A Comparative Effectiveness Analysis of Three Continuous Glucose Monitors

Edward R. Damiano, PhD[1]
Firas H. El-Khatib, PhD[1]
Hui Zheng, PhD[2]
David M. Nathan, MD[3]
Steven J. Russell, MD, PhD[3]

OBJECTIVE—To compare three continuous glucose monitoring (CGM) devices in subjects with type 1 diabetes under closed-loop blood glucose (BG) control.

RESEARCH DESIGN AND METHODS—Six subjects with type 1 diabetes (age 52 ± 14 years, diabetes duration 32 ± 14 years) each participated in two 51-h closed-loop BG control experiments in the hospital. Venous plasma glucose (PG) measurements (GlucoScout, International Biomedical) obtained every 15 min (2,360 values) were paired in time with corresponding CGM glucose (CGMG) measurements obtained from three CGM devices, the Navigator (Abbott Diabetes Care), the Seven Plus (DexCom), and the Guardian (Medtronic), worn simultaneously by each subject. Errors in paired PG–CGMG measurements and data reporting percentages were obtained for each CGM device.

RESULTS—The Navigator had the best overall accuracy, with an aggregate mean absolute relative difference (MARD) of all paired points of 11.8 ± 11.1% and an average MARD across all 12 experiments of 11.8 ± 3.8%. The Seven Plus and Guardian produced aggregate MARDs of all paired points of 16.5 ± 17.8% and 20.3 ± 18.0%, respectively, and average MARDs across all 12 experiments of 16.5 ± 6.7% and 20.2 ± 6.8%, respectively. Data reporting percentages, a measure of reliability, were 76% for the Seven Plus and nearly 100% for the Navigator and Guardian.

CONCLUSIONS—A comprehensive head-to-head-to-head comparison of three CGM devices for BG values from 36 to 563 mg/dL revealed marked differences in performance characteristics that include accuracy, precision, and reliability. The Navigator outperformed the other two in these areas.

Diabetes Care 36:251–259, 2013

Widely accepted clinical standards for accuracy and reliability of the commercially available continuous glucose monitoring (CGM) devices have not yet been established by professional associations or regulatory agencies. To generate such standards, the accumulation of large datasets comparing reference-quality blood glucose (BG) or plasma glucose (PG) measurements with CGM glucose (CGMG) data is needed. Most investigator-initiated studies that have attempted to gather such data have been relatively short in duration (usually several hours), contained low data density, and/or have not included large variations in glucose values or large time rates of change in glucose values that are typical of diabetes (1–3). There is a dearth of studies comparing CGM devices worn simultaneously by the same subject, and those that exist have suffered from the same limitations (3). Data obtained by CGM device manufacturers cannot be directly compared across devices owing to differences in the clinical protocols between studies.

There is a clear and present need to evaluate the relative accuracy and reliability of the commercially available CGM devices over large ranges of BG values and time rates of change in BG values, and over sensor wear periods that are long enough to encompass multiple scheduled calibrations. The present analysis examines the results of a comprehensive study comparing three CGM devices, the Navigator (Abbott Diabetes Care), the Seven Plus (DexCom), and the Guardian (Medtronic). The study was conducted in subjects with type 1 diabetes in a clinical research center setting as part of closed-loop BG control experiments. The three CGM devices were worn simultaneously in each experiment while reference-quality PG levels were measured every 15 min continuously for 48 h. Results were analyzed in point accuracy (including absolute and relative differences), rate-of-change accuracy, and sensor reliability (including variation around mean performance and data reporting percentage).

RESEARCH DESIGN AND METHODS

Subjects

The clinical protocol was approved by the human research committees at Massachusetts General Hospital (MGH) and Boston University. Six subjects with C-peptide–deficient, type 1 diabetes participated. All subjects gave written informed consent. At baseline, subjects were required to be aged ≥18 years, to have had type 1 diabetes for at least 1 year, and to have a stimulated C-peptide level in response to a mixed-meal tolerance test ≤0.1 nmol/L. The study cohort and the closed-loop experiments have been described previously (4). Each subject participated in two separate 48-h experiments (96 h of data for each subject).

Experimental protocol

Subjects were admitted to the MGH Clinical Research Center wearing Navigator, Seven Plus, and Guardian sensors and transmitters, which were inserted the day before the study at ~1500 h according to the respective manufacturers' directions. Upon admission, the three transmitters were linked wirelessly to their respective receiver devices.

Venous PG levels were measured every 15 min with the GlucoScout (International Biomedical) and confirmed

From the [1]Department of Biomedical Engineering, Boston University, Boston, Massachusetts; the [2]Massachusetts General Hospital Biostatistics Center, Massachusetts General Hospital, Boston, Massachusetts; and the [3]Diabetes Unit and Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts.
Corresponding author: Steven J. Russell, sjrussell@partners.org.

Received 11 January 2012 and accepted 3 November 2012.
DOI: 10.2337/dc12-0070. Clinical trial reg. no. NCT01161862, clinicaltrials.gov.
This article contains Supplementary Data online at http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc12-0070/-/DC1.

hourly with a YSI 2300 STAT Plus Analyzer (YSI Life Sciences). The three CGM devices were calibrated according to the manufacturers' instructions, except that venous PG rather than capillary self-monitored BG (SMBG) values were used for calibration. During each 48-h experiment, the Navigator required one scheduled calibration, and the Seven Plus and Guardian required four scheduled calibrations. Beyond the usual scheduled calibrations, any additional calibrations that were requested by any CGM device were also performed. In addition, if the CGMG reading of any device did not meet the International Organization for Standardization standard for accuracy relative to PG at 0600 h daily, then a forced calibration of that device was performed (see Supplementary Data for further details).

Fully automated closed-loop BG control was initiated at 1500 h and ran continuously for 51 h; the last 48 h of each experiment were included in this analysis. Six meals were provided during this period; mean carbohydrate consumption was 78 ± 12 g (range 60–117) per meal. Moderate exercise on a stationary bicycle began 25 h into the experiment and lasted ~30 min.

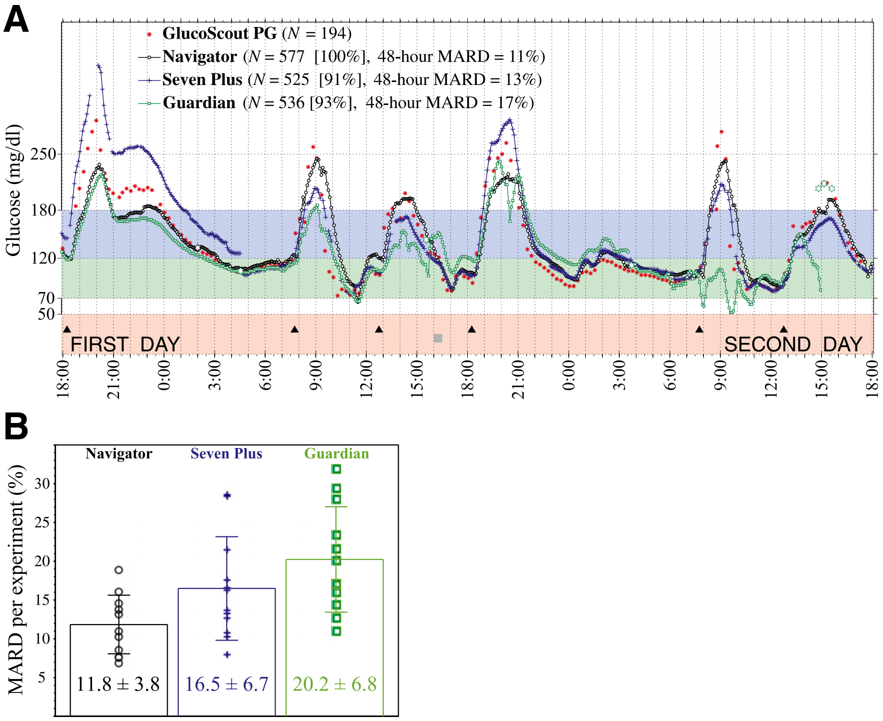


Figure 1—A: *Representative results from one of twelve 48-h closed-loop BG control experiments in one of six subjects showing venous PG concentrations measured every 15 min with the GlucoScout (red symbols) and CGMG values measured approximately every 5 min with the Navigator (black symbols), Seven Plus (blue symbols), and Guardian (green symbols). The timing of six meals is indicated by black triangles. One period of structured exercise at 1600 h (2 h before the fourth meal) is indicated by a gray square. Listed in the legend for each CGM device is the number (*N*) of glucose values measured, the data reporting percentage (in square brackets), and the MARD averaged over the 48-h period, based on 194, 171, and 180 paired PG–CGMG values for the Navigator, Seven Plus, and Guardian, respectively.* B: *The 48-h MARDs computed in each of the 12 experiments are shown for each sensor, with the mean and SD of each of those MARDs superimposed on the data for each device.*

Accuracy, precision, and reliability metrics

The point accuracy of each CGM device is measured in terms of the relative difference (RD), defined as [(CGMG − PG)/PG], and the absolute relative difference (ARD), defined as [(CGMG − PG)/PG]. Negative RD values correspond to an underestimation and positive values to an overestimation of PG by the CGM device. RD provides insight into the extent and direction of bias in the estimation of PG by a CGM device but is not as useful as the ARD is in determining the average error across a set of data because of the cancelation that occurs when summing positive and negative RD values.

The 48-h mean ARD (MARD) relative to PG was calculated for each of the three CGM devices during the 48-h period from 1800 h at the beginning of the first day of each experiment to 1800 h at the end of the second day. The mean and SD of the 48-h MARD across the 12 experiments are shown in Fig. 1*B* for each CGM device. Whereas the average of the 48-h MARD characterizes the mean accuracy of a particular sensor session in a given experiment, the SD of this MARD provides a precision metric of the variation around mean accuracy from one sensor session to another for each device. In essence, the SD quantifies the consistency relative to the device's average performance that can be expected from one sensor to another for a given CGM device.

In addition to assessing point accuracy, we evaluated the rate-of-change accuracy of each of the three CGM devices. Reference rate-of-change data were obtained by taking the difference between two consecutive PG values and dividing by the sampling interval between those two PG measurements (typically 15 min).

Device reliability is measured with the data reporting percentage, defined as the ratio of the number of glucose values reported by the CGM receiver over the 48-h period relative to the total number possible for that period. The three devices were configured to record CGMG values every 5 min. The Seven Plus, which has a rechargeable battery that did not have sufficient capacity to power the device for the entire experiment, was kept plugged into its charging device.

Statistical analyses

Statistical analyses were performed using SAS 9.2 software (SAS Institute Inc., Cary, NC). Repeated-measurements models were used for within-subject repeated measurements on the differences between the paired measurements. This accounted for within-subject correlations and correlations in paired measurements. The repeated-measurements models were fitted with the generalized estimating equation method.

