of methods for statistical analysis of combination studies

Alexander N. Donev

School of Mathematics, University of Manchester, UK

TABLE OF CONTENTS

1. ABSTRACT

This paper is concerned with the statistical analysis of data obtained in studies of the joint action of drugs. The three methods that are compared are illustrated on real data (1), using the statistical package SAS. It is argued that while the results obtained using these methods do not differ substantially, the method allowing for estimating simultaneously all required parameters is to be preferred. It allows for a statistical test for the significance of the joint action of the drug combinations to be carried out.

2. INTRODUCTION

Combinations of drugs have been known to be highly successful in treating different diseases for many years (2). There is a considerable amount of literature discussing related scientific issues as well as the experimental design and analysis of experiments aiming at identifying the existence of synergistic or antagonistic drug effects (2, 3, 4). Such experiments are a great deal more complex than studying the effect of the drugs separately, and typically take a long time to complete, and require substantial resources. Therefore, it is very important to

Table 1. Ratios of the concentrations of compounds B and A. Ray 1 – compound A only, ray 2 – compound B only, rays 3-7 – stack 1; rays 8-14 – stack 2.

Ray	Study 1	Study 2	
3	4.96		49.61
4	2.48		24.80
5	9.92		99.21
6	1.24		12.40
7	19.85		198.43
8	4.96		49.61
9	0.50		4.96
10	49.65		496.07
11	0.20		1.98
12	124.11		1240.18
13	0.10		1.00
14	248.23		2480.36
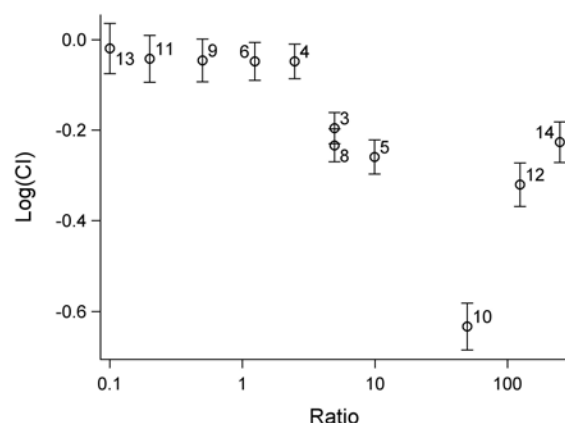
Figure 1. Response values when no compounds have been used.

develop a suitable statistical approach for analyzing such data.

This paper discusses typical methods of analysing such data, demonstrating them on a dataset kindly provided by William Greco and based on his work (1). The dataset contains data from two studies. Greco's description of the experiments is as follows.

"The first drug is trimetrexate (TMQ), a lipophilic inhibitor of the enzyme, dihydrofolate reductase; and the second drug is AG2034, an inhibitor of the enzyme, glycinamide ribonucleotide formyltransferase (GARFT). The first experiment included 2.3 μM folic acid in the medium, and the second experiment included 78 μM folic acid in the medium. All drug concentrations are in μM. The endpoint was the growth of HCT-8 human ileocecal adenocarcinoma cells, in wells of 96-well plates, as measured by the SRB protein stain. Treatments of cells in wells by drugs were randomized across the plates. Each 96-well plate included 8 wells as instrumental blanks (no cells; values are NOT subtracted from treated cell wells); thus 88 wells were used for drug treatments. Five replicate plates were used for each set of 88 treatments. Each of these two large data sets came from two 5-plate stacks (a maximum of 880 treated wells per experiment). There were 176 control wells per experiment (cells, but zero concentrations of both drugs). Each single agent was

studied at 11 dilutions; the two drugs were combined in 12 different fixed ratios per experiment, and also studied in 11 dilutions. The recorded endpoint is an absorbance measurement (ranging about 0.0-2.0), recorded in an automated 96-well plate reader. Complete experimental details and mechanistic implications are included in Faessel et al (1998).

