

# The Geometric Mean Functional Relationship Approach to Linear Regression in Pharmaceutical Studies: Application to the Estimation of Binding Parameters

GEORGIA VALSAMI AND PANOS MACHERAS

*Laboratory of Biopharmaceutics and Pharmacokinetics, Department of Pharmacy, University of Athens, Panepistimiopolis, Athens 15771, Greece*

## Abstract

The aim of this study was to draw attention and apply the geometric mean functional relationship (GMFR) approach to the analysis of data which are currently analysed with the ordinary least-squares method, in spite of the fact that both variables are subject to error.

The method was applied to drug-protein binding data using erroneous simulated data, generated from the Scatchard model with one class of binding sites. For the present study, a computer programme in BASIC was constructed to perform linear regression analysis by means of the "least-triangles" (LT) approach to GMFR and the ordinary least-squares (LS) method.

The least-triangles approach for linear regression analysis was proven to be superior to the ordinary least-squares method when applied to contaminated simulated binding data. The method requires minimum computation and it can be applied to many other types of pharmaceutical studies where linear regression is applied and both variables are subject to error.

In various fields of pharmaceutical sciences it is not uncommon to encounter problems associated with linear regression analysis where both the dependent,  $y$ , and the independent variable,  $x$ , are accompanied by an experimental error. Some typical examples include the linear form of Scatchard equation in binding studies (Jun et al 1975; Löscher 1979; Matsushita et al 1986), the log-log relationships between particle size and the initial rate of dissolution (Kaneniwa & Watari 1974; Watari & Kaneniwa 1976; Farin & Avnir 1987; Avnir 1994), the various forms of in-vitro-in-vitro, in-vitro-in-vivo or in-vitro-in-situ correlations, such as dissolution data and blood level data (El-Yazigi & Sawchuk 1985), log P measurements (Kim et al 1993) or log P-membrane permeability data (Tayar et al 1991). All these studies are analysed with the ordinary least-squares method assuming that the values of the independent variable  $x$  are not subject to error. However, the estimates for the slope and the intercept of the regression line cannot be considered valid since both variables,  $x$  and  $y$  are subject to error.

In statistics literature the geometric mean functional relationship (GMFR) approach to linear regression has been widely applied in fishery studies (Jolicoeur 1975; Sprent & Dolby 1980) for the estimation of growth rates, where both variables ( $x$ ,  $y$ ) are subject to error. The slope estimate,  $b$ , of the GMFR approach is:

$$b = [\text{sign } S_{xy}] \sqrt{(S_{yy}/S_{xx})} \quad (1)$$

where:  $S_{xx} = \sum(x_i - \bar{x})^2$ ,  $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$ ,  $S_{yy} = \sum(y_i - \bar{y})^2$ ,

Correspondence: P. Macheras, Laboratory of Biopharmaceutics and Pharmacokinetics, Department of Pharmacy, University of Athens, Panepistimiopolis, Athens 15771, Greece.

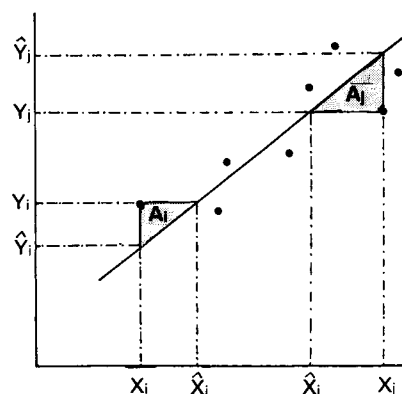


FIG. 1. Schematic representation of the least-triangles approach to GMFR.

the summation being over all observations ( $x_i$ ,  $y_i$ ), i.e. it is the geometric mean of the least-squares regression coefficient for the regression of  $y$  on  $x$  and  $x$  on  $y$ . According to Barker et al (1988) this approach does not require any assumptions concerning the absence of error in either of the variables. The GMFR approach can be interpreted (Barker et al 1988) as the estimate that minimizes an error cost functional based on the sum of the triangular areas formed by connecting the experimental data points to the regression line with lines parallel to the co-ordinate axes (Fig. 1).

This study was undertaken to draw attention and apply the GMFR approach to the analysis of data which are currently analysed with the ordinary least-squares method in spite of the fact that both variables are subject to error. To this end, the method was applied to drug-protein binding

data using erroneous simulated data, generated from the Scatchard model with one class of binding sites (Scatchard 1949). For the needs of the present study, a computer programme in BASIC was constructed to perform linear regression analysis by means of the "least-triangles" (LT) approach to GMFR and the ordinary least-squares (LS) method.

