

Evaluating the Accuracy of Continuous Glucose-Monitoring Sensors

Continuous glucose–error grid analysis illustrated by TheraSense Freestyle Navigator data

BORIS P. KOVATCHEV, PHD¹
LINDA A. GONDER-FREDERICK, PHD¹

DANIEL J. COX, PHD¹
WILLIAM L. CLARKE, MD²

OBJECTIVE — The objective of this study was to introduce continuous glucose–error grid analysis (CG-EGA) as a method of evaluating the accuracy of continuous glucose-monitoring sensors in terms of both accurate blood glucose (BG) values and accurate direction and rate of BG fluctuations and to illustrate the application of CG-EGA with data from the TheraSense Freestyle Navigator.

RESEARCH DESIGN AND METHODS — We approach the design of CG-EGA from the understanding that continuous glucose sensors (CGSs) allow the observation of BG fluctuations as a process in time. We account for specifics of process characterization (location, speed, and direction) and for biological limitations of the observed processes (time lags associated with interstitial sensors). CG-EGA includes two interacting components: 1) point–error grid analysis (P-EGA) evaluates the sensor's accuracy in terms of correct presentation of BG values and 2) rate–error grid analysis (R-EGA) assesses the sensor's ability to capture the direction and rate of BG fluctuations.

RESULTS — CG-EGA revealed that the accuracy of the Navigator, measured as a percentage of accurate readings plus benign errors, was significantly different at hypoglycemia (73.5%), euglycemia (99%), and hyperglycemia (95.4%). Failure to detect hypoglycemia was the most common error. The point accuracy of the Navigator was relatively stable over a wide range of BG rates of change, and its rate accuracy decreased significantly at high BG levels.

CONCLUSIONS — Traditional self-monitoring of BG device evaluation methods fail to capture the important temporal characteristics of the continuous glucose-monitoring process. CG-EGA addresses this problem, thus providing a comprehensive assessment of sensor accuracy that appears to be a useful adjunct to other CGS performance measures.

Diabetes Care 27:1922–1928, 2004

The premise behind increasing research and industrial efforts on the development of continuous glucose sensors (CGSs) is that the accurate assessment of blood glucose (BG) dynamics is a valuable tool for both everyday maintenance of diabetes and long-term effectiveness of glycemic control (1,2). Compared

with a few self-monitoring of BG (SMBG) readings per day, CGSs yield a detailed time series of BG samples (e.g., every 5 min). Thus, CGS technology has the potential to revolutionize diabetes management by providing patients with ongoing online feedback about current BG levels and rate/direction of change, as well as signaling to alert for possible dangerous trends such as rapid BG descents that may lead to hypoglycemia.

Evaluating the accuracy of CGSs, however, is not straightforward, especially if taken in the context of established accuracy measures such as correlation or regression, consensus error grid analysis (EGA) (3), or the clinically based EGA introduced by our group 18 years ago (4,5). The problem is that these measures judge the quality of the approximation of reference BG (RBG) from readings taken at isolated static points in time, regardless of the temporal structure of the data. In other words, a random reshuffling of data in time will not change the accuracy estimates. As such, these measures work well for the evaluation of SMBG devices by providing relatively distant readings in time. However, applying these measures to evaluate the continuous process approximation offered by CGSs is questionable because the time sequence of the data in such a process is of great importance. An analogy of CGSs versus SMBG with camcorders versus still cameras is inevitable and might be helpful; still cameras produce highly accurate snapshots at random sparse points in time, and camcorders generally offer lower resolution of each separate image but capture the dynamics of the action. Thus, it would be inappropriate to gauge the accuracy of still cameras and camcorders using the same static measure of the number of pixels in a single image. Similarly, it is inappropriate to gauge the precision of CGS and SMBG devices using the same measures and to ignore the temporal characteristics of the observed process.

From the ¹Department of Psychiatric Medicine, University of Virginia Health System, Charlottesville, Virginia; and the ²Department of Pediatrics, University of Virginia Health System, Charlottesville, Virginia.

Address correspondence and reprint requests to Dr. Boris Kovatchev, PhD, University of Virginia Health System, Box 800137, Charlottesville, VA 22908. E-mail: boris@virginia.edu.

Received for publication 24 March 2004 and accepted in revised form 16 May 2004.

Additional information for this article can be found in an online appendix at <http://care.diabetesjournals.org>.

B.P.K., L.A.G.-F., D.J.C., and W.L.C. have received consulting fees from TheraSense.

