Understanding Error Grid Analysis

The article entitled "Reservations on the use of error grid analysis for the validation of blood glucose assays" (1) presents a technically accurate description of error grid analysis (EGA), but fails to address the basic idea behind it. EGA is a clinically oriented nonparametric approach to blood glucose (BG) data, based on three assumptions: *1*) BG readings <3.9 mmol should be raised, 2) BG readings >10 mmol should be lowered, and 3) acceptably accurate estimates are within 20% of the reference BG or when both the estimates and reference BG are <3.9 mmol. While the latter was the "standard of the day" (2), the upper and lower limits of acceptable BG were confirmed by the Diabetes Control and Complications Trial (3).

The proposed "more effective" alternative is a generic parametric statistical technique described by Bland and Altman (4), which involves plotting the difference of each couple of values against their mean. It was reported that Moberg et al. (5) had compared the Bland-Altman approach with EGA. Actually, Moberg et al. compared EGA only with linear regression.

We analyzed the data in Fig. 1 of Gough and Botvinick (1), using EGA and the Bland-Altman approach (4). We found that 46% of the estimates were accurate, as defined by the EGA, which is markedly less than the 95% we previously recommended for devices (6). Approximately 4% of hypoglycemic readings were not detected (potentially clinically dangerous overestimates in upper zone D), and ~8% of hyperglycemic readings were not detected (potentially clinically dangerous underestimates in lower zone D). In addition, EGA identified a possible systematic error of the measuring device: it tended to underestimate BG, having 62% of its readings in the lower zones.

The Bland-Altman method correctly assessed the poor agreement between the estimate-reference values because of a 95% confidence interval for the deviations of the measuring device involving overestimates up to 8.6 mmol and underestimates of 6.6 mmol. However, the Bland-Altman approach has two major disadvantages: *1*) it does not provide objective criteria for agreement and 2) it relies on linear clinical equivalence across the entire BG range and suggests a data transformation when this equivalence is not accurate (4). We propose and validate a data transformation specific to the clinical nature of BG readings (34).

The use of EGA was reported in different studies (1). However, this list (5,7–11) is incomplete, with there being many additional studies incorporating EGA (12–31). What is important to point out is that in all of these published studies the authors reported both EGA results and standard statistical results. Neither these, nor our own studies, assume that either EGA or standard statistical approaches are totally adequate. One approach is not more appropriate than the other, but rather they are complementary.

Four specific issues are raised. We would like to address these separately.

1. It was stated that the "exact boundaries separating each region [i.e., zone] of the error grid are somewhat arbitrary . . . . There are problems with the pattern of the grid itself" (1). While it is true that whenever you "draw a line" different arguments can be made for different positioning, the misunderstanding here involves the basic EGA assumptions that we listed above. The example given of a 3.8 mmol estimate with a 0 mmol reference BG would never occur in reality. But if it would occur, the EGA results would be correct; both this estimated and reference BG would lead to the same and appropriate clinical action of elevating BG. In addition, we have discussed that EGA assumptions may need to be altered for different clinical and research purposes (32), and since 1992, the EGA software has allowed users to set their own boundaries. We agree that research to evaluate ideal assumptions would be beneficial—for example, the 10% deviation criteria as proposed by the American Diabetes Association (33).

2. Appropriate criticism is that EGA relies on the assumption that reference readings are "true" or accurate reflections of the actual BG levels. However, all schemes and/or statistics for the evaluation of BG assay accuracy assume accurate reference BG values.

3. "Precision is lost [with EGA] by assigning data to a few broad regions. Statistical methods, in contrast, have useful objective criteria . . . and include criteria to identify outliers. [EGA] provides no such information." The first misunderstanding is that in reality BG readings are not normally distributed. Before parametric statistics can be employed, BG readings need to be normalized. We propose such a procedure (34). Second, EGA can be said to offer clinical, as opposed to statistical, precision. When we applied "objective criteria" to the data in Fig. 1 of Gough and Botvinick (1), the regression model was highly significant ($P < 0.0001$), had a multiple R of 0.77, and a nonsignificant constant term ($P = 0.47$). This implies high accuracy and no systematic errors of the device, in contrast to both EGA and the Bland-Altman approach. The "objective criteria" for outlier used standardized residuals with threshold value of 1.96 (i.e., outlier probability of 0.05). This identified two outliers at reference-estimated BG levels of 13.4–21.1 and 17.8–23.7. EGA identified the former as upper C, leading to a potentially dangerous overtreatment, while the latter (upper B) was considered leading to an appropriate clinical decision to reduce BG. This "objective" approach did not identify the two occurrences of failure to detect hyperglycemia (18–9, 16–8.5) and the failure to detect hypoglycemia (3.5–4.7).

