LINEARLY DIVERGENT TREATMENT EFFECTS IN CLINICAL TRIALS WITH REPEATED MEASURES: EFFICIENT ANALYSIS USING SUMMARY STATISTICS†

LARS J. FRISON1* AND STUART J. POCOCK2
1 Biostatistics, Astra Hässle AB, S-431 83 Mölndal, Sweden
2 Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.

SUMMARY

In many randomized clinical trials with repeated measures of a response variable one anticipates a linear divergence over time in the difference between treatments. This paper explores how to make an efficient choice of analysis based on individual patient summary statistics. With the objective of estimating the mean rate of treatment divergence the simplest choice of summary statistic is the regression coefficient of response on time for each subject (SLOPE). The gains in statistical efficiency imposed by adjusting for the observed pre-treatment levels, or even better the estimated intercepts, are clarified. In the process, we develop the optimal linear summary statistic for any repeated measures design with assumed known covariance structure and shape of true mean treatment difference over time. Statistical power considerations are explored and an example from an asthma trial is used to illustrate the main points. © 1997 by John Wiley & Sons, Ltd.

1. INTRODUCTION

For clinical trials with repeated measurements taken on each subject one often anticipates a gradually increasing divergence between treatments over time. Whatever alternative hypothesis is judged plausible for the treatment difference, the summary statistic approach to the analysis of repeated measures data is simple, informative, requires few assumptions for its validity, and has become increasingly popular lately.1–4 Once an appropriate choice of an individual subject summary statistic is made the subsequent analysis is straightforward, since the statistic may be treated as a single quantitative response.

The choice of alternative hypothesis most commonly has two main options: a constant mean difference over time, or a linear divergence between mean curves over time. The former we addressed previously,5 while here we focus on the latter case of linear divergence.

*Correspondence to: Lars J. Frison, Biostatistics, Astra Hässle, AB, S-431 83 Mölndal, Sweden. E-mail: lars.frison@hassle.se.astra.com

CCC 0277–6715/97/242855–18$17.50
© 1997 by John Wiley & Sons, Ltd.
The most frequently used summary statistic for estimating linearly divergent treatment effects is the individual subject’s linear regression coefficient (SLOPE). This approach entertains wide usage, dating back at least to Wishart, and has the advantage that calculation, interpretation and communication are straightforward. One might also be deceived into thinking it has high statistical efficiency. However, since observations on a given subject are intercorrelated, least squares is not optimal, but merely convenient (see Potthoff and Roy). Our aim is to define alternative summary statistic approaches to analysing linear divergence which have substantially increased statistical efficiency compared with SLOPE.

In Section 2 we define the problem with a simple appropriate model, areas of application, and an illustrative example. In Section 3 we derive the optimal linear summary statistic for any repeated measures design with assumed known covariance structure and shape of true mean treatment difference over time. The notion of asymptotic relative efficiency is introduced to facilitate comparisons between summary statistics. Section 4 defines the four alternative summary statistics for linear divergence, with expected values and variances, and in Section 5 efficiency comparisons are presented. Section 6 discusses the choice of alternative hypothesis, in particular between constant mean difference and linear divergence, and consequent issues of robustness and efficiency are explored. Section 7 is devoted to related design issues, especially statistical power considerations. Section 8 focuses on analyses of the example introduced in Section 2, and Section 9 presents discussion and conclusions.

2. ESTIMATING THE RATE OF LINEAR DIVERGENCE BETWEEN TREATMENTS

The concept ‘rate of change’ may be considered at different stages. It may relate to: the experience of an individual subject; the mean curve for a given treatment group; and finally to a linear divergence between two treatment groups. The latter case is what matters for the power of a subsequent statistical test. Hence, we are concerned both with trials in which linearity occurs in each treatment group, and trials in which linearity only applies to the difference between the two mean curves, as shown in Figure 1. The analysis principles explored in this paper relate to both situations.

For our purpose the following simple model for a repeated measures design, comparing two treatment groups, is appropriate.

Suppose we have response measurements \(x_{ijkt}\), where \(i\) indexes treatment group (\(i = C\) (control) or \(T\) (active treatment)) \(j\) indexes subjects within group \((j = 1, \ldots, n_i)\), \(k\) indexes the repeated measurements \((k = 0, \ldots, r)\), and \(t_k\) the point in time when the \(k\)th measurement was obtained. Let

\[
x_{Cjt_k} = \mu_{t_k} + \varepsilon_{Cjt_k}
\]

and

\[
x_{Tjt_k} = \mu_{t_k} + \beta(t_k - t_0) + \varepsilon_{Tjt_k}.
\]

Our primary interest is in estimating \(\beta\) the rate of linear divergence between treatments per time unit, \(\mu_{t_k}\) represent the underlying true mean response at time \(t_k\) in the control group, and \(\varepsilon_{ijkt}\) are the individual patient variations around these mean effects which are correlated within subject.

