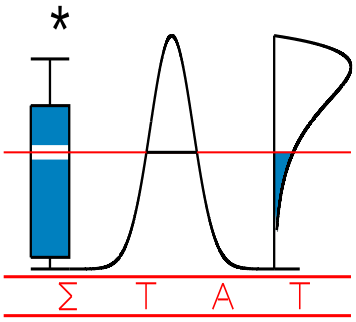
T E C H N I C A L

R E P O R T

06102

EVALUATION OF BIOMARKERS FOR PHARMACOLOGICAL ACTIVITY

ALONSO, A., MOLENBERGHS, G., RENARD, D. and C. DARSTEIN



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Evaluation of Biomarkers for Pharmacological Activity

Ariel Alonso, Geert Molenberghs
Center for Statistics, Hasselt University,
Agoralaan, B–3590 Diepenbeek, Belgium.

Didier Renard
Novartis Pharma AG, Lichtstrasse 35, Basel, Switzerland.

Christelle Darstein
Lilly Services S.A., Lilly Corporate Center, Indianapolis, USA

Abstract

In recent years, the cost of drug development has increased the demands on efficiency in the selection of suitable drug candidates. Biomarkers for efficacy and safety could be a plausible strategy to improve this selection process. In the present work, we focus on the study and evaluation of different physiological variables as biomarkers for pharmacological activity. We proposed three different approaches using multivariate and univariate techniques. We note that even though one could argue that the multivariate procedure is more powerful than the other alternatives, the univariate methods also offer a great flexibility to answer interesting scientific questions. The three approaches were used to analyze a crossover study involving an opioid antagonist.

Keywords: biomarkers, surrogate marker, crossover, optimization.

1 Introduction

The rising costs of drug development and the challenge of facing new and re-emerging diseases are putting considerable demands on efficiency in the selection of suitable drug candidates. An effective strategy in improving this process is the proper selection and application of biomarkers for efficacy and safety during the different stages of the drug development pipeline.

Some authors refer to a biological marker or biomarker as a variety of physiological, pathological, or anatomical measurements that are thought to relate to some aspect of a healthy or pathological process (Temple 1995, Lesko and Atkinson 2001). In the same vein, a biomarker has also been defined as a characteristic that can be measured and evaluated as an indicator of healthy biological processes, pathological processes or pharmacological responses to therapeutic intervention (NIH Biomarker Definitions Working Group, 2001). Other definitions have since emerged and the discussion on what biomarkers should be and where to apply them continues. Biomarkers are currently being used in various areas, including disease identification, target discovery and validation, volunteer/patient inclusion and stratification during clinical studies, drug efficacy and safety and prediction of drug response (Suico *et al* (2006). Such biomarkers include measurements that help identifying the etiology of certain medical problem or the progress of a disease. They also include measurements related to the mechanism of response to treatments and actual clinical responses to therapeutic interventions (Burzykowski, Molenberghs, and Buyse 2005).

From a regulatory perspective, a biomarker is not considered an acceptable endpoint for a determination of efficacy of new drugs unless it has been shown to function as a valid indicator of clinical benefit, i.e., unless it is a valid surrogate. The NIH Bio-

marker Definitions Working Group (2001) also addressed the relationship between biomarkers, clinical endpoints, and surrogate markers. A clinical endpoint is considered the most credible indicator of drug response and is defined as "a characteristic or variable that reflects how a patient feels, functions, or survives". During clinical trials, clinical endpoints should in principle be used, unless a biomarker is available that has risen to the status of a surrogate endpoint and is expected to predict either, clinical benefit, harm, or lack of both benefit and harm. Realistically, the working group points out that probably only a few biomarkers are likely to achieve a consensus surrogate endpoint status.

Biomarkers differ in their closeness to the intended therapeutic response or clinical benefit. Some biomarkers can be thought to be valid surrogates for clinical benefits, such as, for example, blood pressure or cholesterol, while they can also reflect the pathological process and could be considered potential surrogate endpoints, such as, for example, brain appearance in Alzheimer brain infarct size. The evaluation of a biomarker as potential surrogate markers has received considerable attention over the last decade and a detailed discussion of the main contributions in this area can be found in Burzykowski, Molenberghs, and Buyse (2005).

Additionally, other biomarkers have a more uncertain relation to clinical outcome but they can still reflect the drug action, such as, for example, ACE inhibition, degree of binding to a receptor, or inhibition of an agonist. In the present work, we focus on the evaluation of this type of biomarkers. More specifically, emphasis will be on the evaluation of biomarkers of pharmacological activity for a certain compound.

In Section 2, we introduce the motivating case study. Section 3 covers important aspects of the analysis of crossover trials with repeated measurements. Three meth-

ods to evaluate biomarkers for pharmacological activity are introduced in Section 4. In Section 5, the case study is analyzed.

2 Case Study

The case study is a three-period, two-treatment cross-over trial in which 15 male subjects received either Naltrexone or a matching placebo in each period on 3 consecutive days (Suico *et al* 2006).

