

Reproducibility of Glucose Measurements Using the Glucose Sensor

MURIEL METZGER, MD¹
GIL LEIBOWITZ, MD¹
JULIO WAINSTEIN, MD²

BENJAMIN GLASER, MD¹
ITAMAR RAZ, MD¹

OBJECTIVE— Recent studies have confirmed that improved glycemic control decreases the risk of diabetic complications in type 1 and type 2 diabetic patients. The Minimed glucose sensor allows continuous 72-h glucose monitoring and represents a potentially important tool to improve diabetes management. Its use is currently limited to the health care team. Our aim was to evaluate the reproducibility of data provided by the device by comparing data provided by two sensors worn simultaneously by the same subject.

RESEARCH DESIGN AND METHODS— A total of 11 subjects (6 type 1 and 3 type 2 diabetic patients and 2 healthy subjects) agreed to wear two sensors and perform at least four daily finger-stick glucose determinations during 72 h. The simultaneous glucose values provided by the sensors were compared. To determine the clinical implications of the glucose data, each day was divided into eight periods, and for each period the glucose range was rated as satisfactory, too high, or too low by a blinded clinician experienced in interpreting glucose sensor data in the clinical setting. The evaluation of glycemic levels based on the recordings of the two sensors were compared for each paired time interval.

RESULTS— We discarded 18% of the sensor data for technical reasons. Examined as a group, the remaining 3,370 paired data points in all 11 patients were highly correlated ($r = 0.84$). However, when individual pairs were evaluated, large differences in the glucose values were apparent, with differences of $>10\%$ in 70% of the measurements and $>50\%$ in 7% of the measurements. Moreover, clinical evaluation of the glucose range provided simultaneously by two sensors was concordant for only 65% of the evaluation periods.

CONCLUSION— In a real-life setting, the accuracy of data provided by the Minimed glucose sensor may be less than expected. To avoid therapeutic errors, sensor findings should be confirmed by independent means before clinical decisions are made.

Diabetes Care 25:1185–1191, 2002

Recent studies have confirmed that improved glycemic control decreases the risk of diabetic complications in both type 1 and type 2 diabetic patients (1,2). Patients and physicians are currently evaluating and adapting diabetes treatment according to clinical data, laboratory values of HbA_{1c}, and self-monitoring of capillary blood glucose performed by the patient several times a day. However, during large portions of

the day, glucose levels are not determined, permitting the possibility that major excursions of glucose levels occur without the patient's awareness. To alleviate this problem, a new device for continuous glucose monitoring is now available and has received U.S. Food and Drug Administration (FDA) approval for use by the health care team (3). The Minimed glucose sensor, or continuous glucose monitoring system, measures

glucose concentration in the subcutaneous tissue every 5 min (288 times a day) for up to 3 days. In its current configuration, the device stores all determinations in an internal memory. The patient performs finger-stick glucose determinations and enters the data into the device for subsequent calibration. The data are downloaded onto a computer at the end of the 3-day period of monitoring, and glucose levels are calculated automatically for each 24-h cycle. The patient and physician can then review the results retrospectively and decide on therapeutic changes as appropriate. Thus, real-time glucose values are not available to the patient.

This device has been used extensively by several centers during the last 2 years, and several studies emphasized the importance of the information obtained by the device, particularly in identifying clinically unrecognized events of hypoglycemia (4–8).

We initiated a study using this device to evaluate the effect of treatment on glycemic control in a cohort of type 2 diabetic patients, performing ~150 sensor tracings. Surprisingly, the number of hypoglycemic episodes identified by the sensor was exceedingly high in patients at apparent low risk for hypoglycemia, including patients with type 2 diabetes treated with metformin only and healthy volunteers. Some of these episodes occurred while the patient was awake and were not confirmed by glucometer analyses. In a preliminary study, we compared 75 capillary glucose determinations that were made while patients were wearing the sensors but that were not entered into the devices for calibration. The correlation coefficient ($r = 0.74$) was less than expected, and Bland-Altman analysis (9) demonstrated a large random variance that was similar throughout the spectrum of measurements (Fig. 1). These findings suggested the possibility that some unexpected episodes of hypo- or hyperglycemia, previously described using the glucose sensor, may have been spurious.

We therefore embarked on a study aimed at determining the accuracy and reproducibility of tracings obtained by

From the ¹Diabetes Center, Endocrinology and Metabolism Service, Hadassah University Hospital, Jerusalem, Israel; and the ²Diabetes Unit, Wolfson Hospital, Holon, Israel.

