Formation of COPES

On the occasion of the 3rd Joint Meeting of the European Society for Paediatric Endocrinology and the Lawson Wilkins Pediatric Endocrine Society held 29 October–3 November 1989 in Jerusalem, the presidents and the secretaries of the Australasian Paediatric Endocrine Group, the European Society for Paediatric Endocrinology, the International Study Group of Diabetes in Children and Adolescents, the Japanese Society for Pediatric Endocrinology, the Latin American Society for Pediatric Endocrinology, and the Lawson Wilkins Pediatric Endocrine Society decided that a Coordination Office of Pediatric Endocrine Societies (COPES) should be established.

The purpose of COPES is to *1*) create a forum for communication and exchange of information between pediatric endocrine societies; *2*) provide information on the yearly activities of each society (meetings, workshops, and postgraduate courses); *3*) provide information on the possibilities of training programs, research facilities, and fellowships; *4*) provide information on major meetings of endocrine and pediatric societies; and *5*) promote collaborative studies. This information will be published in a newsletter to be issued two or three times yearly and distributed by the secretaries of each society.

The two coordinators elected to head this office for a 4-yr term are Z. Laron, Director of the Institute of Pediatric and Adolescent Endocrinology, Beilinson Medical Center, Petah Tikva, Israel; and G. Werther, President of the Australasian Paediatric Endocrine Group, Deputy Director of the Department of Endocrinology and Diabetes, Royal Children's Hospital, Parkville, Victoria, Australia. Laron and Werther solicit information on meetings, workshops, and proposals for collaborative studies to be published in the newsletter. Material can be sent to Laron or Werther.

ZVI LARON, MD
GEORGE WERTHER, MD

From the Institute of Pediatric and Adolescent Endocrinology, Beilinson Medical Center, Petah Tikva; and Sackler Faculty of Medicine, Tel Aviv University, Israel; and the Department of Endocrinology and Diabetes, Royal Children's Hospital, Parkville, Victoria, Australia.

Address correspondence to Prof. Zvi Laron, MD, Institute of Pediatric and Adolescent Endocrinology, Beilinson Medical Center, Petah Tikva 49 100, Israel; or George Werther, MD, Department of Endocrinology and Diabetes, Royal Children's Hospital, Parkville, Victoria 3052, Australia.

Issues in Use of Intraclass Correlation

Replogle et al. (1) argued that the Pearson product-moment correlation (*r*) should not be used to evaluate the accuracy of the data obtained from self-monitoring of blood glucose (SMBG) and proposed that the intraclass correlation coefficient (ICC) provides a better measure of the level of agreement between SMBG measurements and the corresponding values determined with laboratory procedures (e.g., Beckman Glucose Analyzer II, YSI model 23A glucose analyzer, and other laboratory procedures used to define "true" values). We agree with Replogle et al. that the Pearson *r* should not be used to evaluate the accuracy of SMBG measurements. The Pearson *r* is a statistic designed to measure the extent to which pairs of measurements are associated or consistent not the extent to which pairs of measurements agree or are identical. The implication of this fact is that if measurements from a self-monitoring system differ from their true values systematically (by a constant or proportional amount), the correlation for these data will be 1.0. This perfect correlation, however, does not mean there is an absence of error in the self-monitoring measurements.

We do not agree with Replogle et al., however, that the ICC is the preferred method for evaluating the accuracy of the SMBG measurements. Before presenting our reservations about the ICC, an expansion of Replogle et al.'s description of the ICC is in order. This clarification is necessary because the ICC is more complex conceptually than the Pearson *r* and because there has been some confusion about the use of this statistic in reliability studies (2,3).

The ICC has many different forms. It is derived with analysis of variance (ANOVA) models (e.g., 1- and 2-way random effects and 2-way mixed effects), and, as noted by Shrout and Fleiss (2), "the numerous versions of the ICC can give quite different results when applied to the same data." To determine what form of the ICC should be used, it is necessary to define the research question and experimental design underlying the study. In the typical study examining the accuracy of SMBG data, the main question is whether the two methods for measuring blood glucose are interchangeable (i.e., do the SMBG device and laboratory method produce identical results). To address this question, the experimental design involves selecting a random sample of blood values and testing each blood value with both a self-monitoring device and a laboratory method (matched data). Given these conditions, the appropriate statistical model is a two-way (blood sample and method of measuring blood glucose) random-effects ANOVA model (see Shrout and Fleiss for a discussion of additional designs that are used to compute the ICC).

