Journal of Pharmacy and Bioresources
Vol. 1 no. 1, pp. 17-23 (September 2004)

Journal of
PHARMACY AND
BIORESOURCES

The reliability of randomly selected final year pharmacy students in drug product assessment

Musa A. Ibrahim*, Oluwatobi Ajibola, Andrew O. Ogunleye, Egweye Adaji and Jennifer D. Audu-Peter

Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmaceutical Sciences, University of Jos, P. M. B. 2084 Jos. Nigeria.

Received 14th April 2004; Revised, accepted 2nd July 2004

Abstract

Employing ANOVA, factorial experimental analysis, and the theory of error, reliability studies were conducted on the assessment of the drug product chloroquine phosphate tablets. The G-Study employed equal numbers of the factors for uniform control, and involved three analysts (randomly selected final year Pharmacy students), conducting three tests each (drug content, friability, and hardness), on each of three identical samples taken from a single batch. The results were converted to percentages of "true scores" for uniformity and their collective statistical treatment. The highest source of variance, is the analysts' factor which gave a component variance of 7,221 compared to 333 and 330 for the samples and tests factors respectively. The D-studies showed how to improve assessment reliability (R_A) of the students, and the developed equation below relating the numbers of samples (n_S), tests (n_T) and analysts (n_A), was used to calculate the number of students that would give the acceptable reliability level of 80%, for any given pair of numbers of samples and tests, and tabulated for instant reading off: $3,049.366 = 333.424/n_S + 330.234/n_T + 7,221.525/n_A + 272.84/n_Sn_T + 1,414.4/n_Sn_A + 2,295.13/n_Tn_A + 329.91/n_Sn_Tn_A$. For one sample and one test, the needed six students for achieving the acceptable reliability, may be reduced more efficiently by increasing the number of tests rather than samples. Automation of assessment also improved the reliability of the G-Study from 0.798 to 0.979, and that of the D-Study starting point from 0.5 to 0.906.

Keywords: Reliability; Pharmacy Students; Drug product assessment.

Introduction

In any assessment procedure that is based on judgmental abilities of individuals, problems do arise in reaching or ensuring satisfactory levels of reliability (Smith *et al.*, 1995). In the pharmaceutical world, drug products are often subjected to assessment tests, and the assessment results could be influenced by factors like samples, tests and analysts. Cronbach *et al.* (1963) have pointed out that an investigator asks about the precision or reliability of a measure because he wishes to generalize from the observations in hand to some class of observations to which it belongs. Generalizability studies (G-Studies) are designed to identify the major sources of variance in any assessment procedure. This information is then used to design Decision studies (D-Studies) from which decisions can be made as to the most

* Corresponding author. *E-mail address*: ibrahim@unijos.edu.ng

efficient use of available resources to minimize the identified sources of errors.

The reliability theory as described by Cronbach *et al.* (1963 and 1972) and Levy (1974) is summarized as follows: The classical reliability theory regards an observed test score (x) as being made up of two components, *viz* a true score (π) and an error component (e). The model could be stated as follows:

$$x = \pi + e \text{ ---------- equation 1}$$

Under assumptions concerned with the independence of the error components, it may be shown that the variance of the observed true score (σx^2) is equal to the sum of two component variances, namely the variance of the true score ($\sigma\pi^2$) and the variance of the error components (σe^2),

$$\text{i.e } \sigma x^2 = \sigma\pi^2 + \sigma e^2 \text{ ----- equation 2}$$

The reliability coefficient (R) is equal to the ratio of the true score variance to the observed score variance (Gulliksen, 1950),

$$\text{i.e. } R = \frac{\sigma\pi^2}{\sigma x^2} = \frac{\sigma\pi^2}{\sigma\pi^2 + \sigma e^2} \text{ --------equation 3}$$

A number of modifications in the D-Studies are usually considered and the one that gives an acceptable level of reliability coefficient is selected as been the best procedure that will reduce the errors to a minimum. A reliability coefficient of about 0.8 on a 0 – 1 scale (i.e. 80%) is usually considered acceptable (Smith *et al.*, 1995).

Quite often, Pharmacy students are assigned tasks in research or during industrial attachment that necessitate the assessment of drug products. In this work, quality control tests such as drug content, friability, and hardness will be carried out on samples from a reputable manufacturer using randomly selected final year Pharmacy students as analysts. From the G-Study, various D-Studies will be designed to enable decisions on reliability of drug product assessment by Pharmacy students to be taken.