Address correspondence and reprint requests to Dr Muriel Metzger, Diabetes Unit, Hadassah University Hospital, POB 12000, Jerusalem, Israel. E-mail: muriel@hadassah.org.il.

Received for publication 7 November 2001 and accepted in revised form 10 April 2002.

Abbreviations: FDA, U.S. Food and Drug Administration.

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

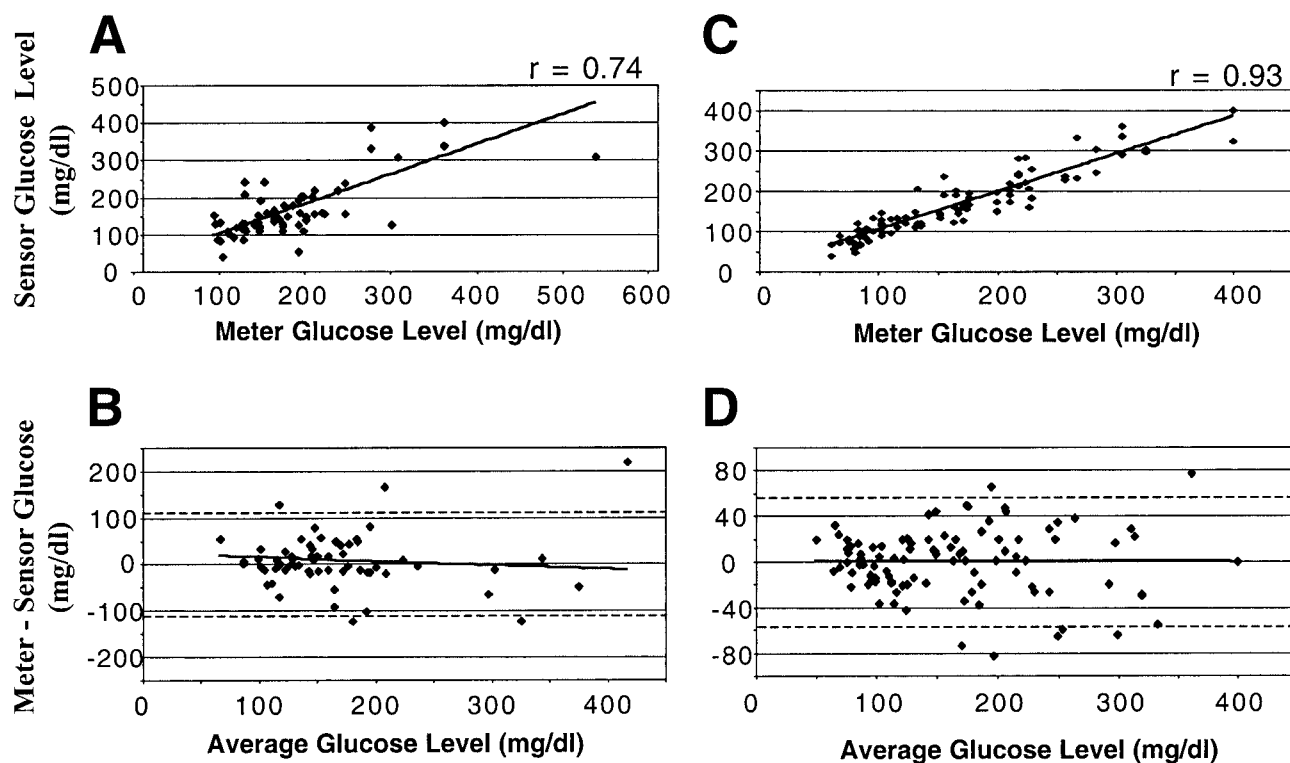


Figure 1—A and C: Graphs demonstrating the correlation between capillary blood glucose determinations and glucose sensor readings. B and D: Bland-Altman plots showing the difference between capillary blood glucose measurements and simultaneous glucose sensor readings at different ambient glucose levels. The x-axis shows the mean of the two readings. The dashed line represents the 95% confidence limits of the differences between the two methods. A regression line is shown on each graph demonstrating that the mean differences between the methods is close to 0 and that there is no systematic difference between the methods at different ambient glucose levels. A and B depict independent capillary blood glucose levels that were measured while patients were wearing the glucose sensor, but that were not entered into the sensor for calibration. These were obtained during a preliminary phase of this study in a separate test population. C and D represent data from the current study population, comparing the capillary blood glucose values that were used for calibration for all of the tracings reported in this study. Values corresponding to tracings that were judged technically inadequate (see RESEARCH DESIGN AND METHODS) were excluded.

this device in real-life situations. A total of 11 patients, including type 1 and type 2 diabetic patients as well as healthy subjects, agreed to wear two glucose sensor devices simultaneously during a 3-day period of normal activity. Our findings raise questions as to the reliability of results obtained using this device in its current configuration and provide a prototype procedure for testing future generations of these devices.

