from two models, one consisting of only the composite and the other consisting of a regression model that included both components. The regression model is a linear combination of the two components in which the weights are chosen to obtain an optimal fit; thus, the regression model itself is a composite, though one in which the fit to the data should be better than $A_1$ months (which is exactly what they found).

Since comparing two composites was not the goal of our study (2), we approached the analyses differently. We developed one regression model including all variables that were significant in the multivariate modeling, including the composite as well as individual components, as candidates for the model. Each partial $R^2$ measures the explanatory value of the corresponding variable beyond the prediction already available from all the other variables in the model. Except for severity of retinopathy at baseline, we found that the composite was consistently the best predictor and that the individual components added little, if anything.

We agree that age at onset and duration added together equal the age of the patient at the time of study, although the appropriate weights for these two time periods in predicting the outcome may differ, and determining whether the weights significantly differ would be of interest. However, this was not a focus of our study.

We also agree that the patient population under study and the choice of outcomes to be analyzed can influence the results and that a continuous neuropathy measure is desirable. Although use of a common outcome measure would assist in comparing our results with those of Orchard et al. (3), such a comparison was not the focus of our study (2). Finally, determining the threshold of chronic glycemia, which induces complications, is a worthy goal, but before we do this we want to include studies of normal subjects and glucose-impaired individuals currently being studied.

PETER J. DYCK, MD[1]
JENNY L. DAVIES, BA[1]
VICKI M. CLARK, CCRP[1]
WILLIAM J. LITCHY, MD[1]
P. JAMES B. DYCK, MD[1]
CHRISTOPHER J. KLEIN, MD[1]
ROBERT A. RIZZA, MD[2]
JOHN M. PACH, MD[3]
RONALD KLEIN, MD[4]
TIMOTHY S. LARSON, MD[5]
L. JOSEPH MELTON, III, MD[6]
PETER C. O'BRIEN, PHD[7]

From the [1]Department of Neurology, Mayo Clinic College of Medicine, Rochester, Minnesota; the [2]Division of Endocrinology, Mayo Clinic College of Medicine, Rochester, Minnesota; the [3]Department of Ophthalmology, Mayo Clinic College of Medicine, Rochester, Minnesota; the [4]Department of Ophthalmology, University of Wisconsin Madison, Madison, Wisconsin; the [5]Division of Nephrology and Hypertension, Mayo Clinic College of Medicine, Rochester, Minnesota; the [6]Division of Epidemiology, Mayo Clinic College of Medicine, Rochester, Minnesota; and the [7]Division of Biostatistics, Mayo Clinic College of Medicine, Rochester, Minnesota.

Address correspondence to Peter J. Dyck, MD, Mayo Clinic College of Medicine, Department of Neurology, 200 First St., SW, Rochester, MN 55905. E-mail: dyck.peter@mayo.edu.

DOI: 10.2337/dc06-2239



**References**

1. Orchard TJ, Costacou T, Miller RG, Prince CT, Pambianco G: Modeling chronic glycemic exposure variables as correlates and predictors of microvascular complications of diabetes (Letter). *Diabetes Care* 30:448, 2007
2. Dyck PJ, Davies JL, Clark VM, Litchy WJ, Dyck PJB, Klein CJ, Rizza RA, Pach JM, Klein R, Larson TS, Melton LJ III, O'Brien PC: Modeling chronic glycemic exposure variables as correlates and predictors of microvascular complications of diabetes. *Diabetes Care* 29:2282–2288, 2006
3. Orchard TJ, Forrest KY, Ellis D, Becker DJ: Cumulative glycemic exposure and microvascular complications in insulin-dependent diabetes mellitus: the glycemic threshold revisited. *Arch Intern Med* 157: 1851–1856, 1997

# A Critical Appraisal of the Continuous Glucose–Error Grid Analysis

## Response to Wentholt et al.

In a recent publication, Wentholt et al. (1) stated that their aim was to critically explore the continuous glucose–error grid analysis (CG-EGA) (2) and to compare it with traditional techniques using data previously reported from two sensors. As developers of the CG-EGA, we hoped that our method might stimulate a discussion on the important problem of the accuracy of continuous monitoring sensors (CGS); therefore, we read this critique with interest.