Let us introduce a practical example. A randomized trial of 46 patients compared to groups for the treatment of asthma during a 42-week period with five assessments obtained of the \(PD_{20}\)
score over time for each patient. One of these measurements was taken pre-entry and four during treatment. Figure 2 shows the mean curves (of log(PD_{20})) over time for the two groups. There is clear indication from the figure of a divergence between mean curves over time and we will return to this example in Section 8.

3. THE OPTIMAL LINEAR SUMMARY STATISTIC

Before exploring the efficiency of different summary statistics for the specific problem of linear divergence, it is useful to tackle the more general problem for any shape of true treatment differences over time.

Consider any choice of repeated measures design, assumed shape of mean treatment divergence over time, and assumed underlying covariance structure. By extending the results of O'Brien and
Pocock et al., it is possible to derive an optimal linear summary statistic in the sense of maximizing the power to detect a treatment effect, as follows.

A linear summary statistic for each individual subject $j$ (based on $p$ pre-randomization and $r$ post-randomization measurements) is defined as $S_{ij} = \sum_{k=-(p-1)}^{c} c_k x_{ijk} = c^T x_{ij}$, where $c$ is a vector of weights to be chosen. The means $S_c$ and $S_t$ are then compared straightforwardly, leading to an asymptotically Normal overall test statistic and corresponding point and interval estimates of treatment difference. For any choice of weights $c$, denoting the underlying mean vector for treatment group $i$ by $q_i$ and the within-subject covariance matrix by $\Sigma$, $E[S_{ij}] = c^T \tau_i$ and $\text{var}[S_{ij}] = c^T \Sigma c$.

**Theorem:**

If $\gamma' = (\tau_T - \tau_C)$ is the known vector of true mean treatment differences over time, and $\Sigma$ is the covariance matrix (assumed known, and identical between treatments), the optimal choice of weights, $c'$, for a linear summary statistic are proportional to $c' \Sigma^{-1}$.

Note $c'$ depends on proportionate differences, that is, the magnitude of $\gamma'$ does not matter but shape does. A proof of this theorem is in the Appendix. The optimality theorem may be viewed either as an extension of O'Brien's generalized least squares procedure, or as an application of results for Fisher’s linear discriminant function (see, for example, Chatfield and Collins\(^\text{10}\)), adapted to repeated measures designs. Indeed, these coefficients were derived by Fisher\(^\text{11}\) (1936) when searching for a linear combination of a set of variables which had maximum between-group difference relative to its within-group standard deviation.

To contrast such an optimal choice with other summary statistics we use the notion of asymptotic relative efficiency (ARE, see Dawson and Lagakos\(^\text{12}\)). Then for any two given summary statistics, $S_1 = c_1^T X$ and $S_2 = c_2^T X$:

$$\text{ARE}(S_1 : S_2) = \frac{(c_1^T [\tau_T - \tau_C])^2 / \text{var}[S_1]}{(c_2^T [\tau_T - \tau_C])^2 / \text{var}[S_2]}.$$

This ARE is equal to the ratio of their squared non-centrality parameters (Cox and Hinkley\(^\text{13}\)). The appropriateness of this formula follows from the asymptotic Normality of linear summary statistics.

The ARE reflects the relative power of one statistical test to another. For example, if the ARE of $S_1$ to $S_2$ in a particular setting is 0.75, this means that using $S_2$ would be more efficient than $S_1$, in that, asymptotically, only 75 per cent as many subjects would be needed for it to have the same power. Thus, AREs that are close to one imply that the two summary statistics have similar power against the treatment effect being considered, while values far from one imply the opposite. For two summary statistics with given $c'$-vectors, and which have equal expected values, the ARE simplifies to a variance ratio.

4. FOUR ALTERNATIVES FOR ANALYSING LINEAR DIVERGENCE

Returning to the specific problem of linearly divergent treatment effects, we now focus on four alternative summary statistics as follows:
(i) SLOPE: that is, each individual subject’s linear regression coefficient of response on time as a summary statistic. For individual \( j \) in treatment group \( i \) the summary statistic is

\[
\text{SLOPE}_{ij} = \frac{\sum_{k=0}^{r} (t_k - \bar{t}) x_{ijk} \kappa}{\sum_{k=0}^{r} (t_k - \bar{t})^2}.
\]

The observed treatment difference in mean SLOPEs is, \( \text{SLOPE}_T - \text{SLOPE}_C \), which has expected value \( \beta \) which represents the mean rate of treatment divergence per unit time.

For example, with one baseline measurement and \( r \) equidistant post-randomization measurements, \( \mathbf{e} \propto [-r \ -r + 2 \ \ldots \ r - 2 \ r] \) is the set of weights for fitting individual patient slopes.

Its variance can be obtained from the general variance formula in Section 3, but is quite complex for a general covariance matrix \( \Sigma \). However, if we assume compound symmetry, that is, constant variance over time and equicorrelation \( \rho \) between all time points is both treatment groups, then

\[
\text{var}[\text{SLOPE}_T - \text{SLOPE}_C] = \left( \frac{1}{n_A} + \frac{1}{n_B} \right) \frac{(1 - \rho)\sigma^2}{\sum_{k=0}^{r} (t_k - \bar{t})^2}.
\]

For more explicit comparisons between summary statistics if is convenient to assume equidistant time-intervals between measurements. For notational convenience we set both the time-intervals and the treatment difference in mean SLOPEs, \( \beta \), to equal unity. This is a straightforward rescaling which changes the unit of measurement, but does not affect the efficiency of analyses in any way.