### Theoretical Section

Since the GMFR approach has not been discussed in the pharmaceutical literature, a brief description of the LT approach (Barker et al 1988) is presented below.

The fitting of the model  $y = a + bx$  to the data is based on the minimization of the sum of the areas of the right-angled triangles formed by the regression line and the lines passing through the data points parallel to the axes as shown in Fig. 1 i.e.:

$$\sum_{i=1}^n A_i = \frac{1}{2} [\text{sign}(b)] [\sum_{i=1}^n (y_i - bx_i - a)(x_i - \frac{1}{b}(y_i - a))] = \min \quad (2)$$

where  $[\text{sign } b] = -1$  if  $b < 0$  and  $[\text{sign } b] = 1$  if  $b > 0$ .

This line passes through the mean,  $(\bar{x}, \bar{y})$ ; equation 2 takes the form of equation 3 after substitution for,  $a$ , from equation,  $a = \bar{y} - b\bar{x}$ :

$$\sum_{i=1}^n A_i = -\frac{1}{2} [\text{sign}(b)] [\sum_{i=1}^n [(y_i - \bar{y}) - b(x_i - \bar{x})] [(x_i - \bar{x}) - \frac{1}{b}(y_i - \bar{y})]] = \min \quad (3)$$

The value of  $b$ , denoted as  $b_0$ , which minimizes the value of  $\sum A_i$  in equation 3, can be obtained by differentiating equation 3 with respect to  $b$  and equating the resulting equation with zero, i.e.:

$$\frac{dF}{db} = -\frac{1}{2} [\text{sign}(b)] [\sum_{i=1}^n \frac{1}{b^2} (y_i - \bar{y})^2 - (x_i - \bar{x})^2] = 0 \quad (4)$$

and

$$b_0 = [\text{sign}(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))] \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (5)$$

This slope, if drawn through the mean  $(\bar{x}, \bar{y})$ , gives the same line whether  $y$  is regressed on  $x$  or vice versa.

A measure of the extent to which the observations are not explained by the model can be obtained by substituting  $b_0$  from equation 5 into equation 3:

$$\Delta L_T = \frac{[\text{sign}(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})) \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]}{+\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

the summation being for all observations.

### Simulation Studies

Simulated data were generated according to the Scatchard equation for a single class of binding sites:

$$r/F = nk - kr \quad (7)$$

where  $r = B/P_i$  is the number of moles of the drug bound per mole of protein,  $B$  and  $F$  are the bound and free molar concentrations of the drug, respectively,  $P_i$  is the total

protein molar concentration,  $n$  is the number of binding sites, and  $k$  is the binding constant for the association of the drug with the protein. Error-free simulated values for  $r$  were calculated according to the equation  $r = nkF/(1 + kF)$ , the nonlinearized form of equation 7, using a BASIC program. The free drug concentration range was assigned, as well as values for the parameters  $n$ , and  $k$ , and  $r/F$  values were calculated for each pair of data  $F$ ,  $r$ . Erroneous simulated data were generated by adding to each pair of error-free values, a pseudorandom normal variate of mean zero and relative standard deviation equal to 1–10% of the error-free value. Finally, equation 7 was fitted to the data by means of both the LT and the LS methods using a BASIC program. The program requires the number of data points and the  $r$ ,  $r/F$  values, and undertakes the calculation of both the GMFR and LS slope and intercept estimators. The results are presented arithmetically and graphically.

### Results and Discussion

The estimates for the intercept and the slope of equation 7 using both methods, are shown in Table 1. In all cases examined the GMFR approach to linear regression gave better parameter estimates than the ordinary LS method. For low noise levels (1–2%) added on both  $x$  and  $y$  variables the two methods were almost equivalent. Using the GMFR approach the % error of the calculated intercept ( $nk$ ) and slope ( $k$ ) values ranged from 1.1 to 17.6 and from 0.6 to 17.4, respectively. The corresponding ranges of % error observed when the LS method was applied were 1.7 to 29.3 for the intercept ( $nk$ ) and from 1.4 to 31.0 for the slope ( $k$ ). Overall, both methods exhibited the lowest % error when a 2% noise was added on both  $x$  and  $y$  variables, while the highest % error was observed when a 2% noise was added on the  $x$  variable and a 10% noise on the  $y$  variable.

