Appl. Statist. (1985)
34, *No.* 3, *pp.* 258–263

Use of the Structural Equations Model in Assessing the Reliability of a New Measurement Technique

By GABRIELLE E. KELLY †

University College, Cork, Ireland

[Received May 1984. Revised February 1985]

SUMMARY

Methods for assessing a new measurement technique are examined. The structural equations model is used to estimate the linear relationship between new and standard methods. The delta method is used to find the variance of the estimated parameters. A comparison with regression analysis is made. This is illustrated by a study on a new method of measuring cardiac output with data drawn from published reports. The estimates from the different reports are combined to give an overall estimate of the relationship between the new and standard methods.

Keywords: Structural equations model; Regression; Delta method; Reliability; Nonparametric maximum likelihood; Mixing distribution

1. Introduction

When a new method of measurement is introduced the question of its comparability with accepted methods must be considered. Here we analyse 13 reports from published data concerned with methods of measuring cardiac output. In recent years a thermal dilution (TD) method was introduced which has become more popular than the standard Fick (F) or dyel dilution (DD) methods. This is so because of its simple technique, immediate results and easily repeated measurements. The purpose of this study is to determine how to interpret a TD determination. Here we define a TD "determination" as the average of a variable number of TD "measurements". There are two critical questions concerning the new method: (1) to what extent does a difference between two cardiac output determinations separated by time actually represent a real change in cardiac output? (reproducibility); (2) what is the relation between a TD determination and an F or DD determination (accuracy)? Many studies in the literature present the results of simultaneous determinations of the TD method with either the F or DD method and examine the relationship of the two methods by regression. Such an analysis might produce high regression coefficients, yet examination of individual data points might show discrepancies of 50 per cent or more between the two methods. Moreover, as all the methods are subject to error there is no compelling choice for which variable should be considered the independent one in a regression equation. Thus the question of accuracy was not adequately answered by these studies. In this paper we have compiled and re-analysed existing data so that it could be more effectively used in assessing the quantitative properties of the TD method.

† *Present address*: School of Public Health, Division of Biostatistics, Columbia University, 600 West 168th Street, New York, N.Y. 10032, USA.

 0035–9254/85/34258 $2.00

2. Statistical Analysis

The reproducibility of the TD method in each study was evaluated by examining the standard deviation of repeated measurements of cardiac output done during a stable period. The achieved reproducibility in each study is expressed in terms of the standard error of the mean, relative to the mean output

$$\text{SEM\%} = \frac{\text{standard error of the mean}}{\text{mean cardiac output}} \times 100.$$

The accuracy of the TD method in comparison with either the F or the DD methods was evaluated for each study by the method of structural equations analysis. Structural analysis is a generalization of regression analysis. The structural model assumes that the correct TD and F or DD levels in the ith person are x_i and y_i respectively, but that these cannot be observed without error. The observable quantities are ξ_i and η_i which are related to x_i and y_i by

$$\xi_i = x_i + \delta_i \quad \text{and} \quad \eta_i = y_i + \epsilon_i. \tag{1}$$

The correct levels x_i and y_i are related through the structural relation

$$y_i = \alpha + \beta x_i. \tag{2}$$

The basic question of whether the two different methods of measurement are equivalent reduces to whether $\alpha = 0$ and $\beta = 1$.

The underlying probability model assumes that the pairs (x_i, y_i) are independently and identically jointly distributed, with x_i having mean μ_x and variance σ_x^2. The errors δ_i and ϵ_i are assumed to be identically distributed each according to its own distribution with mean 0 and variance σ_δ^2 and σ_ϵ^2 respectively and to be independent of each other and the pairs (x_i, y_i). The underlying joint distribution function of all variables will be denoted by F.