RESEARCH DESIGN AND METHODS

A total of 6 type 1 and 3 type 2 diabetic patients and 2 healthy subjects (10 male, 1 female) were selected on the basis of their willingness and ability to perform several finger-stick glucose determinations a day and enter the results into the monitor for calibration. All diabetic subjects were used to performing finger-stick tests, and the two healthy subjects belonged to the diabetes health care team.

Experimental protocol

The glucose sensors (CGMS; Minimed, Northridge, CA) were attached with strict compliance to the manufacturer's instructions. Each sensor was inserted horizontally into the abdominal subcutaneous tissue, 4–5 cm to the right or left of the umbilicus, avoiding areas of tissue scarring, using the Sen-serter automatic device. After connecting the sensor to the monitor, the cable was pasted horizontally with two pieces of adhesive tape, one next to the other, to minimize the risk of traction on the connection or on the sensor itself. The patients were warned against the risk of pulling at the sensor or dropping it. Special attention was paid to the occurrence of any pain or discomfort during or after insertion of the sensor, which may be indicative of subcutaneous edema around the needle and may influence the quality of the results. After initialization of the monitor, the patient was instructed to wait at least 1 hour before

entering the first capillary blood glucose value for calibration. A minimum of four measurements was required for calibration, and the time interval between the measurements was not to exceed 8 h. The patients were requested to perform one test during the night or the early morning hours of the day and to perform some tests either immediately before meals or during the early postprandial period to increase the precision of calibration by covering a wider range of glucose values. Mealtimes were recorded in the patient's diary, and the glucose values were entered into the monitor within 5 min of determination. All capillary blood glucose tests performed during sensor recording were entered for calibration according to the manufacturer's instructions. The patients were asked to wear the sensors during at least 2 entire days in order to identify recurrent glucose patterns. They were requested to try and avoid sleeping on the sensors. When the sensors were removed

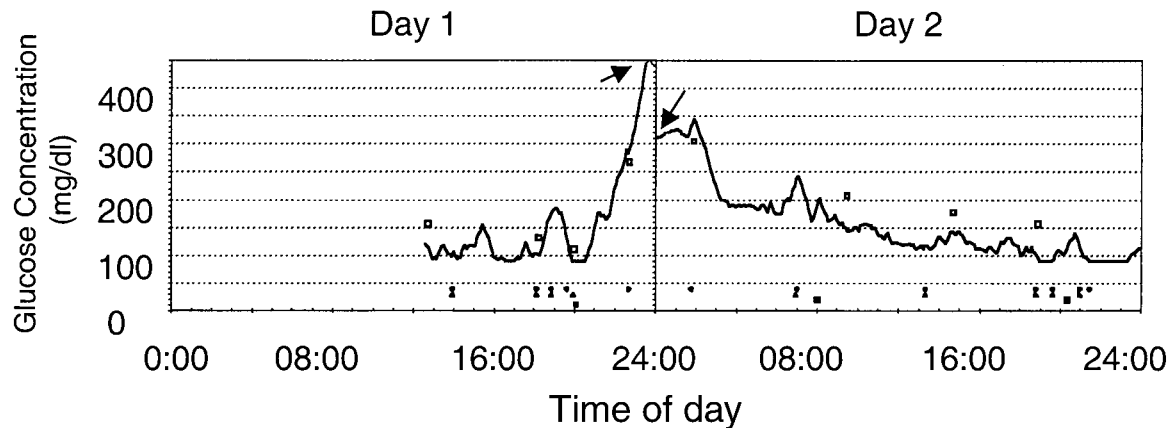


Figure 2—The midnight shift artifact. The graph shows continuous recording of glucose levels during 2 consecutive days. Note that after midnight there is a marked decrease in glucose values (short arrows) that could not be explained by physiological changes in blood glucose.

at the end of the 3-day recording, the tip of each was carefully examined. Any abnormal bending was recorded.

Data analysis

After completion of the 3-day analysis, data from both devices were downloaded. Each monitor was placed inside a Minimed Com-Station, and the sensor glucose values were immediately calculated by the software provided by the manufacturer (Minimed Graphs, version 1.6B).