4. "Error grid analysis was not intended to be used directly in a clinical setting . . . . The actual clinical relevance is therefore limited." In fact, EGA was initially employed in a clinical setting, evaluating the accuracy of patient BG measurements, relative to reference laboratory readings (12), and the accuracy of patient BG estimates relative to their meter readings (14,16,35). EGA was specifically developed to identify the types and frequency of clinically significant errors (i.e., does meter/patient overestimate hypoglycemia [upper zones D and E], underestimate hyperglycemia [lower zones D and E], or misinterpret euglycemia for hyper- or hypoglycemia [C zones]). From a clinical perspective, it is absolutely critical to determine if a measurement device misinterprets hypoglycemia for euglycemia or hyperglycemia.

In conclusion, we agree that EGA should not be used alone and that those employing both EGA and parametric statis-

tical approaches should realize their inherent limitations (34,36) and not extrapolate beyond them. However, these two approaches are neither incompatible nor mutually exclusive, but rather they are complementary. We will continue to employ both approaches, as we encourage others to do so.

DANIEL J. COX, PHD
LINDA A. GONDER-FREDERICK, PHD
BORIS P. KOVATCHEV, PHD
DIANA M. JULIAN, MA
WILLIAM L. CLARKE, MD

From the University of Virginia Health Sciences Center, Charlottesville, Virginia.

Address correspondence to Daniel J. Cox, Behavioral Medicine Center, Box 223, University of Virginia Health Sciences Center, Charlottesville, VA 22908.

Received for publication 7 March 1997 and accepted 12 March 1997.

BG, blood glucose; EGA, error grid analysis.

Acknowledgments — This research was supported by National Institutes of Health Grants DK28288 and RR00847.

