Continuous Glucose Sensors:
Continuing Questions about Clinical Accuracy

William L. Clarke, M.D., and Boris Kovatchev, Ph.D.

Abstract

Continuous glucose sensors (CGS) offer the potential to greatly change the lives of people with diabetes. Even though two of these systems (Guardian RT, Medtronic, Northridge, CA, and DexCom STS, DexCom, San Diego, CA) have been approved by the Food and Drug Administration for use as adjuncts to self-blood glucose monitoring (SBGM), questions remain concerning the accuracy of these devices. When considering accuracy, two distinct approaches should be emphasized: (1) numerical and (2) clinical. Because CGS data are a process in time, each of these two approaches includes two subtypes of accuracy: point and rate. Conventional statistics such as correlation coefficients, mean and median relative absolute differences, and International Standards Organization criteria are measures of numerical point accuracy. A new measure, the R deviation, is introduced to quantify numerical rate accuracy. Error-grid analysis (Clarke EGA) measures clinical point accuracy. The only measure of both clinical point accuracy and rate accuracy is continuous glucose error-grid analysis. This analysis is a combination of two components, P-EGA measuring point accuracy and R-EGA measuring rate accuracy, which are designed to assess the information that distinguishes continuous glucose measurements from intermittent SBGM determinations. Further, a better understanding of the source of the error associated with time lag and its effect on CGS readings may improve sensor output. Finally, the reliability of the CGS sensors, in terms of initial calibration and long-term application, needs to be assessed carefully if current CGS systems are to be used as hypoglycemia monitors or incorporated in the future design of closed loop (artificial pancreas) systems.

Introduction

Continuous glucose sensors (CGS) offer the potential to markedly change our understanding of glucose homeostasis in health and disease and to provide the missing information required to achieve near-normoglycemia among persons with both type 1 and type 2 diabetes. Despite their having appeared on the scene less than 5 years ago, CGS already have been shown to be associated with short-term reductions in glucose variability, time spent in nocturnal hypoglycemia, time spent in hyperglycemia, and lower glycosylated hemoglobin values. Their potential to assist patients and their families in day-to-day decision making, to warn of impending hypoglycemia, and to reduce the fear of its occurrence has yet to be fully appreciated. Indeed, the advent of CGS has stimulated the diabetes research community to reexamine the feasibility of developing a closed loop “artificial pancreas.” Also, preliminary clinical studies of prototype closed loop systems are currently being initiated. Given the enormous therapeutic and safety potential of these systems, it seems reasonable to assume that they have been proven to be numerically and clinically accurate in both the display of glucose values and the tracking of glucose trends and rate of change. Indeed there are numerous publications documenting the accuracy of different CGS systems. Unfortunately, much of the data regarding their accuracy is presented according to criteria identical to those developed decades ago for assessing the accuracy of home blood glucose (BG) monitors providing episodic self-monitoring (SBGM) readings and thus do not include assessments designed to evaluate the “continuous” time-dependent information (in particular rate and direction of glucose change) that is unique to CGS. This article presents current assessments of CGS system accuracy in four broad categories: numerical and clinical point accuracies; numerical and clinical rate accuracies; time lag; and sensor reliability.

Point Accuracy

Numerical Measures

Current U.S. Food and Drug Administration (FDA) accuracy criteria for CGS systems are identical to those for SBGM and include statistics calculated from paired reference and sensor glucose determinations. These criteria include linear regressions and correlation coefficients, International Standards Organization (ISO) criteria, and mean and median absolute and relative absolute differences (MAD, MARD). While MAD gives an indication of the propensity of a particular CGS to read high or low compared to reference, MARD is calculated as the relative (in percentage) deviation of a sensor from reference. ISO criteria are based on the percentage of CGS readings within 15 mg/dl from reference when the reference BG ≤75 mg/dl, or within 20% from reference when the reference BG is >75 mg/dl. Representative data, currently available from either FDA submissions or recent publications, of numerical accuracy of three different CGS systems are shown in Table 1.