The data set for the low (2.3 μM) folic acid case contains 871 data points (9 points were deleted for experimental problems); and the data set for the high (78 μM) case includes 879 data points (1 point was deleted). The former data set includes the complication of synergism dominating for some fixed ratios and antagonism dominating for other ratios. The latter data set includes the phenomenon of very large (super) synergy, and possible differential background effects (recorded absorbances at very high drug concentrations) for the single agents."

We refer to the compounds TMQ and AG2034 as compounds A and B, to the response of interest as Y, and to the data collected at 2.3 μM and 78 μM folic acid as Study 1 and Study 2, respectively. A ray design (4) has been used in each of the studies, where the rays are defined by the constant ratios of the tested compounds for which the response has been measured - see Table 1. Each study is conducted using two stacks of plates. In order to allow for testing the significance of the differences between the stacks within each study, one combination in common was tested on each of the two stacks within each study. These combinations are rays 3 and 8.

3. STATISTICAL METHODOLOGIES

The statistical analysis of data obtained in a combination study should follow the usual steps of any complex statistical investigations. The details of these steps depend on the aim of the study itself. We present some of the available methods and apply them to analyse the data described in Section 2. The computer package SAS is used to analyse the data. SAS has been previously reported to be a powerful tool in such calculations (5), however the code provided by the authors is incomplete and cannot be used. The SAS code implementing the statistical analysis in the current paper is available from the author on request. All figures in this paper were created using general graphical tools provided by ODS Statistical Graphics in SAS Software, Version 9.2.

3.1. Preliminary statistical analysis

Figure 1 shows a histogram of the observations when no compounds have been used. It is clear that there are unusually small observations in the Study 1. Using such values in the subsequent statistical analysis can create various problems. On the basis that these values were not seen as representing genuine variability, they were excluded from the dataset. The remaining observations can be seen as approximately following the normal distribution, seen by the closeness of curves for the kernel estimation and the normal probability density function for the data.

Statistical tests show no sufficient evidence for differences between the observations taken in the two
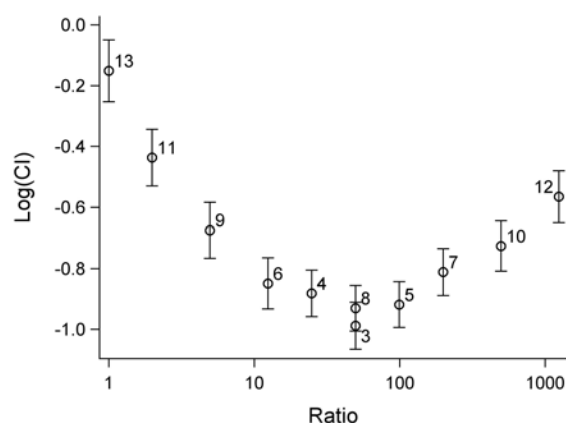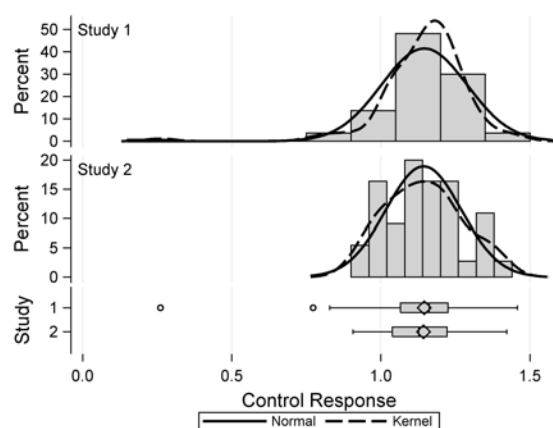
Figure 2. The data for Study 1.

Figure 3. The data for Study 2.

studies and the two stacks of plates. The data for the latter are therefore pooled.

