The Nature of Experimental Error in Enzyme Kinetic Measurements

By ANDREW C. STORER, MARK G. DARLISON* and ATHEL CORNISH-BOWDEN
Department of Biochemistry, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, U.K.

(Received 1 May 1975)

The nature of experimental error in the determination of initial velocities of enzyme-catalysed reactions was investigated. No support was found for the generalization that such velocities should be homogeneous in variance. Instead the variance increased with the velocity in all of the cases studied. The weighting of sets of replicate observations according to their sample variances was tested in simulated experiments and found to give unacceptable results. It was practicable to decide on the best weighting scheme by examining the variability of the deviations from fitted curves, provided that the number of observations was large enough.

All procedures for analysing experimental results incorporate some assumptions about the nature of experimental error. But, as Ottaway (1973) has pointed out, there have been no studies in enzyme kinetics to assess what assumptions are appropriate and how far they can be trusted. Consequently, all published discussions of the statistical properties of enzyme kinetic measurements rest on conjecture and a presumed affinity with such few studies as have been made in other sciences.

The lack of experimental support for analytical methods would be of little importance if the experimental errors in kinetic measurements were usually very small, or if the methods commonly used were robust (i.e. insensitive to failures of the assumptions), or perhaps even if there were general agreement about the nature of experimental error. But in reality none of these conditions is satisfied (Cornish-Bowden & Eisenthal, 1974), and experimental information is therefore desirable.

Common weighting practice is illogical and inconsistent: on the one hand errors are usually given in general statements as percentages, with the implication of a constant coefficient of variation; but this is contradicted by the equally widespread practice of assigning equal weight to each observation in regression analysis. (If regression with equal weights is applied to transformed data, as is often done, the result may be very different from the intention, but this does not deny the intention.) Several authorities (Wilkinson, 1961; Johansen & Lumry, 1961; Cleland, 1967) suggest that it is usually appropriate to assume that velocities observed under constant conditions are distributed with uniform variance. It would be convenient to be able to rely on a simple generalization of this sort, but other discussions (e.g. Dowd & Riggs, 1965; Reich, 1970; Ottaway, 1973) display less certainty about the universality of any one approach,

and suggest that the experimental variance may often increase with the velocity measured.

It seemed to us that this was an unsatisfactory situation, and that an effort ought to be made to obtain some experimental information about weighting in enzyme kinetics. First, a small amount of evidence would be better than none at all for assessing and developing methods of analysing data. Secondly, we needed some information for the special case of reactions catalysed by glucokinase (EC 2.7.1.2), because we were unwilling to embark on an extended series of kinetic experiments without any knowledge of how to weight the results. Thirdly, we wished to determine the amount and complexity of experimentation necessary for gaining such knowledge, so that reliable rules could be formulated.

Another aspect of data analysis in which current practice is inadequately supported by experimental evidence is the assumption that observations are normally distributed. This assumption is critical because departures from it can cause the least-squares method to give poor results, even with correct weighting. This problem can be largely eliminated in fitting data to the Michaelis–Menten equation, by the use of median estimates of the kinetic parameters (Cornish-Bowden & Eisenthal, 1974), with the added benefit that weighting information is not required. But this method cannot readily be applied to complex equations with several parameters, and experimental justification for assuming a normal distribution is therefore desirable, to establish whether correct weighting alone is likely to be a significant precaution with the least-squares method.

Materials and Methods

Materials and methods for the glucokinase experiments were as given by Storer & Cornish-Bowden (1974). Concentrations of MgATP^{2-} were calculated from the total Mg^{2+} and total ATP con-

* Present address: Department of Molecular Sciences, University of Warwick, Coventry CV4 7AL, U.K.

centrations by means of the stability constants given by Phillips *et al.* (1966). In all experiments the total Mg^{2+} concentration was 1 mM greater than the total ATP concentration to ensure that the ATP would exist almost entirely as $MgATP^{2-}$ and the free (hydrated) Mg^{2+} concentration would be constant.