Abbreviations: BG, blood glucose; CG-EGA, continuous glucose–error grid analysis; CGS, continuous glucose sensor; EGA, error grid analysis; P-EGA, point–error grid analysis; RBG, reference blood glucose; R-EGA, rate–error grid analysis; SBG, sensor blood glucose; SMBG, self-monitoring of blood glucose.

A table elsewhere in this issue shows conventional and Système International (SI) units and conversion factors for many substances.

© 2004 by the American Diabetes Association.

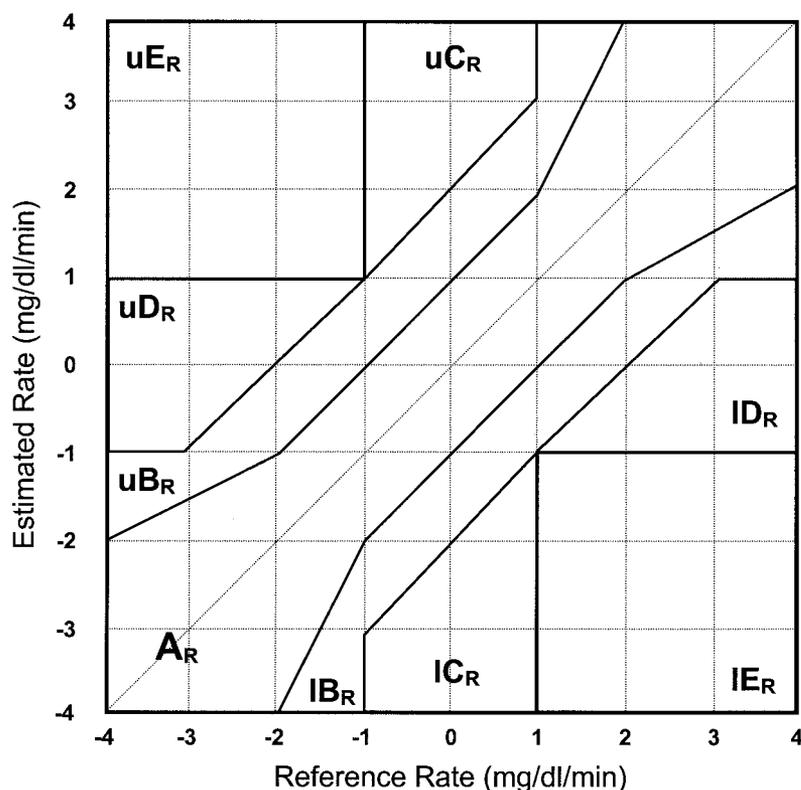


Figure 1—R-EGA divided into A_R , B_R , C_R , D_R , and E_R for SBG versus RBG. *l*, lower; *R*, rate; *u*, upper.

This study introduces the new continuous glucose–error grid analysis (CG-EGA), which was specifically designed to evaluate the clinical accuracy of CGSs in terms of both precision of BG readings and precision of BG rate of change. Unlike the original EGA, the CG-EGA examines temporal characteristics of the CGS data, analyzing pairs of reference and sensor readings as a process in time represented by a bidimensional time series and taking into account inherent physiological time lags. The estimates of point and rate precision are then combined in a single accuracy assessment presented for each one of three preset BG ranges: hypoglycemia ($RBG \leq 70$ mg/dl), euglycemia (70–180 mg/dl), and hyperglycemia (>180 mg/dl). Like the original EGA (5,6), CG-EGA focuses on the clinical implications of measurement errors by addressing the question of what type of clinical outcome might occur if the patient took action based on CGS feedback about BG levels and rate of change. Thus, CG-EGA evaluates the accuracy of CGSs to prompt appropriate clinical action by the patient or by a future automated insulin-delivering device coupled with CGSs. CG-EGA is

not intended to study the accuracy of long-term trends depicted by CGSs. In our conclusions, we discuss this point further, together with ideas for studying long-term trend accuracy using contemporary time series analysis methods. Thus, CG-EGA is only the first step in a series of methods for evaluating the rather complex problem of estimating time series approximation accuracy.

While development of the CG-EGA is based on accepted clinical assumptions and is data independent, we illustrated the application of the CG-EGA by using data from a clinical trial to analyze the accuracy of one new CGS device, the TheraSense Freestyle Navigator (TheraSense, Alameda, CA).