Then, we may redefine our summary statistic SLOPE\(_{ij}\) as

\[
\text{SLOPE}_{ij} = \sum_{k=0}^{r} \frac{(2k - r)x_{ijk}}{(r + 2)\binom{3}{2}}.
\]

Note: the simplifications introduced are only for comparative purposes, SLOPE and the other summary statistics may be used under much more general circumstances.

(ii) RTO: ‘regression through the origin’, may be calculated as

\[
\text{RTO}_{ij} = \frac{\sum_{k=0}^{r} (t_k - t_0)x_{ijk} \kappa}{\sum_{k=1}^{r} (t_k - t_0)}.
\]

However, to obtain the same expected value as SLOPE, for the situation with equidistant unit time-intervals, the following rescaled version will be used:

\[
\text{RTO}_{ij} = \sum_{k=1}^{r} \frac{6k}{r(r + 1)(2r + 1)} x_{ijk}.
\]
There are two reasons behind its introduction, neither of which relates to statistical efficiency. First, it has been put forward as an alternative to SLOPE (Senn\(^14\)) because SLOPE gives negative weights to the earlier post-treatment measurements, where positive treatment differences are anticipated, which is something that RTO avoids. Second, RTO is introduced because it helps to understand the relation between ANCOVA (the mean of each subject’s post-treatment measurements adjusted for the mean of the pre-treatment measurements), the optimal linear summary statistic for a constant mean treatment difference\(^5\) and the corresponding optimal choice under linear divergence (see below).

(iii) SLANC: An obvious way to improve the efficiency of SLOPE, in an ANCOVA spirit, is to use SLOPE as a dependent variable in an analysis of covariance model with the pre-entry measurement (PRE) for each subject as covariate. This we call SLANC (a SLOPE-based ANCOVA).

The summary statistic SLANC may be calculated as follows:

\[
SLANC_{ij} = \frac{\sum_{k=0}^{r} (t_k - \bar{t})x_{ijk}}{\sum_{k=0}^{r} (t_k - \bar{t})^2} - \hat{\gamma}_{\text{PRE,SLOPE}}(x_{ij0} - \bar{x}_{..0}) = \text{SLOPE}_{ij} - \hat{\gamma}_{\text{PRE,SLOPE}}(x_{ij0} - \bar{x}_{..0})
\]

where \(\hat{\gamma}_{\text{PRE,SLOPE}}\) is the estimated regression coefficient of SLOPE on PRE.

Applying the same assumptions and scalings as above, the summary statistic may be obtained from:

\[
SLANC_{ij} = \sum_{k=0}^{r} \left( \frac{(2k - r)x_{ijk}}{(r + 2)} + \frac{r(1 - \rho)x_{ij0}}{(r + 2)} \right).
\]

(iv) SLAIN: the fourth alternative summary statistic is derived using the optimal linear summary statistic theorem from Section 3. As proved in the Appendix the optimal choice under linear divergence and compound symmetry is based on an analysis of covariance model, however, slightly different from the model used for SLANC. Again we have SLOPE as dependent variable, but instead of the pre-entry measurement, we have the estimated intercept for each subject as covariate. This approach has previously been suggested by other authors.\(^15\)~\(^17\) A more heuristic explanation of why SLAIN improves on SLANC is given in the next section.

Denoting the estimated intercept for subject \(j\) in group \(i\) by \(\text{INTER}_{ij}\) we may obtain our summary statistic from \(\text{SLAIN}_{ij} = \text{SLOPE}_{ij} - \hat{\gamma}_{\text{INTER,SLOPE}}(\text{INTER}_{ij} - \hat{\text{INTER}}_\cdot)\), where \(\hat{\gamma}_{\text{INTER,SLOPE}}\) is the estimated regression coefficient of SLOPE on INTER.

Applying, once more, the same assumptions and scalings as earlier, SLAIN\(_{ij}\) may be obtained from

\[
\text{SLAIN}_{ij} = \sum_{k=0}^{r} \left( \frac{12k + 6\rho r(2k - r - 1)}{r(r + 1)[\rho r(r - 1) + 2(2r + 1)]} \right)x_{ijk}.
\]

Note that under the current scalings and assumptions, all the summary statistics estimate the same \(\beta\) and have the same expected value. On average, over all clinical trials, all four summary statistics are unbiased. However, given an inevitable observed pre-treatment mean difference, \(\bar{x}_{A,0} \neq \bar{x}_{B,0}\), the two unconditional summary statistics (SLOPE and RTO) will be biased, while the two conditional summary statistics (SLAIN and SLANC) will be unbiased.
To show the promised correspondence between summary statistics for linear divergence, relative to summary statistics for a stable treatment difference, we have, once more, to rescale RTO. The rescaled version $RTO^*$, gives the same total weight as SLOPE to the post-treatment measurements, that is

$$
\sum_{k=1}^{r} c_k^{RTO^*} = \sum_{k=1}^{r} c_k^{SLOPE}.
$$

As a consequence

$$
RTO^* = \sum_{k=1}^{r} \frac{2k x_{ijk}}{(r + 1)(r + 2)}.
$$

As already pointed out, such rescaling does not affect the performance of the summary statistic, only the units of measurement.