Bovine pancreatic α -chymotrypsin ($4 \times$ crystallized) was obtained from BDH Chemicals Ltd., Poole, Dorset, U.K. Stock solutions were prepared by dissolving the enzyme in 1 mM-HCl.

Benzoyl-L-arginine ethyl ester hydrochloride was obtained from BDH Chemicals Ltd. Stock solutions were adjusted to pH 6.0 before use by the addition of 0.1 M-NaOH, and were used within 2 days of preparation.

The hydrolysis of benzoyl-L-arginine ethyl ester was followed under a stream of N_2 at pH 8.0 and $25^\circ C$ by means of an automatic titrator model TTT2 with an autoburette ABU12 and Titrigraph SBR3, all manufactured by Radiometer, Copenhagen, Denmark. NaOH (0.1 M) was used as titrant and was standardized each day with standard HCl. Each reaction was left to reach the correct temperature before the addition of enzyme (50 μ l).

For both enzymes the conditions were such that the rate was constant throughout the period used for assessing it. Thus the problems inherent in estimating the initial tangent of a progress curve did not arise. Although errors from this source may often be important, they are likely to be largely systematic and hence outside the scope of this paper. Methods for overcoming such errors have been discussed elsewhere (Cornish-Bowden, 1975).

All experiments were carried out in as 'normal' a way as possible, i.e. no advantage was taken of the fact that each day's experiments included many identical reaction mixtures; instead, each reaction mixture was made up from the stock solutions as if it were a new combination. In each series of experiments each assay was set up with the same number of pipetting operations, and so no allowance was necessary for variations in the cumulative effect of several pipetting operations. Since human error is likely to be a component of the experimental error under investigation, it is relevant to record that different experimenters carried out the experiments with glucokinase (A.C.S.) and α -chymotrypsin (M.G.D.).

In the study of glucokinase by Parry & Walker (1967), velocities were measured at three concentrations of enzyme for each glucose concentration. In this study the velocities, which were provided from the original raw data by Professor D. G. Walker, were adjusted to a common enzyme concentration before analysis by assuming that the velocity was proportional to the enzyme concentration.

Simulated experiments were done essentially as described by Cornish-Bowden & Eisenthal (1974).

Results

Investigation of the proper weights for glucokinase

Ten velocities were measured at each of six concentrations of $MgATP^{2-}$, with other concentrations held constant, and the mean and s.d. were calculated for each set of ten values. A second series of experiments was then carried out in the same way with eleven velocities measured at each of six concentrations of glucose. The calculations involve no assumptions about the form of the rate equation, since for any rate equation the variation between repeated measurements should depend solely on experimental error and time-dependent phenomena (such as slow deterioration of stock solutions, instrumental drift, or poor temperature control). No significant time-dependence was detected and so no correction was made. The results (Fig. 1) demonstrate that the experimental standard error is not a constant but increases with the velocity. The two series of experiments were done on different days with different stock solutions, but apart from consistently higher standard deviations in the second series of experiments the results agree very well and indicate that the experimental error depends on the velocity itself rather than on the concentration of either substrate by itself. Although the continuous line through the origin agrees reasonably well with the results, a rather better fit is given by the broken line with a positive

Fig. 1. Dependence of experimental error on velocity for glucokinase

Each point represents the sample mean and s.d. of ten or eleven velocities for the glucokinase-catalysed phosphorylation of glucose by $MgATP^{2-}$, measured at $30^\circ C$ and pH 8.0, at six different concentrations of glucose (○) or of $MgATP^{2-}$ (●). The two sets of data were obtained on different occasions and the general degree of experimental error was somewhat higher in the variable-glucose experiments; in combining the data into one illustration this difference has been corrected by using two different scales for the ordinate, as indicated.