<table>
<thead>
<tr>
<th>FDA criteria</th>
<th>Mean absolute relative difference (%)</th>
<th>Median absolute relative difference</th>
<th>EGA zone A (accurate) (%)</th>
<th>EGA zones A + B (acceptable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guardian RT (19)</td>
<td>0.89</td>
<td>17.32</td>
<td>13.98</td>
<td>61.7</td>
</tr>
<tr>
<td>DexCom STS (20)</td>
<td>0.88</td>
<td>18.75</td>
<td>15.42</td>
<td>49</td>
</tr>
<tr>
<td>Navigator (14)</td>
<td>12.8</td>
<td>9.2</td>
<td>81.7</td>
<td>98.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ISO criteria (within ±15 mg/dl when reference ≤75 mg/dl)</th>
<th>Guardian RT (19) (%)</th>
<th>DexCom STS (20) (%)</th>
<th>Navigator (14) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40–80 mg/dl</td>
<td>68</td>
<td>56</td>
<td>55</td>
</tr>
<tr>
<td>&lt;70 mg/dl</td>
<td></td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>81–120 mg/dl</td>
<td>60</td>
<td>44</td>
<td>72</td>
</tr>
<tr>
<td>70–180 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121–240 mg/dl</td>
<td>62</td>
<td>46</td>
<td>81</td>
</tr>
<tr>
<td>&gt;180 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>241–350 mg/dl</td>
<td>61</td>
<td>65</td>
<td></td>
</tr>
</tbody>
</table>
**Clinical Measures**

In addition to these standard numerical criteria, a measure of clinical accuracy, error-grid analysis (EGA), has been suggested. EGA is a method used for evaluating the clinical significance of treatment decisions based on errors in patient-generated glucose determinations and has been in use since the mid-1980s. Clinically significant EGA data for the entire glucose range are also shown in Table 1. Of note, in contrast to presentations of SBGM data where “clinically accurate” results are limited to zone A values (±20% from reference and <70 mg/dl when reference is <70 mg/dl), clinically accurate CGS data are often presented as the sum of zones A (clinically accurate) and zone B (benign errors) and expressed as “clinically acceptable.” Data presented in Table 1 demonstrate that although CGS systems may generate statistically acceptable results, they are not as clinically accurate as SBGM systems. Clearly such information is important to the surveillance for clinically significant hypoglycemia and for making safe therapeutic decisions. Since all currently approved CGS systems have lower numerical and clinical point accuracies than SBGM, they are restricted for use as adjuncts to rather than replacements for SBGM.

In order to adapt the EGA to the nearly continuous data stream generated by CGS, we have introduced the continuous glucose error-grid analysis (CG-EGA). The CG-EGA combines two components, the point EGA (P-EGA) measuring clinical point accuracy and rate EGA (R-EGA described later) measuring clinical rate accuracy. The difference between traditional Clarke EGA and P-EGA is in the dynamical adjustment of the error-grid zones depending on the momentary rate of change of the reference BG process, which is designed to accommodate a possible time lag between reference and sensor readings. The P-EGA and R-EGA are combined in different ways during hypoglycemia, euglycemia, and hyperglycemia, which reflect differences in the relative significance of point and rate accuracies under these three clinically distinct conditions.

**Rate Accuracy**

Continuous glucose sensor systems provide additional clinically important information than just the measurement of glucose at isolated points in time. These systems record a nearly continuous stream of glucose data never before available to either the patient or the care provider, which allows tracking the direction and rate of glucose fluctuation. Indeed, it can be argued that these additional data may more than compensate for the reduced (as compared to SBGM) sensor point accuracy seen with all current CGS systems. However, rate and direction of change are often ignored when accuracy data are presented. This is unfortunate because these distinct properties distinguish CGS data from intermittent SBGM measurements; permit the display of glucose trends, which would help prevent serious glucose extremes such as severe hypoglycemia; and facilitate the calculations of projected glucose levels, which will enable algorithms being developed to direct insulin administration with an artificial pancreas.

**Numerical Measures**

While numerical metrics of point accuracy are widely accepted (e.g., mean error, mean absolute error, percent error), a numerical accuracy metric for rate of change still needs to be established. Following the concept of mean error/absolute error, we suggest R deviation (RD), a new numerical metric of rate of change accuracy, which is based on a distance (deviation) between the rates of change (derivatives) of two processes (e.g., reference and sensor glucose fluctuations).