RESEARCH DESIGN AND METHODS

Mathematical model

BG fluctuations are a continuous process in time $[BG(t)]$. Each point of that process is characterized by its location, speed, and direction of change. Thus, at any point in time $BG(t)$ is a vector with a specific bearing. CGSs allow the monitoring of this

process in short (e.g., 5- to 10-min) increments, producing a parallel discrete time series that approximates $BG(t)$. CG-EGA has to judge the precision of this process of approximation in terms of both accuracy of BG readings and accuracy of evaluation of BG change. Thus, we introduced a new concept of rate–error grid analysis (R-EGA) as well as modified the traditional EGA into a new point–error grid analysis (P-EGA) that reflects the temporal characteristics of $BG(t)$.

CGS testing protocol

To capture the accuracy of a CGS throughout dynamic BG fluctuation, a device testing procedure would need to record frequent pairs of RBG and sensor BG (SBG) readings. Regardless of the sensor frequency of measurement, the pace of data acquisition for CG-EGA is controlled by the pace of acquisition of reference data points. Since obtaining frequent RBG data is a laborious process, we suggest reference readings taken in 10- to 15-min increments, a sampling frequency that is sufficient enough to capture a representative picture of BG fluctuation, and in several ~ 4 -h blocks that are representative of the life of the sensor and for the testing conditions (e.g., insulin, carbohydrate, exercise challenges, etc.). While the precise frequency of reference readings and testing protocol duration would be established by a future consensus, to construct R-EGA and P-EGA we assume that paired RBG-SBG readings are available through a sufficiently frequent sampling.

R-EGA

For each pair of RBG readings $[RBG(t_1), RBG(t_2)]$ taken at times t_1 and t_2 , the RBG rate is computed as ΔBG divided by the elapsed time. RBG rate of change ($\text{mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$) = $[RBG(t_2) - RBG(t_1)] / (t_2 - t_1)$. Similarly, for each SBG-reported pair $[SBG(t_1), SBG(t_2)]$, SBG rate is computed as SBG rate of change ($\text{mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$) = $[SBG(t_2) - SBG(t_1)] / (t_2 - t_1)$.

SBG rate is then plotted against the RBG rate (Fig. 1). The boundaries of this particular scatterplot are set to -4 to 4 $\text{mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$, which would depict $\sim 99\%$ of the observed rates of change. However, the R-EGA is not limited by these boundaries (any rate of change could be evaluated), as the R-EGA zones theoretically extend to infinity.

The R-EGA scatterplot is divided into zones A through E, which have a clinical

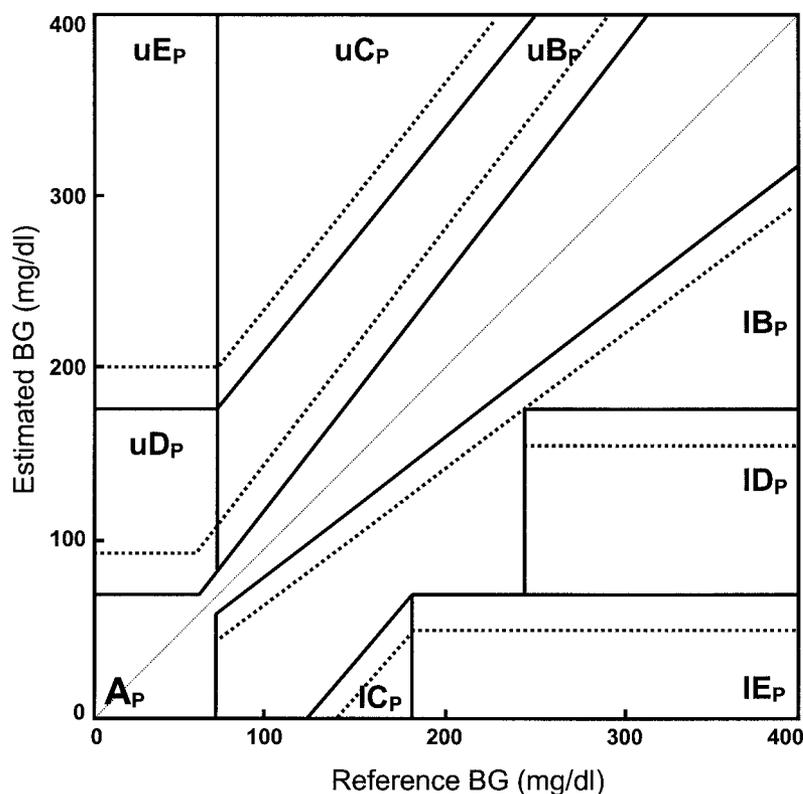


Figure 2—P-EGA divided into A_p , B_p , C_p , D_p , and E_p for SBG versus RBG. *l*, lower; *P*, point; *u*, upper.