All of the graphs provided by the software were coded, printed, and analyzed. Anonymity was guaranteed by erasing the patient's identity and the date. Because some tracings were technically poor during portions of the day, and in order to obtain the maximum amount of information from the tracings available, each day was divided into eight time intervals (night, morning pre- and postprandial, noon pre- and postprandial, evening pre- and postprandial, and bedtime), according to the meal times indicated by the patient. Data for each time interval was evaluated independently and was classified as satisfactory (A) if all glucose values fell between 80 and 150 mg/dl, too high (B) if glucose values were >150 mg/dl during >1 h, too low (C) if glucose values were <70 mg/dl during >30 min, or impossible to evaluate for technical reasons (D). Those tracings classified as D were further subclassified according to the following criteria:

- D1: Low concordance between the glucose values calculated by the sensor and measured by the glucometer. The program usually indicates above the daily

graph or in the summary that the concordance is not satisfactory (correlation coefficient $r < 0.8$), or that the difference between meter and sensor values is too high (mean absolute difference >28%).

- D2: Insufficient number of meter glucose values entered for calibration. When the number of glucose values entered for calibration is less than three tests/day, the calibration slope varies greatly from one day to the next, and the values calculated by the sensor need to be corrected.
- D3: Strong midnight shift, with a significant shift in sensor glucose values immediately after midnight. This problem is usually (but not always) due to an insufficient number of calibration values, which causes great variations in the daily calibration slope. As a consequence, there may be a big difference in the glucose values calculated by the sensor a few minutes before and after midnight, making it very difficult to relate to the absolute glucose values provided by the sensor.

Time intervals for which paired evaluations were available were analyzed by two different observers (I.R. and M.M.), and the concordance rates for the two sensors were determined.

Statistics

Data are presented as the means \pm SD. Simultaneous individual readings from two sensors in the same patients were compared, and correlation coefficient was calculated by simple linear regression. Capillary glucose levels and sensor read-

ings were also compared using the Bland-Altman analysis (9).

RESULTS

Initial evaluation of all sensor data

Altogether, 11 patients wore two sensors for a mean of 60.4 ± 16.9 h each. There were no significant adverse events, and tolerability of the device was excellent. No abnormal bending of the sensors was noted after removal. A total of 432 single time intervals (corresponding to 216 potential paired sets of data) were initially evaluated. Of these, 78 (18%) were discarded for technical reasons: 32 (7%) because the sensor did not record glucose values and 46 (11%) because the data were not interpretable (classified as "D," as described in RESEARCH DESIGN AND METHODS). This was caused by gross discordance between glucometer and sensor glucose values (D1) in 28 time intervals (6%), insufficient calibration (D2) in 11 time intervals (3%), and an unacceptably large midnight shift (D3) in 7 time intervals (2%). An example of a strong midnight shift is shown in Fig. 2. Finally, of 216 initial paired sets of data, 139 were available for between-sensor comparisons. Of the paired sets, 92 came from six type 1 diabetic patients, 30 from three type 2 diabetic patients, and 17 from two nondiabetic subjects. All subsequent analyses were performed on this subset of data.

Comparison between capillary glucose determinations and simultaneous sensor readings

To obtain an estimate of the precision and accuracy of the sensor readings, we compared the capillary glucose levels used to

Table 1—Percent discordance between simultaneously obtained glucose readings in all subjects

Patient	Number of samples	Percent of samples with differences greater than:				
		10%	20%	30%	40%	50%
A	167	83.9	57.7	22.6	13.7	6.0
B	433	70.0	44.6	25.6	10.4	4.6
C	456	79.4	61.0	37.9	19.5	6.6
D	288	71.5	50.0	36.1	21.5	17.0
E	173	76.9	52.0	31.8	17.3	12.1
F	144	31.9	10.4	1.4	0.0	0.0
G	232	59.5	35.8	25.0	14.2	6.0
H	692	69.8	40.3	23.4	15.8	8.5
I	461	67.0	31.0	11.7	0.7	0.0
J	192	68.8	32.8	21.4	10.4	4.7
K	132	63.6	41.7	32.6	16.7	12.1
All patients	3,370	69.3	42.7	24.9	12.9	6.8

calibrate the sensor to simultaneous sensor readings (Fig. 1). An overall correlation coefficient (r) of 0.93 was obtained with a slope of 0.9316 and an intercept of 11. Comparison of the data using a Bland-Altman plot indicated a 95% confidence limit variation of ± 57 mg/dl over most of the spectrum of glucose values. The mean difference between the meter and sensor readings was 0 at all levels, suggesting that there was no consistent bias between the two methods. A somewhat greater variance was seen between 125 and 225 mg/dl, precisely the range that may be of particular clinical importance.

Comparison of paired data

Data from simultaneously obtained tracings for all patients are shown in Table 1 and Fig. 3.

