment effect. Often, only partial information about the variance is provided in the publication of the clinical trial. This paper provides heuristic suggestions for variance imputation based on partial variance information. Both pretest–posttest (parallel groups) and crossover designs are considered. A key idea is to use separate sources of incomplete information to help choose a better variance estimate. The imputation suggestions are illustrated with a data set.

Crossover Meta-analysis Parallel groups Pretest–posttest

INTRODUCTION

Overviews, or meta-analyses, of clinical trials are an important means of summarizing experimental evidence on a specific scientific issue [1–3]. Although many small clinical trials may provide equivocal evidence on the issue, when taken all together, a firm conclusion may be drawn. A major feature of an overview is a quantitative summary or effect estimate, along with uncertainty bounds.

When the primary outcome is continuous, trial specific estimates of the treatment effects as well as their variability are necessary to provide a pooled effect estimate. Variance estimates are required both to combine effect estimates and to provide confidence limits for the overall effect estimate, but may be unreported. A p-value inequality, or some other incomplete information might be given instead. To include such a trial in an effect estimate requires some kind of imputation of the effect variance. This paper provides some suggestions for imputation.

Our research arose from an overview of clinical trials designed to test the impact of sodium reduction on lowering blood pressure [4]. Fewer than half of the trials published a variance estimate or information to allow one to derive it. The absence of variance estimates led us to develop a heuristic variance imputation scheme for both pretest–posttest (parallel groups) and crossover trials.

STATISTICAL NOTATION

Consider a simple clinical trial that compares a treatment to a control in terms of a continuous outcome X. In a pretest–posttest trial, the treatment and control changes in X from baseline to end-of-study are compared. For crossover studies, the difference in X following the treatment

and control phases is evaluated. In symbols, the estimates of treatment effect* are, respectively,

$$\hat{\Delta}_p = \bar{X}_T^{(f)} - \bar{X}_T^{(b)} - [\bar{X}_C^{(f)} - \bar{X}_C^{(b)}] \quad (1)$$

$$\hat{\Delta}_x = \bar{X}_T - \bar{X}_C, \quad (2)$$

where

$\bar{X}_T^{(f)}$ = final response average in the treatment group
$\bar{X}_T^{(b)}$ = baseline response average in the treatment group
$\bar{X}_C^{(f)}$ = final response average in the control group
$\bar{X}_C^{(b)}$ = baseline response average in the control group
$\bar{X}_T$ = response average following the treatment phase of a crossover trial
$\bar{X}_C$ = response average following the control phase of a crossover trial.

Let n_T, n_C be the number of patients in the treatment and control arms of a pretest–posttest trial and n the number of patients in a crossover trial. For simplicity, we will assume that no subjects drop out, and that sample size is always reported. The variances for these effect estimates or *effect variances* are:

$$V[\hat{\Delta}_p] = \sigma_\Delta^2(1/n_t + 1/n_c) \quad (3)$$

$$V[\hat{\Delta}_x] = \sigma_\Delta^2(1/n), \quad (4)$$

where

$$\sigma_\Delta^2 = 2\sigma^2(1-\rho). \quad (5)$$

The parameter σ^2 is the variance of a single measurement or *measurement variance*, σ_Δ^2 is the variance of the difference between two measurements on an individual or *change variance*, and ρ is the correlation between two measurements on an individual. The two measurements are either baseline and end-of-study measurements (for pretest–posttest designs) or treatment phase and control phase measurements (for crossover designs). Note that we are assuming that the variances are the same for both treatment and control groups (or phases) for a given trial. Randomization ensures such equality if treatment has no effect.

IMPUTATION STRATEGIES

If a common treatment effect across all trials is assumed, an overall effect estimate is easily derived. The effect for the ith trial, $\hat{\Delta}_i$, is approximately normally distributed with mean Δ and variance $V[\hat{\Delta}_i]$. If the measurements themselves are normally distributed then the approximation is exact, otherwise approximate normality follows from the Central Limit Theorem. It follows that the maximum likelihood estimate of Δ from such data is

$$\hat{\Delta} = \sum_{i=1}^{m} \frac{w_i\hat{\Delta}_i}{\sum_{j=1}^{m} w_j}, \quad (6)$$

where $w_i = V[\hat{\Delta}_i]^{-1}$, and is given by either equations (3) or (4). Even if normality does not hold, $\hat{\Delta}$ is the best linear unbiased estimate of Δ. The variance of $\hat{\Delta}$ is $1/\Sigma V[\hat{\Delta}_i]^{-1}$. In practice, the $V[\hat{\Delta}_i]$s are replaced with their sample estimates and $\hat{\Delta} \pm 1.96\sqrt{1/\Sigma\hat{V}[\hat{\Delta}_i]^{-1}}$ is used as an approximate 95% confidence interval for Δ. Unfortunately, $\hat{V}(\hat{\Delta}_i)$ is often not reported and it is unclear how to compute $\hat{\Delta}$ and $\hat{V}[\hat{\Delta}]$. Ignoring such studies, though common [7], has little appeal other than expediency. Below we suggest two approaches that allow incorporation of studies with incomplete data.