**Computing R Deviation**

The derivatives of the two processes, reference and sensor fluctuations, are first estimated using numerical approximation. Such estimation is possible via a variety of methods using first- or second-order divided differences, spline, or polynomial approximation, as well as time-series approaches, which could include noise reduction and artifact rejection. To illustrate the idea of R deviation, in this particular application we used first-order divided differences, which is the computationally simplest approach. Figure 1 presents graphically the deviation between reference and sensor rates of change. These rates are computed as

\[
\frac{\Delta R}{\Delta t} = \tan(\alpha)
\]

for reference glucose fluctuations and

\[
\frac{\Delta S}{\Delta t} = \tan(\beta)
\]

for sensor-depicted fluctuations, where \(\Delta R\) and \(\Delta S\) are the reference and sensor differentials over a time period \(\Delta t\) (typically, \(\Delta t\) would be ~15 minutes).

Definition 1: R deviation is defined as the difference between reference and sensor instantaneous rates of change:

\[
RD = \frac{\Delta R - \Delta S}{\Delta t}.
\]
Definition 2: Absolute R deviation (ARD) is defined as the absolute value of R deviation:

\[ ARD = \left| \frac{\Delta R - \Delta S}{\Delta t} \right| \]

Mean RD corresponds to mean error in point accuracy, whereas mean ARD corresponds to mean absolute error in point accuracy. Both are measured in mg/dl/min. The R deviation was tested with 12,000 reference-sensor (Freestyle Navigator) data pairs collected at ~15-minute intervals during steady state and periods of glycemic challenges. A detailed description of this data set has been presented previously. Results show excellent compatibility with results from R-EGA, the component of the CG-EGA measuring the clinical accuracy of rate of change presented later.

Figure 1. Computation of R deviation as the distance between derivatives (rates of change) between the glucose fluctuation processes reflected by reference and sensor measurements.

Figure 2. The empirical distribution of R deviation with superimposed normal density curve. This distribution is symmetric around zero, which permits the use of symmetric confidence intervals.

Clinical Measures
As described earlier, continuous glucose-error grid analysis includes the R-EGA, a separate analysis of the clinical accuracy of rate and direction of change. For each pair of reference BG readings, the rate and direction of change are computed and plotted against simultaneous CGS-generated values. The boundaries of this plot are from -4 to +4 mg/dl/min, and the scatter plot is divided into zones A through E, which are clinically similar to those of the original point EGA.

Thus zone A (clinically accurate) reference vs CGS rates are within ±1 mg/dl/min of each other. Zone D (failure to detect) includes points with CGS-generated rates between -1 and +1 mg/dl/min when reference rates are actually either -1 to -4 mg/dl/min or +1 to +4 mg/dl/min. Zone D represents failures to detect extreme rates of change. Zone E values are erroneous extreme errors in which CGS rates are in the extreme opposite direction from the reference rates. Zone C values are CGS rates at the extremes when the actual reference rates are -1 to +1 mg/dl/min. These represent potential overcorrection errors. The association between R deviation and zones of R-EGA is very strong: the Spearman correlation between the two measures is 0.824, \( p < 0.00001 \), showing excellent agreement between numerical and clinical accuracies.

Combining Point and Rate Accuracies
The CG-EGA combines point and rate accuracies separately for each of the three critical BG ranges: hypoglycemia (BG <70 mg/dl), euglycemia (70 mg/dl > BG ≤180 mg/dl), and hyperglycemia (BG >180 mg/dl) using matrices of point accuracy vs rate accuracy. These
matrices reflect the relative importance of point accuracy vs rate accuracy in different clinical situations. Because of the differences in the relative importance of point and rate in hypoglycemia, euglycemia, and hyperglycemia, we advocate against combining point and rate accuracies uniformly across the entire glucose range—the suggested mode of reporting of results is separate, by clinical region.