RESULTS—Six subjects (three men, three women) each participated in two 51-h closed-loop BG control experiments. Subjects weighed 72 ± 10 (54–85) kg, were aged 52 ± 14 (33–72) years, and had type 1 diabetes for 32 ± 14 (17–50) years.

CGM calibrations

No additional Navigator calibrations were performed other than the scheduled calibrations requested by the device

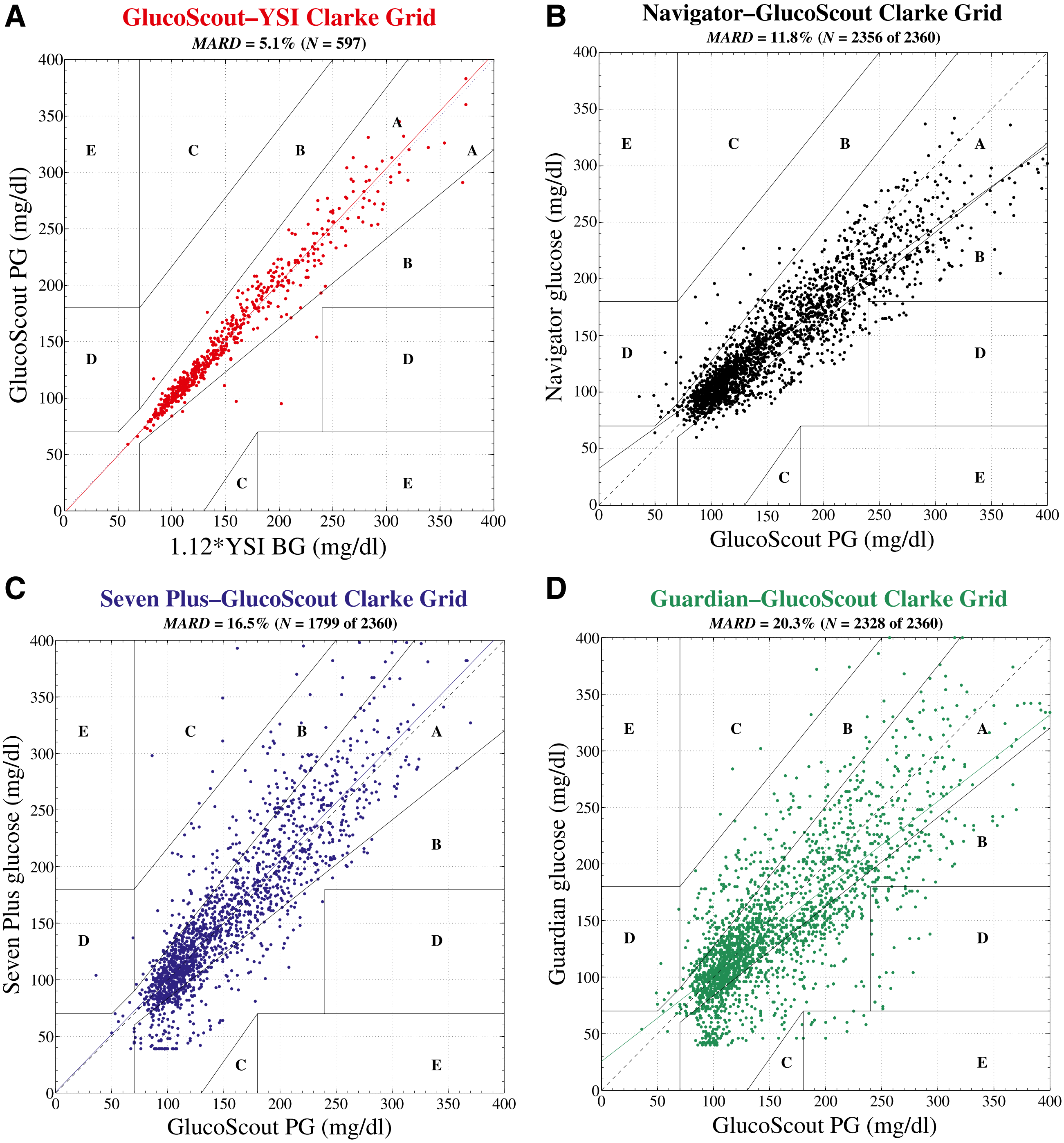


Figure 2—*Clarke error grid analyses of venous plasma glucose (PG) measured by the GlucoScout (*A*), with venous BG measured by the YSI designated as the reference, and CGMG measured by the Navigator (*B*), the Seven Plus (*C*), and the Guardian (*D*), with venous PG measured by the GlucoScout designated as the reference.* A: *Based on a total of 597 GlucoScout–YSI glucose pairs, 98.3% of points fell in zone A and the remaining 1.7% fell in zone B. The slope and intercept of the linear least squares fit to these data (solid red line) were 1.02 and −2 mg/dL, respectively. The MARD was 5.1% between GlucoScout PG and YSI BG (after converting the latter to PG with a multiplicative factor of 1.12).* B: *Based on a total of 2,356 Navigator–GlucoScout pairs, the Navigator achieved 80.6% of points in zone A, 18.3% in zone B, 0% in zone C, and 1.0% in zone D. The slope and intercept of the linear least squares fit to these data (solid black line) were 0.71 and 33 mg/dL, respectively. The Navigator achieved an overall data reporting percentage of 99.8% and a MARD of 11.8 ± 11.1%.* C: *Based on a total of 1,799 Seven Plus–GlucoScout pairs, the Seven Plus achieved 76.2% of points in zone A, 22.7% in zone B, 0.9% in zone C, and 0.1% in zone D. The slope and intercept of the linear least squares fit to these data (solid blue line) were 1.02 and 1 mg/dL, respectively. The Seven Plus achieved an overall data reporting percentage of 76.2% and a MARD of 16.5 ± 17.8%.* D: *Based on a total of 2,328 Guardian–GlucoScout pairs, the Guardian achieved 63.7% of points in zone A, 33.2% in zone B, 0.3% in zone C, and 2.1% in zone D. The slope and intercept of the linear least squares fit to these data (solid green line) were 0.77 and 26 mg/dL, respectively. The Guardian achieved an overall data reporting percentage of 98.6% and a MARD of 20.3 ± 18.0%.*

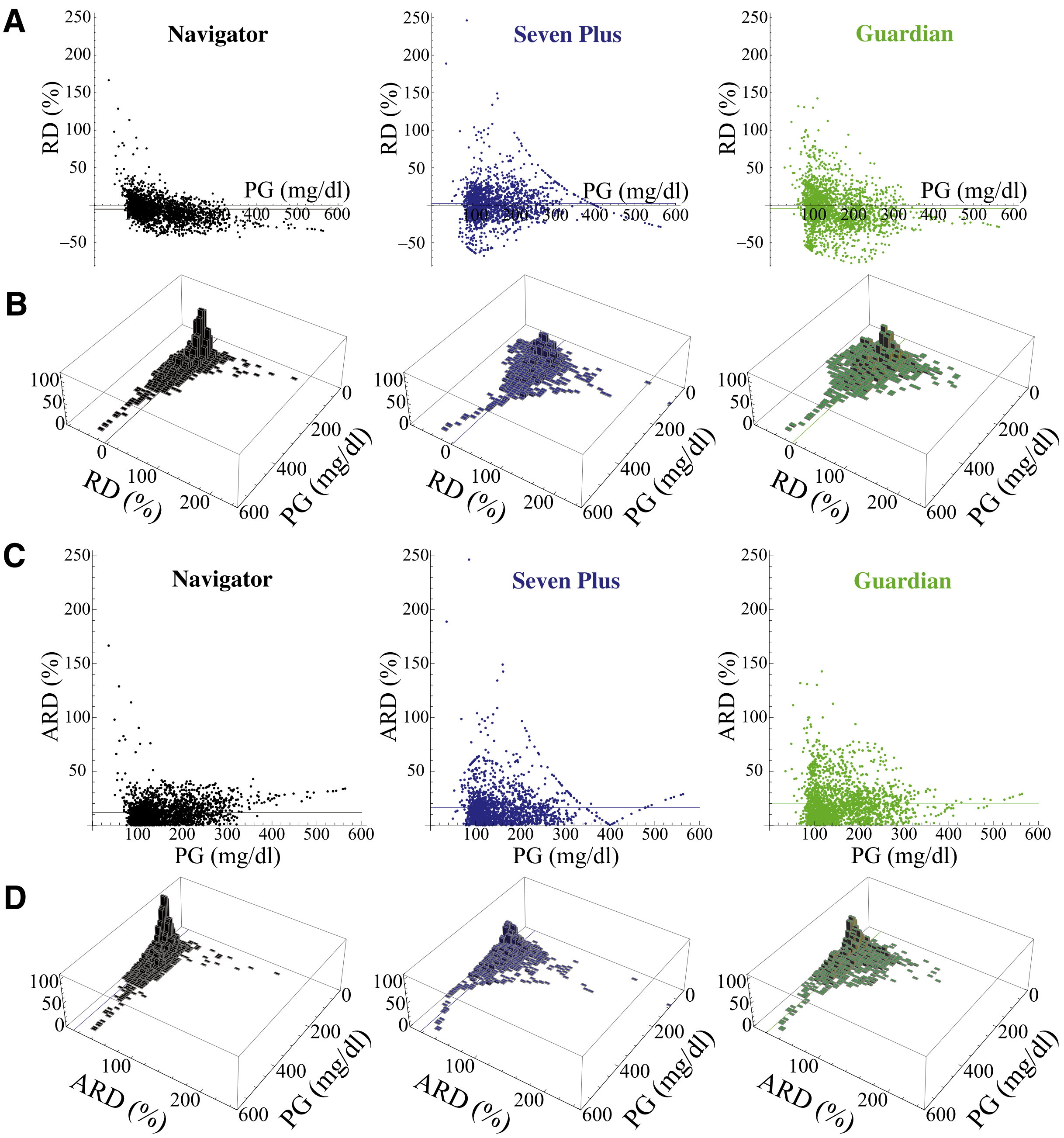


Figure 3—*A: Distribution, as a function of PG, of the RD between each CGMG measurement and its corresponding PG value (measured with the GlucoScout) for the Navigator (black), Seven Plus (blue), and Guardian (green). B: Histograms in the PG–RD plane for each of the datasets shown above in A. The horizontal line in each panel in A and the line in the PG–RD plane in each panel in B correspond to the MRD for each of the three datasets. C: Distribution, as a function of PG, of the ARD between each CGMG measurement and its corresponding PG value (measured with the GlucoScout) for the Navigator (black), Seven Plus (blue), and Guardian (green). D: Histograms in the PG–ARD plane for each of the datasets shown in C. The horizontal line in each panel in C and the line in the PG–ARD plane in each panel in D correspond to the MARD for each of the three datasets. Note, it can be seen that the data in C and D are derivable by reflecting all negatively valued RD data that fall below the PG axis in A and B to their corresponding positive values above the PG axis. The five largest bins for the Navigator had frequencies of 96, 103, 105, 85, and 81 (all between 0 and 7% ARD) corresponding with PG values of 91–98, 98–105, 105–112, 112–119, and 119–126 mg/dL, respectively. These five bins collectively contain 470 of the 2,356 data points (20%). The remaining bins had fewer than 60 hits each. Of the 2,356 data points, 940 (40%) fell in the bins with 0–7% ARD. The five largest bins for the Seven Plus had frequencies of 46, 48, 46, 43 (all between 0 and 7% ARD), and 44 (between 7 and 14% ARD) corresponding with PG values of 98–105, 105–112, 112–119, 119–126, and 98–105 mg/dL, respectively. These five bins collectively contain 227 of the 1,795 data points (12.6%). The remaining bins had fewer than 40 hits each. Of the 1,795 data points, 565 (31.5%) fell in the bins with 0–7% ARD. The five largest bins for the Guardian had frequencies of 50, 58, 61, 53, and 56 (all between 0 and 7% ARD), corresponding with PG values of 91–98, 98–105, 105–112, 112–119, and 119–126 mg/dL, respectively. These five bins collectively contain 278 of the 2,324 data points (12%). The remaining bins had fewer than 50 hits each. Of the 2,324 data points, 569 (24.5%) fell in the bins with 0–7% ARD.*