Naltrexone is an opioid receptor antagonist, i.e., it acts by blocking the activation of opioid receptors. The goal of the study was to identify the best biomarker of pharmacological activity for this kind of compound. Several biomarkers were considered in the study: essentially, a group of variables was measured, under different conditions and at two different days. At day 1, measurements of 5 neurohormones took place: Adrenocorticotropic (Acth), Cortisol, Luteinizing (LH), Follicle-stimulating hormone (FSH), and Prolactin were taken following single dose administration of either Naltrexone or placebo. Additionally ,pupil diameter under three different light conditions: scotopic (low luminosity — 0.04 lux), mesopic Lo (medium luminosity — 0.4 lux), and mesopic Hi (high luminosity — 4.0 lux) was taken. At day 3, measurements of neurohormones and pupil diameter following a two-minute fentanyl dose infusion were recorded. A cold pressor test was also carried out at the third day, following the fentanyl infusion.

Day 1 measurements were taken following a single dose of Naltrexone; they represent the direct pharmacological action. Day 3 measurements followed a short infusion of the opioid receptor agonist fentanyl, that is, a substance that increases the activation of opioid receptors. Therefore, measurements taken at the third day show the ability of Naltrexone, an antagonist, to block the pharmacological effects of the agonist. We

Table 1: *Candidate biomarkers.*

Biomarkers	Day 1	Day 3
Acth	⋆	⋆
Cortisol	⋆	⋆
LH	⋆	⋆
FSH	⋆	⋆
Prolactine	⋆	⋆
Mesopic Hi	⋆	⋆
Mesopic Lo	⋆	⋆
Scotopic	⋆	⋆
Cp		⋆

can then consider that these measurements represent another way of evaluating the pharmacological activity of the antagonist compound.

The cold pressor test (Cp) is typically used to evaluate the analgesic effects of a compound (such as fentanyl). Thereupon, this test will show the ability of Naltrexone to block the effects of fentanyl. The test consisted in rating the pain felt by a subject during 2 minutes following immersion of the subject's hand in warm and cold baths.

The variables recorded for each subject are displayed in Table 1. Eight biomarkers were measured at the first and third days whereas one (Cp) was measured only at day three. Each combination biomarker-day is of scientific interest and hence 17 responses in total will be analyzed.

The main objective of the study is to identify biomarkers for pharmacological activity and therefore, we are primarily interested in determining for which biomarker the difference between Naltrexone and placebo is largest.

3 Crossover Designs and Repeated Measurements: Remarks for the Analysis

This was a crossover study with repeated measurements in which two treatments, three periods and two sequences were considered: ABB and BAA. This design is optimal in the sense that it allows for a minimum variance unbiased estimator for the treatment effect.

Let us denote by $Y_{ih\ell}$ the response observed on the ℓ^{th} subject in period h and in sequence group i. Additionally, we will denote by t, p, and s, the number of treatments, periods, and sequences, respectively. Note that for the ABB–BAA design, $t = 2$, $p = 3$, and $s = 2$. Further, we will define

$$\overline{Y}_{ih.} = \frac{1}{n_i} \sum_{\ell=1}^{n_i} Y_{ih\ell},$$

where n_i is the number of patients in sequence group i. According to Jones and Kenward (2003) one of the issues in modeling crossover data with repeated measurements is how best to handle both the between-period and within-period covariance structure. These authors observed that, in the two sequence design, one can avoid the need to introduce a between-period structure by exploiting the fact that all estimators take the form $\overline{A}_1 - \overline{A}_2$, where $\overline{A}_i = \sum_{h=1}^{p} a_h \overline{Y}_{ih.}$ and, for within-subject estimators, $\sum_{h=1}^{p} a_h = 0$. Conventional repeated measurement methods can then be applied to the derived subject contrast:

$$C_{i\ell} = \sum_{h=1}^{p} a_j Y_{ih\ell}.$$

For the design considered in this study, i.e., ABB–BAA, one can estimate the treatment effect using the contrast $CT_{i\ell} = -2Y_{i1\ell} + Y_{i2\ell} + Y_{i3\ell}$. Note that the treatment effect is then represented by the difference in the mean values of the two-sequence group contrast. In the following section, we introduce three possible methods to determine the biomarker on which this effect is largest.

4 Three Strategies for the Selection of the Best Biomarker

4.1 Approach I: The Ellipsoid Method

Let us now denote by μ_{ij} the mean for the i^{th} sequence at the j^{th} time point, with $i = 1, 2$ and $j = 1, \ldots, m$. Further, let us denote the mean evolution over time for the i^{th} sequence by $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{im})'$.

We shall further denote by $\widehat{\mu}_i$ the maximum likelihood estimator for the previous mean profile, computed based on a saturated linear model. The treatment effect over time $\Delta_T = \mu_1 - \mu_2$ can then be estimated as $\widehat{\Delta}_T = \widehat{\mu}_1 - \widehat{\mu}_2$, where $\widehat{\Delta}_T$ has asymptotic distribution $\widehat{\Delta}_T \sim N(\Delta_T, \Sigma_{\Delta T})$. Note that $\Sigma_{\Delta T}$ can be estimated using the estimate for the variance-covariance matrix of $(\widehat{\mu}_1, \widehat{\mu}_2)$. From multivariate analysis it is known that, asymptotically:

$$(\widehat{\Delta}_T - \Delta_T)' \widehat{\Sigma}_{\Delta T}^{-1} (\widehat{\Delta}_T - \Delta_T) \sim \chi^2_m,$$

producing confidence region:

$$R = \{\Delta_T : (\widehat{\Delta}_T - \Delta_T)' \widehat{\Sigma}_{\Delta T}^{-1} (\widehat{\Delta}_T - \Delta_T) \leq C(\alpha)\},$$

where the constant is chosen so that $P(R) = 1 - \alpha$. Testing the hypothesis $H_0 : \Delta_T = 0$ can now be done by verifying whether $0 \in R$ or, equivalently, using the test $\widehat{\Delta}_T' \widehat{\Sigma}_{\Delta T}^{-1} \widehat{\Delta}_T > C(\alpha)$. Further, let us denote by r the distance between zero and the ellipsoid defined by the frontier of R:

$$\partial R = \{\Delta_T : (\Delta_T - \widehat{\Delta}_T)' \widehat{\Sigma}_{\Delta T}^{-1} (\Delta_T - \widehat{\Delta}_T) = C(\alpha)\}.$$

