

Comparison of the Numerical and Clinical Accuracy of Four Continuous Glucose Monitors

BORIS KOVATCHEV, PHD¹
STACEY ANDERSON, MD¹

LUTZ HEINEMANN, PHD²
WILLIAM CLARKE, MD¹

OBJECTIVE — The purpose of this study was to compare the numerical and clinical accuracy of four continuous glucose monitors (CGMs): Guardian, DexCom, Navigator, and Glucoday.

RESEARCH DESIGN AND METHODS — Accuracy data for the four CGMs were collected in two studies: Study 1 enrolled 14 adults with type 1 diabetes at the University of Virginia (UVA), Charlottesville, Virginia; study 2 enrolled 20 adults with type 1 diabetes at the Profil Institute for Metabolic Research, Neuss, Germany. All participants underwent hyperinsulinemic clamps including 1.5–2 h of maintained euglycemia at 5.6 mmol/l followed by descent into hypoglycemia, sustained hypoglycemia at 2.5 mmol/l for 30 min, and recovery. Reference blood glucose sampling was performed every 5 min. The UVA study tested Guardian, DexCom, and Navigator simultaneously; the Profil study tested Glucoday.

RESULTS — Regarding numerical accuracy, during euglycemia, the mean absolute relative differences (MARDs) of Guardian, DexCom, Navigator, and Glucoday were 15.2, 21.2, 15.3, and 15.6%, respectively. During hypoglycemia, the MARDs were 16.1, 21.5, 10.3, and 17.5%, respectively. Regarding clinical accuracy, continuous glucose–error grid analysis (CG-EGA) revealed 98.9, 98.3, 98.6, and 95.5% zones A + B hits in euglycemia. During hypoglycemia, zones A + B hits were 84.4, 97.0, and 96.2% for Guardian, Navigator, and Glucoday, respectively. Because of frequent loss of sensitivity, there were insufficient hypoglycemic DexCom data to perform CG-EGA.

CONCLUSIONS — The numerical accuracy of Guardian, Navigator, and Glucoday was comparable, with an advantage to the Navigator in hypoglycemia; the numerical errors of the DexCom were ~30% larger. The clinical accuracy of the four sensors was similar in euglycemia and was higher for the Navigator and Glucoday in hypoglycemia.

Diabetes Care 31:1160–1164, 2008

Evaluation of the accuracy of continuous glucose monitors (CGMs) is complex for two primary reasons: 1) CGMs assess blood glucose fluctuations indirectly by measuring the concentration of interstitial glucose but are calibrated via self-monitoring to approximate blood glucose; and 2) CGM data reflect an underlying process in time and therefore consist of ordered-in-time highly interde-

pendent data points. Because CGMs operate in the interstitial compartment, which is presumably related to blood via diffusion across the capillary wall (1,2), there are a number of significant challenges in terms of sensitivity, stability, calibration, and physiological time lag between blood and interstitial glucose concentration (1,3–6). In addition, the temporal structure of CGM data poses sta-

tistical challenges to the direct use of established accuracy measures, such as correlation or regression, or the clinically based error grid analysis (EGA) (7,8), because these measures judge the quality of approximation of reference blood glucose measurements by readings at isolated points in time, without taking into account the temporal structure of the data. In other words, a random reshuffling of the sensor-reference data pairs in time will not change these accuracy estimates. It is therefore imperative to judge the accuracy of CGMs across several dimensions and to use both numerical and clinical metrics to support this judgment.

Defined as the closeness between CGM readings and corresponding in-time reference blood glucose measurements, numerical accuracy is computed by several traditional measures including mean absolute difference (MAD) and mean absolute relative difference (MARD), median absolute difference (MedAD) and median absolute relative difference (MedARD), and ISO (International Standards Organization) criteria. The ISO criteria refer to the percentage of CGM readings within 0.8 mmol/l (15 mg/dl) from reference when the reference blood glucose is ≤ 4.2 mmol/l (75 mg/dl) or within 20% from reference when blood glucose is > 4.2 mmol/l (9). These measures reflect the numerical proximity of CGM-reference data pairs as if these pairs were independent from each other, without taking into account their temporal order and the rate of glucose change. To measure CGM rate accuracy, we have recently suggested the R-deviation, a numerical measure of the proximity between sensor and reference blood glucose rate of change computed similarly to MAD but taking into account the rate of the CGM and reference blood glucose processes, not their values (10).