The method of moments equations for the unknown parameters are

$$\begin{aligned} &\hat{\mu}_x = \bar{\xi}, \quad \hat{\alpha} + \hat{\beta}\,\hat{\mu}_x = \bar{\eta}, \\ &\hat{\sigma}_x^2 + \hat{\sigma}_\delta^2 = s_\xi^2, \hat{\beta}^2\,\hat{\sigma}_x^2 + \hat{\sigma}_\epsilon^2 = s_\eta^2, \\ &\hat{\beta}\,\hat{\sigma}_x^2 = s_{\xi\eta} \end{aligned} \tag{3}$$

where ξ, η, s_ξ^2, s_η^2 and $s_{\xi\eta}$ are the sample means, variances and covariances of the ξ and η variables. For normal populations these equations are also maximum likelihood equations. Without extra information these equations do not have a unique solution because there are five equations in six unknowns. If extra information is available concerning σ_δ^2 and σ_ϵ^2 their ratio $\lambda = \sigma_\epsilon^2/\sigma_\delta^2$ may be estimated. Assuming that λ is known, the solutions for $\hat{\alpha}$ and $\hat{\beta}$ is given by

$$\begin{aligned} &\hat{\alpha} = \bar{\eta} - \hat{\beta}\,\bar{\xi}, \\ &\hat{\beta} = \frac{s_\eta^2 - \lambda s_\xi^2 + \sqrt{\{(s_\eta^2 - \lambda s_\xi^2)^2 + 4\,\lambda s_{\xi\eta}\}}}{2 s_{\xi\eta}} \end{aligned} \tag{4}$$

and the estimates of the other parameters follow from (3). This model has been extensively discussed in the literature. For a review the reader is referred to Kendall and Stuart (1961, pp. 377-382). It follows from the delta method (Rao, 1965, pp. 321-322) that approximate variances for $\hat{\alpha}$ and $\hat{\beta}$ are given by

$$\begin{aligned} \sigma_{\hat{\beta}}^2 = n^{-1}\,\{&\sigma_\delta^2\,(\beta^2 + \lambda)/\sigma_x^2 + \lambda(\sigma_\delta^2/\sigma^2)^2\,(\beta^4 - 4\lambda\beta^2 + \lambda)/(\beta^2 + \lambda)^2 + \\ &\beta^2\,(E_F\epsilon^4 + \lambda^2 E_F\delta^4)/(\beta^2 + \lambda)^2\,\sigma_x^4\}, \end{aligned} \tag{5}$$

and

$$\sigma^2_{\hat{\alpha}} = n^{-1}\{(\sigma^2_\delta \beta^2 + \lambda) - 2\mu_x (\beta E_F \epsilon^3 + \lambda\beta^2 E_F \delta^3)/\sigma^2_x (\beta^2 + \lambda) + \eta\mu^2_x \sigma^2_{\hat{\beta}}\},$$

(see Kelly, 1984), where E_F denotes expectation w.r.t. the distribution function F.

For normal populations the equations (5) reduce to

$$\sigma^2_{\hat{\beta}} = n^{-1} \{\beta^2 \sigma^2_\xi \sigma^2_\eta - \sigma^2_{\xi\eta})/\sigma^2_{\xi\eta}\}, \tag{6}$$

and

$$\sigma^2_{\hat{\alpha}} = n^{-1} \{\sigma^2_\eta - 2\beta\sigma_{\xi\eta} + \beta^2 \sigma^2_\xi + \eta\mu^2_\xi \sigma^2_{\hat{\beta}}\}.$$

as given in Gleser (1981, pp. 37–38). These can be estimated by substituting sample moments for population moments. Since all method of moments estimators are M-estimators by the results of Huber (1977, pp. 20–22) $\hat{\alpha}$ and $\hat{\beta}$ will be consistent and asymptotically normal. Thus tests of hypothesis and confidence intervals can be constructed in the usual way.

It should be noted that the estimators given in equations (4) and (5) make no distributional assumptions.

Simultaneous measurements on both DD and F were not available in any single study. All 13 slope estimators were used to estimate the population distribution G of true slopes in the structural equations model between new and standard methods. We assume the $\hat{\beta}_i$, $i = 1, \ldots, 13$, come from a mixture distribution with density

$$h_i(\hat{\beta}_i/G) = \int_R f_i(\hat{\beta}_i/\beta)\, dG(\beta), \quad i = 1, \ldots, 13,$$

where

$$f_i(\hat{\beta}_i/\beta_i) = N(\beta_i, \sigma^2_i),$$

with known σ^2_i given by estimating equation (5), and

$$\beta_i \sim G$$

The nonparametric maximum likelihood estimator (NPML) of G can be calculated as in Laird (1978). The estimated G is a step function with steps at θ_j with probability Π_j, $j = 1, \ldots, k$, where the number k of steps is also estimated. Using the estimated G the prior mean and variance of the β distribution can be found, thus giving us one overall estimate of β and an associated variance. The posterior means and variances of β for the individual studies can also be calculated. The question of whether there are systematic differences between the F and DD studies may be assessed by examining whether the two types of study give systematically different slopes.