Note that r is the solution of the optimization problem:

$$r = \min_{\Delta_T \in \partial R} \| \Delta_T \|^2 . \tag{1}$$

Similar to the univariate setting, we note that the larger r is, the further the ellipsoid is from the origin and therefore the larger the treatment effect is. The problem is then reduced to finding the solution of the optimization problem given in (1). The following theorem offers an analytic expression for this solution.

Theorem 1 *The solution of the optimization problem (1) is given by*

$$r = \sum_i \left(\frac{q_i \lambda}{\alpha_i + \lambda} \right)^2, \tag{2}$$

where:

a) α_i *are the eigenvalues of* $\Sigma_{\Delta T}$,

b) $q' = (q_1, q_2, \ldots, q_m) = P\Delta_T$ *with* P *an orthogonal matrix so that* $\Sigma_{\Delta T} = P' D_0 P$, *and* $D_0 = (\alpha_i)_{ii}$,

c) λ *is a root of*

$$\sum_i \frac{\alpha_i q_i^2}{(\alpha_i + \lambda)^2} = C(\alpha).$$

An outline of the proof can be found in the appendix. Under Approach I, one can then calculate, for each biomarker, the distance from zero to the corresponding ellipsoid and then choose as the best biomarker the one for which its ellipsoid is furthest away from the origin.

4.2 Approach II: The L_2–Norm Method

Let us start by considering the following model:

$$\begin{cases} Y_1(t) = f_1(t) + \varepsilon_1(t), \\ Y_2(t) = f_2(t) + \varepsilon_2(t), \end{cases}$$

where t denotes time and Y_i the response variable for group i ($i = 1, 2$), f_i is a general function that describes the average time evolution for group i, and $(\varepsilon_1(t), \varepsilon_2(t))$

follows a bivariate Gaussian distribution with mean zero and variance-covariance matrix $\Sigma(t)$.

In the absence of treatment effect, $f_1(t) = f_2(t)$ and therefore it is intuitively appealing to use the distance between f_1 and f_2 as a measure of the effect's magnitude. If we further denote the time interval by $I = [a, b]$, then we can measure the distance between f_1 and f_2 using the L_2 norm:

$$d_2(f_1, f_2)^2 = \| f_1(t) - f_2(t) \|^2 = \int_a^b [f_1(t) - f_2(t)]^2 dt. \tag{3}$$

In practice, $g(t) = f_1(t) - f_2(t)$ is unknown and hence needs to be estimated. We can estimate g, for instance, through fitting a saturated linear model for f_1 and f_2 in such standard software packages as SAS, R, or Splus. Given that we can only consider a fixed set of time points $\{t_1, t_2, \ldots, t_m\}$, fitting a saturated model merely leads to estimates of g at these prespecified values:

$$\mu_j^* = g(t_j) = f_1(t_j) - f_2(t_j) = \mu_{1j} - \mu_{2j},$$

$(j = 1, \ldots, m)$. Using the points (t_j, μ_j^*), we can approximate (3) using the trapezoidal integration formula:

$$\| f_1(t) - f_2(t) \|^2 = \| g(t) \|^2 \approx \upsilon(f_1, f_2) = \sum_{j=1}^{m-1} \frac{\mu_j^{*2} + \mu_{j+1}^{*2}}{2} \Delta_j,$$

where $\Delta_j = t_{j+1} - t_j$. Note that $\upsilon(f_1, f_2)$ can also be written as:

$$\upsilon(f_1, f_2) = \sum_{j=1}^{m} \alpha_j \mu_j^{*2} = \sum_{j=1}^{m} \alpha_j (\mu_{1j} - \mu_{2j})^2,$$

where $\alpha_j = \frac{\Delta_{j-1} + \Delta_j}{2}$ and $\Delta_0 = \Delta_m = 0$. In terms of the original time points, the weights α_j take the form $\alpha_j = (t_{j+1} - t_{j-1})/2$, with $t_0 = t_1$ and $t_{m+1} = t_m$.