meaning similar to the original EGA (5,6). Accurate A_R zone is the main diagonal in Fig. 1 and signifies a perfect fit. An SBG rate within $1 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$ from the diagonal is considered accurate. The accuracy boundaries are expanded to $\pm 2 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$ at extreme BG rates of $\pm 4 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$. Because such rapid rates are rare and cannot be sustained for prolonged periods, a correct recognition of their direction is sufficient for an accurate clinical decision. In the C_R zone (over-correction), the reference rate is -1 to $1 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$ showing no significant BG fluctuation. However, the sensor displays a significant BG fluctuation, which could lead to overtreatment. The C_R zone is divided into overestimation (upper C_R) and underestimation (lower C_R) of the reference rate of change. In the D_R zone (failure to detect), RBG shows significant change, while SBG fails to detect that change, showing readings within -1 to $1 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$. Upper D_R and lower D_R zones signify the failure to detect rapid BG fall or rise. In the E_R zone (erroneous reading), the sensor display readings are opposite the reference rate of change. Upper E_R , which is an actual BG

decline, is estimated as BG rise, whereas lower E_R , which is an actual BG rise, is interpreted as BG fall. The B_R zone (benign errors) shows sensor errors that do not cause inaccurate clinical interpretation, or if they do, treatment action is unlikely to occur or to result in a negative outcome.

P-EGA

This analysis follows the clinical logic of our traditional EGA (5,6). As presented in Fig. 2, P-EGA plots SBG versus RBG divided into A_p , B_p , C_p , D_p , and E_p zones.

To account for the specifics of BG fluctuations interpreted as a process in time, these P-EGA zones are defined depending on the reference rate of BG change as follows. 1) If the RBG rate is within -1 to $1 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$ (e.g., no significant change), P-EGA zones are identical to the zones of the traditional EGA (Fig. 2, solid lines). 2) If the RBG is falling at a rate of -2 to $-1 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$, the upper limits of upper A_p , B_p , and D_p zones are expanded by 10 mg/dl . Similarly, if the RBG is falling faster than $-2 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$, the upper limits of upper A_p , B_p , and D_p zones are expanded

by 20 mg/dl (Fig. 2, dotted lines in uA_p , uB_p , and uD_p). 3) If the RBG is rising at a rate of 1 – $2 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$, the lower limits of lower A_p , B_p , and D_p zones are expanded by 10 mg/dl . If the RBG is rising faster than $2 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$, the lower limits of lower A_p , B_p , and D_p zones are expanded by 20 mg/dl (Fig. 2, dotted lines in lA_p , lB_p , and lD_p).

These adjustments are made dynamically for each upcoming data pair. Thus, appropriate software is needed to compute P-EGA. In addition, these adjustments equate, to a certain extent and in terms of clinical accuracy, the process observation by CGS to the point observation by SMBG. If BG is rapidly falling and the sensor accurately depicts this descent, then a sensor reading right above the A_p zone will be clinically interpreted as a SMBG reading within the A_p zone. For example, if RBG is 68 mg/dl and the sensor reads 75 mg/dl and is falling at $2 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$, the sensor reading will cause a treatment reaction similar to the reference reading. In that sense, the sensor display is clinically accurate, while in the traditional EGA this would be an upper D zone error. Similarly, when BG is rapidly rising, the lower limits of the lower zones are expanded to accommodate the clinical interpretation of the display. The zone-expansion constants 10 and 20 mg/dl correspond to rates of change 1 – $2 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$ and faster than $2 \text{ mg} \cdot \text{dl}^{-1} \cdot \text{min}^{-1}$. This means that, on average, the sensor reading will reach the corresponding traditional EGA zone within 7 min ($1.5 \text{ mg} \cdot \text{dl}^{-1} \cdot 7 \text{ min}^{-1}$ or $3 \text{ mg} \cdot \text{dl}^{-1} \cdot 7 \text{ min}^{-1}$). The constant 7 min was selected on the basis of reported average delays between blood and interstitial glucose (7). In that sense, the zone expansion accounts for time lags inherent to interstitial sensor.