If a common change variance is assumed across trials, the σ_Δ^2s implicit in the numerator and denominator of equation (6) cancel out, so that $\hat{\Delta}$ depends only on the sample sizes and the Δ_is. While $\hat{V}[\hat{\Delta}]$ does require variance estimates for each trial, all depend on a common σ_Δ^2. A good estimate of σ_Δ^2 here is to take the weighted (by $n-1$ or n_T+n_C-2) average of the reported† change variance estimates. This estimate is unbiased and minimizes the variance of any weighted average of $\hat{\sigma}_\Delta^2$s.

If the change variances are assumed to differ across trials, a separate estimate is required for each trial. If an effect or change variance is reported, of course, imputation is not required. With interval or boundary p-values, one can guess the true p-value and invert the guessed p-value to obtain a variance estimate. For interval p-values, a reasonable guess is to take the midpoint of the interval. For upper boundary p-values, e.g. $p < 0.05$ we suggest using the upper bound. The impact is to downweight trials that don't give complete information.

*In the social sciences, effect is commonly taken to be the difference between the treatment and control means divided by the standard deviation of a measurement. Hedges and Olkin [1] discuss combining such effect estimates. In other areas, such as hypertension research, the unstandardized changes are commonly used (see e.g. Cutler and Brittain [5] or Law *et al.* [6]).

†We take reported variance to mean a reported change or effect variance, or an exact p-value or t-statistic. Given any one, the other three can be calculated given the sample size and treatment effect.

Lower boundary p-values, e.g. $p > 0.05$ contain very little information about the variance, and we do not advocate variance imputation based soley on such a p-value.

If measurement variances are included, but the change or effect variance is not reported, we can impute a change variance, and then an effect variance via equation (3) or (4), by assuming a specific correlation since $\sigma_\Delta^2 = 2\sigma^2(1-\rho)$. If the measurement variances for the two periods are the same, then presumably the correlation is at least 0.5, otherwise the pretest–posttest design is less efficient than a test based on just the final measurements. The assumed correlation may be checked by determining ρ for trials which report change and measurement variances.

Some trials will provide both measurement variances and interval or boundary p-values, but not report a change or effect variance. The variance imputed from the p-value or t-statistic bounds are conservative, and the ρ based variance is conservative if ρ is not overguessed. The true variance should be no larger than the minimum of the two. Therefore, if imputation is required for a trial, we suggest use of the minimum of the correlation imputed variance and the p-value/t-statistic imputed variance. If only one of these imputed variances is available it should be used. If a lower boundary p-value or no variance information is provided, one can guess an effect variance, say, based on the weighted average of the change variances in the other trials.

EXAMPLE

This research arose from an investigation of randomized clinical trials that provide information on the effect of dietary sodium reduction on blood pressure. Lower sodium intake is thought to produce lower blood pressue. A complete description of the trials and methods used for abstraction has been published [4]. Twenty-eight clinical trials were included; however, in this paper we drop the two trials that were crossover additions to pretest–posttest trials. Fifteen of the trials followed a pretest–posttest design, while the remaining 11 were crossover trials. Table 1 presents a breakdown of the trials in terms of the amount of variance information for diastolic blood pressure.

Only 10 of the trials provided complete information about the effect variance. Of these 10, 4 also provided measurement variances. Of the 16 trials without complete effect variance information, 10 reported either incomplete p-values only or measurement variances only. The remaining 6 reported both, so that comparison between the variance based on bounded p-values or t-statistics and the variance based on a chosen correlation was possible.

Table 2 presents the complete data. The effect variance estimate $\tilde{V}(\tilde{\sigma}_\Delta^2)$ is obtained either directly from the paper, or imputed via inversion of the incomplete p-value or t-statistic information. The effect variance estimate $\tilde{V}(\rho = 0.5)$ was based on an assumed correlation of 0.5 and the reported measurement variances. The average correlation for trials that reported both measurement and change variances was 0.55, suggesting 0.5 is not an unreasonable assumption. The derived ρ column is the solution, in terms of ρ, to either

$$\tilde{V}(\tilde{\sigma}_\Delta^2) = 2\sigma^2(1-\rho)/n \tag{7}$$

for crossover trials or

$$\tilde{V}(\tilde{\sigma}_\Delta^2) = 2\sigma^2(1-\rho)(1/n_t + 1/n_c) \tag{8}$$

for pretest–posttest trials.