Figure 2 shows all data for each ray of Study 1, while Figure 3 shows all data for each ray of Study 2. Careful inspection of these plots indicates possible problems in the subsequent analyses of the data. For example, it can be seen that the data for Ray 7 of Study 1 includes virtually no observations producing responses in the middle between the maximum and the minimum and without such observations the estimation of important parameters may not be possible. It can be also seen that in all plots the variability of the response increases with the response. This is an important observation which influences the definition of the model in Section 3.2.

3.2. Main Analysis

The statistical analysis of bioassay data for a single compound action requires estimating the mode

$$Y_i = \gamma + \frac{\delta - \gamma}{1 + 10^{(x_i - \alpha)\beta}} + \varepsilon_i, \quad (1)$$

where Y_i is the response observed at a dose whose logarithm to base 10 is x_i, α corresponds to the logarithm to base 10 of the IC50 of the drug; β is the Hill slope, named after Hill (6) who proposed the model; γ is the mean response when no compound has been used and δ is the mean response for maximum possible drug effect. When α is smaller the drug is more potent. In (ε_i is the experimental error which we assume to be normally distributed with zero mean and variance that could be heterogeneous, e.g. $\phi(Y)\sigma_\varepsilon^2$. Based on practical experience, also supported by the data shown in Figures 2 and 3, we assume that $\phi(Y) = Y^\rho$, where ρ is another model parameter that needs to be estimated. There are certainly many different ways to reparameterize model (1) and the right choice depends on the robustness of the numerical procedure used to estimate the model parameters and the scale on which an estimate of the potency is required. Methods for estimating the parameters of nonlinear models like (10) are described in the literature (7-9).

When a ray design has been used to study the joint effect of two drugs, say *A* and *B*, the data for each ray can be treated as a new drug, defined by the ratio in which the two drugs have been combined, and analysed by estimating model (1). Evidence that model (1) is not applicable would indicate that mechanistic considerations different from those assumed by Hill and Michaelis and Menten (10) to justify model (1) are present, and therefore the study of such a drug combination may require a different treatment.

Typically the assessment of the joint action of two compounds is carried out by evaluating the evidence of its deviation from what would be expected if the joint action of the two compounds were additive. The definition of additivity proposed by Loewe (11) is most commonly used, though other definitions (12) also exist. Loewe's definition leads to the so called *combination*, also called *interaction*, *index*

$$CI = \frac{D_a}{D_A} + \frac{D_b}{D_B} \quad (2)$$

where D_A and D_B are doses of the compounds *A* and *B*, respectively, producing the same mean response as a combination of doses D_a and D_b of the two compounds. If $CI < 1$, this provides evidence of synergistic action of the drugs, while if $CI > 1$, this suggests an antagonistic action. Some researchers provide general advice about how smaller or larger than one the CI value should be in order to declare the drug action as synergistic or antagonistic. However, as the combination index is a random variable following approximately lognormal distribution, a better approach is to test *CI* for its significance. Such a statistical test would take into account the variability in the experiment and the number of observations used to estimate the combination index.

We estimate $D_A = IC50_A$ and $D_B = IC50_B$ as well as D_a and D_b for each ray using the following three methods, with SAS/STAT Software, Version 9.1.3.

Table 2. Study 1. Methods M1 and M2. Estimates of the LogIC50 (α). Individual models for each ray.

Ray	Homogeneous variance		Heterogeneous variance	
	Estimate	Standard Error	Estimate	Standard Error
1	-2.8840	0.0277	-2.8836	0.0188
2	-2.2190	0.0232	-2.2141	0.0178
3	-2.6085	0.0156	-2.6157	0.0118
4	-2.5697	0.0226	-2.5707	0.0148
5	-2.5849	0.0187	-2.5945	0.0138
6	-2.6728	0.0218	-2.6772	0.0145
7	-2.4756	11.0986	-2.4806	
8	-2.6410	0.0205	-2.6528	0.0148
9	-2.7952	0.0254	-2.7975	0.0172
10	-2.8111	0.1210	-2.8747	0.0241
11	-2.8706	0.0288	-2.8657	0.0197
12	-2.5501	0.0247	-2.5512	0.0191
13	-2.8685	0.0285	-2.8663	0.0213
14	-2.4418	0.0216	-2.4460	0.0161