Under compound symmetry, with equicorrelation $\rho$ between the repeated measurements, we have the following relationship for our mean summary statistics (ANCOVA was briefly introduced earlier, POST is the mean of all post-treatment measurements for each subject, and CHANGE equals POST minus baseline):

$$
ANCOVA = (1 - \rho)POST + \rho CHANGE.
$$

Under linear divergence we have a corresponding relationship:

$$
SLAIN = (1 - \rho)RTO^* + \rho SLOPE.
$$

5. EFFICIENCY COMPARISONS

In general, for any particular choice of alternative hypothesis and covariance structure we may compare the efficiency of the four alternative summary statistics (and any others) in terms of asymptotic relative efficiencies. For the specific case of true linear divergence and compound symmetry we may make more simple direct comparisons in terms of variances of estimates.

For instance, Table I shows the variances under compound symmetry and equidistant time-intervals. From these formulae it is easy to show that for any choice of $r$ and $\rho$, SLAIN and SLANC are more efficient in terms of precision than SLOPE. It is also evident that the variance for SLAIN is lower than the variance for SLANC. The reason behind this is as follows. PRE and INTER have the same expected value, and the same covariance with SLOPE (which equals $\sigma^2(\rho - 1)$), while the variance for INTER is always less than the variance for PRE. Actually, under compound symmetry,

$$
\text{var}[PRE] = \sigma^2, \quad \text{while} \quad \text{var}[INTER] = \sigma^2 \left(1 - \frac{(r - 1)(1 - \rho)}{(r + 1)(r + 2)}\right).
$$

The relationship between the variance for SLAIN, SLANC, SLOPE, RTO and ANCOVA are illustrated in Figure 3. The reason for incorporating ANCOVA is that we want to convey some feeling for how the efficiency of mean summary statistics decreases when the underlying model is one of linear divergence.
Table I. Variances, assuming a known covariance matrix, for the four summary statistics under linear divergence, compound symmetry, and with equidistant time-intervals between measurements. \((1/n_a + 1/n_b)\sigma^2\) has been factored out from the variances

\[
\begin{align*}
\text{var}[\text{SLOPE}_T - \text{SLOPE}_C] &= \frac{2(1 - \rho)}{r + 2} \\
\text{var}[\text{SLANC}_T - \text{SLANC}_C] &= \frac{2(1 - \rho)}{r + 2} \left[ 1 - \frac{r^2(1 - \rho)}{2} \right] \\
\text{var}[\text{SLAIN}_T - \text{SLAIN}_C] &= \frac{2(1 - \rho)}{r + 2} \left[ 1 - \frac{r^2(1 - \rho)}{2} \right] \left[ 1 - \frac{(r - 1)(1 - \rho)}{(r + 1)(r + 2)} \right] \\
\text{var}[\text{RTO}_T - \text{RTO}_C] &= \frac{3(r + 2)(1 - \rho) + 9r(1 + r\rho)}{r(r + 1)(2r + 1)^2}
\end{align*}
\]

Figure 3(a) reveals that the degree of inferiority for SLOPE relative to SLAIN depends primarily on the correlation. With \(\rho\) in the plausible range 0.5 to 0.7 (see reference 5) the ARE stays around 0.80. This implies that a SLAIN analysis would require about 20 per cent fewer subjects to obtain the same power as a SLOPE analysis. With increasing \(r\) SLOPE gets slightly closer to SLAIN in relative efficiency. From Figure 3(b) we see that, for high correlations, SLANC is almost as powerful as SLAIN; only with correlations below 0.5 will the ARE drop below 0.9. With a substantial number of post-treatment measurements SLANC loses somewhat more in efficiency relative to SLAIN. Figure 3(c) indicates that, with few post-treatment measurements, ANCOVA is relatively powerful even under linear divergence, especially when correlations are low. However, the more \(r\) increases the more superior SLAIN will be relatively to ANCOVA, provided of course that linear divergence really does exist. From Figure 3(d) it is evident that RTO performs very poorly in this kind of comparison. To exemplify this, with 1 + 5 measurements pre- and post-randomization, and a \(\rho\) of 0.5, the ARE relative to SLAIN is 0.47; with a \(\rho\) of 0.7 the ARE drops even further to 0.26. In general, the more post-treatment measurements, and the higher the correlation, the worse the case for RTO. However, note that whereas POST adjusted for PRE, and CHANGE adjusted for PRE, are both called ANCOVA, so RTO adjusted for INTER and SLOPE adjusted for INTER both lead to SLAIN.