(corresponding to one calibration during the 2-day duration of each experiment); the final 41–42 h of each experiment were performed without any calibrations of the Navigator. The scheduled calibrations of the Seven Plus and Guardian occurred approximately every 12 h; therefore, four calibrations occurred during the 2-day duration of each experiment. An average of 4.7 (4–6) calibrations per experiment were performed for the Guardian (see Supplementary Data for further details).

CGM point accuracy

The 48-h MARD for the Navigator was 11.8 ± 3.8% compared with 16.5 ± 6.7% for the Seven Plus (z = −2.05, *P* = 0.040) and 20.2 ± 6.8% for the Guardian (z = −3.14, *P* = 0.002). The 48-h MARDs for the Seven Plus and Guardian were not significantly different (z = −1.17, *P* = 0.240).

The point accuracy for each CGM device is shown as Clarke error grids in Fig. 2 and as RD and ARD distributions in Fig. 3. The aggregate MARD across the 2,356 paired points obtained with the Navigator was 11.8 ± 11.1% compared with an aggregate MARD of 16.5 ± 17.8% for the 1,799 paired points obtained with the Seven Plus (z = −1.62, *P* = 0.110 vs. Navigator), and 20.3 ± 18.0% for the 2,328 paired points obtained with the Guardian (*z* = −3.07, *P* = 0.002 vs. Navigator). The aggregate MARDs for the Seven Plus and Guardian were not significantly different (z = −1.31, *P* = 0.190). To put these MARD values in perspective, we calculated the maximum bound on the MARDs for each of the three devices by randomly shuffling the paired CGMG and PG values for each dataset and recalculating the MARDs (see Supplementary Data for details). This procedure yielded upper bounds on the MARD of 41% for the Navigator, 54% for the Seven Plus, and 47% for the Guardian. We estimate that the lower bound of the possible MARD values would be 5.1%, the MARD of the reference quality GlucoScout device relative to the reference quality YSI Stat Plus Glucose monitor when measuring the same sample.

The RD distributions are shown in Fig. 3*A*. In the case of the Navigator, 13 of 2,356 points had positive RD values >50% (range 51–167%). Of these 13 points (PG range 36–135 mg/dL), 6 corresponded to PG values <70 mg/dL, and all PG values <76 mg/dL had positive RD values. Thus, the Navigator consistently overestimated PG in the hypoglycemic range; conversely, there were no negative Navigator RD values >50% (−43% was the most negative RD value). Data in the hyperglycemic range accounted for the most negative RD values; 92% of points with PG values >250 mg/dL had negative RD values for the Navigator. Thus, the Navigator tends to underestimate PG in the hyperglycemic range. With the Seven Plus, 48 of 1,795 points had RD values >50% (51–247%), but these were distributed over a much broader range of PG values (36–261 mg/dL) than with the Navigator. The Guardian had even more points (78 of 2,324) with RD values >50% (50–143%), and like the Seven Plus, these were distributed over a much broader range of PG values (49–257 mg/dL) than with the Navigator. Like the Navigator, the Guardian also tended to underestimate PG in the hyperglycemic range, with 77% of PG values >250 mg/dL having negative RD values. This was not the case for the Seven Plus, where only 45% of PG values >250 mg/dL had negative RD values, showing essentially no bias in the hyperglycemic range.

This lack of bias in the Seven Plus data is also evident in the near-unity slope (1.02) of the linear least squares fit (Fig. 2*C*). By comparison, the slopes of the linear least squares fit are 0.71 for the Navigator and 0.77 for the Guardian (Fig. 2*B* and *D*), which is consistent with the bias in those two devices to underestimate PG in the hyperglycemic range and, to a lesser extent, overestimate PG in the hypoglycemic range. This bias in the Navigator and Guardian devices is further evident in the underestimation in the mean PG obtained from each device. The mean PG across the 12 experiments as measured by the GlucoScout was 158 ± 20 mg/dL; the Navigator and Guardian underestimated the mean PG by 13 mg/dL (145 ± 17, z = −3.78, *P* = 0.0002) and 12 mg/dL (146 ± 26, z = −3.62, *P* = 0.0003), respectively, whereas the Seven Plus overestimated the mean PG by 5 mg/dL (163 ± 24, *z* = 2.67, *P* = 0.0075).

Owing to the high data density in the distributions shown in Fig. 3*A* and *C*, particularly over the range of PG values between 80 and 160 mg/dL, it is instructive to collect the data into frequency bins in the PG–RD plane of Fig. 3*A* and in the PG–ARD plane of Fig. 3*C* and generate histograms over the PG–RD plane (Fig. 3*B*) and PG–ARD plane (Fig. 3*D*), respectively. For the bin sizes shown in Fig. 3*B* and *D* (which span 7% by 7 mg/dL in the PG–RD and PG–ARD planes), the Navigator had bins with the highest number of PG–RD and PG–ARD pairs. Relative to the data obtained from the Seven Plus and Guardian, the data obtained from the Navigator are much more concentrated in the 0–7% error bins and show much less dispersion over the PG–RD and PG–ARD planes (Fig. 3), demonstrating graphically the greater accuracy and precision of Navigator estimates of PG.

When the MARD is calculated for the clinically relevant PG ranges of 70–120, 120–180, and 180–250 mg/dL (Fig. 4*D*), the Navigator outperformed the other two devices in MARD and SD of the MARD. Its performance was relatively better in the 70 to 120 mg/dL range than in the 180 to 250 mg/dL range, but the Seven Plus and Guardian each showed relatively similar performance across the three PG ranges, with the former outperforming the latter in a mean sense in all three.

CGM rate-of-change accuracy

We also evaluated the time-rate-of-change accuracy for each of the three CGM devices. Rate-of-change measurements from the reference PG data yielded 1,699 slopes from the twelve 48-h experiments. Similarly, time-rate-of-change data corresponding to these 1,699 reference values were extracted from the CGM data. The absolute value of the difference between the PG slopes and each of the corresponding CGM slopes was computed and averaged over the 1,699 paired slopes for all three CGM devices. On average, the time-rate-of-change error (relative to PG) for the Navigator was 0.66 ± 0.96 mg/dL/min compared with 0.86 ± 1.20 mg/dL/min for the Seven Plus (z = −2.94, *P* = 0.003 vs. Navigator) and 0.86 ± 1.26 mg/dL/min for the Guardian (z = −2.60, *P* = 0.009 vs. Navigator). The time-rate-of-change errors for the Seven Plus and Guardian were not significantly different (z = 0.01, *P* = 0.990). Figure 5 shows, for each of the three CGM devices, the absolute value of the time-rate-of-change error sorted into eight bins, where each bin includes all paired points in a particular range of absolute values of the time rate of change in PG. The largest physiological rise and fall in PG that we observed over a 15-min interval was 8.1 and 7.3 mg/dL/min, respectively.

CGM precision and reliability

When the mean and SD of all PG–ARD pairs associated with a particular PG value were computed for each PG value