The methods used by Wentholt et al. (1) unfortunately failed to take into account the basic structure of CGS data, which represent time series (i.e., sequential readings that are ordered in time) (3). This structure leads to two fundamental requirements in their analysis. First, consecutive sensor readings taken from the same subject within a relatively short time are highly interdependent. Therefore, standard statistical analyses such as *t* tests, while appropriate for independent data points, will produce inaccurate results if applied to CGS data. Second, the order of the CGS data points is essential for clinical decision making. For example, the sequences 90 → 82 → 72 mg/dl and 72 → 82 → 90 mg/dl are clinically very different. Standard accuracy measures, such as the mean absolute deviation (MAD) used by Wentholt et al. (1), do not account for the data's temporal order; if reference-sensor data pairs are reshuffled, the MAD remains the same.

As a result, the primary statistical analysis used by Wentholt et al. is flawed, both to demonstrate significant differences between the sensors and to imply that CG-EGA is insensitive. The CGS data from 13 subjects were pooled to compare 2 MADs (15.0 ± 12.2 vs. 13.6 ± 10.2%). The result was reported as significant ($P = 0.013$), but for these highly overlapping MADs to differ statistically required a large number (>1,000) of degrees of freedom, which was calculated by pooling the total number of CGS data points (735 and 1,156) across all subjects. Such an approach led to inaccurate conclusions because there were only 13 independent subjects, and the data points within each subject were highly dependent. If the correct number of degrees of freedom is used, the MADs of the two sensors are not different ($P > 0.5$), which confirms the CG-EGA results showing no differences.

Other conclusions by Wentholt et al. also deserve comment. First, they stated that CG-EGA is time consuming. Indeed, analyses of temporal data are intrinsically more sophisticated than standard time-independent statistics, but such analyses are essential for this type of data. CG-EGA software is available. Second, Wentholt et al. stated that "poor accuracy rate is barely noticeable in the final CG-EGA outcome," implying that this result of the CG-EGA is incorrect. However, this result is not incorrect because better combined (rate and point) accuracy during hypoglycemia is observed with the sensor, showing poorer rate accuracy in this critical region. It is

clinically apparent that when blood glucose is <3.9 mmol/l point accuracy should be given more emphasis than rate accuracy. A strength of CG-EGA is its ability to vary the input of either rate or point accuracy to overall clinical accuracy depending on blood glucose range. Third, the results of CG-EGA vary with time intervals. This is also an intuitive strength of CG-EGA, which is designed to account for increased noise associated with frequent sampling. We advocated (2) adopting a uniform sampling protocol with reference and/or sensor pairs taken every 10–15 min to standardize comparisons of rate accuracy, which is a sampling scheme based on physiological considerations of possible glucose change rates. Fourth, Wentholt et al. (1) questioned the appropriateness of the formulae to shift point EGA based on interstitial time lag. However, the authors reported an average time lag of ~7 min in one of their sensors, which is identical to that assumed for CG-EGA, thus confirming that ~7 min is a reasonable average for blood-to-interstitial diffusion delays. CG-EGA software allows setting this parameter to any value <7 min.

We are pleased that both the discussion regarding CG-EGA and the analysis of time series data have begun, and we look forward to continuing this important dialogue. However, we also recommend careful consideration of basic statistical assumptions when analyzing sensor-generated glucose data; their inherent temporal structure should be taken into account.

**William L. Clarke, MD[1]**
**Linda Gonder-Frederick, PhD[2]**
**Daniel Cox, PhD[2]**
**Boris Kovatchev, PhD[2]**

From the [1]Department of Pediatrics, University of Virginia Health System, Charlottesville, Virginia; and the [2]Department of Psychiatry and Neurobehavioral Sciences, University of Virginia Health System, Charlottesville, Virginia.

Address correspondence to Boris Kovatchev, Department of Psychiatry and Neurobehavioral Sciences, University of Virginia Health System, Box 800137, Charlottesville, VA 22908. E-mail: boris@virginia.edu.