We now study the effect departures from compound symmetry have on the efficiency comparisons. In particular we look at declining correlations with increasing time-intervals between measurements, and on variances that change over time. For this purpose Table II provides several illustrative examples, each with \(r = 3\) post-randomization visits for simplicity.

In each case the summary statistic OPTI in the table refers to the optimal choice as defined in Section 3. The four columns underneath OPTI refer to the weights, \(C_0\), \(C_1\), \(C_2\) and \(C_3\), given to the four repeated measurements by this summary statistic.
In the top half of Table II, the consequence of declining correlations over time is illustrated for a design with one pre- and three post-randomization visits. Here, a banded correlation structure is assumed with correlation coefficients decided in such a way that the overall average correlation, for the within-subjects covariance matrix, in all instances remains at 0·7. The degree of decline in correlation for each further visit apart is 0, 0·05, 0·10 and 0·20, respectively, for the four rows reported here.

In the bottom half of Table II, equicorrelation with a $\rho$ of 0·7 is assumed, now the consequences of departure from homoscedasticity is illustrated. Once again four visits are assumed, and four different scenarios are illustrated, with variance for the four time-points assumed as: 1, 0·9, 0·8, 0·7;
Table II. Optimal linear summary statistics (OPTI) for linearly divergent mean response curves and under covariance structures different from compound symmetry. Asymptotic relative efficiencies compared with other summary statistics (in all instances $\gamma' = [0, 1, 2, 3]$)

<table>
<thead>
<tr>
<th>$\Sigma$ (banded)</th>
<th>OPTI</th>
<th>SLAIN</th>
<th>SLANC</th>
<th>SLOPE</th>
<th>RTO</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.70</td>
<td>-0.18</td>
<td>0.33</td>
<td>0.85</td>
<td>1</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.733</td>
<td>0.683</td>
<td>0.633</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.63</td>
<td>-0.14</td>
<td>0.29</td>
<td>0.85</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.767</td>
<td>0.667</td>
<td>0.567</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.57</td>
<td>-0.10</td>
<td>0.22</td>
<td>0.88</td>
<td>0.98</td>
<td>0.78</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.833</td>
<td>0.633</td>
<td>0.433</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.43</td>
<td>0.05</td>
<td>-0.11</td>
<td>1.06</td>
<td>0.92</td>
<td>0.90</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.58</td>
<td>-0.20</td>
<td>0.29</td>
<td>0.91</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.745</td>
<td>0.567</td>
<td>0.36</td>
<td>0.81</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>-0.80</td>
<td>-0.17</td>
<td>0.36</td>
<td>0.81</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.98</td>
<td>-0.13</td>
<td>0.39</td>
<td>0.74</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.35</td>
<td>-0.07</td>
<td>0.41</td>
<td>0.66</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>$\sigma^2$: 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$: 1</td>
<td>0.767</td>
<td>0.667</td>
<td>0.567</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.82</td>
<td>0.02</td>
<td>0.29</td>
<td>0.69</td>
<td>0.97</td>
<td>0.96</td>
</tr>
</tbody>
</table>

2864 L. J. FRISON AND S. J. POCOCK

1, 1·1, 1·2, 1·3; 1, 1·33, 1·67, 2, and finally 1, 2, 3, 4. In the final row of the table, the joint effect of increasing variances and declining correlations over time is illustrated.

For all departures from compound symmetry investigated, SLAIN stays close to OPTI in relative efficiency. Only with an extreme increase in variance will the ARE drop below 0·9. The difference between SLAIN and SLANC is very small in this table, though with lower correlations, or with more post-treatment measurements, the advantage for SLAIN would increase. ANCOVA gets relatively more powerful as correlations decline over time and/or variances increase with time. SLOPE becomes a more inefficient approach to analysis as correlations and/or variances decrease with increasing time-intervals since randomization. Comprehensive results for increasing $r$ are difficult to convey, since these will depend on whether we assume that an increasing number of post-treatment measurements will imply a prolonged study period or shorter time-intervals between visits. Typically both ANCOVA and SLOPE will lose in efficiency relative to OPTI when $r$ gets larger.

The choice of repeated measures designs above illustrate some plausible scenarios. It should be noted that anyone can usefully 'play these games' as an aid to determining both the best design and the best analysis strategy for their particular trial.

### 6. CONSTANT MEAN DIFFERENCE OR LINEAR DIVERGENCE

The best choice of summary statistic, as far as efficiency and informativeness is concerned, depends on the clinical objectives of the study, the covariance structure, and the hypothesized mean difference between treatment groups over time. Of key importance is the primary objective of the study. This emphasizes one main attraction of the summary statistic approach: the possibility for a tailor-made analysis. With regard to trends over time it is worth noting that the actual shapes of the group mean profiles have no direct influence on the analysis; it is the difference between the mean profiles that matters.

The choice of alternative hypothesis for the difference in time trends between treatment groups is often not clear cut. For instance, it is often anticipated that the difference between the mean curves will increase over the whole study period, but the actual rate of divergence might diminish over time. This might be termed attenuated divergence, and lies between a constant effect and linear divergence. In design situations where it is unclear exactly which shape of alternative hypothesis to expect, it is of interest to know how robust are the different approaches to analysis. To investigate this, Table III presents some hypothetical examples in the same spirit as Table II, again for $r = 3$ post-randomization measurements. It compares the AREs for the summary statistics under investigation under some alternative shapes of difference between mean curves over time.