If we further denote by $\widehat{\mu}_i' = (\widehat{\mu}_{i1}, \widehat{\mu}_{i2}, \ldots, \widehat{\mu}_{im})$ the maximum likelihood estimator of μ_i, $(i = 1, 2)$, then we have that:

$$\widehat{\upsilon}(f_1, f_2) = \sum_{j=1}^{m} \alpha_j (\widehat{\mu}_{1j} - \widehat{\mu}_{2j})^2. \tag{4}$$

Taking into account that $\widehat{\Delta}_T \sim N(\Delta_T, \Sigma_{\Delta T})$ with $\widehat{\Delta}_T = \widehat{\mu}_1 - \widehat{\mu}_2$ we can apply now the delta method to obtain:

$$\widehat{\upsilon}(f_1, f_2) = \sum_{j=1}^{m} \alpha_j \widehat{\Delta}_{Tj}^2 \sim N(\upsilon(f_1, f_2), \sigma^2_{N(f_1,f_2)}), \tag{5}$$

where $\sigma^2_{N(f_1,f_2)} = \delta' \Sigma_{\Delta T} \gamma$ and $\delta = (2\alpha_1 \Delta_{T1}, 2\alpha_1 2\Delta_{T2}, \ldots, 2\alpha_m \Delta_{Tm})'$. Finally, using (5) the following confidence interval is obtained:

$$CI_\alpha[\upsilon(f_1, f_2)] = \left[\widehat{\upsilon}(f_1, f_2) - z_{1-\frac{\alpha}{2}} \sigma_{N(f_1,f_2)}, \widehat{\upsilon}(f_1, f_2) + z_{1-\frac{\alpha}{2}} \sigma_{N(f_1,f_2)}\right]. \tag{6}$$

We should like to point out that $\upsilon(f_1, f_2)$ has been considered under the current Approach II as an approximation to the distance between f_1 and f_2. We are then constructing confidence intervals, not for the parameter of interest $\| f_1(t) - f_2(t) \|^2$, but rather for an approximation of this distance. If (6) contains zero, then the data are not in contradiction with the equal treatment effects hypothesis.

4.3 Approach III: Different Weights Method

In Approach II, $\upsilon(f_1, f_2)$ was considered an approximation for the L_2 distance between f_1 and f_2. However, we could consider this parameter in the following, more general, way:

$$\upsilon(f_1, f_2) = \sum_{j=1}^{m} \alpha_j (\mu_{1j} - \mu_{2j})^2, \tag{7}$$

with $alpha_j > 0$ and $\sum_j \alpha_j = 1$. By using different sets of weights one can study a variety of interesting questions, such as, for instance, for which biomarker the treatment effect is largest at the end of the study. Alternatively, we may be interested in finding the biomarker for which the treatment effect is mainly expressed at the beginning of the study, and so on. All of these situations can be explored using (7), by selecting an appropriate set of weights. Here again, we can construct confidence intervals in a similar way as we did in the previous subsection and finally we could select the biomarker with interval farthest away from the origin.

5 Analysis of Case Study

We will apply the methods introduced in the previous section to the data introduced in Section 2. A logarithmic transformations was used for the neurohormones variables.

5.1 Exploratory Analysis

Let us start by noting that conventional graphical techniques for longitudinal data would ignore the crossover design of our study. For instance, in a mean profile by treatment graph, each patient would contribute to both treatment groups ignoring the between-period association. Hence, it is more appropriate to base our exploratory analysis on the individual contrast introduced in Section 3. The contrast profile per patient is presented in Figure 1.

The difference of the two mean contrast profiles for each sequence gives an estimate of the evolution of the treatment over time. This treatment effect evolution over time is displayed in Figure 2. Note that any deviation from the horizontal zero-line indicates a treatment effect.

It seems from Figure 2 that Prolactine at day3 is the biomarker in which a largest treatment effect is observed. This pattern is also present in Figure 1 where two clearly differentiated groups can be observed for Prolactine at day3.

Another important issue emerging from the exploratory analysis is the difference between the relative behavior of Naltrexone and placebo at days one and three. For the biomarkers on which the treatment effect seems to be largest, the mean evolution of the Naltrexone group lies above the mean evolution of the control group at day one, provoking a positive treatment effect. However, this behavior seems to be reversed at the third day. This could be explained by the infusion of fentanyl