3. Results

Studies reporting the reproducibility of the TD method are listed in Table 1 with the achieved values of SEM%. These results imply that the TD method of determining cardiac output might need to change by between 10.5 to 26.1 per cent ($1.96\sqrt{2}$ SEM%) to establish a significant change at the 5 per cent level.

A complete list of the studies comparing the TD method with one of the reference methods F or DD is given in Table 2. Using scatter plots a rough empirical check on the model was carried out; that is, the linearity and homoscedacity assumptions. These appeared to be plausible. In each study only one reference method was available so we did not have simultaneous measurements on all three methods as in Barnett (1969). Because of the form in which the data was presented in the studies neither did we have available repeated measurements on each (x_i, y_i) as in Chan and Mak (1969). However most studies reported reproducibility estimates as given in Table 1 and estimates of λ, $\hat{\lambda}$ could be obtained directly for these studies. For the remaining studies in which reproducibility estimates were not reported, the average estimate of the variance of individual

TABLE 1
Studies evaluating the reproducibility of the thermodilution (TD) method

Reference	*SEM%*
Branthwaite, 1968 (2)	3.6
Olsson, 1970 (12)	7.4
Forrester, 1972 (5)	4.6
Enghoff, 1973 (3)	5.5
Andreen, 1974 (1)	3.5
Vandermoten, 1977 (15)	6.8
Kohanna, 1977 (11)	8.7
Fischer, 1978 (4)	6.9
Hoel, 1978 (8)	5.0

observations from the studies summarized in Table 1 was used to derive an estimated SEM% using the known numbers of measurements per determination in each study.

The values $\hat{\lambda}$ were then used to solve equation (4). It should be noted that the resulting estimators are still method of moments estimators and so will be consistent and asymptotically normal. This method was also suggested by Chan and Mak (1969, p. 266). The estimates $\hat{\lambda}$ entering into the structural equation and parameter estimates are displayed in Table 2. Equations (5) were used to estimate the variance of the estimated parameters. It was assumed that using an estimate $\hat{\lambda}$ for λ would not invalidate these unduly. There was little difference between these estimates and those provided by equations (6) which require the extra assumption of normality. Except for one study (13) (peri-operative conditions) 99 per cent confidence limits for the slope and intercept include the possibility of the ideal calibration line (slope = 1, intercept = 0).

TABLE 2
Studies evaluating the thermodilution method by comparison with the Fick (F) or the dye-dilution (DD) methods

References	*Number of points*	*Reference method*	*Ratio of error variances* $\hat{\lambda}$	*Parameter estimates* $\hat{\beta} \pm$ (SD)	$\hat{\alpha} \pm$ (SD)
Branthwaite, 1968 (2)	17	F	0.06	1.04 (0.02)	0.18 (0.15)
Olsson, 1970 (12)	17	DD	0.11	1.03 (0.06)	−0.13 (0.30)
Ganz, 1971 (6)	63	DD	0.42	1.02 (0.04)	−0.09 (0.20)
Hodges, 1975 (7)	98	F	0.25	0.72 (0.10)	1.45 (0.70)
Weisel, 1975 (17)	22	DD	0.28	0.9 (0.04)	0.56 (0.21)
Venkataraman, 1976 (16)	15	DD	0.42	1.16 (0.19)	−0.28 (0.57)
Vandermoten, 1977 (15)	12	F	0.22	1.1 (0.13)	−0.43 (0.71)
Kohanna, 1977 (11)	25	DD	1.0	1.05 (0.12)	−0.35 (0.59)
Fischer-A, 1978 (4)	34	DD	1.0	0.83 (0.16)	0.18 (0.58)
Fischer-B, 1978 (4)	32	DD	1.0	1.03 (0.08)	−0.21 (0.39)
Hoel, 1978 (8)	10	F	0.19	0.92 (0.08)	0.37 (0.44)
Stawicki-A, 1979 (13)	11	F	0.19	1.13 (0.02)	−0.02 (0.15)
Stawicki-B, 1979 (13)	10	F	0.19	0.92 (0.08)	0.50 (0.44)

Our calculated structural lines of the TD and other methods, for these studies, are displayed in Fig. 1.