Table 3. Study 1. Methods M1 and M2. Estimates of the Hill's slope (β). Individual models for each ray

Ray	Homogeneous variance		Heterogeneous variance	
	Estimate	Standard Error	Estimate	Standard Error
1	-2.2146	0.2701	-2.2436	0.1729
2	-3.1630	0.4227	-2.6237	0.2552
3	-6.1619	2.7414	-4.0424	0.3811
4	-5.2128	0.8215	-5.0270	0.4789
5	-4.6937	0.7349	-3.5923	0.3881
6	-3.6302	0.4889	-3.6340	0.3312
7	-18.8973	5538.8315	-18.2115	
8	-5.6210	1.5845	-3.4556	0.3497
9	-2.6595	0.3482	-2.2805	0.1477
10	-5.8322	6.0973	-3.4322	0.4789
11	-2.1606	0.2783	-2.0191	0.1383
12	-2.9418	0.4383	-2.6589	0.2676
13	-2.2650	0.2861	-2.0015	0.1565
14	-3.9111	0.6763	-2.9750	0.2358

Table 4. Study 1. Methods M1 and M2. Estimates of the bottom level (γ). Individual models for each ray

Ray	Homogeneous variance		Heterogeneous variance	
	Estimate	Standard Error	Estimate	Standard Error
1	0.1382	0.0222	0.1391	0.0063
2	0.1530	0.0203	0.1433	0.0076
3	0.1415	0.0169	0.1349	0.0045
4	0.1302	0.0204	0.1293	0.0054
5	0.1383	0.0167	0.1329	0.0054
6	0.1275	0.0200	0.1287	0.0047
7	0.1329	0.0169	0.1330	0.0063
8	0.1401	0.0182	0.1305	0.0054
9	0.1320	0.0216	0.1243	0.0047
10	0.1501	0.0161	0.1457	0.0051
11	0.1395	0.0225	0.1338	0.0057
12	0.1594	0.0192	0.1554	0.0068
13	0.1459	0.0207	0.1377	0.0073
14	0.1508	0.0194	0.1429	0.0045

Table 5. Study 1. Methods M1 and M2. Estimates of the top level (δ). Individual models for each ray

Ray	Homogeneous variance		Heterogeneous variance	
	Estimate	Standard Error	Estimate	Standard Error
1	1.1731	0.0087	1.1729	0.0094
2	1.1672	0.0086	1.1687	0.0094
3	1.1683	0.0082	1.1704	0.0090
4	1.1698	0.0090	1.1702	0.0099
5	1.1639	0.0082	1.1660	0.0091
6	1.1605	0.0086	1.1607	0.0094
7	1.1686	0.0085	1.1692	0.0090
8	1.1678	0.0090	1.1707	0.0101
9	1.1787	0.0090	1.1805	0.0099
10	1.1679	0.0087	1.1706	0.0099
11	1.1766	0.0089	1.1770	0.0098
12	1.1722	0.0090	1.1729	0.0100
13	1.1661	0.0082	1.1673	0.0089
14	1.1594	0.0093	1.1613	0.0104

Table 6. Study 1. Method M2. Estimates of ρ

	Heterogeneous variance	
Ray	**Estimate**	**Standard Error**
1	1.3333	0.1596
2	1.1932	0.1677
3	1.4122	0.1368
4	1.2969	0.1378
5	1.2688	0.1475
6	1.4600	0.1420
7	0.9910	0.1301
8	1.3603	0.1479
9	1.5776	0.1506
10	1.3194	0.1404
11	1.4420	0.1524
12	1.2115	0.1476
13	1.1815	0.1584
14	1.6164	0.1509