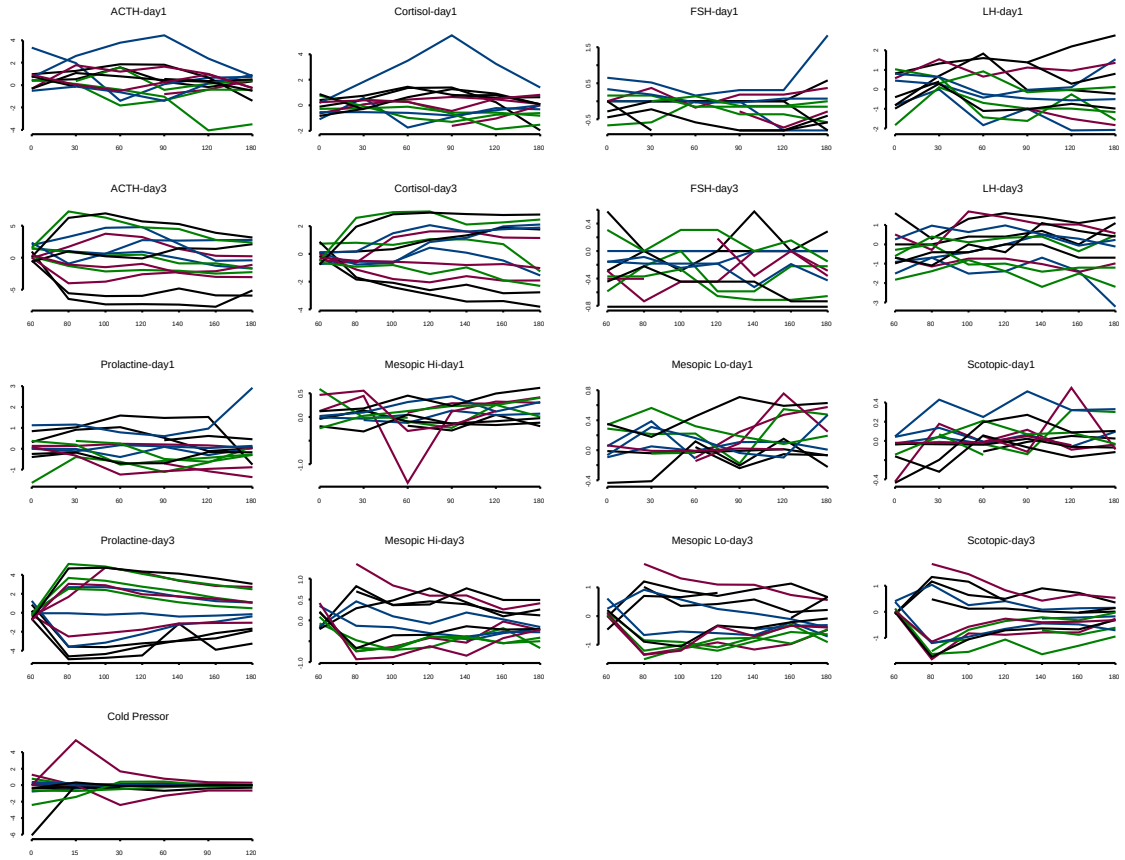


Figure 1: *Contrast profiles.*

administrated to the patients at day 3 before the measurements were taken.

Fentanyl is an opioid receptor agonist, i.e., it increases the activation of opioid receptors. This could explain the lower mean evolution of the treatment group relative to the placebo. Nevertheless, Figure 2 illustrates that, in spite of the initial time decrease in the treatment effect, a tendency to recover over time appears towards the end of the time interval.

5.2 Assessment of Biomarker Quality

Let us first note that all methods are based on confidence regions and on the assumption that they actually contain the true value of the parameter. Given the

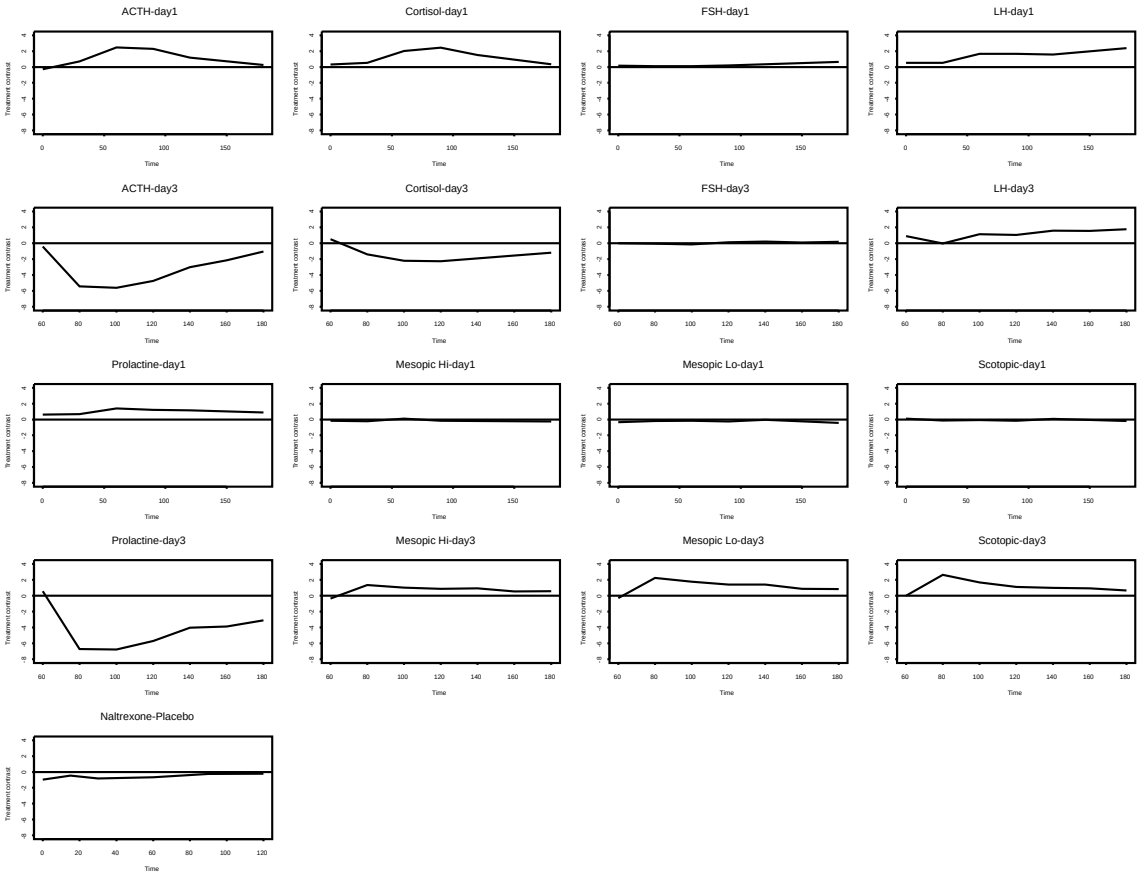


Figure 2: *Mean treatment effect over time: Naltrexone-Placebo.*

large number of biomarkers we are considering here, in all of the following analyzes a Bonferroni correction was applied to account for the multiple comparisons effect. Table 2 displays the results obtained after applying the procedure described in Section 4.1 to the data. Using Theorem 1, we calculated the distance from the origin to each of the ellipsoids defined by each of the biomarkers. For Mesopic Lo-day3, the saturated model did not converge and therefore this biomarker was not included in the analysis.