The 13 slope estimates were then used to obtain the NPML estimator of G, the population distribution of slopes between new and standard methods. The estimated G had four distinct steps: $\theta = (0.73, 0.92, 1.03, 1.13)$ with probability $\pi = (0.05, 0.29, 0.53, 0.13)$. The last three components of θ were not considered to be different from a practical viewpoint, but the step at 0.73 indicates that 5 per cent of the time the TD measurement will be smaller than the standard

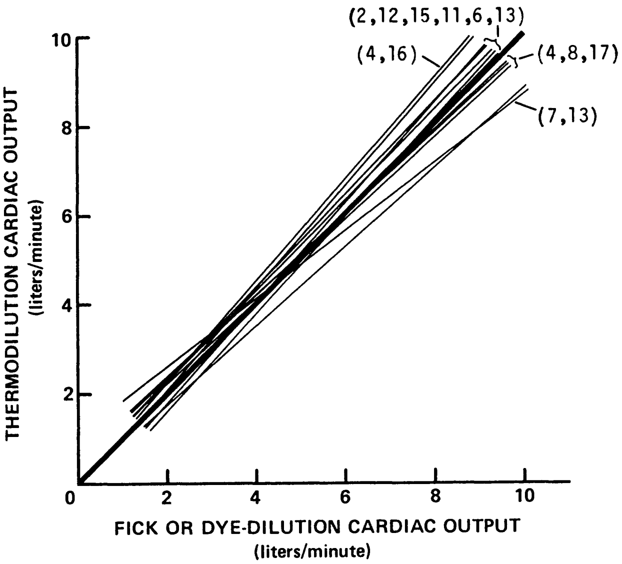

Fig. 1. Calculated structural lines for each clinical study comparing thermodilution to simultaneously obtained Fick or dye-dilution methods of determining cardiac output. Reference numbers are enclosed in parentheses. The ideal calibration line is identical to the average of the structural lines and is shown in heavy print.

methods. The mean and standard deviation of the estimated G were 1.00 and 0.09 respectively, i.e. the overall estimate of β is 1.00 with standard error 0.09. The posterior (i.e. revised) means and variances of β were then calculated, but the smaller values of θ were not systematically related to a particular method-F or DD. Thus all three methods TD, F and DD can be regarded as comparable. This answers the question of accuracy while the question of reproducibility is answered by Table 1.

To study the effect of assumptions about λ, for each study the two regression lines η on ξ and ξ on η corresponding to $\lambda = +\infty$ and $\lambda = 0$, respectively were also calculated. The resulting estimators of slope $\hat{\beta}(\lambda = 0)$ and $\hat{\beta}(\lambda = +\infty)$ were similar to each other and to that of the structural analysis $\hat{\beta}(\lambda)$. It is easily verified that

$$\beta(\lambda = +\infty) \leqslant \beta(\lambda) \leqslant \beta(\lambda = 0) \tag{7}$$

with equality between $\beta(\lambda = +\infty)$ and $\beta(\lambda = 0)$ iff $\eta = c\xi$ with probability one for some constant c. Since the results of the studies reflect this equality it is not surprising that all estimated slopes are practically equal. However, the precision of these estimators differs according to which analysis is used. The standard errors given by regression analyses are smaller than those of the structural analysis. This is to be expected since the structural method takes into account the increased uncertainty due to not knowing the x variable exactly.

4. Further Remarks

Because this was a retrospective study, we did not have first-hand knowledge of how each experiment was done. It was unfortunate that several studies required translation of data in the figures into numerical form to allow analysis. We have tried, however, to extract all relevant information from each study so that studies could be directly compared.

Acknowledgements

The author wishes to acknowledge that this was joint work with Christian W. Stez, Rupert G. Miller and Thomas A. Raffin.