Table 7. Study 1. Method M2. Estimates of σ_ε^2

	Heterogeneous variance	
Ray	Estimate	Standard Error
1	0.0934	0.0054
2	0.0949	0.0056
3	0.0894	0.0053
4	0.0993	0.0058
5	0.0906	0.0054
6	0.0941	0.0055
7	0.0940	0.0056
8	0.0998	0.0060
9	0.0952	0.0056
10	0.0964	0.0059
11	0.0949	0.0056
12	0.0986	0.0058
13	0.0884	0.0052
14	0.1013	0.0060

Table 8. Study 1. Method M3. Estimates of the model parameters

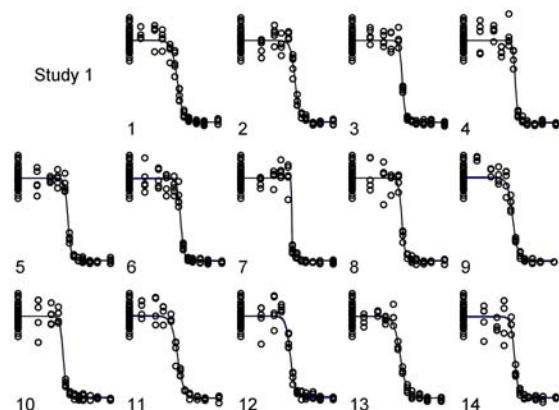| **Ray**	**Parameter**	**Estimate**	**Standard Error**
1	α	-2.8813	0.0187
1	β	-2.2302	0.1622
2	α	-2.2116	0.0170
2	β	-2.4692	0.1797
3	α	-2.6157	0.0124
3	β	-4.1880	0.4397
4	α	-2.5725	0.0141
4	β	-5.1806	0.4564
5	α	-2.5969	0.0141
5	β	-3.6828	0.3606
6	α	-2.6812	0.0151
6	β	-3.7907	0.3451
7	α	-2.4794	
7	β	-18.8972	
8	α	-2.6538	0.0137
8	β	-3.6961	0.3470
9	α	-2.7958	0.0172
9	β	-2.5329	0.1793
10	α	-2.8750	0.0214
10	β	-3.1808	0.3416
11	α	-2.8630	0.0196
11	β	-2.0748	0.1398
12	α	-2.5449	0.0183
12	β	-2.2819	0.1849
13	α	-2.8689	0.0217
13	β	-1.9669	0.1367
14	α	-2.4447	0.0157
14	β	-2.9758	0.2378
	γ	0.1366	0.0016
	δ	1.1698	0.0026
	ρ	1.2810	0.0394
	σ_ε	0.0956	0.0015

Figure 4. Study 1. Method M3. Logarithm to base 10 of the combination indices and 95% confidence intervals for the rays (ray number given in brackets).
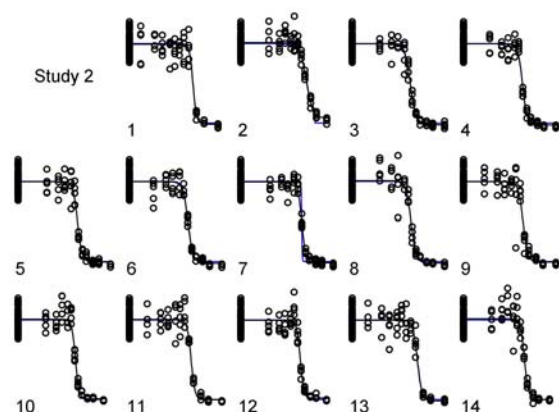

Figure 5. Study 2. Method M3. Logarithm to base 10 of the combination indices and 95% confidence intervals for the rays (ray number given in brackets).