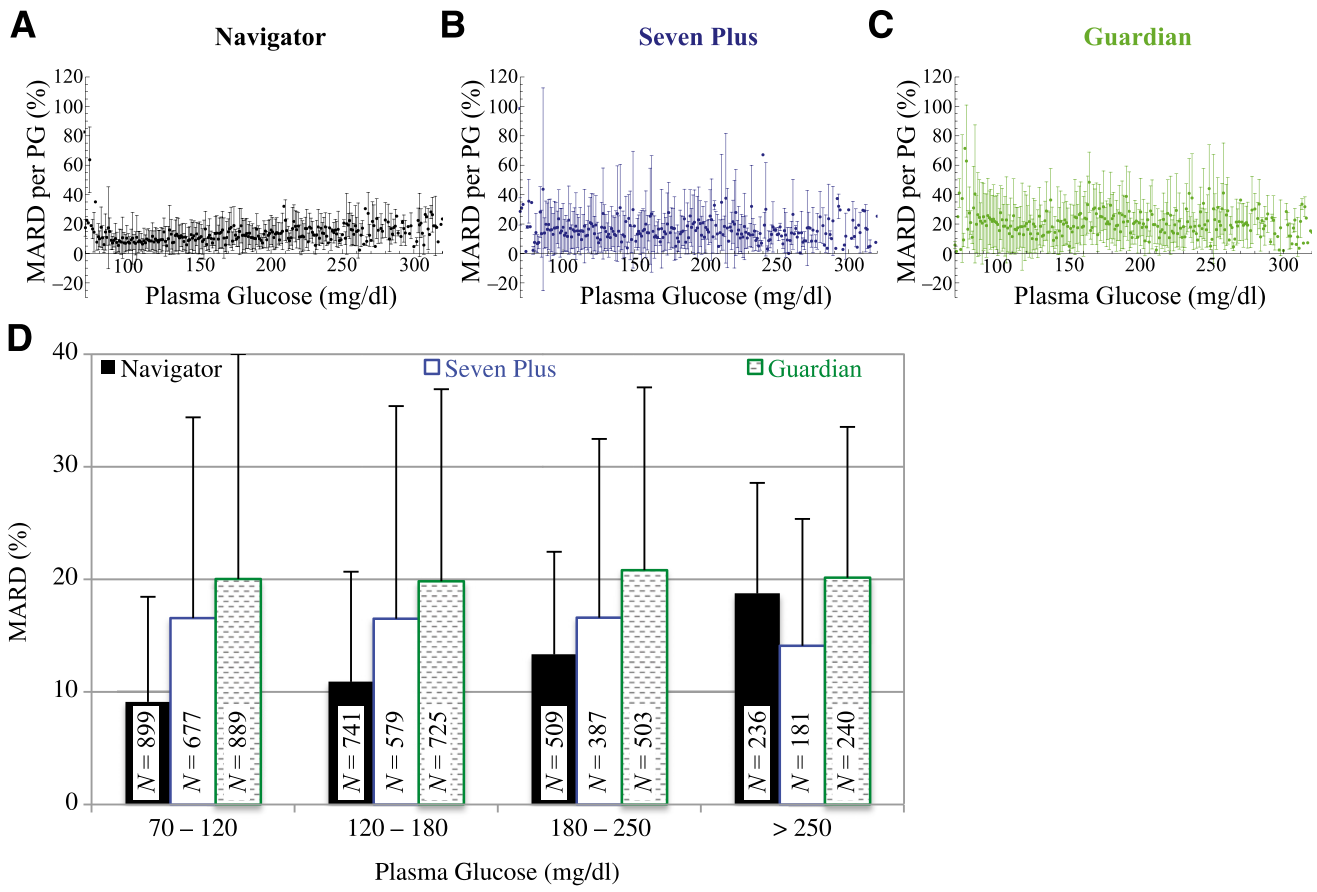


Figure 4—A–C: *The MARD and SD in the MARD corresponding to each PG value from 70 to 320 mg/dL for the Navigator, Seven Plus, and Guardian, respectively. Data points without error bars represent sole values for that particular PG value.* D: *The MARD and SD in the MARD corresponding to the clinically relevant PG ranges from 70–120, 120–180, 180– 250, and ≥250 mg/dL for the Navigator, Seven Plus, and Guardian. The number* (N) *of data in each PG range is shown in the corresponding bar for each device. For PG values in the normoglycemic range, from 70 to 120 mg/dL, the MARDs were 9.1 ± 9.3%* (N = 899)*, 16.5 ± 17.8%* (N = 677)*, and 20.0 ± 19.9%* (N = 889)*, for the Navigator, Seven Plus, and Guardian, respectively. Much less reliable, because of the small sample size obtained, are the data corresponding to PG values in the moderate-to-mild hypoglycemic range from 50 to 70 mg/dL (not shown here); in this range, the MARDs were 46 ± 33%* (N = 14)*, 31 ± 25%* (N = 11)*, and 36 ± 40%* (N = 14)*, for the Navigator, Seven Plus, and Guardian, respectively.*

between 70 and 320 mg/dL (Fig. 4*A*–*C*), the average SD at each PG value was much smaller for the Navigator (8.8 ± 3.9) than for the Seven Plus (13.9 ± 8.7) and the Guardian (15.1 ± 7.4), indicative of higher precision of the Navigator relative to the other two devices.

Occasionally, glucose values will be skipped during online operation of a CGM. When averaged across the 12 experiments, the data reporting percentages were 99.8 ± 0.6% for the Navigator, 75.9 ± 20.7% for the Seven Plus, and 98.5 ± 2.5% for the Guardian. Data reporting percentages for the three CGM devices for each experiment are provided in the legends of Supplementary Figs. 1–12.

CONCLUSIONS—Results from this head-to-head-to-head analysis of three commercially available CGM devices worn simultaneously by each subject for 2 days under the same experimental conditions are in remarkably good agreement with the results of studies reported in the manufacturers' own labeling. However, the direct comparison of the three CGM devices over a broad range of PG values is unique and, we think, compelling. The manufacturers' user guides for each of these devices report MARDs (relative to YSI BG measurements) of 12.8 ± 13.6% for the Navigator for 20,362 paired glucose values ≥20 mg/dL (compared with 11.8 ± 11.1% for 2,356 paired points in the current study), 16% for the Seven Plus for 1,827 paired glucose values between 40 and 400 mg/dL (compared with 16.5 ± 17.8% for 1,799 paired points in the current study), and 19.7 ± 18.4% for the Guardian for 3,941 paired glucose values ≥40 mg/dL (compared with 20.3 ± 18.0% for 2,328 paired points in the current study). Thus, the manufacturers' labeling for point accuracy is consistent with our findings in a direct comparison study, despite inevitable differences in study populations and conditions between the different manufacturers studies.

The clinical utility of the CGM devices, especially when applied to closed-loop BG control, depends not only on device accuracy but also on reliability. Interruption in the glucose data stream under open-loop glucose management requires the user to revert to SMBG therapy without trend information until data reporting resumes. Under automated closed-loop BG control, such interruptions would take the closed-loop system offline. Of the three CGM devices studied here, only the Seven Plus seemed prone to gaps in data reporting. Another metric of reliability is precision, as measured by the variability

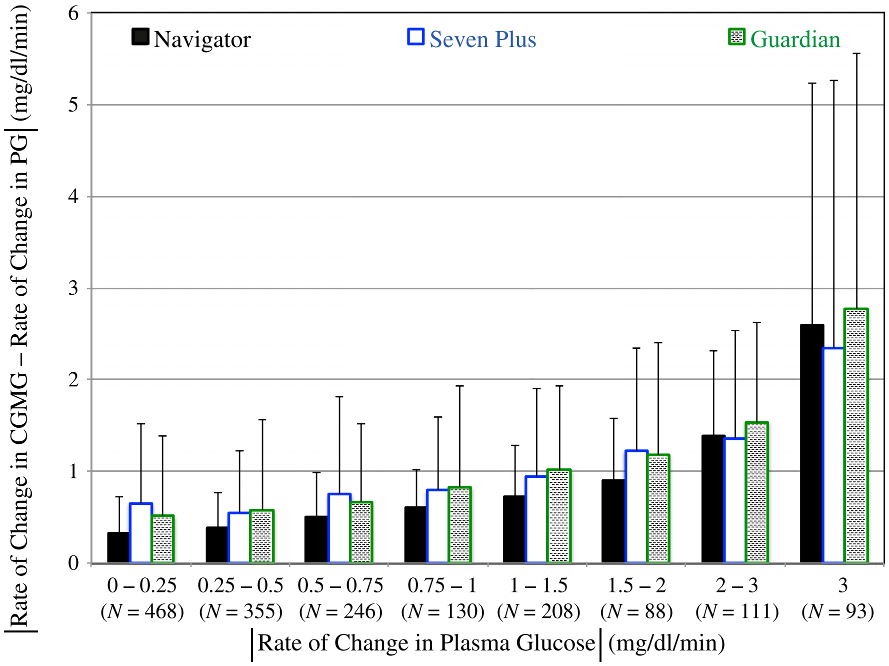


Figure 5—*The absolute value of the difference between the time rate of change in CGMG and the time rate of change in PG corresponding to eight different ranges in the absolute value of the time rate of change in PG (0–0.25, 0.25–0.5, 0.5–0.75, 0.75–1, 1–1.5, 1.5–2, 2–3, and ≥3 mg/dL/min) for the Navigator, Seven Plus, and Guardian. The number* (N) *of data in each range is shown below the range label.*

around mean performance. This was quantified here by the SD around the aggregate mean of all ARD values and around the mean of ARD values from each individual experiment from each CGM device. The latter confers information about the variation in performance of a CGM device from one sensor session to another and may be a more clinically useful concept than the SD around the aggregate mean. The Navigator variability was approximately one-half that of the other two CGM devices for both metrics.

In essentially every respect (aggregate MARD, MARD per experiment, precision, distribution of relative errors in the PG–RD plane, rate-of-change errors, and data reporting frequency), the results of the current study point to the Navigator as having the best performance in the normoglycemic and hyperglycemic range. Although the Guardian had comparable performance to the Navigator in data reporting frequency, it had numerically the worst performance for most of the metrics analyzed. Our conclusions about performance in the normoglycemic range are qualitatively and quantitatively different from those of a previous study that directly compared the accuracy of the Navigator and Guardian devices in the setting of a short-term glucose clamp study (3). That study concluded that the accuracy of the Guardian and Navigator was comparable in the normoglycemic range (3). All of their data were collected when the BG was clamped at 100 or 45 mg/dL, or during a slow transition (at a rate of 1 mg/dL/min) between these two BG values. Thus, no data were collected in the hyperglycemic range, and the effect of physiologic lag on accuracy was minimized by the negligible or low time rates of change in BG during the measurement period. In contrast, our data include many comparisons in the hyperglycemic range, and we sampled a much broader range of rates of change of BG (up to −7 and +8 mg/dL/min for short periods of time). Further, the number of paired BG-CGMG points in the normoglycemic range was at least threefold greater than in the study of Kovatchev et al. (3), and each experiment was much longer, allowing us to observe sensor inaccuracies associated with sensor drift, which is a common source of error for CGM devices. Finally, the study of Kovatchev et al. (3) did not show, as we did, the degree of variability in the accuracy of each sensor, a critical determinant of its reliability. Therefore, the results of our study are more informative regarding the suitability of each CGM device as the input sensor for closed-loop BG control.

One of the purposes of this head-to-head-to-head comparison study was to determine whether the Seven Plus and/or Guardian could substitute for the Navigator in closed-loop BG control. In the closed-loop experiments from which these data are derived (4), the Navigator served as the sole input to a fully autonomous system that successfully regulated BG continuously over a 2-day period (average PG of 158 mg/dL, with PG <70 mg/dL <0.7% of the time) (4). Other closed-loop studies have used the Seven Plus or Guardian and reported MARDs for those devices that were much better than the MARDs we found for those devices in this study and were comparable to the MARD we found for the Navigator (5–9). However, some of these studies report switching between multiple sensors based on reference data and/or calibrating sensors more frequently than recommended by the manufacturer (on average every 4–6 h) (7–9), while the others report inserting two sensors on each subject (5,6) without providing details about when and if switching between sensors occurred.

High-frequency calibrations are impractical in outpatient usage, and switching between multiple sensors based on frequently sampled BG undermines system autonomy; results of experiments using these strategies will not be representative of system performance in routine outpatient usage. Thus, it is not clear whether the Seven Plus or Guardian devices are accurate or reliable enough to serve as the sole input to an autonomous closed-loop BG control system when calibrated at a practical clinical frequency and operating without the benefit of frequently sampled BG to monitor their accuracy. Evidence that the Seven Plus or Guardian devices may not meet this standard was apparent in several of the experiments conducted in this study. There were repeated instances during which the Seven Plus and Guardian devices showed aberrant behavior that would likely have led a control algorithm to severely underdose insulin on some occasions and to severely overdose on others. There were three occasions each for the Seven Plus and Guardian when the devices overestimated the subject's glucose by ≥70 mg/dL for a period of 1–5 h (Supplementary Figs. 2 and 12 for the Seven Plus and Supplementary Figs. 1 and 12 for the Guardian), which would have resulted in overdosing insulin. Conversely, we observed one occasion for the Seven Plus (Supplementary Fig. 1) and seven for the Guardian (Supplementary Figs. 1, 2, 7, 9, and 10) when the devices failed to detect large postprandial glucose excursions around meals, and underdosing insulin would have resulted. Finally, we observed one occasion for the Seven Plus (Supplementary Fig. 1) and three occasions for the Guardian (Supplementary Figs. 1, 2, and 6) when the devices falsely predicted severe hypoglycemia for ≥3 h.