Combining R-EGA and P-EGA

CG-EGA recognizes that the clinical meaning of CGS rate accuracy depends greatly on the location of absolute BG, with different BG levels requiring different interpretations of the combination R-EGA + P-EGA. For this reason, CG-EGA computes combined R-EGA + P-EGA accuracy in three clinically relevant regions: hypoglycemia defined as $\text{BG} < -70 \text{ mg/dl}$ (3.9 mmol/l) and euglycemia and hyperglycemia both defined as $\text{BG} > 180 \text{ mg/dl}$ (10 mmol/l). Thus, a CGS gets three estimates of its performance com-

Downloaded from <http://diabetesjournals.org/care/article-pdf/27/8/1922/6511646/zdc00804001922.pdf> by guest on 16 August 2022

| | | Point Error-Grid Zones | | | | | | | | | | |
|-----------------------|------|------------------------|-------|------|---------------------|-------|------|----------------|------|----|------|----|
| | | Hypoglycemia | | | Euglycemia | | | Hyperglycemia | | | | |
| | | BG ≤ 70 mg/dl | | | 70 < BG ≤ 180 mg/dl | | | BG > 180 mg/dl | | | | |
| | | A | D | E | A | B | C | A | B | C | D | E |
| Rate Error-Grid Zones | A | 57.4% | 18.8% | 0% | 53.2% | 23.4% | 0.1% | 55.9% | 9.4% | 0% | 0.1% | 0% |
| | B | 13.5% | 5.1% | 0% | 11.6% | 5.7% | 0.1% | 20.1% | 3.7% | 0% | 0.1% | 0% |
| | uC | 0% | 0% | 0% | 0.6% | 0.5% | 0% | 1.5% | 0.6% | 0% | 0% | 0% |
| | IC | 1.2% | 0.4% | 0% | 0.4% | 0.5% | 0% | 1.1% | 0.3% | 0% | 0% | 0% |
| | uD | 0.6% | 1.4% | 0% | 1.0% | 0.7% | 0% | 2.0% | 0.8% | 0% | 0% | 0% |
| | ID | 0.6% | 0% | 0% | 1.0% | 0.4% | 0% | 2.4% | 0.3% | 0% | 0% | 0% |
| | uE | 0% | 0.2% | 0% | 0.1% | 0.2% | 0% | 0.6% | 0.3% | 0% | 0% | 0% |
| IE | 0.8% | 0% | 0% | 0.4% | 0.1% | 0% | 0.7% | 0.1% | 0% | 0% | 0% | |



Accurate Readings



Benign Errors



Erroneous Readings

Figure 3—Error matrix combining R-EGA and P-EGA zones.

puted according to the error grid matrix presented in Fig. 3. The premise behind the definition of the error grid matrix is similar to the clinical idea of our original EGA. We sought to determine the extent at which the sensor reading would result in accurate treatment or benign or significant error. In the next section, we discuss this combined error grid matrix in detail, together with data from TheraSense Navigator sensors.

Clinical trial procedure

The application of CG-EGA is illustrated by data from a 3-day inpatient protocol used to test a total of 48 Navigator sensors on 30 subjects (8 women and 22 men, aged 20–85 years) with type 1 diabetes. The previous publication on this clinical trial provides a detailed description of the study population, the sensor insertion and calibration, and the RBG measurement (2). The study included two hypoglycemic and two hyperglycemic challenges and recorded RBG-SBG pairs every 15 min.

RESULTS

P-EGA

Overall, 75% of the Navigator point estimates fell into the A_P zone, 23.7% fell into the B_P zones, 0.1% fell into the C_P zones, and 1.2% fell into the D_P zones. There were no E_P zone errors. Compared with the traditional EGA (without dynamic zone adjustment), this A zone accuracy was higher and the percentage of benign errors was lower. The traditional EGA resulted in 70.7% A zones and an additional 27.6% benign errors. Stratified by BG range, the A_P zone accuracy of the sensor was 74.1% in the hypoglycemic range, 68.3% at euglycemia, and 84.2% at hyperglycemia ($P < 0.01$ using Kruskal-Wallis). Additional 25.9, 31.5, and 15.5% benign errors were observed in these BG ranges. The performance of the sensor did not decrease significantly with higher BG rates of change. The percentage of readings in the traditional EGA A + B zones (which are not dynamically adjusted and therefore useful for across-rate comparisons) was 98.7% when BG

fluctuations were slow (within ± 1 mg · dl⁻¹ · min⁻¹), 98.6% when BG fluctuations were moderate (absolute BG change between 1 and 2 mg · dl⁻¹ · min⁻¹), and 96.3% when BG fluctuations were rapid (absolute BG change > 2 mg · dl⁻¹ · min⁻¹).