Each double set of sensor data showed significant and prolonged differences between the values provided by the two sensors, sometimes during several hours. These differences were noted in all subjects. The correlation between 3,370 individual measurements obtained simultaneously by the two sensors was high ($r = 0.84$); however, 69% (range 32–84) of all measurements were discordant, with a difference between values of $>10\%$. Variability of $>50\%$ was observed in 7% (0–17) of the total number of measurements (Table 1; Fig. 4). Bland-Altman analysis demonstrated that the difference between the readings of the two sensors was similar at all ambient glucose levels, although some increase in scatter was seen at glucose levels >150 mg/dl.

Clinical interpretation of data

Although the correlation between simultaneously obtained data points is important, the more critical question is how the variability of the measurements might affect the clinical decisions. The data from each tracing was interpreted as described above, and the results of simultaneously obtained tracings were compared. The percentage of time intervals available for comparison was similar in all subjects. The clinical interpretations were concordant for only 65% of the periods. In 25% of the time intervals, one sensor showed that glucose levels were too high, whereas the other suggested satisfactory glycemic control. In 9% of time intervals, one tracing indicated that glucose levels were too low, whereas the other showed satisfactory glucose levels. In one case, one sensor showed high glucose, whereas the other showed too-low glucose values for the same time interval. There was no difference between the nondiabetic subjects and the type 1 and type 2 diabetic patients. Some of these regions of clinically relevant discrepancy can be seen in Fig. 3.

CONCLUSIONS — We have studied the reliability of data obtained from glucose sensors used in the setting of the patient's everyday life. The data show that under these circumstances, the accuracy and reproducibility of the sensor is considerably lower than previously believed. All subject groups tested (control subjects and type 1 or type 2 diabetic patients) yielded a similar percentage of interpretable data and showed a similar degree of correlation for all parameters measured.

After removing technically inadequate tracings, as described in RESEARCH DESIGN AND METHODS, the correlation between capillary glucose levels entered for calibration and simultaneous sensor readings ($r = 0.93$) was similar to that previously reported using this sensor. Gross and Mastrototaro (10) analyzed the data provided by 415 sensors and compared the sensor and the glucometer glucose values, obtaining a median daily correlation of $r = 0.92$, with only two cases of extreme disagreement between the sensor and the glucometer results. Outpatient studies (including type 2 and type 1 diabetic patients, children, and pregnant women) with subjects wearing a single sensor also showed apparently similar correlation coefficients ($r = 0.9$) when all technical precautions were taken into account (11–13). However, when we analyzed these same data using the Bland-Altman plot, it became obvious that the random variation between the two methods of glucose determination was much greater than would be considered clinically acceptable. The 95% confidence limits of the differences between the two methods was ± 57 mg/dl across the entire range of glucose values. In fact, this difference tended to be greatest between 125 and 225 mg/dl, a range that is of particular importance in intensively treated patients.

The correlation of the glucose values obtained by the two sensors ($r = 0.84$) was lower than that obtained between the capillary levels and the sensor readings. Moreover, the variability in the simultaneously obtained values was large. Specifically, 69% of the measurements varied by $>10\%$, and differences of $>50\%$ were seen in 7% of measurements overall and in 12–17% of readings in three patients. In one healthy subject, one of the sensors recorded a prolonged episode of hypoglycemia not evident in the second sensor. Reevaluation of the measurements after the completion of the recordings did not reveal any evidence for sensor failure during this episode.

The problem of reliability of tracings cannot be due to lack of experience of the research team, because before this study, this team had performed >150 tracings as part of a separate study. Of the sensor data, 18% were discarded for technical reasons. This compares favorably to previous reports, in which only 50–60% of the sensor readings were usable after fil-