- M1: Model (1) is fitted for each ray assuming $\rho = 0$. This simplifies the analysis. This method requires the use of PROC NLIN.
- M2: Model (1) is fitted for each ray, obtaining also the maximum likelihood estimate of ρ. This method requires the use of PROC NLMIXED.
- M3: A global model - fitting simultaneously (1) for each ray including the mono therapies, estimating common values for γ, δ, ρ and σ_ε^2, and calculating the combination index for each ray and testing it's significance. This method requires the use of PROC NLMIXED.

In all cases the default setting for the SAS procedures are used.

4. COMPARISON OF METHODOLOGIES

4.1. Study 1

Tables 2-5 show the estimates of the parameters of model (1) α, β, γ and δ when methods M1 and M2 are used. Not surprisingly the standard errors of the parameters' estimates when method M2 is used are smaller. However, the differences between the estimates for α for the studied combinations are relatively small. Both methods incurred difficulties in estimating α for ray 7, although this is shown in the computer output differently: when method M1 is used, the standard error is reported much higher than the other standard errors, while for method M2 a standard error for the estimate of α is not produced. The inspection of the data (Figure 2) discussed in Section 3.1 indicated the problem with the data for this drug combination - there are not enough points in the middle range of values for the response.

The close similarity of the estimates of γ and δ (Tables 4 and 5) provides further evidence that there is no notable variability in the results between plates and raises a question whether separate estimates of the parameters for each ray in this study is indeed necessary. This concern can also be extended to the estimates of ρ and σ_ε^2 obtained using method M2 provided in Tables 6 and 7. For example, nothing would justify the use of different estimates for ρ for each drug combination, as the variance of the response depends on its mean and not on the compound that has been studied.

Method M3 does not have any of the shortcomings that methods M1 and M2 have. The model parameters estimates when this method is used are shown in Table 8. The estimates of α are similar to those obtained previously by methods M1 and M2, and their standard errors are in general smaller then those obtained with method M1, but they are larger then those obtained using method M2. The latter could be explained by the over-fitting that methods M1 and M2 do, i.e. providing different estimates for the parameters γ, δ, ρ and σ_ε^2. On the contrary, using method M3 allows for obtaining common estimates for γ, δ, ρ and σ_ε^2.

Table 9 shows the estimates of the logarithms at base 10 of the combination index (2) for all rays obtained using method M3. The tests for significance show strong evidence of synergistic action for rays 3, 5, 8, 10, 12 and 14; some evidence for synergy for rays 4 and 6, but no evidence for synergy for rays 11 and 13. Furthermore, the strength of synergy appears to increase with the ratio of the drugs up to a ratio of 49.65, and then decreases as the ratio increases further. Clear illustration of this phenomenon is shown on Figure 4 which shows the 95% confidence intervals for the logarithm to base 10 of the combination index. If the distribution of the combination index, rather than of its logarithm, is assumed to be approximately normal, the tests for significance provide very similar results to those presented in Table 9. This may not be the case in the analysis of other data.

4.2. Study 2

The data for this study are shown in Figure 3. When their analysis is carried out in the same way as that used for Study 1, similar results are obtained. Therefore

complete details are not provided. All methods failed to fit a sensible model to the data for ray 14 and the results for this drug combination are omitted. Method M2 also had difficulties estimating β for rays 7 and 10. Table 10 lists the logarithm to base 10 of the combination indices for the studied drug combinations, while Figure 5 shows their 95% confidence intervals. There is clear evidence for synergistic action of the drugs, and it appears to be largest at the drug ratio 49.65, the same ratio for which the synergy was largest in Study 1. Therefore, it seems the increase of the folic acid increases the strength of synergy between the two drugs, but does not appear to change the drug ratio at which the largest synergy occurs. However, this observation has to be checked experimentally.