References

1. Gough DA, Botvinick EL: Reservations on the use of error grid analysis for the validation of blood glucose assays. *Diabetes Care* 20:1034–1036, 1997
2. American Diabetes Association: Bedside blood glucose monitoring in hospitals (Position Statement). *Diabetes Care* 9:89, 1986
3. The Diabetes Control and Complications Trial Research Group: Diabetes Control and Complications Trial (DCCT): results of feasibility study. *Diabetes Care* 10:1–19, 1989
4. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 8:307–310, 1986
5. Moberg E, Lundblad S, Lins PE, Adamson U: How accurate are home blood-glucose meters with special respect to the low glycemic range? *Diabetes Res Clin Pract* 19:239–243, 1993
6. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL: Evaluating the clinical accuracy of self-blood glucose monitoring systems. *Diabetes Care* 10:622–628, 1987
7. Cox DJ, Moll ME, Gonder-Frederick LA, Savory J, Garrison S: Accuracy of time-delayed filter paper blood glucose measurements. *Diabetes Care* 11:517–518, 1988
8. Clarke WL, Becker DJ, Cox DJ, Santiago JV, White NH, Betschart J, Eckenrode K, Levandoski LA, Prusinski EA, Simineiro LM, Snyder AL, Tideman AM, Yaeger T: Evaluation of a new system for self blood glucose monitoring. *Diabetes Res Clin Pract* 4:209–214, 1988
9. Ross D, Heinemann L, Chantelau EA: Short-term evaluation of an electro-chemical system (ExacTech) for blood glucose monitoring. *Diabetes Res Clin Pract* 10:281–285, 1990
10. Fischer U: Continuous in vivo monitoring in diabetes: the subcutaneous glucose concentration. *Acta Anesthesiol Scand* 104:21–29, 1995
11. Tamada JA, Bohannon NJV, Potts RO: Measurement of glucose in diabetic subjects using noninvasive transdermal extraction. *Nat Med* 1:1198–1201, 1995
12. Pohl SL, Gonder-Frederick LA, Cox DJ, Evans WS: Self-management of blood glucose concentration: clinical significance of patient-generated measurements. *Diabetes Care* 8:617–619, 1985
13. Cox DJ, Herrman JJ, Gonder-Frederick LA, Snyder A, Reschke J, Clarke W: Stability of reacted Chemstrip BG. *Diabetes Care* 11:288–291, 1988
14. Cox DJ, Carter WR, Gonder-Frederick LA, Clarke WL, Pohl S: Blood glucose discrimination training in insulin-dependent diabetes mellitus patients. *Biofeedback Self Regul* 13:201–217, 1988
15. Gonder-Frederick L, Snyder A, Clarke W: Accuracy of blood glucose estimation by children with IDDM and their parents. *Diabetes Care* 14:565–570, 1991
16. Cox DJ, Gonder-Frederick LA, Julian D, Cryer P, Lee JH, Richards FE, Clarke WL: Intensive versus standard blood glucose awareness training (BGAT) with insulin-dependent diabetes: mechanisms and ancillary effects. *Psychosom Med* 53: 453–462, 1991
17. Davidson MR: Another look at self-monitoring of blood glucose (Letter). *Clin Diabetes* 9:67, 1992
18. Biocontrol diasensor 100 non-invasive monitor needs more studies for FDA clearance, panel says; redesigned 50–100 patient study recommended. *MMDI Reports*, 24 March 1996
19. Matthews DR, Burton SF, Bown E, Chusney G, Dornan T, Gale EA, McKinnon G, Steemson J: Capillary and venous blood glucose measurements using a direct glucose-sensing meter. *Diabet Med* 8:875–880, 1991
20. Odenthal C: Evaluation of an electro-enzymatic system (Satellite G) for blood glucose monitoring in hospital use. *Diabete Metab* 17:534–537, 1991
21. Muhlhauser I, Heinemann L, Karinganamattom J, Schuwirth W, Berger M: Accuracy of blood glucose self monitoring in elderly insulin treated diabetic patients. *Diabete Metab* 17:476–482, 1991
22. Atkin SH, Dasmahapatra A, Jaker MA, Chorost MI, Reddy S: Fingerstick glucose determination in shock. *Ann Intern Med* 114:1020–1024, 1991
23. Nurick MA, Johnson SB: Enhancing blood glucose awareness in adolescents and young adults with IDDM. *Diabetes Care* 14:1–7, 1991
24. Hunt JA, Alojado NC: A new, improved test system for rapid measurement of blood glucose. *Diabetes Res Clin Pract* 7:51–55, 1989
25. Tieszen KL, Burton SF, Dornan TL, Matthews DR, McMurray JR: Evaluation of a second-generation electrochemical blood glucose monitoring system. *Diabet Med* 12:173–176, 1995
26. Bustamante MA, Hennessey JV, Teter ML, Stachler RJ, Warner B: Clinical accuracy of capillary blood glucose monitoring in hospitalized patients with diabetes. *Diabetes Educ* 20:212–215, 1994
27. Kabadi UM, O'Connell KM, Johnson J, Kabadi M: The effect of recurrent practice at home on the acceptability of capillary blood glucose readings: accuracy of self blood glucose testing. *Diabetes Care* 17:1110–1123, 1994
28. Devreese K, Leroux-Roels G: Laboratory assessment of five glucose meters designed for self-monitoring of blood glucose concentration. *Eur J Clin Chem Clin Biochem* 31:829–837, 1993
29. Schwingshandl J, Borkenstein M: Accuracy of blood glucose self-monitoring in children with type-I diabetes. *Wien Klin Wochenschr* 105:382–384, 1993
30. Poitout V, Moatti-Sirat D, Reach G: Calibration in dogs of a subcutaneous miniaturized glucose sensor using a glucose meter for blood glucose determination. *Biosens Bioelectron* 7:587–592, 1992
31. Kovatchev BP, Cox DJ: Statistical distributions related to blood glucose fluctuations of subjects with insulin dependent diabetes mellitus (IDDM). *Health Psychol.* In press
32. Cox DJ, Richards F, Gonder-Frederick LA, Julian DM, Carter WR, Clarke WL: Clarification of error grid analysis. *Diabetes Care* 12:235–236, 1989
33. American Diabetes Association: Clinical practice recommendations. *Diabetes Care* 19 (Suppl. 1), 1996
34. Kovatchev B, Cox D, Gonder-Frederick L, Clarke W: Transforming the blood glucose scale: statistical and clinical implications (Abstract). *Diabetes* 46 (Suppl. 1):268A, 1997
35. Cox DJ, Clarke W, Pohl S, Gonder-Frederick L, Hoover C, Zimbelman, Pennebaker JW: Accuracy of perceiving blood glucose in IDDM. *Diabetes Care* 8:529–536, 1985
36. Dedrick RF, Davis WK: What do statistics really tells about the quality of the data from self-monitoring of blood glucose? *Diabet Med* 6:267–273, 1991