The plots shown in Fig. 2A correspond to a data set with 2% noise added on the  $x$  variable and a 5% noise added on the  $y$  variable. Even at this low level of noise in the  $x$  variable, the GMFR approach describes the binding phenomenon more validly if one compares the two regression lines with the theoretical line (solid line in Fig. 2A). Quite often in drug-protein binding studies, experimental data points with high  $r/F$  values are missing due to analytical limitations. The lower free drug concentrations cannot be measured when the drug is extensively bound to proteins. To mimic this situation the initial four data points of the simulated contaminated data used in Fig. 2A were excluded from the regression and the results obtained are shown in Fig. 2B. The superiority of the LT to LS approach in the analysis of the limited data is apparent. The overall performance of the two methods in the analysis of limited data with various levels of noise on  $x$  and  $y$  variables is presented in Table 1. As can be seen, in all cases studied, the estimates derived from the LT approach were found to be remarkably superior to those derived from the LS approach.

In conclusion, the LT approach for linear regression analysis was proven to be superior to the ordinary least-squares method when applied to contaminated simulated binding data. The method requires minimum computation and it can also be applied to many other types of

Table 1. Estimates for  $k$  (slope) and  $nk$  (intercept) calculated with the GMFR and the LS methods using contaminated data<sup>a,b,c</sup>.

Noise (%)		GMFR		LS	
$x$	$y$	$nk \times 10^4$	$k \times 10^4$	$nk \times 10^4$	$k \times 10^4$
1	2	14.76 (98.4)	4.94 (98.8)	14.66 (97.7)	4.90 (98.0)
		15.03 (100.2)	5.03 (100.6)	14.30 (95.3)	4.77 (95.4)
2	2	14.84 (98.9)	4.97 (99.4)	14.75 (98.3)	4.93 (98.6)
		15.04 (100.3)	5.04 (100.8)	14.37 (98.3)	4.79 (95.8)
2	5	14.07 (93.8)	4.72 (94.4)	13.51 (90.1)	4.50 (90.0)
		13.31 (88.7)	4.44 (88.8)	10.66 (71.1)	3.48 (69.6)
2	7.5	13.25 (88.3)	4.44 (88.8)	12.11 (80.7)	4.00 (80.0)
		11.44 (76.0)	3.78 (75.6)	7.30 (48.7)	2.26 (45.2)
2	10	12.36 (82.4)	4.13 (82.6)	10.61 (70.7)	3.45 (69.0)
		9.77 (65.1)	3.19 (63.8)	4.88 (32.5)	1.39 (27.8)
5	5	14.30 (95.3)	4.80 (96.0)	13.80 (92.0)	4.61 (92.2)
		13.37 (89.1)	4.46 (89.2)	10.97 (73.1)	3.59 (71.8)
5	7.5	13.46 (89.7)	4.51 (90.2)	12.40 (80.7)	4.11 (82.2)
		11.47 (76.5)	3.79 (75.8)	7.61 (50.7)	2.38 (47.6)
5	10	12.54 (83.6)	4.19 (83.8)	10.88 (72.5)	3.55 (71.0)
		9.78 (65.2)	3.19 (63.8)	5.16 (34.4)	1.49 (29.8)

<sup>a</sup>True values of the parameters:  $k=5.0 \times 10^4$ ,  $nk=15.0 \times 10^4$ ,  $n=3.0$ . <sup>b</sup>Parameter estimates in the second row for each  $x$ ,  $y$  data set have been calculated from limited data; the four data points with the higher  $r/F$  values have been excluded (see also Fig. 2B). <sup>c</sup>Numbers in parentheses are percentages of true values.