R-EGA

While it is inappropriate to consider R-EGA separately from absolute BG values and P-EGA, information about rate accuracy might be useful for evaluating a sensor's dynamic characteristics. Overall, 72.1% of the Navigator rate of change estimates fell into the A_R zone, 20.1% fell into the B_R zones, 2.5% fell into the C_R zones, 4% fell into the D_R zones, and 1.3% fell into the E_R zones. Stratified by BG range, the A_R zone accuracy of the sensor was approximately the same in the hypoglycemic (76.3%) and the euglycemic ranges (76.7%) and was significantly lower in the hyperglycemic range (65.4%) ($P < 0.01$).

Combined CG-EGA results

As presented in Fig. 3, a more complex pattern in accuracy emerges when both rate and point are considered concurrently. CGS estimates can be clinically accurate in terms of BG location but inaccurate in terms of rate of change and vice versa. In Fig. 3, estimates are considered to be clinically accurate when they fall into the A or B zones of both the P-EGA and the R-EGA. Clinically benign errors are those with acceptable point accuracy (A or B P-EGA zones) and significant errors in rate accuracy (C, D, or E R-EGA zones), which would unlikely lead to clinical action or negative clinical consequences. Clinically significant errors are those that could lead to negative clinical action and outcome.

As Fig. 3 illustrates, the zones considered clinically benign depend on absolute BG level and therefore differ across the three BG ranges. For example, when CGS feedback accurately indicates that hypoglycemia is occurring, treatment to raise BG is likely to be needed regardless of rate information. For this reason, most rate errors (upper and lower C_R , lower D_R , and lower E_R zones) in this case are likely benign. However, even if hypoglycemia is accurately detected, failure to detect that BG is continuing to fall rapidly (upper D_R zone) or showing that BG is increasing rapidly when it is actually continuing to fall (lower E_R zone) is a clinically significant error. When CGS accurately indicates euglycemia, treatment is typically not needed and is unlikely to occur regardless of feedback about BG change, so tolerance for rate errors is greater. Only those estimates that indicate a rapid change in the wrong direction are considered to be clinically significant (upper and lower E_R zones), whereas the other rate error zones (upper and lower C_R and D_R zones) are considered benign. In contrast, with hyperglycemia, failure to detect a rapid increase in BG (lower D_R zone) could lead to negative clinical consequences and is considered a clinically significant error along with upper and lower E_R zone estimates.

According to Fig. 3, the percentage of Navigator readings that were clinically accurate or resulted in benign errors was 73.5% at hypoglycemia (70.9% accurate + 2.6% benign), 99% at euglycemia (93.9% accurate + 5.1% benign), and 95.4% at hyperglycemia (89.1% accurate + 6.3% benign). Clinically significant er-

rors occurred for 26.5% of hypoglycemic, 1.0% of euglycemic, and 4.6% of hyperglycemic reference values, indicating that the device is extremely accurate when BG is near normal but less accurate at BG extremes. Accuracy was poorest during hypoglycemia due to a high rate of upper D_P zone errors, indicating that the device failed to detect 25.9% of the hypoglycemic readings even after the P-EGA was dynamically adjusted to account for interstitial time lag of the data.

Sensor placement and reliability over time

CG-EGA showed no statistically significant difference in the accuracy of the Navigator when used for arm versus abdominal testing. There were also no significant performance differences during test days 1, 2, and 3 of the study.

CONCLUSIONS— While the first Minimed continuous glucose-monitoring system was approved only for evaluation of BG trends (and does not display momentary data), new generations of CGSs aim at replacing SMBG in terms of prompting immediate treatment, signaling events such as upcoming hypoglycemia and, most importantly, automating glycemetic control by closed-loop insulin infusion algorithms. To achieve such goals, CGS would need to pass the test of momentary accuracy at any point in time and in terms of both BG value and direction of change. This study introduces a new approach to analyzing such a performance. We would like to emphasize, however, that the purpose of this report is not to set a method in stone but to take a fresh look at and initiate a debate on a problem that is current and important.

CG-EGA has three advantages over other CGS evaluation procedures. 1) CG-EGA considers the sensor information as a process in time, as opposed to correlation, regression, consensus error grid (3), or the traditional EGA (5), all of which ignore the temporal structure of the BG fluctuations. 2) CG-EGA judges CGS performance separately at hypoglycemia, euglycemia, and hyperglycemia, “utilizing stratified data according to the magnitude of glycemia,” as advocated by Klonoff (8). 3) CG-EGA preserves the clinical assumptions of the established EGA, thus facilitating the transition from SMBG to CGS accuracy estimation.