The premise behind evaluation of clinical accuracy is to assess the impact of sensor errors on treatment decisions based on CGM output. Previously proposed solutions to such an assessment include the Clarke EGA (7) and consensus error grid (11), both of which were designed before the advent of CGMs. We

From the ¹University of Virginia, Charlottesville, Virginia; and the ²Profil Institute for Metabolic Research, Neuss, Germany.

Corresponding author: Boris Kovatchev, PhD, University of Virginia Health System, Box 800137, Charlottesville, VA 22901. E-mail: boris@virginia.edu.

Received for publication 17 December 2007 and accepted in revised form 10 March 2008.

Published ahead of print at <http://care.diabetesjournals.org> on 13 March 2008. DOI: 10.2337/dc07-2401. B.C., S.A., and W.C. have received grant support from Abbott Diabetes Care, Alameda, CA. L.H. has received research support from A. Menarini Diagnostics S.r.l., Florence, Italy.

Abbreviations: CG-EGA, continuous glucose–error grid analysis; CGM, continuous glucose monitor; EGA, error grid analysis; ISO, International Standards Organization; MAD, mean absolute difference; MARD, mean absolute relative difference; MedAD, median absolute difference; MedARD, median absolute relative difference; P-EGA, point–error grid analysis; R-EGA, rate–error grid analysis; UVA, University of Virginia.

© 2008 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

have proposed the continuous glucose–error grid analysis (CG-EGA), which was specifically designed to assess the clinical accuracy of CGMs (12). The CG-EGA has two components: the point–error grid analysis (P-EGA) assessing clinical point accuracy and the rate–error grid analysis (R-EGA) assessing clinical rate accuracy. Both P-EGA and R-EGA preserve the premise of the Clarke EGA, dividing the glucose or glucose rate ranges into clinically meaningful zones: zone A, corresponding to clinically accurate reading; zone B, corresponding to benign errors; zone C, signifying overcorrection errors; zone D, indicating failure to detect clinically significant blood glucose or rate of change; and zone E, indicating an erroneous reading. The difference between the traditional Clarke EGA and P-EGA is in the dynamic adjustment of the error grid zones depending on the rate of change of the reference blood glucose process, which is designed to accommodate a possible time lag between reference and sensor readings. The CG-EGA combines point and rate accuracy separately for each of the three critical blood glucose ranges: hypoglycemia (blood glucose ≤ 3.9 mmol/l), euglycemia, and hyperglycemia (blood glucose > 10 mmol/l) using distinct matrices of point versus rate accuracy. These matrices reflect the relative importance of point versus 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 accuracy uniformly across the entire blood glucose range (12).

In summary, the metrics of CGM accuracy can be classified into a 2×2 (numerical-clinical) \times (point-rate) accuracy table. In this article we use all four components of this table to compare the numerical and clinical performance of four CGMs: Guardian (Medtronic, Northridge, CA), Freestyle Navigator (Abbott Diabetes Care, Alameda, CA), DexCom STS (DexCom, San Diego, CA), and Glucoday (A. Menarini Diagnostics, Florence, Italy). The first three are needle-type sensors providing real-time glucose readings at a frequency of 5–10 min. The Glucoday is a microdialysis device measuring interstitial glucose every 3 min (13,14).

RESEARCH DESIGN AND METHODS

Two clinical trials were performed at the University of Virginia (UVA), Charlottesville, Virginia, and at

the Profil Institute for Metabolic Research, Neuss, Germany. The studies were approved by the review boards of their respective institutions. The UVA study recruited 14 and the Profil study recruited 20 adults with type 1 diabetes. All subjects gave written informed consent and had a physical examination before the beginning of the study protocol, including review of medical history and laboratory tests.

At UVA, subjects were admitted to the General Clinical Research Center in the evening before testing. Three continuous monitoring sensors, Guardian, Freestyle Navigator, and DexCom STS (3-day sensor), were inserted and used simultaneously during the testing. The sensors were calibrated according to the manufacturers' instructions, and their clocks were adjusted to match a master clock in the room, which allowed for further synchronization of the data. In the morning of the study the participants underwent hyperinsulinemic glucose clamps including 1.5–2.0 h of maintained euglycemia at a target level of 5.6 mmol/l followed by gradual (45–60 min) descent into hypoglycemia with a target level of 2.5 mmol/l, sustained hypoglycemia for 30 min, and recovery to normoglycemia. Reference glucose sampling was performed every 5 min using a YSI blood glucose analyzer (YSI, Yellow Springs, OH). The hand and forearm were warmed to provide arterialized venous samples. Reference blood glucose and CGM data were synchronized with a precision of 30 s. The participants in the Profil study arrived at the research institute in the morning. After admission, they were connected to an artificial pancreas (Biostator) and to the subcutaneous minimally invasive glucose sensor Glucoday. The euglycemic and hypoglycemic glucose targets of the Profil trial were identical to these at UVA: during a run-in phase of 120 min the blood glucose concentration of the patients was stabilized by intravenous infusion of insulin and/or glucose solution at 5.6 mmol/l. In this time period the glucose sensors were also calibrated for the first time. Then, hypoglycemia was induced with a target level of 2.5 mmol/l, which was maintained for ~ 30 min. However, the rates of descent into hypoglycemia and recovery were substantially higher than in the UVA study (6.2 vs. 3.4 mmol \cdot $l^{-1} \cdot$ h^{-1}). Reference blood glucose sampling was performed every 5 min and synchronized with the readings of the Glucoday.

Accuracy metrics

Numerical point accuracy was evaluated using MAD, MARD, MedAD, and MeddARD. MAD and MedAD are computed as the average/median of the absolute values of the differences between sensor readings and reference blood glucose values. MARD and MeddARD are the absolute differences expressed as a percentage of the reference blood glucose values. The ISO criteria include the percentage of CGM readings within 0.8 mmol/l (15 mg/dl) from reference when the reference blood glucose is ≤ 4.2 mmol/l (75 mg/dl) or within 20% from reference when blood glucose is > 4.2 mmol/l (9). Numerical rate accuracy was measured by the recently suggested absolute R-deviation (10). The absolute R-deviation was computed similarly to MAD, but taking the first-order divided differences of the CGM and reference blood glucose time series, e.g., as the average of the absolute values of $(\Delta S - \Delta R)/\Delta t$, where ΔS and ΔR are the differentials of sensor and reference blood glucose over a time period of Δt . Clinical point and rate accuracy were computed using the two components, P-EGA and R-EGA, of the previously introduced CG-EGA (12).

RESULTS

Overall sensor reliability

During the UVA study all three CGM sensors experienced periods of transient loss of sensitivity, particularly during hypoglycemia, identified as sensor readings holding steady at a very low glucose value (e.g., 2.1 mmol/l), whereas blood glucose was higher and fluctuating. The percentage of such unreliable data points was 6.9% for the Guardian, 29.8% for the DexCom, and 16.8% for the Navigator. These unreliable data were not considered in the accuracy analysis of the sensors presented in the following sections. There were no missing data in the study of Glucoday.

Numerical point and rate accuracy

Table 1 presents the numerical point accuracy of the four CGMs during maintained euglycemia and induced hypoglycemia and their ability to follow the trend down into induced hypoglycemia and up to recovery (rate accuracy using absolute R-deviation). To account for nonsymmetric distributions, we present both mean and median absolute and relative differences, which lead to analogous conclusions: During euglycemia, the

Table 1—Numerical accuracy

	Guardian	DexCom	Navigator	Glucoday
Point accuracy: euglycemia (blood glucose = 3.9–10 mmol/l)				
MAD (mmol/l)	0.91	1.24	0.89	0.87
MARD	15.2	21.2	15.3	15.6
MedAD (mmol/l)	0.82	1.06	0.85	0.60
MedARD	13.3	18.4	11.8	10.7
ISO: % readings within 20% from reference when reference >4.2 mmol/l	73.2	52.2	72.2	76.9
MAD (mmol/l)	0.55	0.73	0.36	0.47
MARD	16.1	21.5	10.3	17.5
MedAD (mmol/l)	0.42	0.64	0.24	0.40
MedARD	13.8	22.5	7.4	15.6
ISO: % readings within 0.8 mmol/l from reference when reference ≤4.2 mmol/l	76.5	52.9	79.4	83.0
Rate accuracy: descent into hypoglycemia Absolute R-deviation (mmol · l ⁻¹ · h ⁻¹)	2.9	2.4	2.2	5.8
Rate accuracy: ascent from hypoglycemia Absolute R-deviation (mmol · l ⁻¹ · h ⁻¹)	3.0	3.3	3.3	9.3

Data are percent unless otherwise indicated.

MARD and MedARD of Guardian, Navigator, and Glucoday were similar. During hypoglycemia, the mean and median differences of the Navigator were lower than those of Guardian and Glucoday. The DexCom registered ~30% larger errors during both euglycemia and hypoglycemia. The numerical rate errors of Guardian, Navigator, and DexCom were comparable. Glucoday had higher rate errors, which could be explained by the higher overall rates of blood glucose change achieved in the Profil study compared with the UVA study (6.2 vs. 3.4 mmol · l⁻¹ · h⁻¹ on average).

Because sequential CGM data points are highly interdependent, standard statistical analyses would produce inaccurate results. However, a previously reported 1-h block-aggregation of the data produces composite readings that are suitable for statistical analyses (15). Thus, to apply statistical tests, we aggregate the data beginning at time 0 in sequential 1-h blocks. Then we use ANOVA with contrasts to compare the MAD of each pair of sensors. The three significant contrasts observed were for Guardian versus DexCom ($F = 104.9$, $P < 0.001$), Navigator versus DexCom ($F = 55.1$, $P < 0.001$), and Glucoday versus DexCom ($F = 65.2$, $P < 0.001$). The contrasts between all other pairs of sensors were not significant.

Clinical point and rate accuracy

Table 2 presents the clinical point and

rate accuracy of the four CGMs using the CG-EGA and its two components, P-EGA and R-EGA. The percentages of readings of the four sensors in zones A + B of the CG-EGA were similar during euglycemia. However, this overall similar clinical accuracy was achieved by different means as revealed by separate P- and R-EGA: the P-EGA showed highest zone A score for the Glucoday, whereas the R-EGA showed lower rate accuracy of Glucoday compared with the other three sensors. During hypoglycemia Navigator and Glucoday had the highest CG-EGA accuracy

scores. As mentioned in the previous section there were insufficient DexCom data to perform the analysis.

For statistical analysis of clinical accuracy we face the problem of dependence of adjacent CGM points, which may cause inaccurate interpretation of the P level. Thus, we use nonparametric comparisons and a normal approximation of the resulting statistics, which is less vulnerable to data dependence (i.e., does not use degrees of freedom). The significant P-EGA differences observed were for Guardian versus DexCom ($Z = 7.0$, $P < 0.001$),

Table 2—Clinical accuracy

Zone	Guardian	DexCom	Navigator	Glucoday
P-EGA				
A	76.3	55.4	76.3	88.5
A + B	97.5	100	99.7	98.7
Errors (C, D, E)	2.5	0	0.3	1.3
R-EGA				
A	73.4	70.7	74.4	64.1
A + B	92.1	91.7	94.3	86.1
Errors (C, D, E)	7.9	8.3	5.7	13.9
CG-EGA: hypoglycemia				
A	81.9	—*	95.5	87.3
A + B	84.4	—	97.0	96.2
Errors (C, D, E)	15.6	—	3.0	3.8
CG-EGA: euglycemia				
A	91.3	91.1	93.7	84.4
A + B	98.9	98.3	98.6	95.5
Errors (C, D, E)	1.1	1.7	1.4	4.5

Data are %. *There were insufficient Dexcom data to perform CG-EGA during hypoglycemia.

Navigator versus DexCom ($Z = 5.0, P < 0.001$), and Glucoday versus DexCom ($Z = 8.2, P < 0.001$), which is consistent with the numerical results from the previous section. In addition, the contrast between the Navigator and Guardian CG-EGA results during hypoglycemia was significant ($Z = 2.7, P = 0.007$).

CONCLUSIONS—CGMs provide detailed time series of consecutive observations upon the underlying process of glucose fluctuations. Because CGMs are able to track these fluctuations, time-dependent measures of numerical and clinical accuracy must be considered in addition to traditional accuracy assessment methods that reflect only the static proximity between CGM and reference blood glucose values. Knowing solely the accuracy of CGM point approximation of the process of glucose fluctuation is insufficient. It is also important to evaluate how closely the CGM follows the rate and direction of blood glucose change, i.e., its trend or rate accuracy. Rate accuracy is particularly important when CGM data are used for prediction of acute glycemic events such as hypoglycemia, for hypo-/hyperglycemia alarms, or in algorithms for closed-loop control. Mathematically, numerical rate accuracy is assessed by the closeness between the first derivatives of the process of blood glucose fluctuation and its CGM representation, a property that is reflected by the recently introduced R-deviation (10). However, the R-deviation is only the first step in evaluation of the dynamics of glucose fluctuations. Higher-order dynamic properties and long-term trends may provide additional valuable information about sensor performance.

There are two general approaches to measuring proximity between time series, e.g., temporal performance of CGM. The first is purely numerical, relying on mathematical “distances” between the true blood glucose values and trends and their estimates. The second approach is clinical, the device is judged by the clinical accuracy of the clinical message it sends. We suggest that CGMs be evaluated using the entire array of numerical and clinical metrics of point and rate accuracy because such a multidimensional assessment would reveal a more comprehensive picture of sensor performance.

In this article we present a comparison of the accuracy of four sensors currently manufactured in the U.S. and in Europe: Guardian, Freestyle Navigator,

DexCom, and Glucoday. The data for the comparison of these devices were collected in two clinical trials. The first trial at the UVA tested Guardian, Navigator, and DexCom simultaneously. To the best of our knowledge, this is the first study to assess the accuracy of three devices worn by the participants at the same time. The data collected in Germany at the Profil Institute for Metabolic Research assessed the accuracy of Glucoday. Because the two studies had similar design and glycemic goals, a comparison of the results was possible.

In terms of numerical metrics, the accuracy of Guardian, Navigator, and Glucoday was comparable, with advantage to the Navigator in hypoglycemia, whereas the numerical errors of the DexCom were ~30% larger. Specific data aggregation allowed these conclusions to be supported by statistical tests. The comparison of clinical accuracy using CG-EGA showed comparable results for all sensors during euglycemia but substantially higher accuracy of the Navigator and the Glucoday during hypoglycemia. Here we have to acknowledge that the DexCom device tested in this study used the older model 3-day sensor. Recent data collected with the new 7-day DexCom sensor showed improved accuracy of the device, namely MARD of 15.8% and MedARD of 13.3% (16).

One limitation to the presented comparisons was the higher rate of glucose change induced in the Profil study, which may be the reason for poorer rate accuracy of the Glucoday compared with the other sensors. The higher rate of change, however, did not affect the point accuracy of the Glucoday, leading to overall comparable clinical performance. Thus, similar overall clinical accuracy can be achieved by different routes and only a detailed point and trend (rate) accuracy analysis can reveal its specific components. We should also note that reference blood glucose during the studies was measured using venous samples, which would differ from the capillary samples used for sensor calibration. Because of this difference and the induced high rates of glucose change, the sensor errors observed during these clamp studies may be larger than the errors that would be observed in everyday use.

In summary, the numerical accuracy of Guardian, Navigator, and Glucoday was comparable, with advantage to the Navigator in hypoglycemia; the numerical errors of the 3-day DexCom sensor were ~30% larger. The clinical accuracy

of the four sensors reflected by the CG-EGA was similar during euglycemia and was higher for the Navigator and Glucoday during hypoglycemia.

Acknowledgments—The UVA study was supported by National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases Grant R01 DK 51562, by the UVA General Clinical Research Center, and by material support from Abbott Diabetes Care (Alameda, CA). The Profil study was supported by a grant from A. Menarini Diagnostics (Florence, Italy).

References

1. Boyne M, Silver D, Kaplan J, Saudek C: Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor. *Diabetes* 52:2790–2794, 2003
2. Steil GM, Rebrin K, Hariri F, Jinagonda S, Tados S, Darwin C, Saad MF: Interstitial fluid glucose dynamics during insulin-induced hypoglycaemia. *Diabetologia* 48: 1833–1840, 2005
3. Cheyne EH, Cavan DA, Kerr D: Performance of continuous glucose monitoring system during controlled hypoglycemia in healthy volunteers. *Diabetes Technol Ther* 4:607–613, 2002
4. Stout PJ, Racchini JR, Hilgers ME: A novel approach to mitigating the physiological lag between blood and interstitial fluid glucose measurements. *Diabetes Technol Ther* 6:635–644, 2004
5. Kulcu E, Tamada JA, Reach G, Potts RO, Lesho MJ: Physiological differences between interstitial glucose and blood glucose measured in human subjects. *Diabetes Care* 26:2405–2409, 2003
6. King CR, Anderson SM, Breton MD, Clarke WL, Kovatchev BP: Modeling of calibration effectiveness and blood-to-interstitial glucose dynamics as potential confounders of the accuracy of continuous glucose sensors during hyperinsulinemic clamp. *J Diabetes Sci Technol* 1:317–322, 2007
7. 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
8. Chen E, Nichols J, Duh S, Hortin G: Performance evaluation of blood glucose monitoring devices. *Diabetes Technol Ther* 5:749–768, 2003
9. 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 publ. no. 15197)
10. Clarke WL, Kovatchev BP: Continuous

- glucose sensors—continuing questions about clinical accuracy. *J Diabetes Sci Technol* 1:164–170, 2007
11. 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
 12. Kovatchev BP, Gonder-Frederick LA, Cox DJ, Clarke WL: Evaluating the accuracy of continuous glucose monitoring sensors: continuous glucose error-grid analysis (CG-EGA) illustrated by Therasense Free-style Navigator data. *Diabetes Care* 27:1922–1928, 2004
 13. Maran A, Crepaldi C, Tiengo A, Grassi G, Vitali E, Pagano G, Bistoni S, Calabrese G, Santeusano F, Leonetti F, Ribaud M, Di Mario U, Annuzzi G, Genovese S, Riccardi G, Previti M, Cucinotta D, Giorgino F, Bellomo A, Giorgino R, Poscia A, Varalli M: Continuous subcutaneous glucose monitoring in diabetic patients: a multicenter analysis. *Diabetes Care* 25:347–352, 2002
 14. Ricci F, Caprio F, Poscia A, Valgimigli F, Messeri D, Lepori E, Dall'Oglio G, Palle-schi G, Moscone D: Toward continuous glucose monitoring with planar modified biosensors and microdialysis: study of temperature, oxygen dependence and in vivo experiment. *Biosens Bioelectron* 22:2032–2039, 2007
 15. Kovatchev BP, Clarke WL, Breton M, Brayman K, McCall A: Quantifying temporal glucose variability in diabetes via continuous glucose monitoring: mathematical methods and clinical application. *Diabetes Technol Ther* 7:849–862, 2005
 16. Zisser H, Shwartz S, Rather R, Wise J, Bailey T: Accuracy of a seven-day continuous glucose sensor compared to YSI blood glucose values. In *Proceedings of the 27th Workshop of the AIDPIT Study Group, Innsbruck, Austria, 2008*. AIDPIT p. S03. Available from <http://www.aidpit.org>. Accessed 27 January 2008