The CG-EGA has been used to compare the performance of difference CGS during clamped euglycemia and in the critical hypoglycemic range. It is worth noting that presenting only point accuracy at various rates of change cannot substitute for the description of accuracy of rate and direction trends presented by the R-EGA. For example, point accuracy data of CGS (Navigator, Abbott Diabetes Care, Alameda, CA), using the original EGA in six different rates of change ranges, have been published. As might be anticipated, point accuracy was greatest when rate of change was the least. However, these data do not provide information to what extent the CGS was accurate in following reference rate and direction of change. The clinical importance of rate and direction accuracies can be illustrated with a close-up photograph of an automobile on a road. The photograph is an accurate but an incomplete depiction of the activity occurring as there is no information regarding direction or speed of travel of the automobile. Such information could be critical to preventing serious consequences, such as a fatal crash. Rate and direction of change information will be critical to direct insulin infusions with closed loop systems. To date, there have been no published studies of the accuracy of rate and direction of glucose change other than those described here.

**Time Lag**

Boyne and colleagues have provided important information regarding the physiologic time lag between blood and interstitial glucose concentrations. Current CGS systems measure interstitial glucose and thus have an inherent lag, which can be influenced by the rate of glucose change. Because CGS systems should not be penalized for not measuring what is not there, point accuracy reports should at least partially accommodate possible interstitial lag, which could be responsible for some of the inaccuracy observed. The CG-EGA includes modifications of the original EGA to take into account the time lag. The point EGA upper zones are expanded when BG is falling and the lower zones are expanded when the BG is rising. Adjustments are made dynamically for each pair of consecutive data points. As an illustration, if the CGS reads 75 mg/dl and the reference reads 68 mg/dl and the rate of change shows the BG to be falling at 1–2 mg/dl/min, the A zone upper limit is expanded by 10 [1.5 mg/dl/min × 7 minutes (physiologic time lag estimate)] and the point pair becomes clinically accurate. The treatment decision would be the same as if the reference and CGS values were identical.

Another method for accounting for the time lag in CGS sensors has been proposed by Wentholt and associates. They have used a novel method of curve fitting based on least-squares regression to assess the delay (or horizontal shift) between sensor and reference glucose values and used this method, along with conventional mean absolute deviation MAD data, to compare differences in sensor accuracy during hypo- and hyperglycemia using two different types of CGS systems.

Our research group has attempted to separate BG vs CGS error into two components using mathematical modeling of an approximation of interstitial glucose (IG). Using such an approximation, one can evaluate the accuracy of BG vs IG and IG vs CGS separately. Such an analysis of retrospectively collected data points during clamped euglycemia and hypoglycemia demonstrated that BG vs IG clinical accuracy during euglycemia was 96% and IG vs CGS clinical accuracy was 90%. However, during hypoglycemia BG vs IG was 95% while IG vs CGS accuracy was 62%, suggesting that the error associated with at least one CGS system during hypoglycemia remains an engineering and not a physiologic problem. Thus advances in sensor technology may reduce the error associated with time lag.

**Reliability and Performance in Distinct Glucose Ranges**

Currently there is very little information available regarding sensor failure, i.e., number of times a CGS must be removed and replaced because of an inability to calibrate or failure to generate or transmit a signal. Anecdotal reports suggest that CGS technology has not progressed to the level of reliability associated with less sophisticated SBGM systems. Because diabetes is often characterized by widely fluctuating BG levels and because early reports of continuous monitoring with CGS have demonstrated that far more persons with type 1 diabetes spend far more time with low BG levels, especially during the night, than previously imagined, the performance of CGS systems in the hypoglycemic range is critical. Clinical point accuracy of three CGS systems during hypoglycemia, as reported by the manufacturers, is shown in Table 2. These data contrast with those generated by two of these systems during...
gradual descent into hypoglycemia as analyzed using CG-EGA. Although both of these systems performed equally well during euglycemia, one was significantly more accurate at tracking glucose levels below 70 mg/dl (Table 2).