One of the limitations of our analysis was that the data were collected as part of a closed-loop study and, therefore, contained relatively few points <70 and >250 mg/dL. Glucose values were thus concentrated in a narrower range than typically arises in standard-of-care type 1 diabetes therapy. In particular, our data do not allow us to assess the accuracy of the three sensors in the hypoglycemic range (BG <70 mg/dL). When comparing the accuracy of the Guardian and Navigator, Kovatchev et al. (3) concluded that the Navigator had better accuracy in the hypoglycemic range. However, that study censored data in the hypoglycemic range whenever the CGMG was at the low threshold for that CGM device and was not changing with respect to time (3). Furthermore, different percentages of the data were censored for the two sensors (3). This approach undermines the applicability of their analysis to closed-loop control because a control decision must be rendered at each time step under closed loop.

Another limitation of our work is that although the timing of calibrations was strictly followed according to the manufacturers' specifications, the calibrations were done using reference-quality PG rather than capillary SMBG measurements. These factors could have led us to overestimate the accuracy of the CGM devices when used as a part of current standard-of-care therapy.

An additional limitation is that our dataset, although containing a large number of BG-CGMG pairs, was collected from 12 experiments in six subjects and therefore may not sample as much biological variability as a study in which fewer measurements were collected from each of a larger number of participants. A post hoc analysis revealed that there was nearly as much variation in the accuracy ranking of the three CGM devices between experiments in a single subject as there was between subjects, suggesting that the results were not due to the chance inclusion of subjects who idiosyncratically were capable of achieving better performance with one sensor than another (data not shown).

Although the performance differences we observed between the Seven Plus and Guardian are not as pronounced as between the Navigator and Seven Plus, the Seven Plus demonstrates consistently better point accuracy and comparable rate-of-change accuracy compared with the Guardian. However, an evident disadvantage of the Seven Plus relative to the Guardian lies in its lower data reporting frequency. This weakness is less critical under open-loop than under closed-loop control. According to Dexcom representatives, leaving the receiver device plugged in to its charger during the experiments might have contributed to the poor reporting frequency. However, we observed that gaps in reporting were not randomly distributed but tended much more often to occur during times when the Seven Plus CGMG was changing rapidly (typically >2 mg/dL/min), suggesting that loss of reporting may be related to filters in the BG estimation algorithm.

The results of this head-to-head-to-head comparative effectiveness study reveal the Navigator was the most accurate and precise of the current generation of CGM devices, followed by the Seven Plus and the Guardian. Integration of the Navigator into a truly autonomous closed-loop BG control system provides a demonstration of the clinical utility of the Navigator in driving that system (4). Combining those findings with results of the current study provides quantitative benchmarks for accuracy and reliability for a CGM device to serve as sole input for a closed-loop BG control system. Further study is required to determine whether the Seven Plus or Guardian, calibrated according to manufacturer's directions, would be sufficiently accurate and reliable for effective closed-loop BG control in a clinical protocol that does not undermine the autonomy of the system by acting on the knowledge of frequently sampled PG values. In light of the results of our analysis, it is unfortunate that the manufacturer has recently withdrawn the Navigator from the North American market. We are currently using the same methodology described in this report to compare the performance of the next-generation Navigator with the next-generation devices from DexCom and Medtronic.

Acknowledgments—This study was supported by National Institutes of Health (NIH) Grant R01-DK085633 to E.R.D., grants M01-RR01066 and UL1-RR025758 through the General Clinical Research Center and Clinical and Translational Science Center programs from the NIH National Center for Research Resources, Clinical Investigations Research Grant 22-2009-798 to E.R.D. from the Juvenile Diabetes Research Foundation, and a grant to D.M.N. from the Charlton Fund for Innovative Research in Diabetes.

No potential conflicts of interest relevant to this article were reported.

E.R.D. designed the study, performed the analysis, interpreted the data, wrote the first draft of the manuscript, and participated in revision of the manuscript for important intellectual content. F.H.E.-K. and D.M.N. designed the study, performed the analysis, interpreted the data, and participated in revision of the manuscript for important intellectual content. H.Z. performed the analysis, interpreted the data, and participated in revision of the manuscript for important intellectual content. S.J.R. designed the study, supervised the human studies, performed the analysis, interpreted the data, and participated in revision of the manuscript for important intellectual content. S.J.R. 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.

Parts of this study were presented at the 72nd Scientific Sessions of the American Diabetes Association, Philadelphia, Pennsylvania, 8–12 June 2012.

The authors thank the volunteers for their time and enthusiasm; the diabetes care providers who referred potential subjects for the study; the nurses and laboratory staff of the Massachusetts General Hospital Clinical Research Center, especially Kathy Hall and Kathy Grinke, and the study staff at the Diabetes Research Center, including Kendra Magyar, Kerry Grennan, Richard Pompei, Cathy Beauharnais, and Laurel Macey, for their dedicated effort and careful execution of the experimental protocol; Mary Larkin, Camille Collings, and Nancy Kingori, Diabetes Research Center, Massachusetts General Hospital, for organizational and logistical support; John Segars and Jennifer Isenberg, International Biomedical, for providing GlucoScout monitors and technical assistance in their use; and John Godine, Deborah Wexler, and Carl Rosow for serving on the data safety and monitoring board for the study, and the members of the Partners Human Research Committee and Boston University Medical Campus Institutional Review Board for their oversight of the study.

References

1. Maran A, Crepaldi C, Tiengo A, et al. Continuous subcutaneous glucose monitoring in diabetic patients: a multicenter analysis. Diabetes Care 2002;25:347–352

2. Wentholt IM, Vollebregt MA, Hart AA, Hoekstra JB, DeVries JH. Comparison of a needle-type and a microdialysis continuous glucose monitor in type 1 diabetic patients. Diabetes Care 2005;28:2871–2876
3. Kovatchev B, Anderson S, Heinemann L, Clarke W. Comparison of the numerical and clinical accuracy of four continuous glucose monitors. Diabetes Care 2008;31: 1160–1164
4. Russell SJ, El-Khatib FH, Nathan DM, Magyar KL, Jiang J, Damiano ER. Blood glucose control in type 1 diabetes with a bihormonal bionic endocrine pancreas. Diabetes Care 2012;35:2148–1215
5. Steil GM, Rebrin K, Darwin C, Hariri F, Saad MF. Feasibility of automating insulin delivery for the treatment of type 1 diabetes. Diabetes 2006;55:3344–3350
6. Weinzimer SA, Steil GM, Swan KL, Dziura J, Kurtz N, Tamborlane WV. Fully automated closed-loop insulin delivery versus semiautomated hybrid control in pediatric patients with type 1 diabetes using an artificial pancreas. Diabetes Care 2008;31: 934–939
7. Castle JR, Engle JM, El Youssef J, et al. Novel use of glucagon in a closed-loop system for prevention of hypoglycemia in type 1 diabetes. Diabetes Care 2010;33: 1282–1287
8. Steil GM, Palerm CC, Kurtz N, et al. The effect of insulin feedback on closed loop glucose control. J Clin Endocrinol Metab 2011;96:1402–1408
9. Hovorka R, Allen JM, Elleri D, et al. Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial. Lancet 2010; 375:743–751

Supplementary Data

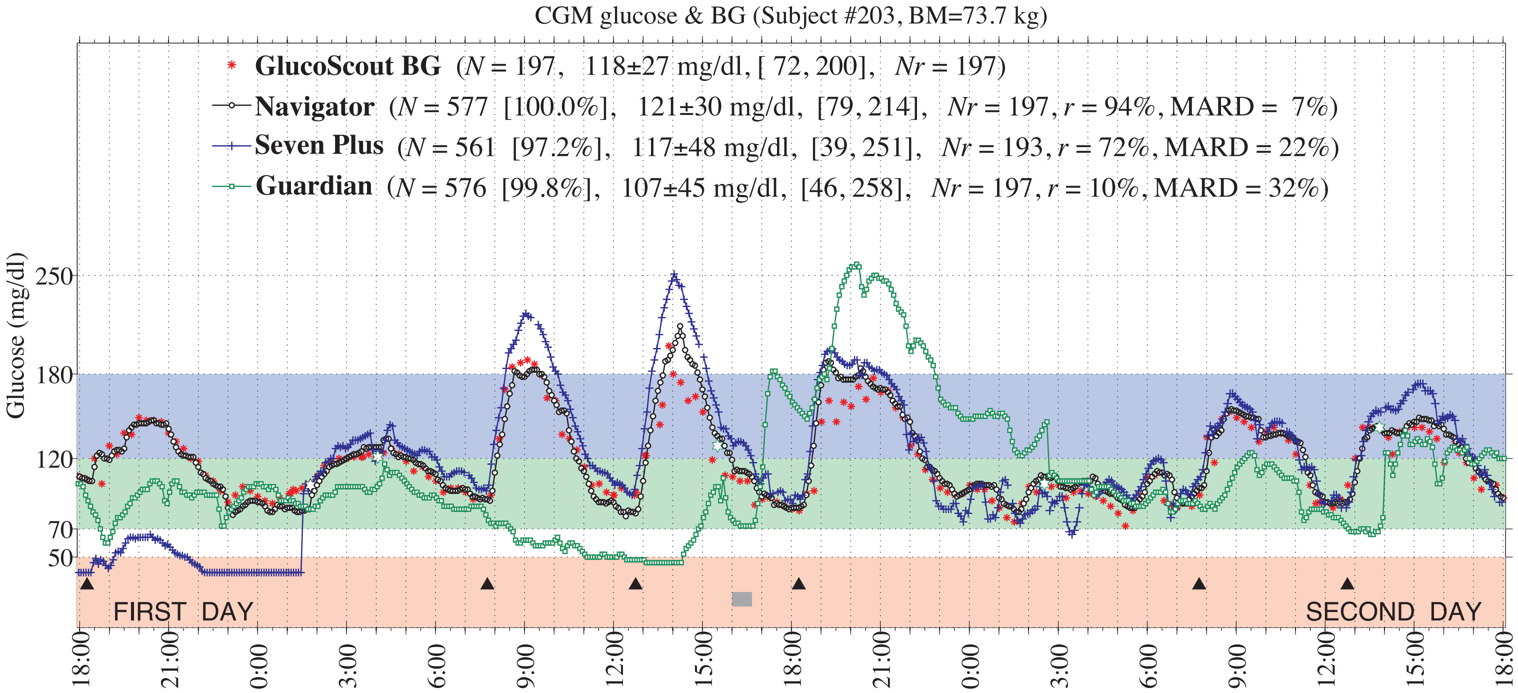
Damiano *et al.* – A comparative effectiveness analysis of three continuous glucose monitors in type 1 diabetes

1 CGM Calibrations

The manufacturer's calibration schedule for the Navigator requires calibrations at approximately 0, 2, 14, and 62 hours from the first calibration request, whereas the Seven Plus and Guardian require calibrations approximately every 12 hours from the first calibration. All three CGM devices were inserted the day before the experiment. The first two calibrations of the Navigator, and the first calibration of the Seven Plus and the Guardian were typically performed in the two to three hour period leading up to the experiment. Therefore, in each experiment, one scheduled calibration was requested by the Navigator and four scheduled calibrations were requested by the Seven Plus and the Guardian. Occasionally, a CGM device would request an additional calibrations beyond the usual scheduled calibrations; these were performed whenever requested. Additionally, there was a provision in the protocol to force a calibration of any of the three CGM devices at 6:00 daily if the glucose level displayed by a device was not within the International Organization for Standardization (ISO) standard compared with PG; namely within 20% of PG if the PG $>$ 75 mg/dl or within 15 mg/dl of PG if the PG $<$ 75 mg/dl. According to these criteria, of 24 possible occasions, this calibration was required and performed twice for the Guardian and once for the Seven Plus. There were two occasions when the Seven Plus was not reporting data at 6:00, and therefore these criteria could not be evaluated. There was one occasion when the Navigator and Seven Plus did not meet ISO standards at 6:00; according to protocol forced calibrations should have been performed at that time, but were omitted in error.