DOI: 10.2337/dc06-1901



**References**

1. Wentholt IM, Hoekstra JB, DeVries JH: A critical appraisal of the continuous glucose–error grid analysis. *Diabetes Care* 29:1805–1811, 2006
2. Kovatchev BP, Gonder-Frederick LA, Cox DJ, Clarke WL: Evaluating the accuracy of continuous glucose-monitoring sensors: continuous glucose–error grid analysis illustrated by Therasense Freestyle Navigator data. *Diabetes Care* 27:1922–1928, 2004
3. Chatfield C: *The Analysis of Time Series: An Introduction*. 6th ed. Boca Raton, FL, Chapman and Hall/CRC, 2004, p. 1

# A Critical Appraisal of the Continuous Glucose–Error Grid Analysis

## Response to Clarke et al.

We thank Clarke et al. (1) for their thought-provoking response to our article (2). With their comments (1), they not only took on the important issue of how to optimally assess the accuracy of continuous glucose monitors (CGMs); they moved the discussion one step further.

In our study (2), we did indeed take the statistical liberty of deriving degrees of freedom from all pooled data points—in contrast to the proposal by Clarke et al. (1) who compared the accuracy of two sensors using one average mean absolute deviation (MAD) value per patient. The latter approach may be too rigid because not all readings are interdependent. For example, postprandial glucose sensor readings at lunch and at night depend little on each other, if at all. It is common practice to derive degrees of freedom from pooled data in the sensor field. In a previous study, Clarke et al. (3) compared the accuracy of two CGMs in 16 type 1 diabetic patients by using the continuous glucose–error grid analysis (CG-EGA). The difference in pooled readings in the hypoglycemic area that ended up in zones A and B was reported to be highly significant between both sensors (88 vs. 62.8%, respectively) ($P < 0.0005$). This level of significance implies that degrees of freedom were derived from all data pairs in the hypoglycemic range (250 mg/dl) rather than from the actual amount of participants ($n = 16$). Even with a strict statistical policy, the better MAD for the microdialysis sensor in the hypoglycemic area in our study (2) (12.0% for the 7-min corrected microdialysis sensor vs. 25.2% for the needle-type sensor, calculated per patient [df = 12], $P = 0.036$ by Wilcoxon's signed-rank test) and the larger sensitivity for hypoglycemia associated with this sensor (75.0 [75 data pairs] vs. 55.9% [56 data pairs], $P = 0.018$ by Pearson's $\chi^2$, with 16 of 16 and 12 of 15 hypoglycemic episodes detected by the microdialysis and needle-type sensor, respectively, $P = 0.06$ by Pearson's $\chi^2$) contrasted with the CG-EGA that noted no difference (51.5 vs. 60.0% accurate readings and benign errors in the hypoglycemic range [df = 42], $P = 0.841$ by Pearson's $\chi^2$ for the microdialysis and the needle-type sensor, respectively). Therefore, even with a mild statistical approach (i.e., deriving degrees of freedom from 43 data pairs rather than 13), CG-EGA could not confirm the different accuracy of the sensors in the hypoglycemic range.

As to the order of CGS data points, the sensor's ability to follow the rate and direction of glucose changes is nicely reflected by the MAD: A sequence of glucose values that has been incorrectly reported by a given sensor (e.g., 90 → 82 → 72 mg/dl instead of 72 → 82 → 90 mg/dl) will result in a worsened MAD.

In reaction to the comment by Clark et al. (1) in regards to time consumption, we were happy to learn that the software for CG-EGA has become available. Nevertheless, the laborious collection of frequent blood samples on fixed intervals (in addition to the construction of a rate, a point accuracy plot, and, finally, a combining matrix) will remain inevitable drawbacks of CG-EGA.

With the attempt to standardize the length of the time intervals, Clark et al. clearly tried to improve the CG-EGA methodology. Nevertheless, a time interval that can vary by 5 min (10–15 min) still leaves the door open for interobserver variability.

As to our finding in a previous study (4) of a 7-min delay that was inherent to the microdialysis instrument itself and not seen in the needle-type sensor, Clarke et al. (1) alluded to a (much-disputed) constant 7-min physiological delay resulting from the relationship between interstitial and blood glucose. This physiological delay has been reported to be anywhere between 0 and 30 min, so the 7-min assumption made for the CG-EGA is questionable. Fortunately, Clarke et al. have now implemented into the software the possibility of setting the delay <7 min.

Currently, the optimal way to assess a CGM seems to be the combination of MAD calculated per glucose range, com-