It is apparent that SLAIN is robust to moderate departures from linear divergence between mean treatment curves, for example, when attenuated or exponential divergence applies. Even for a constant mean difference SLAIN performs better than SLOPE, RTO or SLANC. Robustness against misspecifications for the underlying covariance structure is also noted. ANCOVA, the optimal linear summary statistic under a constant mean difference and compound symmetry, is seen to possess a similar robustness from it underlying assumptions. For attenuated divergence there is little to choose between ANCOVA and SLAIN. Should a treatment effect be transient, only ANCOVA preserves any reasonable power. It should be noted that the robustness alluded to above relates to statistical power. Estimated intercepts are sensitive to model misspecification, as noted by Laird and Wang. Expected values

Table III. Optimal linear summary statistics for some different classes of vectors of mean treatment differences, and for two different correlation structures. AREs compared with some other summary statistics

<table>
<thead>
<tr>
<th>Σ (banded)</th>
<th>OPTI</th>
<th>SLAIN</th>
<th>SLANC</th>
<th>SLOPE</th>
<th>RTO</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential divergence:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(γ’ = [0, 0.5, 2, 5])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ²: 1 0.7 0.7 0.7</td>
<td>-0.70</td>
<td>-0.49</td>
<td>0.13</td>
<td>1.37</td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>ρ: 1 0.767 0.667 0.567</td>
<td>-0.40</td>
<td>-0.66</td>
<td>-0.31</td>
<td>1.97</td>
<td>0.80</td>
<td>0.85</td>
</tr>
<tr>
<td>Linear divergence:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(γ’ = [0, 1, 2, 3])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ²: 1 0.7 0.7 0.7</td>
<td>-0.70</td>
<td>-0.18</td>
<td>0.33</td>
<td>0.85</td>
<td>1</td>
<td>0.98</td>
</tr>
<tr>
<td>ρ: 1 0.767 0.667 0.567</td>
<td>-0.57</td>
<td>-0.10</td>
<td>0.22</td>
<td>0.88</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>Attenuated divergence:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(γ’ = [0, 1, 1.5, 1.75])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ²: 1 0.7 0.7 0.7</td>
<td>-0.70</td>
<td>0.03</td>
<td>0.39</td>
<td>0.58</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td>ρ: 1 0.767 0.667 0.567</td>
<td>-0.64</td>
<td>0.21</td>
<td>0.34</td>
<td>0.45</td>
<td>0.87</td>
<td>0.80</td>
</tr>
<tr>
<td>Constant mean difference:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(γ’ = [0, 1, 1, 1])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ²: 1 0.7 0.7 0.7</td>
<td>-0.70</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.66</td>
<td>0.54</td>
</tr>
<tr>
<td>ρ: 1 0.767 0.667 0.567</td>
<td>-0.71</td>
<td>0.60</td>
<td>0.22</td>
<td>0.18</td>
<td>0.51</td>
<td>0.41</td>
</tr>
<tr>
<td>Transient effect:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(γ’ = [0, 2, 1, 0.5])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ²: 1 0.7 0.7 0.7</td>
<td>-0.70</td>
<td>1.07</td>
<td>0.19</td>
<td>-0.26</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>ρ: 1 0.767 0.667 0.567</td>
<td>-0.82</td>
<td>1.24</td>
<td>0.03</td>
<td>-0.28</td>
<td>0.02</td>
<td>0.00</td>
</tr>
</tbody>
</table>
for the intercepts may differ between treatment groups under truly non-linear underlying models.

7. DESIGN CONSIDERATIONS

As in the conventional approach to power calculation we define \( \alpha \) and \( \beta \) as the type I and type II errors for the test of our hypothesis. It is convenient to assume that sample sizes are large enough that the Normal approximation to the \( t \)-distribution can be applied. In that case, for two equal sized treatment groups of size \( n \), for a general summary statistic \( S_{ij} = \sum_{k=-(p-1)}^0 c_k x_{ij,k} = c' x_{ij} \), under a general shape for the alternative hypothesis \( (\tau_T - \tau_C = \gamma) \) and with a general \( \Sigma \), we require that

\[
n = \frac{2\epsilon' \Sigma \epsilon}{(c' \gamma)^2} f(\alpha, \beta)
\]

where \( f(\alpha, \beta) = [\Phi^{-1}(1 - \alpha/2) + \Phi^{-1}(1 - \beta)]^2 \), \( \Phi \) being the cumulative distribution of a standardized Normal deviate.