Fig. 2. Dependence of experimental error on velocity for α -chymotrypsin

Each point represents the sample mean and s.d. of ten velocities for the α -chymotrypsin-catalysed hydrolysis of benzoyl-L-arginine ethyl ester at 25°C and pH 8.0, at five different concentrations of substrate. The concentration of enzyme was 0.167 mg/ml.

intercept on the ordinate. This can be explained by supposing that there are (at least) two additive components in the experimental error, of which one is independent of the velocity and the other is proportional to the velocity.

α -Chymotrypsin

As a check on the generality of the results obtained with glucokinase, a similar series of measurements was made of the rate of the α -chymotrypsin-catalysed hydrolysis of benzoyl-L-arginine ethyl ester. (This substrate was chosen in preference to a more specific one to permit working at a conveniently high substrate concentration.) The results (Fig. 2) also indicate that the standard deviation increases with the velocity, but in this case the intercept is either zero or negative. Since a negative intercept is difficult to rationalize it seems most reasonable to interpret the data to mean that the true intercept is close to zero, so that in the range of velocities studied the only significant component of the error is proportional to the velocity.

Validity of the normal distribution

To assess the actual distribution of errors in glucokinase-catalysed reactions the velocity was measured for each of 100 identical reaction mixtures containing 100 mM-glucose and 4.65 mM-MgATP²⁻, and for each of 80 identical reaction mixtures containing 5 mM-glucose and 0.464 mM-MgATP²⁻. The results (Fig. 3) are rather inconclusive: there is some evidence for long-tailed (leptokurtic) deviations from normality for the set of velocities at low concentrations, but the set of velocities at high concentrations is normal to within sampling error. Similar experiments were

Fig. 3. Distribution of errors for glucokinase

The Figure shows cumulative distribution plots of 100 determinations of the velocity of the glucokinase-catalysed phosphorylation of glucose by MgATP²⁻ at 30°C and pH 8.0 with 100 mM-glucose and 4.65 mM-MgATP²⁻ (●) and of 80 determinations of the velocity with 5 mM-glucose and 0.464 mM-MgATP²⁻ (○). The scale of the abscissa is such that points from a normal distribution of errors would lie on a straight line.

carried out for the α -chymotrypsin-catalysed hydrolysis of benzoyl-L-arginine ethyl ester at five different substrate concentrations, with similarly inconclusive results: one distribution appeared to be short-tailed (platykurtic), one was normal and three were moderately long-tailed.

Although these results support the view that long-tailed deviations from normality are likely enough to be worth guarding against, they are not incontrovertible. But it is clear that no more precise information can be obtained without using much larger samples than are practicable when only a few reactions can be assayed simultaneously, when each must be followed for several minutes, and when stock solutions are not indefinitely stable (so that it would not be valid to continue an experiment for several days or weeks).

Weighting and distribution of errors in completed experiments

For one reason or another it may be inconvenient to carry out studies of the type we have described, but

Fig. 4. Determination of appropriate weights in a completed experiment

Data of Parry & Walker (1967) for the glucokinase-catalysed phosphorylation of glucose by MgATP^{2-} were fitted to the equation $v = V[g/K_1 + g^2/K_1 K_2]/[1 + 2g/K_1 + g^2/K_1 K_2]$, (a) assigning equal weight to each velocity, and (b) assigning each velocity a weight proportional to $1/\hat{v}^2$. When the weighting scheme is correct the points should be scattered in a parallel band about the v axis.

one may nonetheless require some usable information about the weighting and distribution of errors before deciding on the most appropriate method of analysis. To test the feasibility of getting such information from completed experiments, we examined some data of Parry & Walker (1967) for the glucokinase-catalysed phosphorylation of glucose by MgATP^{2-} at constant MgATP^{2-} concentration and various glucose concentrations. Since the dependence of the velocity v on the glucose concentration g appeared to be of the form $v = V[g/K_1 + g^2/K_1 K_2]/[1 + 2g/K_1 + g^2/K_1 K_2]$, where V , K_1 and K_2 are kinetic constants, the data were fitted to this equation by the method of least squares, by using the program described by Cornish-Bowden & Koshland (1970) modified as described by Wharton *et al.* (1974). When the observations were equally weighted, i.e. assuming each velocity to have the same standard deviation, the differences $(v - \hat{v})$ between observed and calculated velocities displayed a clear dependence on the calculated velocity, $\hat{v}$ (Fig. 4a). But when the calculations were repeated assuming each velocity to have the same coefficient of variation, the differences $(v - \hat{v})/\hat{v}$ between observed and calculated values were

much less dependent on $\hat{v}$ (Fig. 4b). Thus the second assumption appears to be more nearly correct than the first, though the truth may well be intermediate between the two, as in our more recent studies of glucokinase described above.