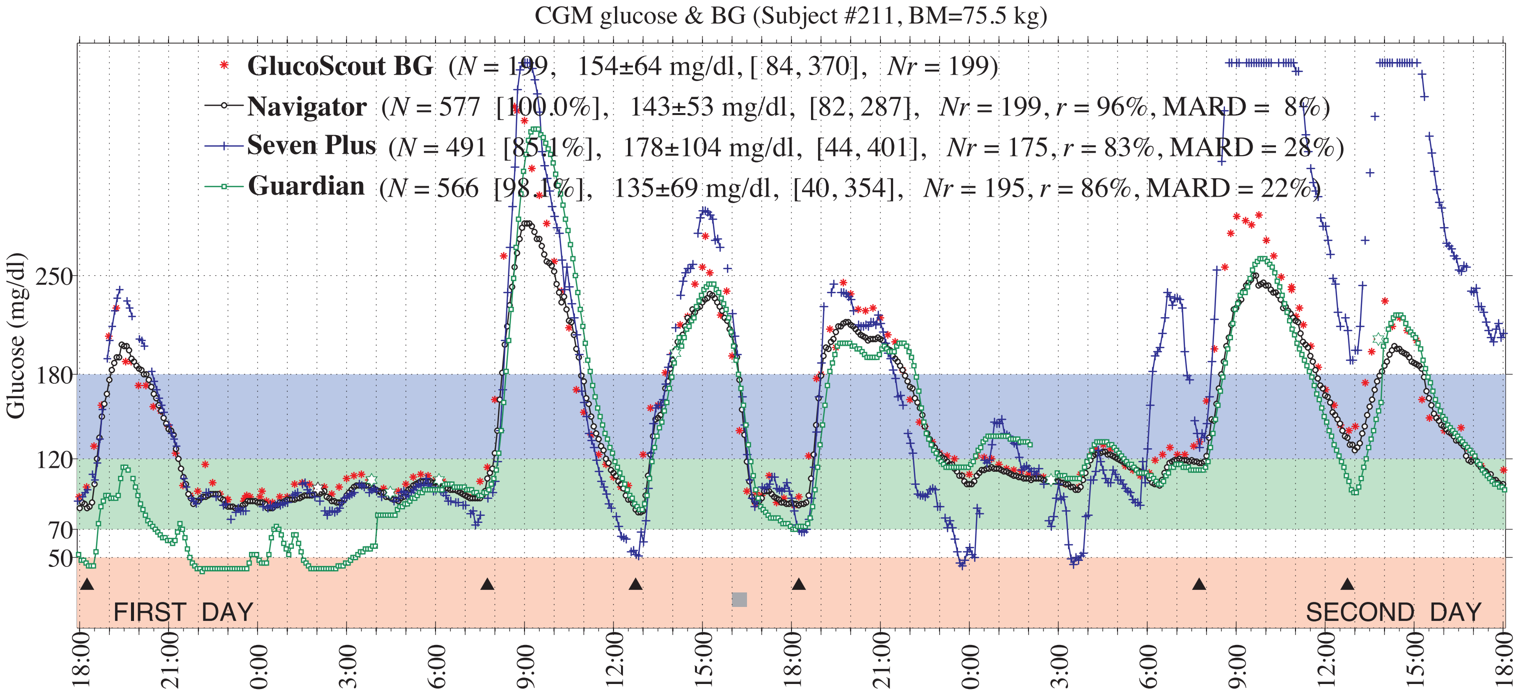
2 Boundedness of Device MARDs

The aggregate MARD of all paired points obtained for each CGM device (Fig. 2) is a quantity that is bounded from above and below. When the 2356 PG data points used to test the Navigator accuracy are randomly shuffled and then paired with the 2356 Navigator data points, the new paired data set is found to consistently produce an aggregate MARD of $\sim$ 41% through many random shuffling trials. This represents an upper bound on the aggregate MARD that the Navigator could achieve if there were no relationship between the Navigator CGMG and the reference PG. When a similar random shuffling is performed on the 1799 paired data points obtained for the Seven Plus and the 2328 paired data points obtained for the Guardian, the upper-bound aggregate MARDs for these data sets are found to be 54% and 47%, respectively. Note that the Clarke error grid for the GlucoScout (Fig. 2*A*) shows an aggregate MARD of 5.1%. Since the GlucoScout and YSI measured the glucose concentration of the same blood sample from the same IV, this MARD is arguably the best that can be achieved with reference-quality glucose monitors. Thus, the aggregate MARD of 11.8 ± 11.1% for the Navigator falls in the range of possible MARDs of between 5 and 41%, the aggregate MARD of 16.5 ± 17.8% for the Seven Plus falls in the range of possible MARDs of between 5 and 54%, and the aggregate MARD of 20.3 ± 18.0% for the Guardian falls in the range of possible MARDs of between 5 and 47%.



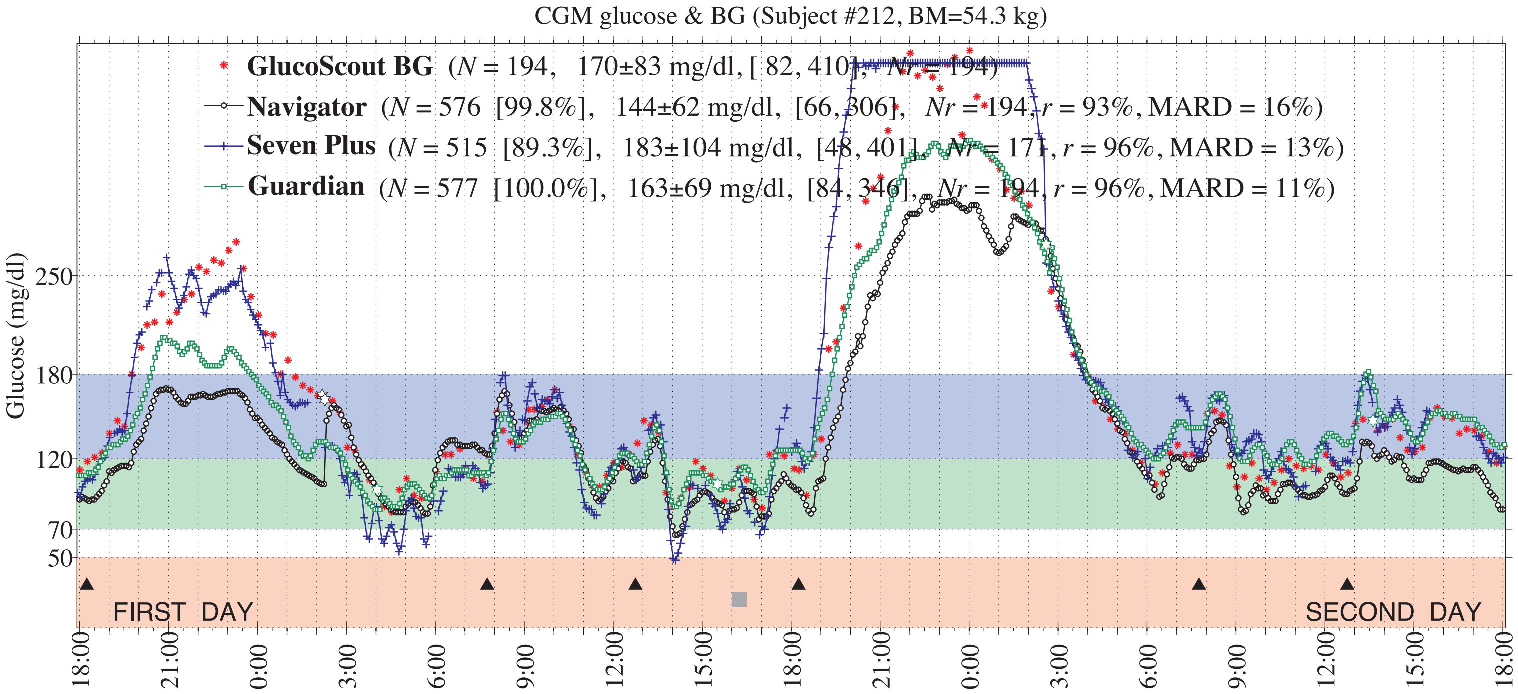
Supplementary Figure 1.

Results obtained during 48 hours of continuous closed-loop control in Subject #203 showing venous BG concentrations measured every 15 minutes with the GlucoScout (red symbols) and CGMG values measured with the Navigator (black symbols), Seven Plus (blue symbols), and Guardian (green symbols). Meals are indicated along the timeline by black triangles. A 30–40 minute period of structured exercise occurred at 16:00 hours at the end of the First Day and is indicated along the timeline by the grey rectangle. Note the 7.5-hour period from 18:00–1:30 hours in which the Seven Plus essentially missed the hyperglycemic excursion associated with the first meal, and falsely predicted severe hypoglycemia during most of this period. Note the 20-hour period from 7:30–3:00 hours in which the Guardian essentially missed the hyperglycemic excursions associated with the second and third meals, and the 7.5-hour period from 19:00–2:30 hours in which the Guardian severely over-estimated glucose during the fourth meal.