Correspondingly, given the sample sizes, the approximate power may under this general scenario be calculated from

\[
1 - \beta = \Phi \left( \frac{c' \sqrt{(n/2)}}{\sqrt{(c' \Sigma c)}} + \Phi^{-1}(\alpha/2) \right).
\]

As illustration, consider the alternative hypothesis of linear divergence resulting in a mean difference of \( \delta = 0.5 \sigma \) at the end of the study period (irrespective of the number of measurements, the length of study is considered fixed). Further let the type I error \( \alpha = 0.05 \) and the type II error \( \beta = 0.2 \), also let \( \rho = 0.6 \), often a realistic value for practical use. Figure 4 shows the required sample size, \( n \), in each group for a variety of study designs and analysis approaches: for \( r = 1, \ldots, 8 \) post-treatment measurements, and for SLOPE, SLANC, SLAIN, RTO and ANCOVA.

It is worth pointing out that other choices of \( \delta \) simply require sample sizes in Figure 4 to be multiplied by \( (0.5\sigma/\delta)^2 \).

8. AN EXAMPLE: THE ASTHMA TRIAL

We now return to the illustrative example introduced in Section 2. From Figure 2 an approximately linear divergence over time was observed between mean curves for the log-transformed \( \text{PD}_{20} \) measurements, the observed vector of mean differences for the five visits being \( [0.19, 0.49, 1.03, 1.24, 1.90] \). It may also be noted that the variance remained stable over time. The average correlation between repeated measurements was 0.66, with some indication of a decline in correlation with increasing time-intervals between measurements. The correlation between visits adjacent in time averaged 0.73, while the correlation between baseline and the 42-week measurement was 0.51. In line with the results derived in previous sections we would expect SLAIN to be the most efficient summary statistic analysis for this example. Indeed this was the case, as shown in Table IV.

All approaches are seen to produce very similar estimates of the mean gain in rate of change over time for the active treatment over the control treatment. However the precision is seen to be better for the two conditional approaches. An illustration of where the increase in precision comes
From is provided by Figure 5. The estimated SLOPE for each patient is given on the y-axis, and the estimated intercepts on the x-axis, with different symbols identifying treatment groups. The regression lines for SLOPE on INTER are given, and assumed to be parallel between groups. The quite strong correlation between SLOPE and INTER ($r = -0.43$) explains why the covariate adjustment is beneficial.

9. DISCUSSION AND CONCLUSIONS

Sound statistical advice is that ‘the more insight you put into an analysis, the more you can expect out’. This applies particularly to the choice of summary statistic analysis of a repeated measures design.

The more information that is available from past experience and medical knowledge about expected treatment effects and within-subject dependencies, the more sensible the choice can be.

---

Table IV. SLAIN, SLANC, SLOPE and RTO analysis of the log PD$_{20}$ data: $n = 46$ patients in total, $r = 4$ measurement post-treatment, $p = 1$ measurement pre-treatment

<table>
<thead>
<tr>
<th></th>
<th>Estimated difference in mean SLOPE</th>
<th>Standard error</th>
<th>$t$-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLAIN</td>
<td>0.045</td>
<td>0.00765</td>
<td>5.89</td>
</tr>
<tr>
<td>SLANC</td>
<td>0.043</td>
<td>0.00782</td>
<td>5.51</td>
</tr>
<tr>
<td>SLOPE</td>
<td>0.045</td>
<td>0.00836</td>
<td>5.35</td>
</tr>
<tr>
<td>RTO</td>
<td>0.044</td>
<td>0.01044</td>
<td>4.19</td>
</tr>
</tbody>
</table>
made for a summary statistic to increase both validity and sensitivity. The extent to which such information exists will vary depending on the therapeutic area, the type of variable being measured, and the clinical phase the treatment is in. Of course, there will never be perfect insight, but some experience will usually be available.

Unless more refined knowledge is available, the general advice is: use ANCOVA in situations where a reasonably stable mean treatment difference over time is expected, and use SLAIN when a divergence between mean curves seems plausible. This will almost always result in reasonably valid and efficient inferences.

Not to use an analysis of covariance, when a pre-entry measurement is available, will always be a mistake. Apart from a highly probable loss of efficiency, there is also risk that the results may be biased. For instance, CHANGE will overcorrect for a chance mean pre-treatment difference, and so will SLOPE. POST and RTO, on the other hand, simply ignores any such imbalance.

Our comparisons between summary statistics for the analysis of rate of change have, for simplicity, been done assuming one baseline recording only. Additional increases in precision may be gained by allowing for multiple pre-entry measurements, though, not as impressive as in the mean summary statistics case.\(^5\) However, the relative ordering of the summary statistics will be the same, and adjusting the estimated slope for the estimated intercept will, under compound symmetry, still be the optimal option.

The poor performance reported for RTO may be improved by adjusting for pre-entry measurements. As noted in Section 5, adjusting for the estimated intercept we ‘re-invent’ SLAIN. Adjusting RTO for the pre-entry measurements its relative efficiency will be comparable to SLOPE. It will be more efficient than the other slope-based statistics under a constant mean difference model, and better than all summary statistics in Table III for attenuated divergence. In summary, RTO adjusted for baseline recordings constitutes a robust, though never optimal, choice.
It should be kept in mind that all efficiency comparisons are asymptotic. No correction factors allowing for estimation of regression coefficients, and for the loss of a degree of freedom, have been adjusted for. Similarly, all statements of optimality have been made conditional on knowledge of the true covariance structure.