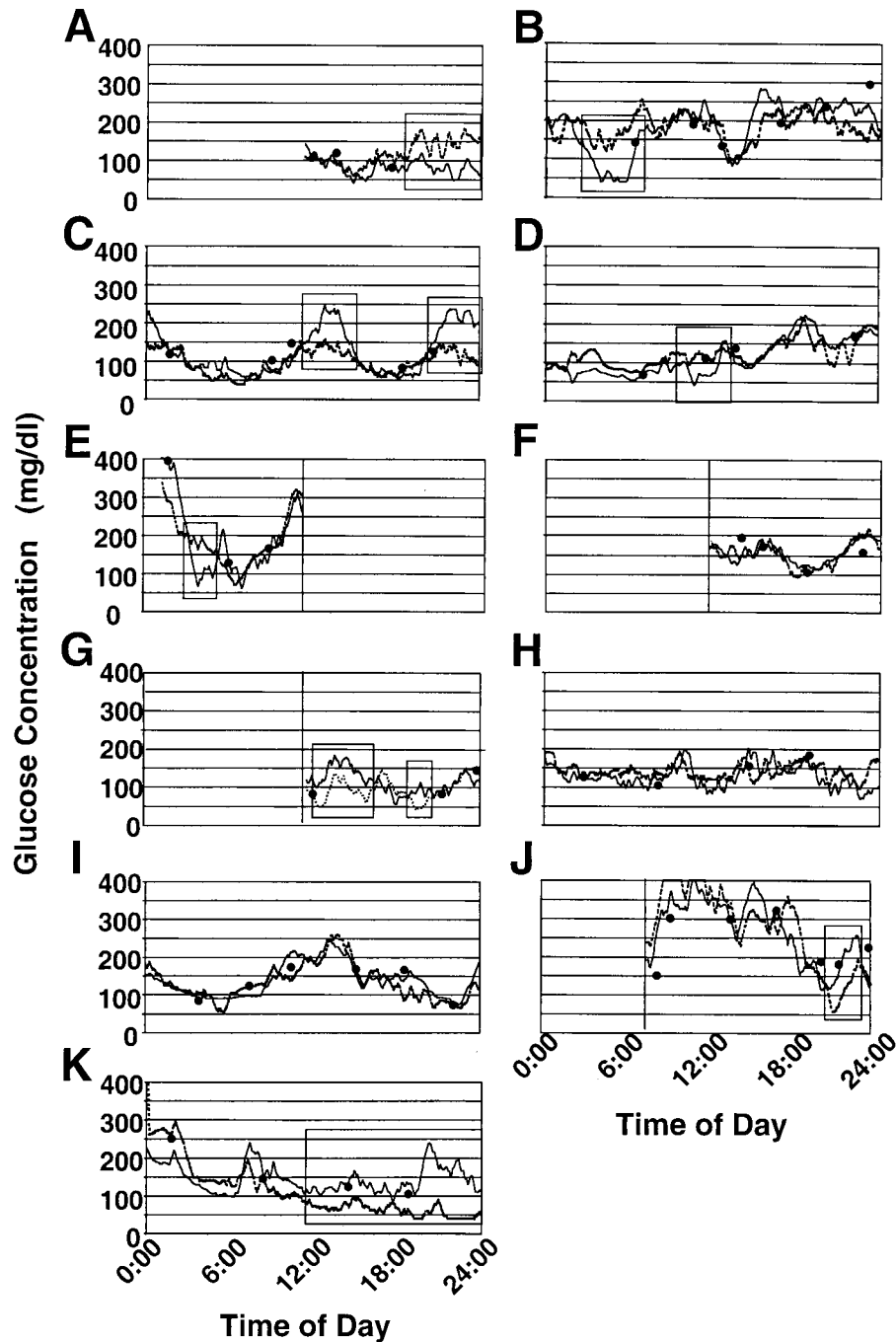


Figure 3—A–K: Continuous monitoring of glucose levels using the Minimed glucose sensor. Each graph shows a sample of the glucose values obtained from two sensors worn simultaneously by the same patient (x-axis: time; y-axis: sensor glucose values [in mg/dl]). ●, corresponding capillary glucose values. Some graphs show <24-h recording because the portions considered technically uninterpretable were excluded. The graphs were obtained from six type 1 (B, C, E, H, I, and K) and three type 2 (D, F, and J) diabetic patients and two healthy volunteers (A and G). Some specific periods of clinically relevant discrepancy are indicated in rectangles.

tering for technical problems (5,7). This attests to the high level of compliance among the subjects chosen. Discordant recordings were found in all patients, indicating that the differences are not the result of poor handling of the sensor by any specific patient.

The high frequency (35%) of important discrepancies between the clinical interpretation of two simultaneously obtained sensor tracings is indeed worrisome. If one assumes a best-case scenario in which concordant tracings are always correct, and one of the two discordant

tracings is correct, then had patients been given only a single sensor, incorrect clinical advice might have been given in 17% of the cases. The actual discordance may be greater if, in fact, both tracings were inaccurate in any of the patients. The potential clinical consequences may be seri-