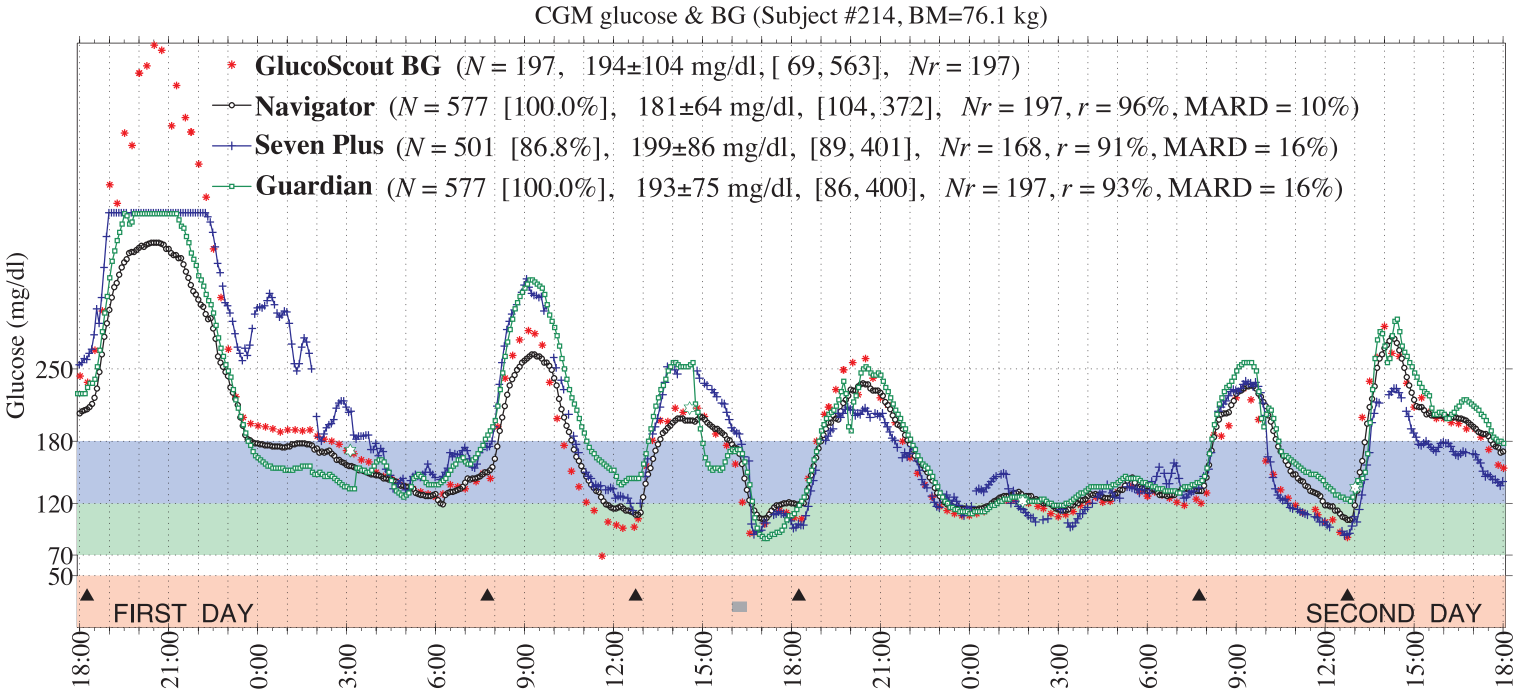
Supplementary Figure 2.

Results obtained during 48 hours of continuous closed-loop control in Subject #211. Note the 12-hour period from 6:00–18:00 hours in which the Seven Plus severely over-estimated glucose during the fifth and sixth meals. Note the 10-hour period from 18:00–4:00 hours in which the Guardian essentially missed the hyperglycemic excursion associated with the first meal, and falsely predicted severe hypoglycemia during some of this period.



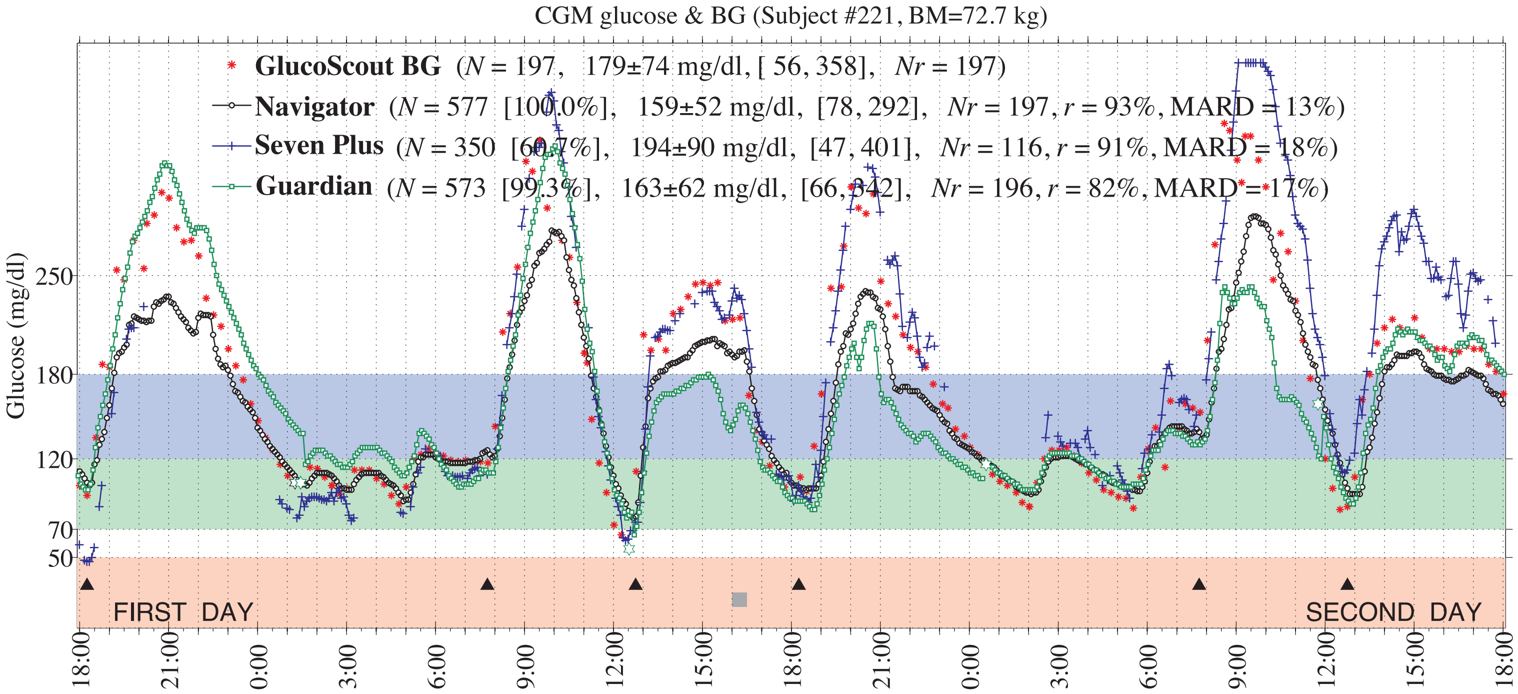
Supplementary Figure 3.

Results obtained during 48 hours of continuous closed-loop control in Subject #212.



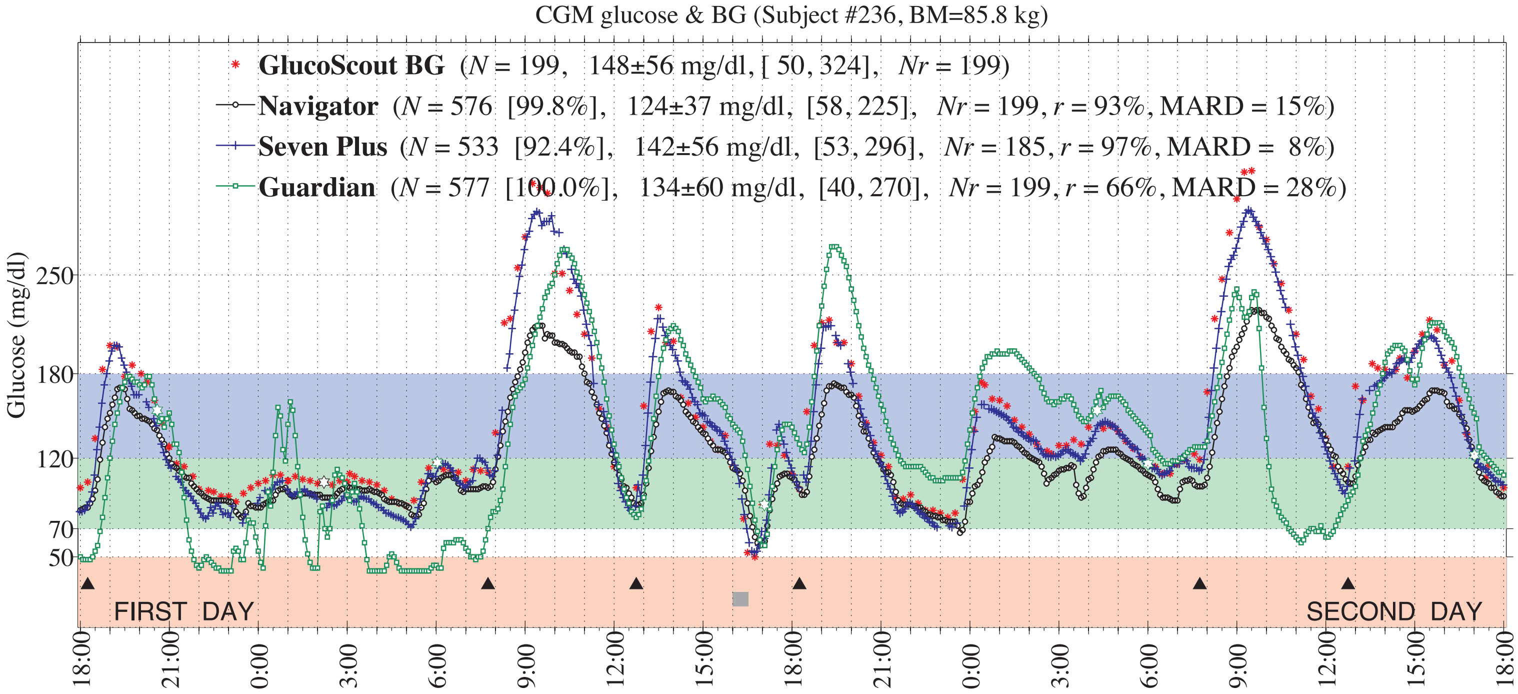
Supplementary Figure 4.

Results obtained during 48 hours of continuous closed-loop control in Subject #214.