Although these results obtained by examining data long after the event are as clear-cut as those measured directly, the interpretation includes an extra assumption not required for the direct measurements, namely that the data have been fitted to the right equation. If the wrong equation is fitted at least part of the difference between observed and calculated values is due to lack of fit rather than experimental error (Reich, 1970; Reich *et al.*, 1972). But if the observations are sufficiently numerous (as in the example given in Fig. 4) any lack of fit should be clearly evident from systematic variations in the observed differences. Since no systematic trends are perceptible in Fig. 4 it is reasonable to use the results as the basis of a weighting scheme. In fact we now consider the equation given above to be an oversimplification (A. C. Storer & A. Cornish-Bowden, unpublished work), but it does appear to describe the earlier data to within experimental error.

Calculation of weights from replicate observations

Some authors (e.g. Reich *et al.*, 1972; Ottaway, 1973) have suggested that uncertainty about weighting can be overcome by making all observations in replicate, because then the reciprocal of the sample variance of each set of replicates can be used as a weight for the sample mean. But this is likely to be a dangerous procedure if the number of replicates in each group is small, as is clear from the results in Fig. 4. It is quite possible for a triplicate group of observations to possess a small sample variance but a significantly deviant mean (e.g. the values at $\hat{v} = 1.08$); and it is equally possible for a group to possess a large sample variance but a mean that is almost correct (e.g. the values at $\hat{v} = 1.00$). In the first case a weight estimated from the sample variance is too large, and in the second case it is too small. In the limit a group of observations may possess a zero-sample variance, particularly if the values are rounded to a small number of significant figures, corresponding to an infinite weight. The fitted line is then constrained to pass through that group of observations, regardless of any contrary information contained in the rest of the data.

To test the practical performance of this method of weighting, we carried out simulated experiments in which sets of data obeying the Michaelis-Menten equation were generated with normal pseudo-random errors at each of ten substrate concentrations in the range $0.2-2K_m$. In each experiment the standard deviation was assumed (a) to be the same for each observation ('simple errors'), (b) to be proportional to

Fig. 5. Effect of estimating weights from replicate observations

Each point represents the sample S.D. of 500 determinations of K_m by various methods in tests in which velocities were simulated for the Michaelis-Menten equation with (a) simple errors, (b) relative errors, at ten substrate concentrations in the range $0.2-2K_m$. The true values of K_m and V were 1.0 in all cases, and the S.D. of velocities at a substrate concentration equal to K_m was $0.02V$ in all cases. At each concentration the number of velocities simulated in replicate ranged from 2 to 6 as indicated on the abscissa. The filled symbols represent determinations with various assumptions about the proper weights, namely: ●, equal weight for each velocity (correct in case a); ▲, weight proportional to $1/v^2$ (correct in case b); ■, weight proportional to $1/v^4$ (equivalent to an unweighted fit to a double-reciprocal plot). The open symbols represent determinations according to alternative responses to presumed lack of knowledge of the proper weights, namely: ○, weights calculated from sample variances of replicate values (connected by broken lines for added emphasis); △, median estimates. Experiments were also simulated with complex errors as described in the text, but the results are not shown because they were straightforwardly intermediate between the two cases shown.

the true velocity ('relative errors'), or (c) to contain two additive components, one constant and the other proportional to the true velocity (Fig. 5). With weights calculated from the sample variances of duplicate observations the results are very bad, almost as bad as those from the double-reciprocal plot in the case of simple errors, and even worse with the other two assumptions. Although there is a great improvement on changing from duplicate to triplicate observations, and smaller improvements on further increasing the number of replicate values in each sample, the

method is never significantly better than methods based on sensible (though not necessarily correct) assumptions about weights. The results also confirm the conclusion of Cornish-Bowden & Eisenthal (1974) that arithmetic-median estimates are the most generally reliable unless the proper weights for the least-squares method are accurately known.