Experimental

Materials. Chloroquine phosphate tablets (three containers representing 3 samples, from the same batch from a reputable manufacturer), chloroquine phosphate powder (Bayer, W. Germany), hydrochloric acid (FSA Laboratory Ltd, Poole England).

Methods. Analysts took tablets randomly from the containers for the following tests:

(a) Drug Content determination: The drug contents of the various samples of the tablets were determined spectrophotometrically at 343nm in 0.01M HCl on a UV/Vis spectrophotometer (Spectronic 200 – 20 double beam Hitachi, Japan) with the aid of a calibration curve.

(b) Physical tests of tablets: The friability, and hardness tests were carried out using the friability tester (Erweka – apparatebau GMBH, Germany), and the Stokes-Monsanto hardness tester (CT 40 Engineering systems Ltd., England) respectively, by the methods described by Rawlins (1980).

(c) Unifying the scales of tests scores: Each analyst's mean score was expressed as a percentage of a "true score" which was taken as the mean of several scores. This made the score scales of the different tests uniform, to enable the collective statistical treatment of the results. Negative and positive readings in Table 1, indicate less than or greater than 100% of the "true score", i.e. 100 was subtracted from each entry to reduce the data and simplify the arithmetic.

(d) Calculation of component variances: Using the methods of Pantony (1961) and Davies and Goldsmith (1970), 3–way ANOVA and factorial experimental analysis were employed in calculating the total sum of squares (SSQ_T), sum of squares (SSQs) for the individual factors and their interactions. By the same methods, the mean squares and component variances were calculated from the numbers of each variable (n),

remainder/random error (σ_0^2), SSQs and degrees of freedom (ϕ). The results are shown in Table 2.

$$R_A = \frac{\beta}{\beta + eev} \quad \text{---------- Equation 5}$$

where β is the total component variance contributing to the variance of the true score. In this analysis β has a value of 12,197.463 as given in Table 2.

(b) Generalizability study (G–Study)

All three analysts, each scoring all three samples on all three tests.

$$eev = \frac{333.424}{3} + \frac{330.234}{3} + \frac{7,221.525}{3} + \frac{272.84}{9} + \frac{1,414.4}{9} + \frac{2,295.13}{9} + \frac{329.91}{27}$$

$\therefore eev = 3083.104$

$$R_A = \frac{12,197.463}{12,197.424 + 3,083.104} \qquad \therefore R_A = 0.798$$

(c) Decision studies (D–Studies)

(i) Starting point of D-Studies. This considered one analyst conducting one test on one sample:

$$eev = \frac{333.424}{1} + \frac{330.234}{1} + \frac{7,221.525}{1} + \frac{272.84}{1} + \frac{1,414.4}{1} + \frac{2,295.13}{1} + \frac{329.91}{1}$$

$\therefore eev = 12,197.463$

$$R_A = \frac{12,197.463}{12,197.463 + 12,197.463} \qquad \therefore R_A = 0.5$$

(ii) Subsequent D-Studies. These were designed such that the number of one factor was varied at a time while keeping the numbers of the other two factors in c(i) above constant at one. This enabled the effect of each factor on R_A to be determined as shown in Figure 1.

(d) Effect of automation.

With automation the variance due to analysts is minimized and may be neglected:

(i) On G-Study

$$eev = \frac{333.424}{3} + \frac{330.234}{3} + 0 + \frac{272.84}{9} + 0 + 0 + \frac{329.91}{27} \qquad \therefore eev = 263.759$$

$$R_A = \frac{12,197.463}{12,197.463 + 263.759} \qquad \therefore R_A = 0.979$$

(ii) On D-Studies starting point :

$$eev = \frac{333.424}{1} + \frac{330.234}{1} + 0 + \frac{272.84}{1} + 0 + 0 + \frac{329.91}{1} \qquad \therefore eev = 1,266.408$$

$$R_A = \frac{12,197.463}{12,197.463 + 1,266.408} \qquad \therefore R_A = 0.906$$

(e) Development of formula relating the numbers of samples, tests and analysts for achieving a reliability of 0.8 (used for generating Table 3).

Working backwards from equation 5, R_A is made equal to 0.8

$$0.8 = \frac{12,197.463}{12,197.463 + eev}$$

$eev = 3,049.366$, which when substituted in equation 4 gives:

$$3,049.366 = \frac{333.424}{n_S} + \frac{330.234}{n_T} + \frac{7221.525}{n_A} + \frac{272.84}{n_S n_T} + \frac{1414.4}{n_S n_A} + \frac{2295.13}{n_T n_A} + \frac{329.91}{n_S n_T n_A} \quad \text{(Equation 6)}$$

Table 3: The number of analysts (n_A) that will ensure a reliability of 0.8 for any given pair of numbers of samples and tests.