5. CONCLUSIONS

The results obtained during all stages of the statistical analysis are important for the understanding of the joint action of drugs. For example, the plots of the raw data (Figure 2 and Figure 3) help the motivation and understanding of the subsequent statistical analysis. The three compared methods showed consistent results. However, results obtained using method M1 ignored the heterogeneity of the variance of the response of interest, while those obtained with method M2 appear to underestimate the standard errors of the estimates of α and β for all drug combinations. Method M3 allows for all required estimates to be obtained using all data. A statistical test for the significance of the drug combinations is easy to carry out too.

The data that were used to compare methods M1, M2 and M3 have some limitations and therefore various possible complications in the statistical analysis of combination data could not be illustrated. For example, no information is available about which data are collected on the same plates and whether the data are collected on one or on different occasions. Often many compounds are studied and the variability due to using different plates and running experiments on different occasions cannot be ignored. Evidence that such variability exists in the two studies analysed in Section 4 is obvious from the comparison of the results for rays 3 and 8 which correspond to identical drug combinations tested on different batches of plates. There is not enough data in this experiment to obtain estimates of the variability due to this and other sources of variability. Naturally this leads to the need to carefully consider the experimental design for a combination study. However, this issue will be addressed in a separate publication.

There are situations when the distribution of the response may not be approximately normal, e.g. the response could be a binary or categorical variable. In such cases, method M3 is still capable of producing correctly the required statistical analysis. This also applies to the case when one or all of the drugs are not full inhibitors. Furthermore, the approach that we recommend, based on programs written in SAS, is flexible and can be tailored to the specific features of any particular combination study. Alternative approaches based on software specifically developed for analysis of combination studies may be a useful tool in the hands of an experienced statistician but should not replace him or her.

7. REFERENCES

1. Faessel, H. M. , H. K. Slocum, R. C. Jackson, T. J. Boritzki, Y. M. Rustum, M. G. Nair, and W. R. Greco: Super *in vitro* Synergy between Inhibitors of Dihydrofolate Reductase and Inhibitors of Other Folate-requiring Enzymes: The Critical Role of Polyglutamylation. *Cancer Res.* 58, 3036-3050 (1998)

2. Chou, T.-C.: Theoretical Basis, Experimental Design, and Computerized Simulation of Synergism and Antagonism in Drug Combination Studies. *Pharmacol. Rev.*, 58, 621-681 (2006)

3. Greco, W. R., G. Bravo and J. C. Parsons: The Search for Synergy: A Critical Review from a Response Surface Perspective. *Pharmacol. Rev.*, 47, 331-385 (1995)

4. Tallarida, R. J.: Drug Synergism and Dose-Effect Data Analysis. Chapman & Hall/CRC (2000)

5. Straetemans, R., T. O'Brien, L. Wouters, J. V. Dun, M. Janicot, L. Bijnens, T. Burzykowski, and M. Aerts: Design and Analysis of Drug Combination Experiments, *Biom. J.*, 47, 299-308 (2005)

6. Hill, A. V.: The combinations of haemoglobin with oxygen and with carbon monoxide, *Biochem. J.,* 7, 471-480 (1913)

7. Bates, D. M. and D. G. Watts: Nonlinear Regression Analysis and Its Application. Wiley: New York (2007)

8. McCulloch, C. E. and S. R. Searle: Generalized, Linear, and Mixed Models. Wiley: New York (2001)

9. Seber, G. A. F. and C. J. Wild: Nonlinear regression. Wiley: New York (2005)

10. Michaelis, L. and M. L. Menten: Die Kinetik der Invertinwirkung, *Biochemische Zeitschrift*, 49, 333-369 (1913)

11. Loewe, S.: The problem of synergism and antagonism of combined drugs, *Arzneim. Forsch.* 3, 285-320 (1953)

12. Webb, J. L.: Enzyme and metabolic inhibitors. New York: Academic Press, 1, 55-79 and 488-512 (1963)

Send correspondence to: Alexander N. Donev, School of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, UK; Tel: 44-0-1613699, Fax: 44-0-161275 5819, E-mail: a.n.donev@manchester.ac.uk

http://www.bioscience.org/current/vol2E.htm