Calculation of weights by fitting the progress curve to a polynomial

Elmore *et al.* (1963) have described a method of calculating initial velocities by fitting progress curves to polynomials. This method also yields estimates of the variances of the velocities, and they suggest that these be used to calculate weights for further analysis. But the variances calculated in this way are determined solely by the 'noise' within each progress curve and take no account of any errors that affect the curve as a whole. Since the latter may well be at least as important as internal 'noise', we doubt whether weights calculated from the internal 'noise' alone are likely to be reliable.

A convenient test is provided by the worked example given by Elmore *et al.* (1963), which is shown, with some additional information, in Table 1. By using the values of K_m and V given by Elmore *et al.* (1963) one may readily calculate the best-fit velocities $\hat{v}$ and the differences $(v-\hat{v})$ between observed and calculated values. When this is done six of the seven differences are numerically greater than the putative standard errors, in one case eleven times greater and in another five times greater. The probability that this might occur by chance alone is vanishingly small, and one must therefore conclude that standard errors calculated from progress curves do not give realistic estimates of the actual errors and so cannot give valid weights.

Discussion

The main conclusion to be drawn from the results is that weighting of observations presents a more difficult problem than has commonly been recognized. In the first place it is clear that no trust can be placed in facile generalizations that purport to predict proper weights *a priori*: the expectation that velocities measured at constant enzyme concentration should be homogeneous in variance was contradicted by all of the three experimental cases that we examined. Instead the standard deviation clearly increased with the velocity in all three, though not necessarily in proportion. So one must conclude either that we examined three exceptional cases and no typical ones, or that the generalization is incorrect.

It also seems rather hazardous to rely on *a posteriori* weights calculated either from the sample

Table 1. *Example of the use of the method of orthogonal polynomials for estimating initial velocities and their standard errors*

The Table is constructed from data used by Elmore *et al.* (1963) to illustrate the use of the method of orthogonal polynomials for obtaining initial velocities v with their standard errors $\sigma(v)$ from seven progress curves. By using the reciprocals of the estimated variances as weights they applied the method of Wilkinson (1961) to get best-fit estimates $\hat{K}_m = 3.105 \text{ mM}$, $\hat{V} = 0.5527 \text{ mM} \cdot \text{min}^{-1}$. In this Table these estimates were used to obtain the calculated velocities $\hat{v}$ and the differences $v - \hat{v}$. All velocities are expressed in $\text{mM} \cdot \text{min}^{-1}$.

$s \text{ (mm)}$	v	$\sigma(v)$	$\hat{v}$	$v - \hat{v}$	$(v - \hat{v})/\sigma(v)$
20	0.48242 ± 0.00217		0.47842	0.00398	1.83
12	0.40094 ± 0.00336		0.43909	-0.03815	-11.36
8	0.41194 ± 0.00273		0.39816	0.01378	5.05
6.4	0.37397 ± 0.00133		0.37215	0.00182	1.37
4.8	0.32908 ± 0.00185		0.33561	-0.00653	-3.53
3.2	0.28715 ± 0.00600		0.28051	0.00664	1.17
1.6	0.18940 ± 0.00268		0.18795	0.00145	0.54

variances of replicate observations or by fitting polynomials to individual progress curves. In the first case the sample variance of a sample with fewer than five members is a very imprecise estimator of the required population variance and cannot be trusted. In the second there is no certainty that the measured variance accounts for the major part of the actual error, and in the example we studied it clearly did not.