Note that Cp, Mesopic Lo at day 1, and ACTH at day 1 have negative values for the r-distance; this is merely a convention to point out that the origin is inside their ellipsoids. Table 2 is in complete agreement with the findings of the exploratory

Table 2: *Ellipsoid method.*

Biomarker	r-distance
ACTH-day1	-0.02
ACTH-day3	1.75
Cortisol-day1	0.00
Cortisol-day3	0.28
LH-day1	0.17
LH-day3	0.38
FSH-day1	0.00
FSH-day3	0.00
Prolactine-day1	0.03
Prolactine-day3	23.87
Mesopic Hi-day1	0.04
Mesopic Hi-day3	1.25
Mesopic Lo-day1	-0.01
Scotopic-day1	0.00
Scotopic-day3	3.03
Cp	-0.01

analysis. Prolactine at day 3 was clearly the biomarker with ellipsoid furthest away from zero, followed by Scotopic at day 3 and ACTH at day 3.

Additionally, we estimated the distance between f_1 and f_2, as described in Section 4.2. The results are summarized in the first three columns of Table 3, where LL and UL denote the lower and upper limits of the corresponding confidence interval, respectively.

Here again, the Prolactine at day 3 is the clear winner, followed by the Scototopic at day 3 and LH at day 1. Note that for other biomarkers, like ACTH at day 3 or Cortisol at day 3, the confidence interval contains the origin and therefore the hypotheses of no treatment effect could not be rejected in these cases. This seems to contradict the results found with the ellipsoid method with which these biomarkers

Table 3: *L_2 and different weights results.*

Biomarker	L_2	LL	UL	Eq	EqLL	EqUL	Be	BeLL	BeUL	End	EndLL	EndUL
ACTH-d1	22.46	-3.14	48.06	2.09	-0.25	4.44	2.05	-0.20	4.30	2.14	-0.37	4.647
ACTH-d3	128.75	-69.18	326.69	14.12	-8.87	37.11	20.30	-10.48	51.08	12.65	-10.25	35.549
Cort.-d1	24.34	-4.94	53.62	2.14	-0.37	4.65	1.93	-0.22	4.07	2.36	-0.55	5.262
Cort.-d3	22.54	-17.79	62.86	2.84	-2.65	8.34	3.59	-2.75	9.93	3.04	-3.45	9.538
LH-d1	28.62	5.41	51.82	2.79	0.53	5.06	2.51	0.32	4.69	3.08	0.63	5.525
LH-d3	9.96	-1.70	21.61	1.61	-0.30	3.52	1.61	-0.47	3.69	2.14	-0.25	4.530
FSH-d1	1.07	-1.41	3.55	0.10	-0.15	0.36	0.08	-0.11	0.26	0.13	-0.18	0.453
FSH-d3	0.10	-0.19	0.39	0.02	-0.04	0.07	0.02	-0.03	0.26	0.02	-0.06	0.105
Prol.-d1	10.59	2.09	19.09	1.02	0.19	1.86	0.97	0.17	1.77	1.08	0.18	1.970
Prol.-d3	206.03	103.42	308.63	23.62	11.40	35.84	32.18	16.21	48.14	22.94	10.33	35.540
Mes. Hi-d1	1.58	-1.03	4.19	0.24	-0.22	0.71	0.31	-0.33	0.95	0.17	-0.12	0.465
Mes. Hi-d3	6.41	3.61	9.21	0.74	0.43	1.05	0.98	0.56	1.41	0.74	0.43	1.044
Mes. Lo-d1	0.59	-0.14	1.32	0.08	-0.02	0.17	0.08	-0.03	0.18	0.08	-0.02	0.173
Scot.-d1	0.14	-0.06	0.33	0.01	0.00	0.03	0.01	0.00	0.02	0.02	0.00	0.039
Scot.-d3	16.64	9.55	23.73	1.75	0.97	2.54	2.51	1.45	3.57	1.58	0.79	2.372
Cp	38.75	-34.54	112.05	0.39	-0.47	1.24	0.46	-0.65	1.58	0.31	-0.29	0.913

LL,UL: Lower and upper limits of the 95% confidence interval

Eq: Weights distributed equally over the whole sequence.

Be: 67% of the weight at the beginning.

End: 67% of the weight at the end.

L_2: L_2–norm method.

produced an ellipsoid that did not contain the origin.

Some comments are in place. Note that the L_2-norm method constructs a confidence interval for $\upsilon(f_1, f_2)$, which is a summary statistic for the mean differences Δ_{Ti}, whereas the ellipsoid approach constructs a confidence region for the Δ_{Ti} in a multivariate fashion. Arguably, the loss of information derived from using a summary statistics implies a reduction of power that could explain this discrepancy.

Note also that some of these biomarkers produced very small values for the r-distance, which can also help to explain the results found when the L_2-norm approach is used. The ACTH at day 3, which was ranked third in the previous approach, produces here a very large point estimate for $\upsilon(f_1, f_2)$ but with a very wide confidence interval that contains zero.