		Number of tests (n_T)									
		1	2	3	4	5	6	7	8	9	10
Number of samples (n_S)	1	*5.33*	*4.12*	*3.78*	*3.62*	*3.53*	*3.50*	*3.43*	*3.40*	*3.37*	*3.35*
	2	*4.30*	*3.46*	*3.21*	*3.09*	*3.02*	*2.97*	*2.94*	*2.92*	*2.90*	*2.88*
	3	*4.01*	*3.26*	*3.04*	*2.93*	*2.86*	*2.82*	*2.79*	*2.77*	*2.75*	*2.74*
	4	*3.88*	*3.17*	*2.95*	*2.85*	*2.79*	*2.75*	*2.72*	*2.70*	*2.68*	*2.67*
	5	*3.80*	*3.11*	*2.90*	*2.80*	*2.75*	*2.71*	*2.68*	*2.66*	*2.64*	*2.63*
	6	*3.75*	*3.08*	*2.87*	*2.77*	*2.72*	*2.68*	*2.65*	*2.63*	*2.61*	*2.60*
	7	*3.71*	*3.05*	*2.85*	*2.75*	*2.70*	*2.66*	*2.63*	*2.61*	*2.60*	*2.58*
	8	*3.68*	*3.03*	*2.83*	*2.74*	*2.68*	*2.64*	*2.62*	*2.60*	*2.58*	*2.57*
	9	*3.66*	*3.02*	*2.82*	*2.72*	*2.67*	*2.63*	*2.60*	*2.58*	*2.57*	*2.56*
	10	*3.65*	*3.01*	*2.81*	*2.71*	*2.66*	*2.62*	*2.60*	*2.58*	*2.56*	*2.55*

N.B.: Each n_A value (italicized) above should be rounded to the next whole number.

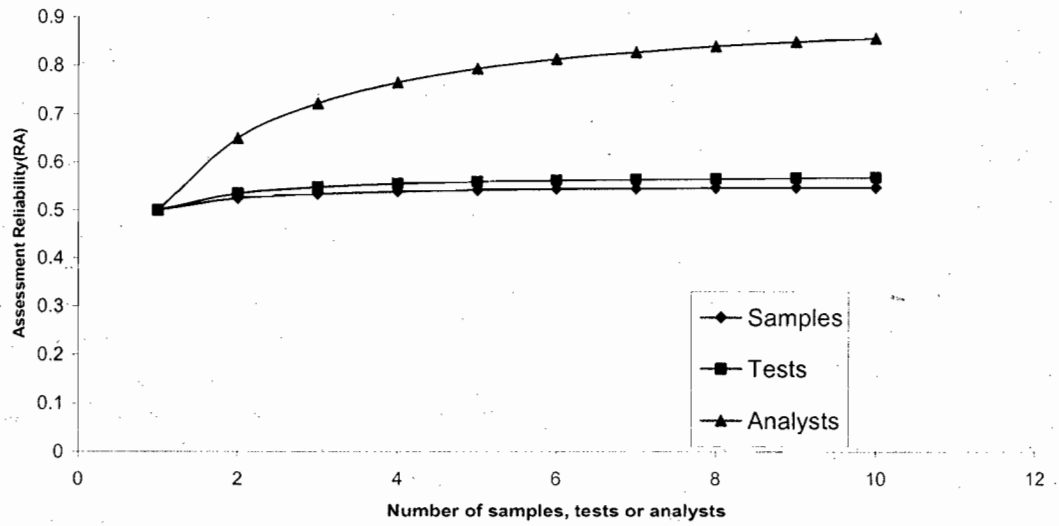

Figure 1: The Effects of increasing the number of one variable factor at a time from the starting point of D-studies on assessment reliability.

Pharmacy students may often be engaged in drug product assessment during research or industrial attachment, and their limited experience will expectedly affect the reliability of the assessment exercise. The G-Study ensured equality in the numbers of samples, tests and analysts (3 each), to give comparable control of the factors, while the D–Studies started with one analyst conducting one test on one sample. Subsequent D-Studies varied the number of one factor at a time while keeping the numbers of the other two factors constant at one, to see the effect of each factor on assessment reliability (R_A). Also the samples were taken from a single batch, thus differences in assessment readings should be due to errors (from samples, tests, analysts, indeterminate or random errors i.e. remainder, etc.). The G-Study enabled the calculations of the magnitude of variance caused by each factor, their interactions, and other random errors (i.e. remainder) as given in Table 2. The information was then used to calculate the assessment reliabilities (R_A) for the D-Studies.