The ICC derived from this design is expressed as a ratio of the variance associated with the blood glucose values over the sum of the variance components associated with the blood glucose values, method (SMBG device or laboratory method), interaction between

TABLE 1
Analysis of variance summary

Source of variation	df	Sum of squares	Mean square
Between-blood sample	4	61,760	15,440
Within-blood sample	5	1000	200
Between method	1	1000	1000
Residual	4	0	0

TABLE 2
Self-monitoring of blood glucose (SMBG) and laboratory measurements of blood glucose

	A		B		C	
Sample	SMBG	Lab	SMBG	Lab	SMBG	Lab
1	60	80	60	80	60	80
2	80	100	80	100	61	81
3	100	120	100	120	62	82
4	120	140	120	140	63	83
5	280	300	140	160	64	84

Intraclass correlation coefficient: A, 0.97; B, 0.83; C, 0.01.

blood glucose value and method, and error. Table 1 illustrates the ANOVA summary table that would be used to compute the ICC for the hypothetical data contained in Table 2A. The formula for estimating the ICC is

$$\frac{BMS - EMS}{BMS + (k - 1)EMS + k(MMS - EMS)/n}$$

where BMS is the between-blood sample mean square, EMS is the residual mean square, MMS is the between-method mean square, k is the number of methods used to measure blood glucose (2 in this case), and n is the number of blood samples examined. The resulting ICC for the data in Table 2A is 0.97, which indicates a high level of agreement between the SMBG and laboratory measurements.

One of our reservations with the ICC is that the size of the coefficient is influenced greatly by the amount of variability in the blood glucose measurements. The data in Table 2, B and C, illustrate how a reduction in the range of blood glucose values leads to a reduction in the ICC. (Note that these are extreme examples used only to illustrate what happens when a restricted range of blood values is examined; studies examining the performance of SMBG devices should include a representative range of blood glucose values.) The ICCs for the data in Table 2, B and C, are 0.83 and 0.01, respectively, despite the fact that the laboratory measurements are 20 U higher than the SMBG measurements in each example.

The second problem with the ICC is that it provides little information about the clinical significance of the errors associated with SMBG. The issues of statistical and clinical significance have been discussed by various investigators in reference to the use of the Pearson r and regression analyses and apply equally to the use of the ICC (4–7). The crux of the argument against the use of the ICC to evaluate the clinical significance of any error in an SMBG system is that two SMBG measurements that deviate by the same amount from their true values may have very different clinical meanings depending on the true values. For example, a 50% underestimate of the true value of 2.8 mM blood glucose will not make any difference in terms of the treatment actions taken by a patient. Under these circumstances, the patient will take corrective action to raise blood glucose. However, a 50% underestimate when the true value is 16.8 mM will result in the patient not taking action to lower a clinically elevated blood glucose level. In view of these comments, we do not recommend the ICC as the primary method for evaluating the quality of the measurements from self-monitoring systems. Currently, the most useful method for evaluating the data from self-monitoring systems is the error-grid analysis graphic display technique developed by Cox et al. (4–6).

ROBERT F. DEDRICK, PhD
WAYNE K. DAVIS, PhD

From the Office of Educational Resources and Research, Department of Postgraduate Medicine, The University of Michigan Medical School, Ann Arbor, Michigan.

Address correspondence and reprint requests to Robert F. Dedrick, PhD, University of Michigan, Office of Educational Resources and Research, Towsley Center, Room G1100, Box 021, Ann Arbor, MI 48109-0201.

REFERENCES

1. Replogle WH, Sanders SD, Eicke FJ, Phillips DM: Potential error in assessing accuracy of blood glucose monitoring systems. *Diabetes Care* 11:603–604, 1988
2. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–28, 1979
3. Lahey MA, Downey RG, Saal, FE: Intraclass correlations: there's more there than meets the eye. *Psychol Bull* 93: 586–95, 1983
4. Cox DJ, Clarke WL, Gonder-Frederick L, Pohl S, Hoover C, Snyder A, Zimbelman L, Carter WR, Bobbitt S, Pennebaker J: Accuracy of perceiving blood glucose in insulin-dependent diabetes mellitus. *Diabetes Care* 8:529–36, 1985
5. Pohl SL, Gonder-Frederick L, Cox DJ, Evans WS: Self-measurement of blood glucose concentration: clinical significance of patient-generated measurements. *Diabetes Care* 8:617–19, 1985
6. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL: Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 10:622–28, 1987
7. Dedrick RF, Davis WK: What do statistics really tell us about the quality of the data from self-monitoring of blood glucose? *Diabetic Med* 6:267–73, 1989

Additional Issues in Use of Intraclass Correlation

We thank Dedrick and Davis for their astute observations concerning the intraclass correlation coefficient (ICC) and appreciate the opportunity to clarify some issues. First, there are numerous versions of the ICC, just as there are numerous versions of other statistical techniques (e.g., independent vs. correlated t test, 1-way vs. 2-way χ^2-test). Each form of the ICC "is appropriate for specific situations defined by the experimental design and the conceptual intent of the study" (1). Inappropriate application by the researcher of any form of the ICC or any other inferential statistical technique will yield erroneous results. This is clearly an error on the part of the researcher and in no way condemns the ICC.

Dedrick and Davis correctly note that the ICC is influenced by the amount of variability in the data matrix,