The six trials in the middle of Table 2 provided both measurement variances and incomplete p-value/t-statistic information. The two variances, $\tilde{V}(\tilde{\sigma}_\Delta^2)$ and $\tilde{V}(\rho = 0.5)$, were compared and the minimum used. In four of these trials the derived ρ was less than 0.5 and $\tilde{V}(\rho = 0.5)$ was used rather than $\tilde{V}(\tilde{\sigma}_\Delta^2)$. In one case the derived ρ was negative suggesting a substantial error would have been made if $\tilde{V}(\tilde{\sigma}_\Delta^2)$

Table 1. Breakdown of 26 clinical trials of sodium reduction in terms of effect variance information

Number of studies	Type of information		
	Exact or derivable effect or change variance	Interval or boundary p-values or t-statistics	Measurement variances
4	Yes		Yes
6	Yes		No
6	No	Yes	Yes
3	No	Yes	No
7	No	No	Yes

Table 2. Data from 26 clinical trials examining the effect of sodium reduction on diastolic blood pressure change

Study type	Number		Effect estimate	Effect Variance estimate		$\tilde{V}(\tilde{\sigma}_\Delta^2)$ obtained ?	Derived ρ
	Control	Treatment		$\tilde{V}(\tilde{\sigma}_\Delta^2)$	$\tilde{V}(\rho = 0.5)$		
P	44	50	−3.4	2.89	4.14	Reported	0.80
P	53	55	−2.8	0.72	1.06	Reported	0.68
P	37	38	−0.2	2.00	1.58	Reported	0.29
X	113	113	−1.4	1.14	1.02	Reported	0.44
P	10	15	−6.3	19.43		Reported	
P	48	52	−4.2	0.81		Reported	
P	174	177	0.2	0.64		Reported	
X	18	18	−0.3	0.64		Reported	
X	35	35	1.2	0.86		Reported	
X	31	31	1.4	0.81		Reported	
P	15	19	0.5	7.14	5.60	Imputed	0.33
P	15	15	−6.7	20.66	9.40	Imputed	−0.11
P	20	21	−5.9	10.90	8.48	Imputed	0.34
X	15	15	3.2	18.70	12.04	Imputed	0.22
X	19	19	−5.0	3.17	4.00	Imputed	0.60
X	20	20	−5.0	2.64	4.00	Imputed	0.67
P	31	31	−6.9	4.50		Imputed	
P	6	6	−4.0	10.86		Imputed	
P	6	6	−8.0	16.66		Imputed	
P	19	19	−1.1		4.13		
P	18	12	2.0		8.08		
P	17	17	−1.8		12.12		
X	20	20	−3.0		4.13		
X	12	12	−1.8		12.61		
X	172	172	−1.9		0.81		
X	40	40	−0.8		2.41		

$\tilde{V}(\tilde{\sigma}_\Delta^2)$ is either the reported effect variance estimate or an imputation based on the inversion of an inexact p-value or t-statistic. It is blank when not reported. $\tilde{V}(\rho = 0.5)$ is the effect variance estimate based on $\sigma_\Delta^2 = 2\sigma^2(1-\rho)$ with $\rho = 0.5$ and the reported $\tilde{\sigma}^2$. It is blank if $\tilde{\sigma}^2$ was not reported. The derived ρ column is the solution, in terms of ρ, to either $\tilde{V}(\tilde{\sigma}_\Delta^2) = 2\sigma^2(1-\rho)(1/n)$ or $\tilde{V}(\tilde{\sigma}_\Delta^2) = 2\sigma^2(1-\rho)(1/n_t + 1/n_c)$. It is blank when either $\tilde{V}(\tilde{\sigma}_\Delta^2)$ or $\tilde{\sigma}^2$ are not reported. Study type is P for pretest–posttest trials and X for crossover trials.

had been used. In two cases, $\tilde{V}(\tilde{\sigma}_\Delta^2)$ was preferred to $\tilde{V}(\rho = 0.5)$.

Using imputed inverse variance weights, the 95% confidence interval for the treatment effect is -1.41 ± 0.53. Using sample size weights (n for crossover trials of $n_t n_c/(n_t + n_c)$ for pretest–posttest trials) the confidence interval is -1.60 ± 0.53. The effect variance for the latter confidence interval was based on the weighted average of the change variances over the 10 trials that reported change variance estimates. In this example, both sample size weights and inverse variance weights provide similar estimates.

The results differ somewhat by study type. Effect confidence intervals for the crossover trials are -0.70 ± 0.71 and -1.34 ± 0.65, respectively, for imputed variance weights and sample size weights. The latter uses the weighted average of change variances as the common variance estimate. The analogous intervals for the parallel trials are -2.26 ± 0.78 and -2.09 ± 0.90. The effect of sodium reduction on blood pressure appears larger in parallel trials, perhaps because of carryover effects.