### Table 2. Clinical Accuracy of CGS during Hypoglycemia

<table>
<thead>
<tr>
<th>Original (point) error grid analysis</th>
<th>BG</th>
<th>Zone A</th>
<th>Zones A+B</th>
<th>Zones C+D+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guardian RT (19)</td>
<td>40–80 mg/dl</td>
<td>60%</td>
<td>76.1%</td>
<td>23.9%</td>
</tr>
<tr>
<td>DexCom STS (20)</td>
<td>40–80 mg/dl</td>
<td>52%</td>
<td>69%</td>
<td>31%</td>
</tr>
<tr>
<td>Navigator (14)</td>
<td>&lt;70 mg/dl</td>
<td>54.5%</td>
<td>54.5%</td>
<td>45.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuous glucose error-grid analysis</th>
<th>BG</th>
<th>Zone A</th>
<th>Zones A+B</th>
<th>Zones C+D+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGMS (13)</td>
<td>&lt;70 mg/dl</td>
<td>61.6%</td>
<td>63.8%</td>
<td>37.2%</td>
</tr>
<tr>
<td>Navigator (13)</td>
<td>&lt;70 mg/dl</td>
<td>82.4%</td>
<td>88%</td>
<td>12%</td>
</tr>
</tbody>
</table>

### Discussion

Continuous glucose sensor systems are in their infancy, yet their growth has been prodigious. Sensor design, calibration, signal transmission, and data displays have all undergone significant changes, and reports of clinical trials describing each new change are being published regularly. It is anticipated that CGS use will increase dramatically over the next few years and possibly suggest new therapeutic regimens. Failure to make use of the full gamut of data available from these devices can only hinder their implementation and use.

The accuracy of CGS systems will need to be evaluated both in terms of numerical proximity and in terms of impact on clinical decisions made on the basis of CGS data. Because CGS trace a process of glucose fluctuations developing in time, the quality of approximation of both the location (point accuracy) and the speed/direction (rate accuracy) of this process become important. A number of well-accepted measures of numerical and clinical point accuracy exist, including MAD, MARD, ISO standards, and the Clarke EGA in the clinical field. Rate accuracy is a new paradigm that was first introduced by the CG-EGA, in particular by its rate analysis component R-EGA. This article introduced the R deviation—a numerical metric of rate accuracy that gauges the proximity between sensor and reference rates of change. This measure involves numerical estimation of the first derivatives of these processes, which can be done by a variety of standard methods. When evaluating numerical and clinical point and rate accuracies, two aspects become important. First, any accuracy evaluation study is vulnerable to random errors in both sensor and reference data. In addition, sensors are vulnerable to calibration errors as well. Thus, the data analysis should attempt to separate the influence on accuracy of calibration, interstitial time lag, and random noise. Second, combining point and rate accuracies is not straightforward. The reason is that these complementary characteristics of the process of glucose fluctuation have different relative importance in different stages of glycemia. It is therefore suggested that accuracy estimates are provided and performance goals are discussed separately in the hypoglycemic, euglycemic, and hyperglycemic ranges.

Accuracy standards should evolve as time lag is reduced, CGS reliability is improved, and new technology is being developed. For example, taking advantage of both glucose values and their trends and rates of change is particularly important for the development of closed loop “artificial pancreas” systems for the automated control of glycemia in insulin-requiring individuals. Indeed some centers have initiated inpatient trials in both adults and children. Such systems have three basic components: (1) a glucose sensor, (2) an automated insulin delivery device (insulin pump), and (3) a set of control algorithms that utilize sensor data and other information to direct the delivery of insulin. The level and the type of accuracy required for closed loop systems to be safe and viable are not clear, but certainly will depend on the projected glucose target. For instance, if the projected glucose target range is between 80 and 200 mg/dl, there is little question that current CGS systems are capable of measuring such levels accurately. If, however, outcomes will include significant reductions in glycemic variability, better glucose accuracy including rate and direction of change analyses may be required. If closed loop systems are to detect and prevent projected hypoglycemia, then the CGS will need to measure glucose accurately in the hypoglycemic range.

In conclusion, it is of paramount importance that researchers in the field of CGS technology suggest and support methods for describing the numerical and clinical accuracies of these systems, which take into account the full gamut of data being generated. Relying on methods to document accuracy that do not recognize the characteristics of continuous data streams and the statistical methods required to describe these data may prevent current CGS systems from fulfilling their clinical potential.
Acknowledgement:
This work was supported in part by NIH NIDDK Grant ROI DK 51562.

References: