The Accuracy Benefit of Multiple Amperometric Glucose Sensors in People With Type 1 Diabetes

Jessica R. Castle, md¹ Amy Pitts, rn¹ Kathryn Hanavan, anp-c¹ Rhonda Muhly, bs¹ Joseph El Youssef, mbbs¹ Colleen Hughes-Karvetski, phd² Boris Kovatchev, phd² W. Kenneth Ward, md¹

OBJECTIVE—To improve glucose sensor accuracy in subjects with type 1 diabetes by using multiple sensors and to assess whether the benefit of redundancy is affected by intersensor distance.

RESEARCH DESIGN AND METHODS—Nineteen adults with type 1 diabetes wore four Dexcom SEVEN PLUS subcutaneous glucose sensors during two 9-h studies. One pair of sensors was worn on each side of the abdomen, with each sensor pair placed at a predetermined distance apart and 20 cm away from the opposite pair. Arterialized venous blood glucose levels were measured every 15 min, and sensor glucose values were recorded every 5 min. Sensors were calibrated once at the beginning of the study.

RESULTS—The use of four sensors significantly reduced very large errors compared with one sensor (0.4 vs. 2.6% of errors \geq 50% from reference glucose, *P* < 0.001) and also improved overall accuracy (mean absolute relative difference, 11.6 vs. 14.8%, *P* < 0.001). Using only two sensors also significantly improved very large errors and accuracy. Intersensor distance did not affect the function of sensor pairs.

CONCLUSIONS—Sensor accuracy is significantly improved with the use of multiple sensors compared with the use of a single sensor. The benefit of redundancy is present even when sensors are positioned very closely together (7 mm). These findings are relevant to the design of an artificial pancreas device.

Diabetes Care 35:706-710, 2012

The development of amperometric glucose sensors has advanced the treatment of type 1 diabetes. Glucose sensors are now commercially available as compact, minimally invasive devices that measure interstitial glucose levels in subcutaneous tissue. Recent studies have shown that the use of glucose sensors improves glucose control in children and adults with type 1 diabetes when used alone (1) or with insulin pump therapy (2) and reduces hypoglycemia (3).

Glucose sensor technology has spurred research efforts into methods of automated glycemic management. The basics of an artificial pancreas system, more accurately known as a closed-loop system, consist of a glucose sensor, a mathematic algorithm, and an insulin-delivery device. To the extent that it is accurate, a glucose sensor that serves as the input for the insulin-delivery controller is capable of minimizing the frequency of hypoglycemia and hyperglycemia.

The accuracy of present day sensors is generally good, but remains imperfect. For this reason, the U.S. Food and Drug Administration does not allow sensor data to be used as a replacement for blood glucose values. Numerous factors may adversely affect sensor accuracy, including calibration error, sensor delay, and sensor drift (4). Glucose sensors require the periodic

From the ¹Department of Medicine, Division of Endocrinology, Diabetes & Clinical Nutrition, Oregon Health & Science University, Portland, Oregon; and the ²Department of Psychiatry and Neurobehavioral Sciences and the Department of Systems and Information Engineering, Center for Diabetes Technology, University of Virginia, Charlottesville, Virginia.

input of a blood glucose value. This calibration procedure allows for the electrical current detected by a sensor to be equated into a sensed glucose level. An inaccurate blood glucose value, due to operator error or an inaccurate glucose meter, will cause a calibration error (5). Sensor delay relates to the delay between changes in glucose levels in the blood and the interstitial fluid. Delay may also be imparted by the algorithms used to smooth sensor data (6). Sensor drift is not as well understood and is likely due to a host of factors, among them the foreign body response that attracts leukocytes, including macrophages, that consume glucose and oxygen and produce peroxide and thus interfere with the accurate measurement of glucose (7).

The use of redundant sensors should not improve sensor inaccuracy caused by calibration error or sensor delay. Redundancy may, however, reduce error caused by sensor drift or sensor signal dropout, because sensor signals may inappropriately drift above or below the reference blood glucose, or the sensor signal may dropout to generate an inaccurately low glucose signal. Whether placing sensors very near one another will reduce the benefit of redundancy is not known. Entrainment of sensors placed closely together is plausible. Sensors that are close together may be exposed to the same microenvironment, causing them to drift in the same direction and to a similar degree.

In this study of adults with type 1 diabetes, we compared the accuracy of multiple sensors worn simultaneously with the accuracy of a single sensor. By evaluating sensor pairs with different intersensor distances, we also evaluated whether short distances decreased any benefit of redundancy.

RESEARCH DESIGN AND

METHODS—Patients with type 1 diabetes were recruited from Oregon Health & Sciences University (OHSU) outpatient clinics in Portland. Patients who were pregnant, had uncontrolled concurrent illnesses, had physical or visual impairment preventing the issue of a continuous glucose monitoring system, or needed uninterrupted acetaminophen use were excluded. The

Corresponding author: Jessica R. Castle, castleje@ohsu.edu.

Received 4 October 2011 and accepted 11 December 2011.

DOI: 10.2337/dc11-1929

^{© 2012} by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See http://creativecommons.org/ licenses/by-nc-nd/3.0/ for details.

Castle and Associates

research protocol was approved by the OHSU institutional review board, and all subjects provided written informed consent. This study was conducted according to the principles expressed in the Declaration of Helsinki.

A total of 36 studies in 19 subjects were performed. The subjects were a mean age of 37.4 ± 3.3 years, with a duration of diabetes of 22.4 ± 2.9 years. Mean HbA_{1c} was $7.9 \pm 0.2\%$, and mean BMI was 25.9 ± 1.0 kg/m².

Study procedures

Subjects wore four Dexcom SEVEN PLUS glucose sensors for each of two studies. Sensors were inserted 8-24 h before the study began. Two sensors were placed to the right of the umbilicus and two to the left. The pair on the right was positioned 20 cm apart from the pair on the left. Each pair was placed at a predetermined distance on the surface of the abdominal skin, with the goal of achieving a fixed distance between the tips of the glucose sensors in the subcutaneous tissue. The study was designed so that the intersensor distances in the subcutaneous tissue would be approximately 2, 10, 20, and 30 mm. The order of distances was randomized using a random number generator. Subjects wore the four sensor receivers in a small pack to keep them within 5 feet of the transmitters and were instructed to calibrate the sensors 2 h and 8 h after insertion.

The following morning, subjects were admitted to the Oregon Clinical & Translational Research Institute at OHSU. An intravenous catheter was placed in a forearm vein. The forearm was warmed with a heating pad to arterialize the venous blood. At the beginning of the 9-h study, the mean of two simultaneous measured venous blood glucose values was used to calibrate the four sensors. Venous glucose was subsequently measured every 15 min for 9 h using a HemoCue Glucose 201 Analyzer. Sensor glucose readings were recorded from the receivers every 5 min.

Subjects were fed two standardized meals and given premeal insulin based on their typical insulin-to-carbohydrate ratio along with additional insulin for correction of hyperglycemia according to their usual outpatient regimen. Subjects were treated with fruit juice for hypoglycemia if the venous glucose value fell below 70 mg/dL. At the end of each study, coneddown oblique X-ray images of the sensor sites were taken to measure the actual distance between the tips of the glucose sensors. Magnification was taken into account by multiplying the distance between the X-ray source to the sensor divided by distance between the X-ray source and the X-ray plate.

Statistical analysis

Data are expressed as mean \pm SEM or mean (95% CI). Sensor accuracy was calculated by comparing sensor glucose with reference glucose values (8). Data were analyzed using generalized estimating equations, which took into account correlated data and repeated measures. The measured distance between the sensor tips was used in place of the expected distance. Calculations were performed using STATA 10.1 software.

Also presented is a continuous glucoseerror grid analysis (CG-EGA), an outcome metric that addresses the problem of evaluating the temporal characteristics of the continuous glucose sensor process by considering pairs of reference and sensor readings as a process in time that takes into account inherent physiologic time lags (9). The CG-EGA classifies errors in three clinically relevant categories—accurate readings (region A), benign errors (region B), and erroneous readings (regions C, D, and E)—with clinically relevant classification cutoffs based on the potential error in treatment given the sensor inaccuracy.

Principal component analysis (PCA) is a multivariate data analysis method that was used here to flag and remove potentially aberrant sensor values to assess if removing these values would improve overall sensor accuracy (10,11). In the case of two sensors, the methodology can be used to detect a potential inconsistency between the sensor signals. When three or more sensor measurements are available, PCA can be used to detect inconsistencies and identify the inaccurate or faulty sensor from the group. In practice, our data set at each time step consisted of three or more variables (sensor sources), each with a set of observations, the number of which depended on the frequency of available sensor measurements and the predetermined length of the measurement history, set to 15 and 75 min, respectively, in this case.

Performing PCA on these data reveals a new set of uncorrelated variables called principle components (PCs), with weights assigned to each of the original sensor sources to indicate the contribution of each original sensor source to the PC. The most influential sensor signal carries the highest weighting coefficient in the first PC, whereas a sensor signal that is considered an outlier will carry the highest weighting coefficient in the last PC. This information, in conjunction with the percent of variability of the original data set that is captured by the first PC, was used retrospectively to identify an inaccurate or faulty sensor; in practice, the method is designed to be used in realtime. The sensor chosen as the inaccurate sensor updates at each time step. In the case of missing data, only the values available are used. For example, if two of three sensors are available, then the two-sensor implementation of the PCA is used.

RESULTS—Eleven women and 8 men with type 1 diabetes participated in 36 studies, each lasting 9 h. The data from one study were excluded due to a diluted venous blood glucose value that caused incorrect calibration of the sensors, leaving 35 studies for inclusion in the data analysis.

Benefit of four sensors

The benefit of multiple sensors was analyzed by comparing how often the mean and median of all four sensor values deviated 25% or more and 50% or more from the reference blood glucose compared with the values of each single sensor. Errors of these degrees may cause inappropriate treatment decisions, and in particular, errors of 50% or higher might cause harm if these sensor values were used as an input into an artificial pancreas algorithm.

Large errors, defined as the sensor value being 25% above or below the reference venous blood glucose values, were significantly reduced by using the mean values (9.5% of values [95% CI 5.8-15.4]) or median values (10.0% [6.2–16.1]) from four sensors compared with one sensor (17.5% [13.3-23.0], P < 0.001). Very large errors, defined as the sensor value being 50% or more above or below the reference blood glucose value, were also significantly reduced by using the mean (0.4% [0.1–1.2]) or median (0.5% [0.2– 1.2]) of four sensor values compared with one sensor (2.6% [1.6-4.0], P < 0.001). In many cases, there were no large errors. However, in those experiments with a substantial number of very large errors, which reached as high as 61% of values, use of the mean of four sensors dramatically reduced the frequency of very large errors (Fig. 1).

The mean absolute relative difference (ARD) was significantly improved by use of the mean (11.6 \pm 1.0%) or median (11.8 \pm 1.0%) of four sensors compared with one sensor (14.8 \pm 1.0%, *P* < 0.001). There was no advantage to selecting the sensor with the lowest ARD at the time of



Figure 1—Summary of very large errors, defined as sensor values \geq 50% away from the reference venous blood glucose. Each study visit is depicted separately. The percentages of very large errors when the four sensors are averaged are shown by black open circles, and values for each single sensor are shown by gray Xs. Note the significant decrease in very large errors with the use of four sensors.

calibration at the study start (mean ARD 15.0 \pm 1.4%, *P* = NS). The mean absolute difference, the metric for reference venous blood glucose values <75 mg/dL, also improved significantly by use of the mean of four sensors compared with a single sensor (mean absolute difference 9.6 \pm 1.0 vs. 14.1 \pm 1.3 mg/dL, *P* < 0.001).

Benefit of two sensors

Large errors, as defined above, were also significantly reduced by using the mean of the sensor pairs (12.9% [95% CI 9.9–19.2]) compared with a single sensor (17.5% [13.3–23.0], P < 0.01). Very large errors were reduced by a mean of 62% by the use of sensor pair means versus one sensor (1.0% [0.5–1.9] vs. 2.6% [1.6–4.0], P < 0.001).

There was a reduction in the mean ARD when comparing the mean of the sensor pairs with a single sensor (13.5 \pm 1.1 vs. 14.8 \pm 1.0%, *P* < 0.01). There was a trend in improvement in sensor accuracy with using the mean of three sensors compared with two sensors (0.9 percentage point ARD reduction, *P* < 0.06), with a clear added benefit of four (1.6 percentage point ARD reduction, *P* < 0.01).

A summary of errors for one, two, and four sensors, categorized by glucose levels, is presented in Table 1.

CG-EGA

In the euglycemic range (70–180 mg/dL), the percentage of erroneous readings

decreased significantly with additional sensor signal information (1.1, 0.8, and 0.5% of readings in the CG-error grids C, D, and E zones for the mean of 1, 2, and 4 sensors, respectively). The finding was similar in the hyperglycemic range above 180 mg/dL (3.5, 2.8, and 1.7% in C, D, and E zones for a mean of 1, 2, and 4 sensors). The number of accurate readings also significantly increased in the hyperglycemic range with increasing number of sensors (90.2, 91.7, and 94.7%). There were no significant differences in readings in the A region in the euglycemic range. There were also no significant differences in erroneous or accurate readings in the limited number of values in the hypoglycemic range.

Clarke Error Grid

There was a significant increase in the A zone of the Clarke Error Grid using multiple sensors (68.4, 70.7, and 75.4% for 1, 2, and 4 sensors, respectively). There were more values in the B zone with use of a single sensor (23.6, 18.6, and 11.5%). There was nonsignificant trend toward reducing the percentage of the values in the C, D, and E zones (2.4, 1.3, and 0.4%).

Voting scheme using PCA

Voting schemes may improve upon using sensor value averaging alone, because if one sensor is highly accurate and the second is highly inaccurate, the mean is worse than use of the single highly accurate sensor. There was a significant reduction in the mean ARD of the mean of three sensors in the case when data flagged by the PCA was removed compared with all data (mean ARD 12.0 \pm 1.1 vs. $12.3 \pm 1.1\%$, P = 0.04). As expected, the mean ARD of the flagged data (16.5 \pm 2.4%) was higher compared with nonflagged data alone (P = 0.02) and also compared with all data (P = 0.01). In the case of two sensors, removing values flagged by PCA did not significantly change the mean ARD versus all data (12.9 \pm 0.9 vs. $13.0 \pm 0.9\%$, *P* = NS), and there was a nonsignificant increase in the mean ARD of the flagged data (15.5 \pm 1.5, P = 0.06 compared with all data). Similar results for the PCA methodology were obtained when four sensors were used.

Intersensor distances for sensor pairs

The intersensor distances as measured by X-ray imaging were 7 ± 1 , 13 ± 1 , 21 ± 1 , and 28 ± 2 mm (Fig. 2).

The effect of intersensor distance

If positioning sensor pairs closely together caused entrainment, then values from both sensors would drift in the same direction at similar rates of change. In such a case, one would expect a correlation between the intersensor distance and the difference between the signed differences (bias) of each sensor in the pair compared with the reference blood glucose value. In other words, if sensors positioned very closely to one another were entrained, they would have similar readings and minimal differences in bias values. However, no such correlation was found between the intersensor distance and the difference between the signed differences ($r^2 = 0.004$, Fig. 3). Furthermore, there was no significant relationship between distance between sensor pairs and sensor accuracy measured by mean ARD (P = NS).

CONCLUSIONS—The concept of redundancy is well established and is commonly used when errors may lead to unacceptable consequences, such as on NASA spacecraft (12). The use of two or more sensors in a closed-loop system is appealing for multiple reasons. First, the presence of redundant sensors provides a reserve for instances of sensor or telemetry failure. Currently available sensors need time for signal stabilization, so if the only

	<70 mg/dL			70–180 mg/dL			>180 mg/dL		
Error (%)	(n = 156)	(<i>n</i> = 78)	(<i>n</i> = 39)	(n = 2,938)	(n = 1,478)	(n = 739)	(<i>n</i> = 2,000)	(<i>n</i> = 1,012)	(n = 506)
	Single	Average of 2	Average of 4	Single	Average of 2	Average of 4	Single	Average of 2	Average of 4
<10	26.9	41.0	38.5	41.9	46.5	50.6	46.6	48.3	54.0
10 to <25	38.5	33.3	38.5	38.0	37.9	37.6	39.3	42.7	39.3
25 to 50	29.5	21.8	17.9	17.1	14.3	11.2	12.6	8.5	6.5
>50	5.1	3.8	5.1	3.1	1.3	0.5	1.6	0.5	0.2

Table 1—Summary of sensor errors categorized by degree of error and venous blood glucose levels

Data are shown as percentages. Note the increase in smaller errors and decrease in larger errors with the use of two or four sensors when the reference blood glucose, listed in the top row, is 70–180 and >180 mg/dL. The number of sensor values, listed in parentheses, indicates there were few values in the <70 mg/dL range.

sensor in place fails, hours elapse before the newly placed sensor is ready for use.

In addition to signal averaging, there are several other potential ways of using data from more than one sensor. Another option is to compare the two sensor values and disregard sensor data when the two values are discrepant beyond a specified criterion (13). In an artificial pancreas setting, this method becomes problematic when no sensor readings are available for anything but a very short time period. Voting schemes can be used when three or more sensors are worn simultaneously. Sensor signals that are quite similar to others in the array are usually more accurate than outliers, and the values that are discrepant can be voted out. Although wearing three or more sensors is impractical with current technology, sensor arrays with multiple sensing units contained in one device may be available in the future. Here we demonstrated that a voting scheme based on PCA is effective, when using three sensors, in detecting potential sensor errors and providing an alert

flag indicating that the sensor mean may be an inaccurate estimate of the patient's blood glucose level.

The PCA methodology can be used in various settings. In conjunction with a multisensor filtering scheme, an inconsistent sensor measurement detected through PCA is removed from the computation of the filtered sensor output. In another setting, the PCA methodology may simply provide an alert to the user that a finger stick measurement is recommended before treatment and to provide a calibration point to resolve inconsistencies in sensor signals. This voting scheme did not significantly improve upon the mean of two sensors, likely due to insufficient data to determine which values to flag. Four sensors provide sufficient data, but PCA may not have improved upon averaging because there are enough values that the mean is not as greatly affected by an erroneous signal as when there are only three sensors.

The 75-min measurement history for the PCA methodology was determined



Figure 2—An example of two X-ray images taken during one of the studies. Note that the sensors are positioned very closely on the right side of the subject's abdomen and are much farther apart on the left side.

retrospectively and needs to be validated by application of the method prospectively. Other methods to improve sensor accuracy by our group have included mathematical correction of background current, which is the current detected by a sensor in the absence of glucose (14). Others have proposed using models to predict glucose levels, which could assist in identifying sensor errors when the sensor values stray greatly from the predicted glucose (15). Enhanced calibration techniques may also improve accuracy (16).

One limitation of this study was that sensors could not be placed closer than 7 mm because of the size of the housing around the sensors. It is unknown whether placement of sensors closer than 7 mm apart would have decreased the benefit of redundancy. Sensors are not approved to be worn in magnetic resonance imaging machines because of potential safety concerns, including heating of the surrounding tissue and possible migration of the sensor, and therefore were imaged by X-ray.

This study was done only with Dexcom SEVEN PLUS sensors, and so the applicability to other types of sensors is unknown. Medtronic Guardian REAL-Time Glucose sensors are not readily apparent on X-ray images and thus were not used. FreeStyle Navigator sensors were not available for purchase in our area when the study was conducted.

We conclude that when four sensors are used simultaneously, there is an accuracy benefit compared with use of one sensor. There is also a benefit of using the mean of three or even two sensors. The benefit of redundancy is present even when sensors are positioned very close together, as close as 7 mm. These findings will be useful in the design of a small, integrated artificial pancreas device and suggest that sensors in such a device can be positioned very closely to one another.

Glucose sensor redundancy



Figure 3—Graph of intersensor distance versus the difference between the signed differences of each sensor in a pair compared with the reference blood glucose. If close proximity caused sensor entrainment, one would expect the value on the y axis to be low. Note that there is no significant correlation between intersensor distance and the difference between the signed differences of each sensor in a pair.

Acknowledgments—This work was supported by grants from The Leona M. and Harry B. Helmsley Charitable Trust, HEDCO Foundation, and by the Juvenile Diabetes Research Foundation (JDRF) Artificial Pancreas Project. B.K. received material support for studies from Insulet Corporation, Dexcom, and Animas Corporation and advisory board/consulting for Dexcom and Animas Corporation. No other potential conflicts of interest relevant to this article were reported.

J.R.C. designed the study, participated in the human studies, and wrote the manuscript. A.P., K.H., and R.M. participated in the human studies. J.E.Y. participated in the human studies and assisted with data analysis. C.H.-K. assisted with data analysis and reviewed and edited the manuscript. B.K. reviewed and edited the manuscript. W.K.W. designed the study, participated in the human studies, analyzed data, and reviewed and edited the manuscript. J.R.C. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

This study was presented at the 71st Scientific Sessions of the American Diabetes Association, San Diego, California, 24–28 June 2011.

The authors thank the staff and research subjects who carried out these studies at the Oregon Clinical and Translational Research Institute (OCTRI), which is supported by grant number UL1 RR024140 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. The authors also thank Michael Lasarev, MS, of OHSU for his statistical assistance, and Matthew Breen of OHSU for his technical support.

References

- 1. Tamborlane WV, Beck RW, Bode BW, et al.; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Continuous glucose monitoring and intensive treatment of type 1 diabetes. N Engl J Med 2008;359:1464– 1476
- 2. Bergenstal RM, Tamborlane WV, Ahmann A, et al.; STAR 3 Study Group. Effectiveness of sensor-augmented insulin-pump therapy in type 1 diabetes. N Engl J Med 2010;363:311–320
- Battelino T, Phillip M, Bratina N, Nimri R, Oskarsson P, Bolinder J. Effect of continuous glucose monitoring on hypoglycemia in type 1 diabetes. Diabetes Care 2011;34:795–800
- 4. Castle JR, Ward WK. Amperometric glucose sensors: sources of error and potential

benefit of redundancy. J Diabetes Sci Tech 2010;4:221–225

- Brauker J. Continuous glucose sensing: future technology developments. Diabetes Technol Ther 2009;11(Suppl. 1):S25–S36
- 6. Wentholt IM, Hart AA, Hoekstra JB, Devries JH. Relationship between interstitial and blood glucose in type 1 diabetes patients: delay and the push-pull phenomenon revisited. Diabetes Technol Ther 2007; 9:169–175
- Wisniewski N, Moussy F, Reichert WM. Characterization of implantable biosensor membrane biofouling. Fresenius J Anal Chem 2000;366:611–621
- Clinical and Laboratory Standards Institute. Performance metrics for continuous interstitial glucose monitoring; approved guideline. CLSI document POCT05-A 2008; 28:1-55
- Kovatchev BP, Gonder-Frederick LA, Cox DJ, Clarke WL. Evaluating the accuracy of continuous glucose-monitoring sensors: continuous glucose-error grid analysis illustrated by TheraSense Freestyle Navigator data. Diabetes Care 2004;27:1922–1928
- Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol 1933;24:417–441, 498–520
- Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag 1901;6:559–572
- Paté-Cornell ME, Dillon RL, Guikema SD. On the limitations of redundancies in the improvement of system reliability. Risk Anal 2004;24:1423–1436
- Schmidtke DW, Pishko MV, Quinn CP, Heller A. Statistics for critical clinical decision making based on readings of pairs of implanted sensors. Anal Chem 1996; 68:2845–2849
- 14. El Youssef J, Castle JR, Engle JM, Massoud RG, Ward WK. Continuous glucose monitoring in subjects with type 1 diabetes: improvement in accuracy by correcting for background current. Diabetes Technol Ther 2010;12:921–928
- 15. Gani A, Gribok AV, Lu Y, Ward WK, Vigersky RA, Reifman J. Universal glucose models for predicting subcutaneous glucose concentration in humans. IEEE Trans Inf Technol Biomed 2010;14:157–165
- Facchinetti A, Sparacino G, Cobelli C. Enhanced accuracy of continuous glucose monitoring by online extended kalman filtering. Diabetes Technol Ther 2010;12: 353–363