So far as the use of replicate observations is concerned our conclusions broadly agree with those of Jacquez *et al.* (1968), who studied the properties of a similar weighting scheme in straight-line regression. It seems that if such a weighting scheme is to be used then there should be at least five replicates in each group. This is considerably more than is usual in enzyme kinetic experiments, but it may nonetheless be a realistic suggestion in contexts where the experimental error is inevitably very large and unpredictable, as for example in radioactive-tracer experiments (Ottaway, 1973).

It would be wrong to imply that replicate measurements have no place in the statistical analysis of kinetic data. On the contrary, all studies should include at least some replicate observations, but unless they are unusually numerous they cannot be reliably used for calculating weights directly. Instead they provide a valuable check on the plausibility of the statistical assumptions made independently. In the context of a least-squares approach, replicate observations permit an independent calculation of the sum of squares, which may be compared by standard methods (see Draper & Smith, 1966) with the sum of squares calculated from the deviations from the fitted equation.

It might perhaps be argued that our results support a generalization whereby the experimental error be assumed to be proportional to the true velocity unless there is evidence to the contrary. This would have the practical advantage of permitting linear methods to

be used for minimizing the sum of squares in a single step, instead of the tedious iterative methods required with even the simplest equations when all velocities are assumed to have the same standard deviation. But it is much safer to avoid generalizations altogether, and to treat each series of experiments as a separate problem, using the weighting scheme that gives the most consistently satisfactory results. If there are not enough observations in any one experiment to indicate a clear preference for one weighting scheme it is possible to combine the results of several experiments, provided that day-to-day variations in the magnitude of the experimental error are negligible or can be corrected for. It is probably unrealistic to make a weighting decision for each experiment in isolation, because then one would in effect assume that the qualitative character of the experimental error was liable to vary from day to day.

We thank the Medical Research Council for financial assistance, the University of Birmingham for a University Research Scholarship (to A. C. S.), Professor D. G. Walker for providing us with the raw data from some of his published work, and Dr. J. H. Ottaway for valuable comments on the manuscript.

References

- Cleland, W. W. (1967) *Adv. Enzymol. Relat. Areas Mol. Biol.* **29**, 1-32
- Cornish-Bowden, A. (1975) *Biochem. J.* **149**, 305-312
- Cornish-Bowden, A. & Eisenthal, R. (1974) *Biochem. J.* **139**, 721-730
- Cornish-Bowden, A. & Koshland, D. E., Jr. (1970) *Biochemistry* **9**, 3325-3336
- Dowd, J. E. & Riggs, D. S. (1965) *J. Biol. Chem.* **240**, 863-869
- Draper, N. R. & Smith, H. (1966) *Applied Regression Analysis*, pp. 26-32, Wiley, New York

- Elmore, D. T., Kingston, A. E. & Shields, D. B. (1963) *J. Chem. Soc.* 2070–2078
- Jacquez, J. A., Mather, F. J. & Crawford, C. R. (1968) *Biometrics* **24**, 607–626
- Johansen, G. & Lumry, R. (1961) *C. R. Trav. Lab. Carlsberg* **32**, 185–214
- Ottaway, J. H. (1973) *Biochem. J.* **134**, 729–736
- Parry, M. J. & Walker, D. G. (1967) *Biochem. J.* **105**, 473–482
- Phillips, R. C., George, P. & Rutman, R. J. (1966) *J. Am. Chem. Soc.* **88**, 2631–2640
- Reich, J. G. (1970) *FEBS Lett.* **9**, 245–251
- Reich, J. G., Wangermann, G., Falck, M. & Rohde, K. (1972) *Eur. J. Biochem.* **26**, 368–379
- Storer, A. C. & Cornish-Bowden, A. (1974) *Biochem. J.* **141**, 205–209
- Wharton, C. W., Cornish-Bowden, A., Brocklehurst, K. & Crook, E. M. (1974) *Biochem. J.* **141**, 365–381
- Wilkinson, G. N. (1961) *Biochem. J.* **80**, 324–332