Since CG-EGA is built on clinically

accepted assumptions and thresholds, no data were involved in its design. To validate and illustrate the analytical options provided by CG-EGA, we present data from a clinical trial of the TheraSense Navigator. This first test of the CG-EGA demonstrates that this analysis provides more comprehensive and detailed information regarding the clinical accuracy of CGS compared with other procedures often used in similar studies. For example, we find that the combined accuracy of the Navigator is greatest during euglycemia and poorest during hypoglycemia, its point accuracy is relatively stable over a wide range of BG rates of change, and its rate accuracy decreases significantly at high BG levels. To facilitate the summary presentation of such information, we propose a convenient standardized presentation of CGS evaluation results (see online appendix [available at <http://care.diabetesjournals.org>]).

The CG-EGA has two interacting parts, P-EGA and R-EGA, which are combined into a single error matrix representing CGS performance across three BG ranges. Although P-EGA and R-EGA may provide separate information about sensor performance, we suggest that their results be considered only in combination. The reason for this is the inherent interaction between point and rate analyses. The zones of P-EGA are dynamically adjusted (for each data pair) according to the RBG rate of change at that point. Conversely, the interpretation of errors in rate is heavily dependent on RBG. For example, an upper D_R zone error would likely have minimal impact on self-treatment when absolute BG is 150 mg/dl. If BG is 85 mg/dl and rapidly falling, this error could result in a failure to take action and prevent hypoglycemia. This said, we would like to outline several points that would benefit from further discussion before reaching a consensus finalizing the parameters of the CG-EGA.

Formalizing the procedure

BG fluctuations are a continuous process in time where $BG(t)$ is observed at discrete time points less frequently and, in general, irregularly by SMBG or more frequently and regularly by CGSs. CGSs, despite their name and regardless of their testing frequency, produce discrete time series readings that reflect the underlying $BG(t)$. To test CGS performance, a clinical trial needs to record a time series of CGS

readings and a parallel time series of RBG values. The latter is almost always the limiting factor setting the temporal resolution of the (reference and sensor) BG pairs. Thus, a CGS evaluation procedure is based on two parallel time series, one of which (reference) is believed to more accurately reflect the underlying continuous process. Consequently, evaluating CGS performance is equivalent to evaluating the proximity between these time series. The resolution of the time series (e.g., the elapsed time between two sequential data pairs) is critical for the evaluation of CGS accuracy. Therefore, a standard protocol needs to be established that would allow future devices to be tested under comparable conditions. We recommend obtaining pairs of reference sensor readings at 15-min intervals as a balance between capturing sufficiently detailed pictures of BG fluctuations and burden on study subjects. In addition, the testing protocol should ensure a realistic coverage of low and high BG range through insulin and glucose challenges and a sufficient cover of the life of the sensor.

Alternative CGS performance analyses

There are two general approaches to measuring proximity between time series (e.g., precision of CGS). The first is purely mathematical and relies on “distances” between the true BG values and their estimates (for example, regressions utilize Euclidean distance). The second approach is clinical. The device is judged by the clinical accuracy of the message it sends. EGA, the consensus error grid (3), and now CG-EGA use a clinical approach. Each of these approaches has its advantages and limitations. A clinical approach would inevitably involve the use of (more or fewer) clinical thresholds that introduce “rough edges” in the analysis. For example, the traditional EGA is designed around preset thresholds of 70 mg/dl for hypoglycemia and 180 mg/dl for hyperglycemia. The consensus error grid has multiple thresholds determined by expert opinion. More thresholds would generally result in a smoother view of the analysis, yet introduction of more or different error zones does not resolve the problem of discontinuity.

The ultimate way to completely eliminate rough edges is through the use of mathematical reference-to-estimate distances. However, as pointed out in a re-

cent review of the state of the art, “The notion of proximity of complex objects, such as time series, is not trivial and is specific to the application domain and also to the nature of the tasks” (10). Several types of mathematical distances can be used. For example, Euclidean distance (10), wavelet transformation (11), or Kullback-Leibler information distance (12) have been used to assess the proximity between time series. The problem with such approaches is that their outcome is not directly clinically interpretable. Although we considered using a mathematical definition of distance between reference and sensor time series as a base for CG-EGA, we opted for a clinical approach that is “specific to the application domain and also to the nature of the tasks” (10). In general, we believe that a comprehensive evaluation of a device should include both clearly clinical (such as EGA or CG-EGA) and clearly mathematical (correlation, regression, and information) approaches.