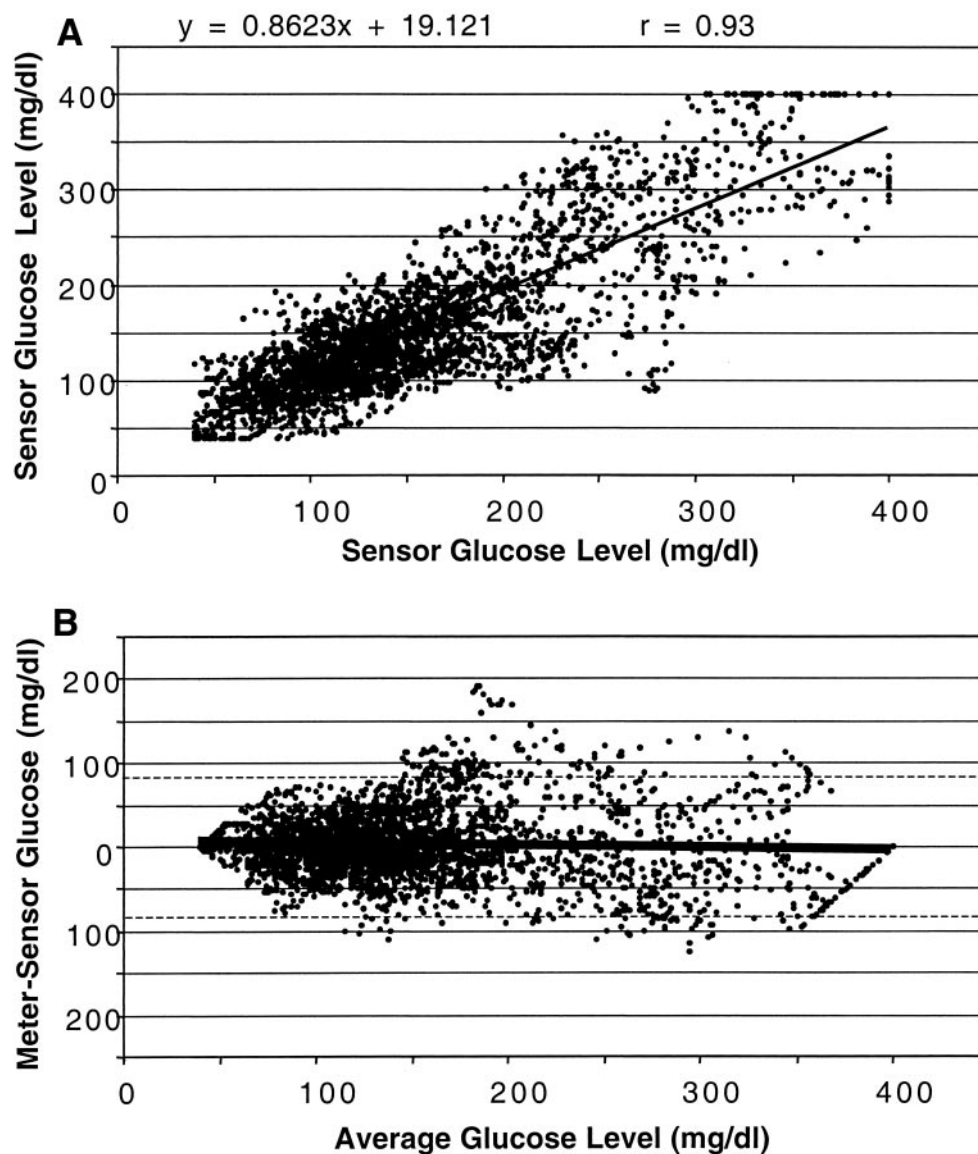


Figure 4—A: Correlation between the data obtained simultaneously by two sensors on the same patient. Only technically acceptable tracings were included in this evaluation. The correlation coefficient was calculated by simple linear regression. B: Bland-Altman plot of the same data showing the difference between the two meter readings on the y-axis, and the mean of the two readings on the x-axis. The regression line is shown and indicates random differences between the two sensors at all levels. The dashed lines show the 95% confidence limits.

ous if high glucose values are recorded by the sensor in a patient with normal or even low true glucose values.

All of the pitfalls in the clinical interpretation of the sensor data must be clearly identified to reduce the risk of making clinical recommendations based on inaccurate tracings. Despite a careful retrospective reevaluation of all discordant tracings, in most cases no satisfactory explanation could be found for discrepancies, and parallel simultaneous tracings of apparent good technical quality still showed markedly different glucose values.

To our knowledge, all studies published so far comparing two or more glucose sensors worn simultaneously were performed under “laboratory” conditions, for instance in hospitalized patients receiving a glucose infusion (14,15), and no results have been published providing such data during everyday life conditions. Thus, the reproducibility of the results obtained by the sensor under real-life conditions appears to be insufficiently documented.

The reproducibility of measurements may also differ according to the variability in fat content in the subcutaneous tissue

of different patients. This question needs to be addressed in further studies.