Finally, we analyzed the data following the approach introduced in Section 4.3. In

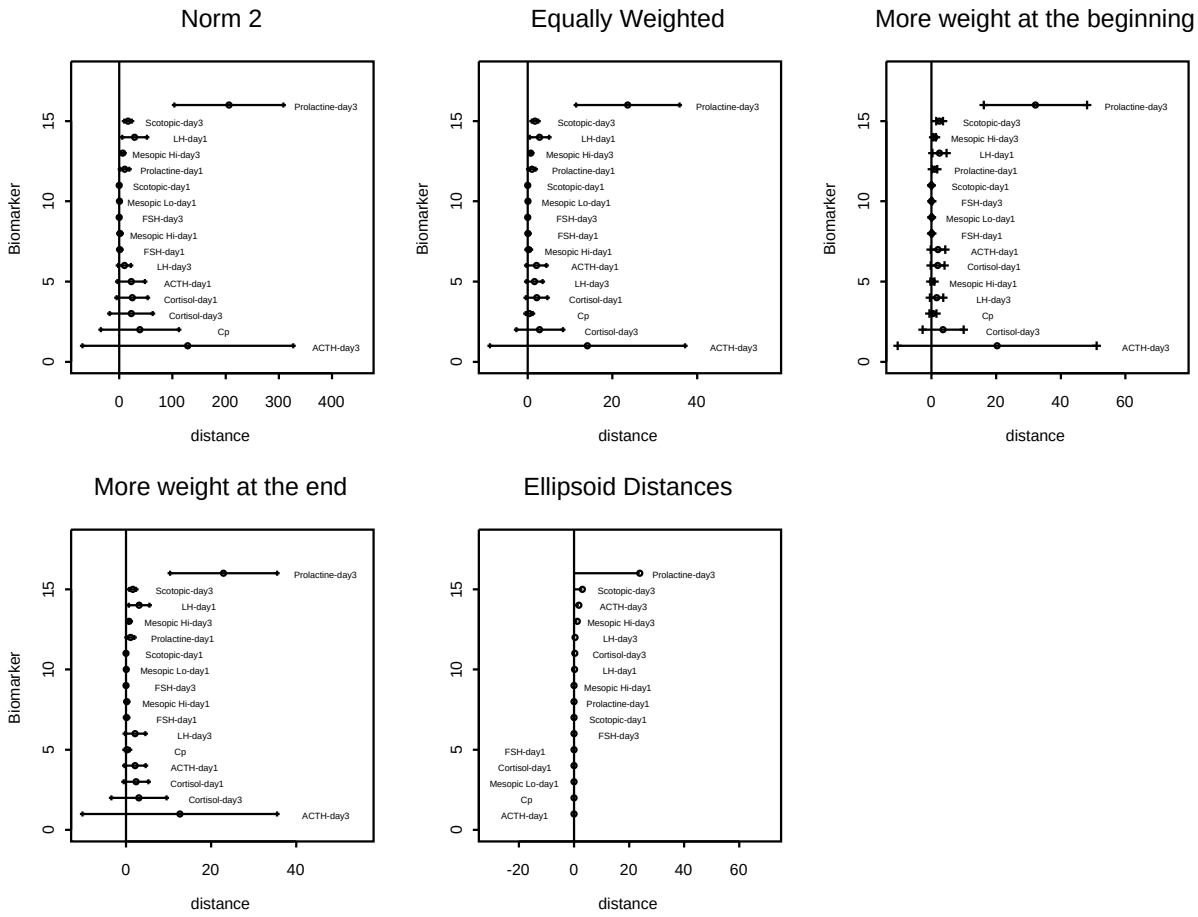


Figure 3: *Summary.*

this analysis three different sets of weights were considered: (a) equal weights at all time points, denoted by 'Eq' in Table 3; (b) 67% of the weight equally assigned to the first half of the longitudinal sequence, 33% equally assigned to the second half, and denoted by 'Be' in Table 3; (c) 33% of the weight equally assigned to the first half of the longitudinal sequence, 67% equally assigned to the second half, and denoted by 'End' in Table 3. The same notation as before was used for the confidence interval limits.

Note that, regardless of the set of weights used in the analysis, the Prolactine at day 3 always produced the best results, followed by the Scotopic at day 3. However,

some variation is seen in the third position. In the cases in which we used equal weights or where we assigned more weight at the end of the sequence, LH at day 1 ranks first. On the other hand, if more weight is assigned at the beginning of the sequence, then Mesopic Hi at day 3 ranks third. This is in total agreement with our findings in the exploratory analysis. Indeed, a closer look at Figure 2 clearly shows that, whereas Mesopic Hi at day 3 seems to have an early response to the treatment that decreases on time, LH at day 1 shows a lower reaction at the beginning that then consistently increases with time.

Figure 3 summarizes the results of all previous analyzes. In all panels the biomarkers, have been ranked in decreasing order, starting by the one in which pharmacological activity is expressed most. For the ellipsoid approach the biomarkers for which the origin is an internal point of their ellipsoid are listed at the left side.

6 Concluding Remarks

Biomarkers are playing an increasingly important role, not only in the study and development of new drugs and therapies, but also in the diagnostics of a medical condition or in improving our understanding of several medical conditions. The recent developments in genetics will likely further increase their utility and use in the near future. Even though considerable research has been done in recent years to study the potential of biological markers as surrogate endpoint, other possible uses have received less attention from a statistical point of view.