References

Andreen, M. (1974) Computerized measurement of cardiac output by thermodilution: methodologic aspects. *Acta Anaesthesiol. Scand.*, **18**, 297–305.

Barnett, V. (1969) Simultaneous pairwise linear structural relationships. *Biometrics*, **22**, No. 1.

Branthwaite, M. A. and Bradley, R. D. (1968) Measurement of cardiac output by thermal dilution in man. *J. Appl. Physiol.*, **24**, 434–438.

Chan, L. and Mak, T. (1979) Maximum likelihood estimation of a linear structural relationship with replication. *J. R. Statist. Soc.* B, **41**, 263–268.

Enghoff, E. and Sjogren, S. (1973) Thermal dilution for measurement of cardiac output in the pulmonary artery in man in relation to choice of indicator volume and injectate time. *Uppsala J. Med. Sci.*, 78, 33–37.

Fischer, A. P., Benis, A. M., Jurado, R. A., Seely, E., Teinstein, P. and Litwak, R. S. (1978) Analysis of errors in measurement of cardiac output by simultaneous dye and thermal dilution in cardiothoracic surgical patients. *Cardiovascular Res.*, **12**, 190–199.

Forrester, J. S., Ganz, W., Diamond, G., McHugh, T., Chonette, D. W. and Swan, H. J. C. (1972) Thermodilution cardiac output determination with a single flow directed catheter. *Amer. Heart J.*, **83**, 306–311.

Ganz, W., Donoso, R., Marcus, H. S., Forrester, J. S. and Swan, H. J. C. (1971) A new technique for measurement of cardiac output by thermodilution in man. *Amer. J. Cardiol.*, **27**, 392–396.

Gleser, L. J. (1981) Estimation in a multivariate "errors in variables" regression model: large sample results. *Ann. Statist.*, **9**, 24–44.

Hodges, M., Downs, J. B. and Mitchell, L. A. (1975) Thermodilution and Fick cardiac index determinations following cardiac surgery. *Critical Care Medicine*, **3**, 182–184.

Hoel, B. L. (1978) Some aspects of the clinical use of thermodilution in measuring cardiac output. *Scand. J. Clin. Lab. Investigation*, **38**, 383–388.

Kelly, G. E. (1984) The influence function in the errors in variables problem. *Ann. Statist.*, **12**, 87–100.

Kendall, M. G. and Stuart, A. (1961) *The Advanced Theory of Statistics, Vol. 2.* New York: Hafner.

Kohanna, F. H. and Cunningham, J. N. Jr. (1977) Monitoring of cardiac output by thermodilution after open-heart surgery. *J. Thoracic Cardiovascular Surgery*, **73**, 451–457.

Laird, N. (1978) Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Ass.*, **73**, 805–811.

Olsson, B., Poole, J., Vandermoten, R., Varnasuskas, E. and Wassen, R. (1970) Validity and reproducibility of determination of cardiac output by thermodilution in man. *Cardiology*, **55**, 136–148.

Rao, C. R. (1965) *Linear Statistical Inference and its Applications.* 1st edition. New York: Wiley.

Stawicki, J. J., Holford, F. D., Michelson, E. L. and Josephson, M. E. (1979) Multiple cardiac output measurements in man. *Chest*, **76**, 193–197.

Stetz, C. W., Miller, R. G., Kelly, G. E. and Raffin, T. A. (1982) Reliability of the thermodilution method in the determination of cardiac output in clinical practice. *Amer. Rev. Respiratory Disease*, **126**, 1001–1004.

Vandermoten, P., Bernard, R., De Hemptinne, J., Gillet, J. M. and Lenaers, A. (1977) Cardiac output monitoring during the acute phase of myocardial infarction: accuracy and precision of the thermodilution method. *Cardiology*, **62**, 291–295.

Venkataraman, K., DeGuzman, M. F. C., Khan, A. H. and Haywood, L. J. (1976) Cardiac output measurement: a comparison of direct Fick, dye dilution and thermodilution methods in stable and acutely ill patients. *J. Nat. Med. Ass.*, **68**, 281–284.

Weisel, R. D., Vito, L., Dennis, R. C., Berger, R. L. and Hechtman, H. B. (1975) Clinical applications of thermodilution cardiac output determinations. *Amer. J. Surgery*, **129**, 449–454.