The procedure of dropping studies with unreported variances leaves only 10 of the 26 studies for analysis. The two effect confidence intervals from these 10 are -0.94 ± 0.72 and -0.90 ± 0.62, respectively, for sample size weights and imputed variance weights. The analagous intervals for the 16 studies which did not report variances are -2.67 ± 0.98 and -2.29 ± 0.76. Dropping studies without variances results in different estimates. Though the confidence intervals do not quite overlap, whether this difference is due to chance or an association between effect size and data completeness is unclear.

As a simple check of the validity of imputation, we replaced the exact variances with imputed variances for the first four trials of Table 2. The resulting confidence interval based on imputed variance weights, -1.33 ± 0.53, is similar to the confidence interval using the exact values for these trials. This sensitivity check is limited by the number of exact variances that can be switched to imputed values and would be more definitive in a meta-analysis with many trials with both exact and imputable variances.

In general, examination of assumptions is important.

DISCUSSION

A quantitative summary of a collection of clinical trials with a continuous response outcome requires an estimate of the effect variance for each trial. Often, sample effect variances are not provided in the trial's written report and a meta-analyst needs to impute a variance estimate, both to weight the effect estimates of the different trials properly, and to calculate confidence intervals. This paper has suggested two possible approaches to variance imputation. One is to assume that the variance of change is the same across all trials, weight the effect estimate by sample size, and estimate the common variance of change from the trials which report it exactly. The second approach is to use the (often incomplete) information from a given trial to impute the effect variance for that trial.

Choice of the approach to use will not always be clear; however some guidelines can be offered. If most of the effect variances are reported, based on large samples, and equality of change variances is suspect, variance imputation is suggested. This approach is less assumption dependent than sample size weighting, but more dependent on accurate variance estimates. If, on the other hand, equality of change variances is not unreasonable, few trials report variances, and variance estimates are based on small samples, sample size weighting should be preferred. A practical compromise is to report both along with a preference.

While this paper has used imputed variances for a specific definition of effect and a specific overall estimate of treatment effect, the results are more widely applicable. If effect is taken to be the standardized difference between treatment and control, a variance is needed to standardize. Also, though other methods of combining effect estimates are available, such as Empirical Bayes or random effects, most require estimates of effect variance either to properly weight trials or to calculate a confidence interval.

Although the absence of variance estimates can lead to complications in calculating a pooled effect estimate, a more serious problem is unreported trials. Publication bias, or the tendency for small negative trials to be unreported in the literature can lead to an effect estimate with positive bias. While the presence of publication bias can be examined based on only the reported trials, methodology for mitigating bias is not straightforward. An alternative to the statistical approach to bias is to obtain data from unreported trials and incorporate these trials in the overview. This suggests that a final approach to variance imputation is analogous to the solution of unreported trials: contact the investigators. When feasible, investigator contact should be preferred to imputation.

Acknowledgements—We thank our colleagues at the Epilepsy Branch of the National Institute of Neurological Disorders and Stroke for generously allowing us the use of their VAX computer. We also appreciate the advice of Michael Hills of the London School of Hygiene and Tropical Medicine, who suggested sample size weighting.

REFERENCES

1. Hedges LV, Olkin I. **Statistical Methods for Meta-Analysis.** San Diego, CA: Academic Press; 1985.
2. Yusuf S, Simon R, Ellenberg S, Eds. Proceedings of the workshop on methodologic issues in overviews of randomized clinical trials. May 1986. **Stat Med** 1987; 6: 217–409.
3. Wachter KW, Straf ML. **The Future of Meta-Analysis.** New York: Russell Sage Foundation; 1990.
4. Cutler JA, Follmann D, Elliott P, Suh I. An overview of randomized trials of sodium restriction and blood pressure. **Hypertension** 1991; 17: 27–33.
5. Cutler JA, Brittain E. Calcium and blood pressure: an epidemiologic perspective. **Am J Hypertension** 1990; 3: 137S–146S.
6. Law MR, Frost CD, Wald NJ. By how much does dietary salt reduction lower blood pressure? I—Analysis of observation data among populations. **Br Med J** 1991; 301: 811–815.
7. Hedges LV. Directions for future methodology. In Wachter KW, Straf ML, Eds. **The Future of Meta-Analysis.** New York: Russell Sage Foundation; 1991.
8. Colton T, Freedman LS, Johnson AL, Eds. (1987) Proceedings of the workshop of methodologic issues in overviews of randomized clinical trials. **Stat Med** 1987; 6: 217–403.