This device and its related software are under continued development and are being improved. Recently, we have been informed that the software has been upgraded to resolve the problem of the “midnight shift” (personal communication, J. Mastroianni, Minimed, Northridge, CA). Despite this, at its current stage of development, specific clinical decisions should not be made on the sole basis of the Minimed glucose sensor data because this may lead to incorrect clinical decisions.

Another device for continuous glu-

cose monitoring, the Glucowatch Biographer, was recently approved by the FDA for clinical use. Thirty-six patients wore two Biographer devices simultaneously on the same arm in a clinic setting. The correlation coefficient between sensors was significantly better than that found in our study ($r = 0.92$). However, these results were not reevaluated in a home setting, and the clinical significance of variations in readings was not tested. Furthermore, this device shows real-time glucose readings, allowing the patient to perform a capillary glucose determination in case of unexpectedly high or low glucose readings reported by the Biographer (16,17).

Clearly, the development of a reliable device for continuous glucose self-monitoring is an important goal that will have great impact on the clinical management of diabetes. Future generations of this and other continuous glucose monitors should be rigorously evaluated for accuracy and reproducibility in real-life clinical settings before release for general use.

References

1. The Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986, 1993
2. U.K. Prospective Diabetes Study (UKPDS) Group: Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 652:837–853, 1998
3. Summary of Safety and Effectiveness Data for the Minimed Continuous Glucose Monitoring System (CGMS). Food and Drug Administration, 1999 (publ. no. PMA P980022)
4. Hirsch IB: Hypoglycemia and the hypoglycemic unawareness syndrome. *Diabetes Technol Ther* 2:S81–S87, 2000
5. Chase HP, Kim LM, Owen SL, MacKenzie TA, Klingensmith GJ, Murfeldt R, Garg SK: Continuous subcutaneous glucose monitoring in children with type 1 diabetes. *Pediatrics* 107:222–226, 2001
6. Bode BW, Gross TM, Thornton KR, Mastrototaro JJ: Continuous glucose monitoring used to adjust diabetes therapy improves glycosylated hemoglobin: a pilot study. *Diabetes Res Clin Pract* 46:183–190, 1999
7. Boullu-Sanchis S, Jeandidier N, Meyer L, Busch M-S, Ott F, Pinget M: Use of continuous glucose monitoring system (CGMS) in 23 insulin treated diabetic patients: feasibility, reliability and efficacy for the diagnosis of unnoticed hypoglycemia (glycemia ≤ 3 mmol/l) (Abstract). *Diabetes* 50 (Suppl. 2):A447, 2001
8. Boland E, Monsod T, Delucia M, Brandt CA, Fernando S, Tamborlane WV: Limitations of conventional methods of self-monitoring of blood glucose: lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes. *Diabetes Care* 24:1858–1862, 2001
9. Bland JM, Altman DG: Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160, 1999
10. Gross TD, Mastrototaro JJ: Efficacy and reliability of the continuous glucose monitoring system. *Diabetes Technol Ther* 2:S19–S26, 2000
11. Bode BW: Clinical utility of the continuous glucose monitoring system. *Diabetes Technol Ther* 2:S35–S41, 2000
12. Gross TM, Ter Veer A: Continuous glucose monitoring in previously unstudied population subgroups. *Diabetes Technol Ther* 2:S27–S34, 2000
13. Kaufman FR: Role of the continuous glucose monitoring system in pediatric patients. *Diabetes Technol Ther* 2:S49–S52, 2000
14. Johnson KW, Mastrototaro JJ, Howey DC, Brunelle RL, Burden-Brady PL, Bryan NA, Andrew CC, Rowe HM, Allen DJ, Noffke BW, MacMaham WC, Morff RJ, Lipson D, Nevin RS: In vivo evaluation of an electroenzymatic glucose sensor implanted in subcutaneous tissue. *Biosens Bioelectron* 7:709–714, 1992
15. Silver DM, Boyne MS, Saudek CD: Effect of recalibration on agreement between blood glucose and subcutaneous glucose sensor readings (Abstract). *Diabetes* 50 (Suppl. 2):A131, 2001
16. Garg SK, Potts RO, Ackerman NR, Fermi SJ, Tamada JA, Chase HP: Correlation of fingerstick blood glucose measurements with Gluco Watch biographer glucose results in young subjects with type 1 diabetes. *Diabetes Care* 22:1708–1714, 1999
17. Pitzer KR, Desai S, Dunn T, Jayalakshmi Y, Kennedy J, Tamada JA, Potts RO: Detection of hypoglycemia with the Gluco-Watch biographer. *Diabetes Care* 24:881–885, 2001