In the present work, we focused on the study and evaluation of different physiological variables as biomarkers for pharmacological activity. This type of studies are typically carried out following a crossover design and include a relatively small group of patients. The use of a crossover design in a longitudinal context will re-

quire special analysis considerations. In all cases, we decided to use a saturated linear model, guaranteeing the necessary flexibility to model the time evolution of a relatively large number of biomarkers. Further, we proposed three different approaches using multivariate and univariate techniques. Note that even though one could argue that the multivariate ellipsoid method is more powerful than the other alternatives, the L_2-norm and weighted procedures also offer a great flexibility to answer interesting scientific questions. Choosing the right set of weights, we could explore not only on which biomarker the treatment effect is expressed most but also on which biomarkers the treatment effect acts in a specific way.

Acknowledgment

Ariel Alonso and Geert Molenberghs gratefully acknowledge financial support from the Interuniversity Attraction Pole Research Network P6/03 of the Belgian Government (Belgian Science Policy). All authors are grateful to Eli Lilly & Company for graciously providing the data.

7 References

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology and Therapy*, **69**, 89–95.

Burzykowski, T., Molenberghs, G., and Buyse, M. (2005). *The Evaluation of Surrogate Endpoints*. New York: Springer.

Jones, B. and Kenward, G.M. (2003). *Design and Analysis of Cross-over tTrials*. London: Chapman & Hall/CRC.

Lesko, L.J.S. & Atkinson, A.J. (2001). Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annals Review of Pharmacology and Toxicology*, **41**, 347-366.

Suico, J., Renard, D., Carter, M.K., Diringer, K., Cantrell, A., Chavers, J., and Chalon, S. (2006). Pharmacodynamic effects of opiate receptor antagonism: an agonist challenge study. *Submitted for publication.*

Temple R.J. (1995) A regulatory authority's opinion about surrogate endpoints. In: W.S. Nimmo and G.T. Tucker G.T. (Eds), *Clinical Measurement in Drug Evaluation*, New York: John Wiley & Sons, pp. 3–22.

Appendix

Proof of Theorem 1

To simplify the notation we will denote $\Delta_T = x$, and $\widehat{\Delta}_T = d$ and $\Sigma = \widehat{\Sigma}_{\Delta T}$. We can then rewrite the previous expressions as:

$$r = \min \parallel x \parallel^2 = x'x$$

$$\text{st:} \quad (x-d)'\Sigma^{-1}(x-d) = C(\alpha).$$

Using Lagrange's method, our problem is reduced to minimizing the following function: $F(x,\lambda) = x'x + \lambda(x-d)'\Sigma^{-1}(x-d) - \lambda C(\alpha),$. Equivalently, we have to solve the simultaneous equations:

$$\frac{\partial F}{\partial x} = 2x + 2\lambda\Sigma^{-1}(x-d) = 0, \tag{8}$$

$$\frac{\partial F}{\partial \lambda} = (x-d)'\Sigma^{-1}(x-d) - C(\alpha) = 0. \tag{9}$$

It is not difficult to show that (8) leads to $x = \lambda(\Sigma + \lambda I)^{-1}d$. Additionally, we have that there exist an orthogonal matrix P so that $\Sigma = P^t D_0 P$ with $P^tP = PP^t = I$. D_0 is a diagonal matrix, i.e., $D_0 = (\alpha_i)_{ii}$, where α_i is the i^{th} eigenvalue of Σ. Using this orthogonal decomposition we see that $x = \lambda P^t(D_0 + \lambda I)^{-1}Pd$.

If we now denote $D_1(\lambda) = \lambda(D_0 + \lambda I)^{-1} = \text{diag}\left(\frac{\lambda}{\alpha_i+\lambda}\right)$, then $x = P^t D_1(\lambda)Pd$. Combining this last expressions for x with 90 we obtain

$$(x-d)'\Sigma^{-1}(x-d) = q^t\left[D_3(\lambda) - 2D_2(\lambda) + D_0^{-1}\right]q,$$

where $D_2(\lambda) = \text{diag}\left(\frac{\lambda}{\alpha_i(\alpha_i+\lambda)}\right)$, $D_3(\lambda) = \text{diag}\left(\frac{\lambda^2}{\alpha_i(\alpha_i+\lambda)^2}\right)$, and $q = Pd$.

The matrix of the previous quadratic form is symmetric with diagonal elements equal to $D_3(\lambda) - 2D_2(\lambda) + D_0^{-1} = \text{diag}\left(\frac{\alpha_i}{(\alpha_i+\lambda)^2}\right)$. If $q' = (q_1, q_2, \ldots, q_m)$ then

$$(x-d)'\Sigma^{-1}(x-d) = \sum \frac{\alpha_i q_i^2}{(\alpha_i+\lambda)^2}, \tag{10}$$

The previous expression clearly illustrates that (9) is equivalent to the equation (2) defined in theorem 1

$$\sum \frac{\alpha_i q_i^2}{(\alpha_i + \lambda)^2} = c(\alpha). \tag{11}$$

Using (11) we calculate λ, and finally we just have to calculate the distance

$$r \quad = \quad x'x = q' D_1(\lambda)^2 q = \sum \left(\frac{\lambda q_i}{\alpha_i + \lambda} \right)^2 .$$