Table 2 shows that the component variance due to analysts (about 7,221) is over twenty one times greater than those due to samples or tests (about 333 and 330 respectively), indicating analysts to be the greatest source of error. However, the variation due to analysts may be eliminated by the use of automated test systems, as shown in the improvement of R_A from 0.798 to 0.979 by automating the G-Study, and from 0.5 to 0.906 by automating the starting point D-Study. An R_A of 0.5 for one analyst conducting one test on one sample suggests a 50-50 chance of the result being correct or wrong. Increasing the sample number only slightly improved R_A, while increasing the number of tests had a slightly greater impact on R_A. On the other hand, increasing the number of analysts profoundly increased R_A (Figure 1). Thus increasing the number of analysts is much more efficient in improving R_A, due to the reduction of the high analysts' component variance. Using equation 6, the number of student analysts that would give a reliability of 0.8 for any given pair of number of samples and number of tests was determined, and tabulated (Table 3) for instant reading off. The number of analysts in Table 3 would have to be upgraded to whole numbers, which will further improve reliability.

Sometimes, only one test may be required e.g. the concentration of prohibited substances in samples taken from athletes, while on the other hand a host of tests may be required for seized suspected dangerous drugs, and the number of samples available may vary in each case. For one sample and one test, six student analysts are needed to achieve the acceptable reliability level of 0.8 (Table 3). Table 3 shows that increasing the number of tests (i.e. moving along the horizontal rows) exerts greater reduction in the number of analysts, than increasing the number of samples (i.e. moving along the vertical columns). Looking at Table 2, it would mean that reducing the higher test – analyst component variance (2,295) is more influential than reducing the smaller sample – analyst component variance (1,414). Thus in trying to reduce the number of analysts, it would be preferable to increase the number of tests, rather than the number of samples. Table 3 further shows that the minimum number of analysts is 3 (after upgrading), for an R_A of 0.8 even for 10 samples and 10 tests. However, the use of 4 samples, 3 tests and 3 analysts or 3 samples, 4 tests and 3 analysts appear to be the most suitable combination and the best utilization of resources.

Conclusions

From the observations derived from the G- and D-Studies, the following generalizations were made:

1) In drug product assessment by Pharmacy students, the analysts' factor is the main source of error. The use of automated assessment procedures can eliminate errors due to analysts. It would be interesting to carry out reliability studies using automated test equipment.
2) The number of student analysts that will ensure 80% reliability may be reduced more efficiently by increasing the number of tests rather than samples.
3) This study enabled a formula to be developed from which a useful table was generated. With this table, the number of student analysts that would ensure the acceptable reliability level of 0.8 for any given pair of numbers of samples and tests can easily be obtained.

Acknowledgement

The correspondence author is thankful to the University of Jos, Jos Nigeria, for providing the research grant to undertake the study.

References

Cronbach L.J., Gleser G.C., Nanda H. and Rajaratnam N. (1972). The dependability of behavioural measurements: Theory of generalizability for scores and profiles. New York: Wiley.

Cronbach L.J., Rajaratnam N. and Gleser G.C. (1963). Theory of generalizability: a liberalization of reliability theory. *Br. J. Statis. Psychol.*, 16: 137–163.

Davies O.L. and Goldsmith P.L. (1970). Statistical methods in research and production 4th edition, Oliver and Boyd.

Gulliksen H. (1950). Theory of mental tests. New York: Wiley.

Levy P. (1974). Generalizability studies in clinical settings. *Br. J. Soc. Clin. Psychol.*, 13, 161--172.

Pantony D.A. (1961). A Chemist's introduction to statistics, theory of error, and design of experiments. R.I.C. lecture series No. 2.

Rawlins E.A. (1980). Tablets and Capsules, in: Bentley's Textbook of Pharmaceutics, 8Th edition, Bailliere Tindal, London, pp. 298 – 299.

Smith F.J., Jolly B.C. and Dhillon S. (1995). The application of generalizability theory to assessment in a practice-based diploma in clinical pharmacy. *The Pharmaceutical Journal*, 254, 198 - 199.