Parameters of the CG-EGA

As with any analysis, CG-EGA uses basic assumptions and preset parameters such as boundaries of hypoglycemia and time lags. Since CG-EGA is intended for software application, most of these parameters could be user selectable. For example, the time lag between blood and interstitial glucose has a default value of 7 min, based on literature data (7). If a device has a longer technical lag (or no inherent lag at all, as with implantable sensors), then the software would allow the time lag used by the P-EGA to be changed. The same is true for setting one or more of the thresholds for hypoglycemia. The concept of CG-EGA is even flexible enough to accommodate other approaches to point accuracy. For example, the P-EGA has alternatives, such as the consensus error grid (3) or point accuracy, which could be assessed using International Organization for Standardization standards (13). The latter is particularly useful in the hypoglycemic range, where it is more restrictive than EGA. Even though it is entirely possible to substitute P-EGA with another analysis in the construct of the overall CG-EGA, at this time we would argue against such a substitution because it would introduce a discord between the currently coherent zone definitions of P-EGA and R-EGA. The downside is that such flexibility could

have unwanted consequences, such as artificially improved (or decreased) accuracy. In addition, no two devices would be comparable if different parameters are chosen for testing each device. This latter point emphasizes the need for a standardized testing protocol. Such a protocol would result from a consensus that would set, in particular, the parameters of the CG-EGA.

Acknowledgments—CG-EGA is based on methods developed by the National Institutes of Health (National Institute of Diabetes and Digestive and Kidney Diseases Grant RO1 DK 51562).

The authors thank Holly Kulp, Geoff McGarraugh, and Tim Goodnow (TheraSense, Alameda, CA) for sharing their data and persuading this research and Dr. Leon Farhy for his insightful comments on the CG-EGA.

References

1. Gross TM, Bode BW, Einhorn D, Kayne DM, Reed JH, White NH, Mastrototaro JJ: Performance evaluation of the MiniMed continuous glucose monitoring system during patient home use. *Diabetes Technol Ther* 2:49–56, 2000
2. Feldman B, Brazg R, Schwartz S, Weinstein R: A continuous glucose sensor based on wired enzyme technology: results from a 3-day trial in patients with type 1 diabetes. *Diabetes Technol Ther* 5:769–778, 2003
3. Parkes JL, Slatin SL, Pardo S, Ginsberg BH: A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care* 23:1143–1148, 2000
4. Cox DJ, Clarke WL, Gonder-Frederick LA, Pohl SL, Hoover C, Snyder A, Zimbleman L, Carter WR, Bobbitt S, Pennebaker J: Accuracy of perceiving blood glucose in IDDM (Letter). *Diabetes Care* 8:526–528, 1985
5. Clarke WL, Cox DJ, Gonder-Frederick LA, Carter WR, Pohl SL: Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 10:622–628, 1987
6. Cox DJ, Gonder-Frederick LA, Kovatchev BP, Julian DM, Clarke WL: Understanding error grid analysis (Editorial). *Diabetes Care* 20:911–912, 1997
7. Boyne MS, Silver DM, Kaplan J, Saudek CD: Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor. *Diabetes* 52:2790–2794, 2003
8. Klonoff DC: The need for separate performance goals for glucose sensors in the hypoglycemic, normoglycemic, and hyperglycemic ranges (Editorial). *Diabetes*

- Care 27:834–836, 2004
9. 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
 10. Kalpakis K, Gada D, Puttagunta V: Distance measures for effective clustering of ARIMA time-series (article online). In *Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, 29 November to 2 December 2001*. p. 273–280. Available from <http://csdl.computer.org/comp/proceedings/icdm/2001/1119/00/1119toc.htm>.
 11. Struzikand Z, Sibes A: Measuring time series similarity through large singular features revealed with wavelet transformation (article online). In *Proceedings of the 10th International Workshop on Database and Expert Systems Applications, 1999*. p. 162–166. Available from <http://csdl.computer.org/comp/proceedings/dexa/1999/0281/00/0281toc.htm>.
 12. Kullback S, Leibler RA: On information and sufficiency. *Annals of Mathematical Statistics* 22:79–86, 1951
 13. International Organization for Standardization: *In Vitro Diagnostic Test Systems: Requirements for Blood-Glucose Monitoring Systems for Self-Testing in Managing Diabetes Mellitus*. Geneva, International Organization for Standardization, 2003 (ISO 15197)