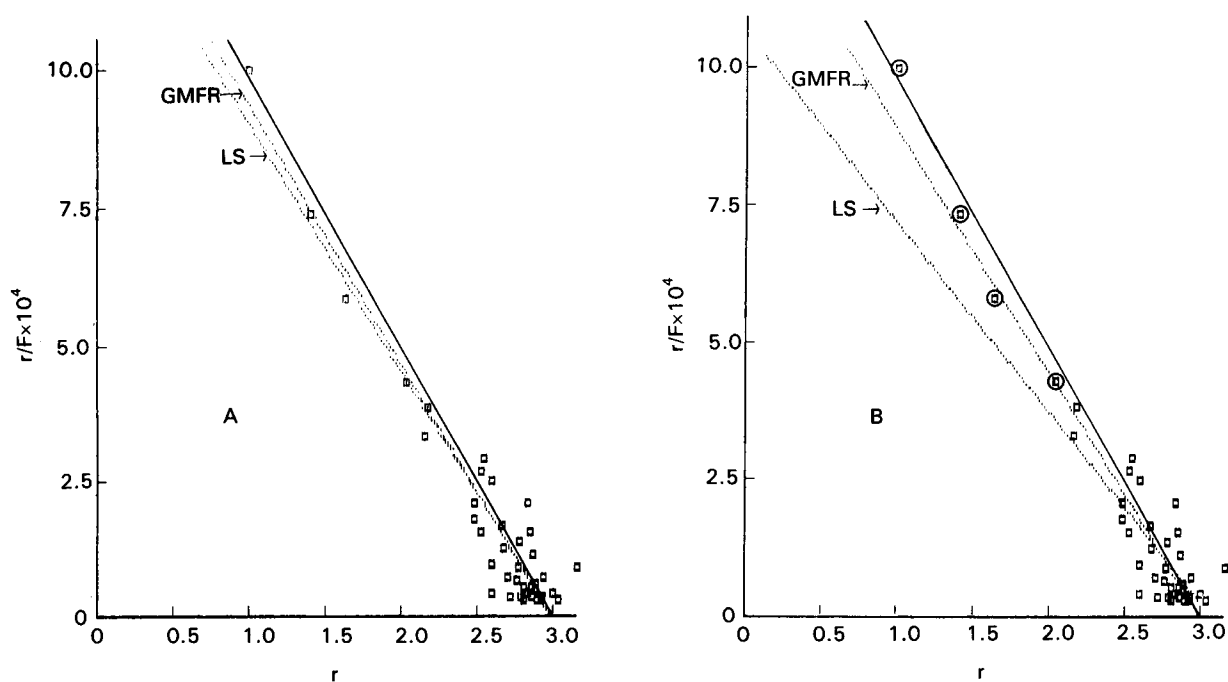


FIG. 2. Scatchard plots of a simulated data set with 2 and 5% noise added on  $x$  and  $y$  variable, respectively. Dotted lines are fitted lines obtained with the GMFR and LS regression methods (A) using all data points and (B) excluding the encircled data points. The solid line corresponds to the theoretical equation  $r/F=15.0 \times 10^4 - 5.0r$ .

pharmaceutical studies where linear regression is applied and both variables are subject to error.

### References

- Avnir, D. (1994) Applications of fractal geometry to pharmaceutical problems. *Proceed. Intern. Symp. Control. Rel. Bioact. Mater.* 21 Controlled Release Society Inc.
- Barker, F., Soh, Y. C., Evans, R. J. (1988) Properties of the geometric mean functional relationship. *Biometrics* 44: 279–281
- El-Yazigi, A., Sawchuk, R. (1985) In vitro-in vivo correlation and dissolution studies with oral theophylline dosage forms. *J. Pharm. Sci.* 74: 161–164
- Farin, D., Avnir, D. (1987) Reactive fractal surfaces. *Phys. Chem.* 91: 5517–5521
- Jolicoeur, P. (1975) Linear regressions in fisheries research: some comments. *J. Fish. Res. Board Can.* 32: 1491–1494
- Jun, H. W., Luzzi, L. A., Ma, J. K. H. (1975) Drug-biomolecule interactions: fluorescence studies on interaction of aminonaphthalenesulfonic acid derivatives with serum albumins. *J. Pharm. Sci.* 64: 493–497
- Kaneniwa, N., Watari, N. (1974) Dissolution of slightly soluble drugs. I. Influence of particle size on dissolution behavior. *Chem. Pharm. Bull.* 22: 1699–1705

- Kim, D.-Ch., Burton, Ph., Borchardt, R. (1993) A correlation between the permeability characteristics of peptides using an in vitro cell culture model (Caco-2) and those using an in situ perfused rat ileum model of the intestinal mucosa. *Pharm. Res.* 10: 1710-1714
- Löscher, W. (1979) A comparative study of the protein binding of anticonvulsant drugs in serum of dog and man. *J. Pharmacol. Exp. Ther.* 208: 429-435
- Matsushita, Y., Takahashi, M., Moriguchi, I. (1986) Binding of fluorescent 7-amino-4-nitro benzoxadiazole derivatives to bovine serum albumin. *Chem. Pharm. Bull.* 34: 333-339
- Sprent, P., Dolby, G. R. (1980) The geometric mean functional relationship. *Biometrics* 36: 547-550
- Scatchard, G. (1949) The attractions of proteins for small molecules and ions. *Ann. NY Acad. Sci.* 51: 660-672
- Tayar, N. El., Tsai, R. S., Testa, B., Carrupt, P. A., Hansch C., Leo, A. (1991) Percutaneous penetration of drugs: a quantitative structure-permeability relationship study. *J. Pharm. Sci.* 80: 744-749
- Watari, N., Kaneniwa, N. (1976) Dissolution of slightly soluble drugs. II. Effect of particle size on dissolution behavior in sodium lauryl sulfate solutions. *Chem. Pharm. Bull.* 24: 2577-2584