Missing values emerge in all clinical trials, in particular for repeated measurements designs, and will imply complications. The summary statistics will often still be estimable for each individual, but assumptions (for example, of equal weights) may be inappropriate and statistical properties somewhat affected. Should there be a vast amount of missingness other methods might be called for, and multi-level models constitute one promising approach under such circumstances (see Beacon and Thompson18).

While a pre-specified analysis strategy is always a good idea, it is worth considering allowance of a cautious flexibility in the choice of summary statistic if the pre-defined $\gamma'$ is clearly wrong. An example was this occurred is given in the childhood asthma trial reported by Van-Essen Zandvliet et al.19, where an assumed linear divergence between mean curves for FEV$_1$ turned out to be a stable treatment effect over the study period. Restricting oneself to always stick rigidly to prespecified summary statistics under all circumstances would be a poor scientific approach, though one also needs to guard against post hoc manipulative choices of analysis aimed at squeezing out extra significance.

In conclusion, our intention has been to present readily applicable ideas on the efficient analysis of repeated measures data in clinical trials using a summary statistic approach. We hope this will enhance the relatively straightforward presentation of meaningful results of such studies and also improve the statistical basis for their design.

APPENDIX

Proof of the optimal linear summary statistic theorem from Section 3

This result follows from the extended Cauchy–Schwarz inequality, which states: let $\beta$ and $\gamma$ be any two $p + r$ (where $p$ and $r$ are the number of pre- and post-treatment measurements, respectively) vectors and let $\Sigma$ be a positive definite $(p + r)(p + r)$ matrix. Then $(\beta'\gamma)^2 \leq (\beta'\Sigma\beta)(\gamma'\Sigma^{-1}\gamma)$ with equality if and only if $\beta = k\Sigma^{-1}\gamma$ (or $\gamma = k\Sigma\beta$) for some constant $k$. A proof is given on page 64 in Johnson and Wichern.20

Then, for an arbitrary non-zero $p + r$ vector $c$,

$$\max_{c \neq 0} \frac{(c'\gamma)^2}{c^T\Sigma c} = \gamma'\Sigma^{-1}\gamma$$

with the maximum attained when $c' = k\gamma\Sigma^{-1}$ for any constant $k \neq 0$.

Proof of optimality for SLAIN under linear divergence and compound symmetry

We will give a proof, based on a derivation by Schouten,21 of the optimality of SLAIN under linear divergence and compound symmetry, for the special case of equidistant unit time-intervals between measurements and one pre-entry evaluation.

From the optimal linear summary statistic theorem we known that, in general the optimal choice is $\gamma'\Sigma^{-1}$. First, we rewrite this optimal choice in the form of the general linear summary statistic, as given in Section 3, then OPTI = $\sum_{i=0}^{r} c_i y_i$. Utilizing results from Rao,22 dealing with
Fisher’s linear discriminant function, the weights, the $c_i$'s, may be written as $c_i = \sum_{j=0}^{r} (\tau_{Tj} - \tau_{Cj}) \sigma^{ij}$, where the $\sigma^{ij}$ are the elements of the inverse $\Sigma^{-1}$. Dividing all measurements by $\sigma^2$ (arbitrary scaling), the covariance matrix $\Sigma$ has elements 1 on the main diagonal and $\rho$ off the main diagonal.

The inverse matrix $\Sigma^{-1}$ has elements

$$
\sigma^{ii} = \frac{1 + (r - 1)\rho}{1 + (r - 1)\rho - \rho^2}
$$
on the main diagonal, and

$$
\sigma^{ij} = \frac{-\rho}{1 + (r - 1)\rho - \rho^2}
$$
off the main diagonal. Then, the optimal summary statistic has weights

$$
c_i = [1 + (r - 1)\rho - \rho^2] \sum_{j=0}^{r} (\tau_{Tj} - \tau_{Cj}) \sigma^{ij}.
$$

With equidistant unit time-intervals this simplifies to

$$
c_i = [1 + (r - 1)\rho - \rho^2] \sum_{j=0}^{r} j \sigma^{ij}
$$

$$
= [1 + (r - 1)\rho - \rho^2] \left( \left[ \frac{1 + (r - 1)\rho}{1 + (r - 1)\rho - \rho^2} \right] + \left( \frac{r(r + 1)}{2} - i \right) \left[ \frac{-\rho}{1 + (r - 1)\rho - \rho^2} \right] \right)
$$

$$
= i[1 + (r - 1)\rho] - \rho \left[ \frac{r(r + 1)}{2} - i \right].
$$

After multiplication by

$$
\frac{12}{r(r + 1)[\rho \cdot r(r - 1) + 2(2r + 1)]}
$$

(to conform to our consistent scaling) this may be rewritten as

$$
c_i = \frac{12i + 6\rho r(2i - r - 1)}{r(r + 1)[\rho \cdot r(r - 1) + 2(2r + 1)]}
$$

which for one pre-entry measure and under compound symmetry is identical to the summary statistic labelled SLAIN above.

REFERENCES