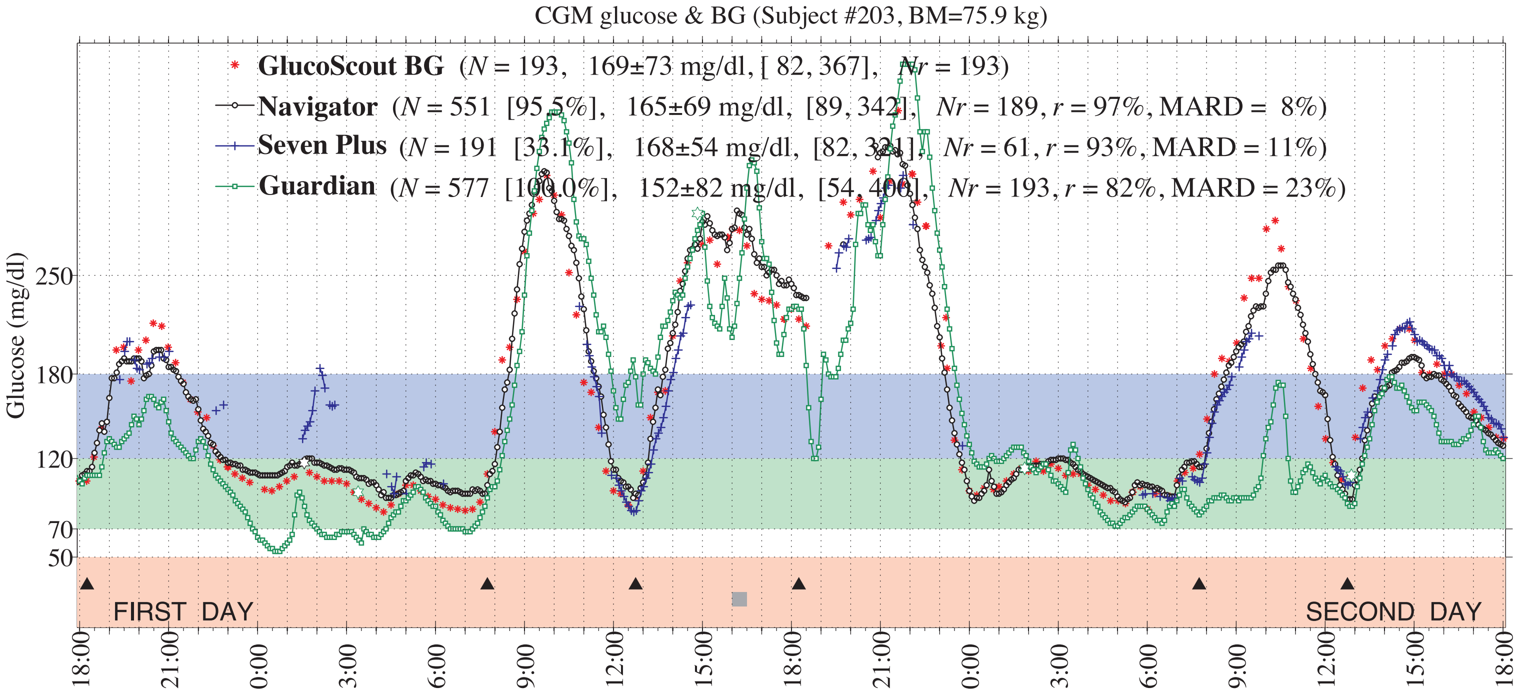
Supplementary Figure 5.

Results obtained during 48 hours of continuous closed-loop control in Subject #221.



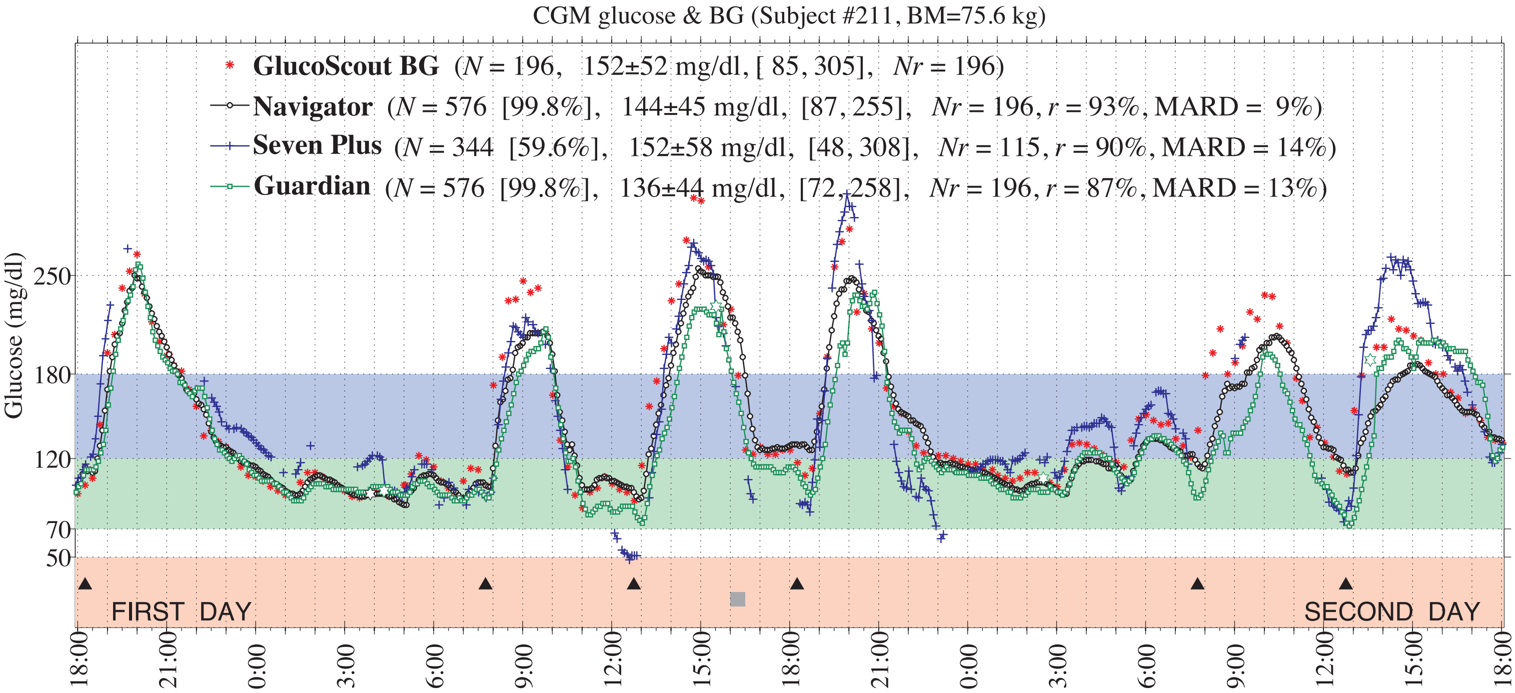
Supplementary Figure 6.

Results obtained during 48 hours of continuous closed-loop control in Subject #236. Note the 9-hour period of normoglycemia from 22:00–7:00 hours on the first night in which the Guardian falsely predicted severe hypoglycemia during most of this period.



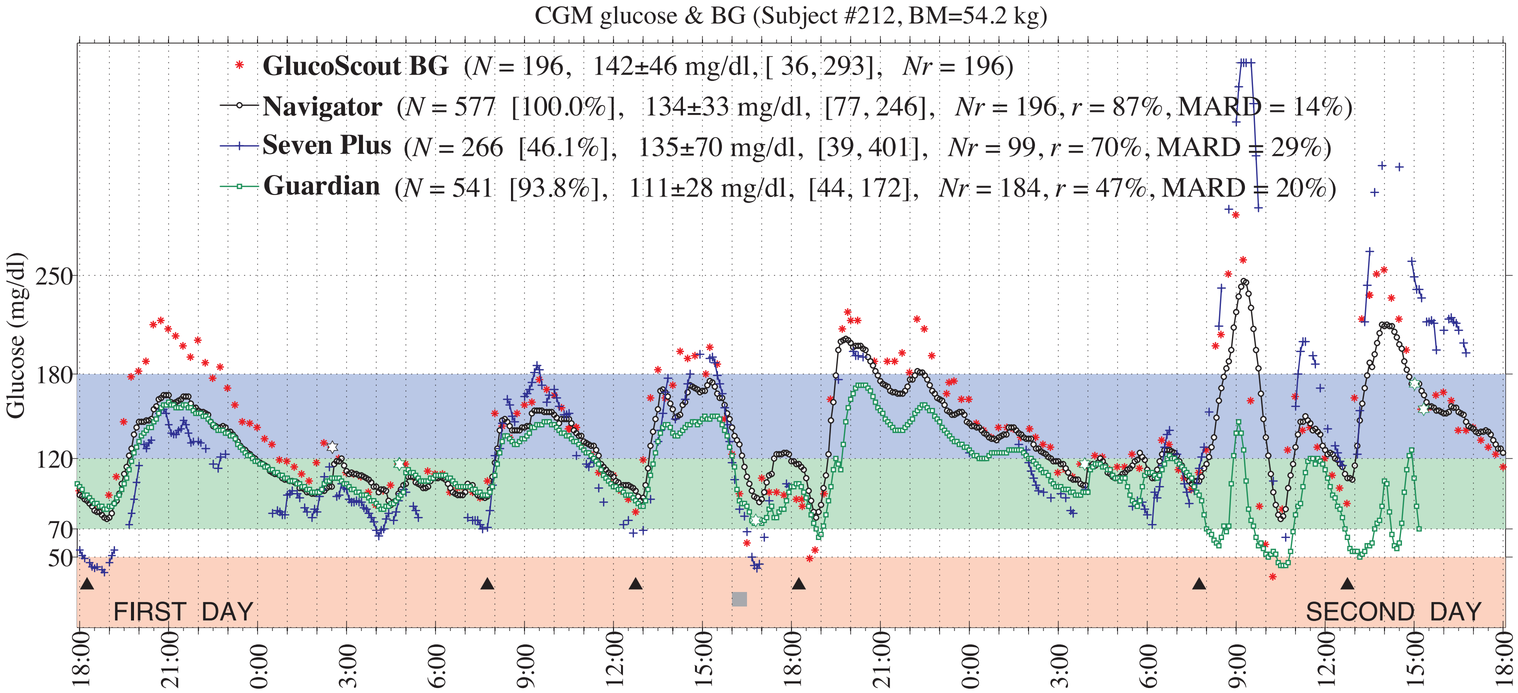
Supplementary Figure 7.

Results obtained during 48 hours of continuous closed-loop control in Subject #203. Note the 5-hour period from 7:00–12:00 hours in which the Guardian essentially missed the hyperglycemic excursion associated with the fifth meal. The closed-loop system had to be restarted between 18:30–20:45 as the fourth meal was ending. Because the Navigator data was streamed to the closed-loop system, it was not recorded during this interval. Consequently, the reporting percentage of the Navigator in this experiment is a lower-bound estimate.



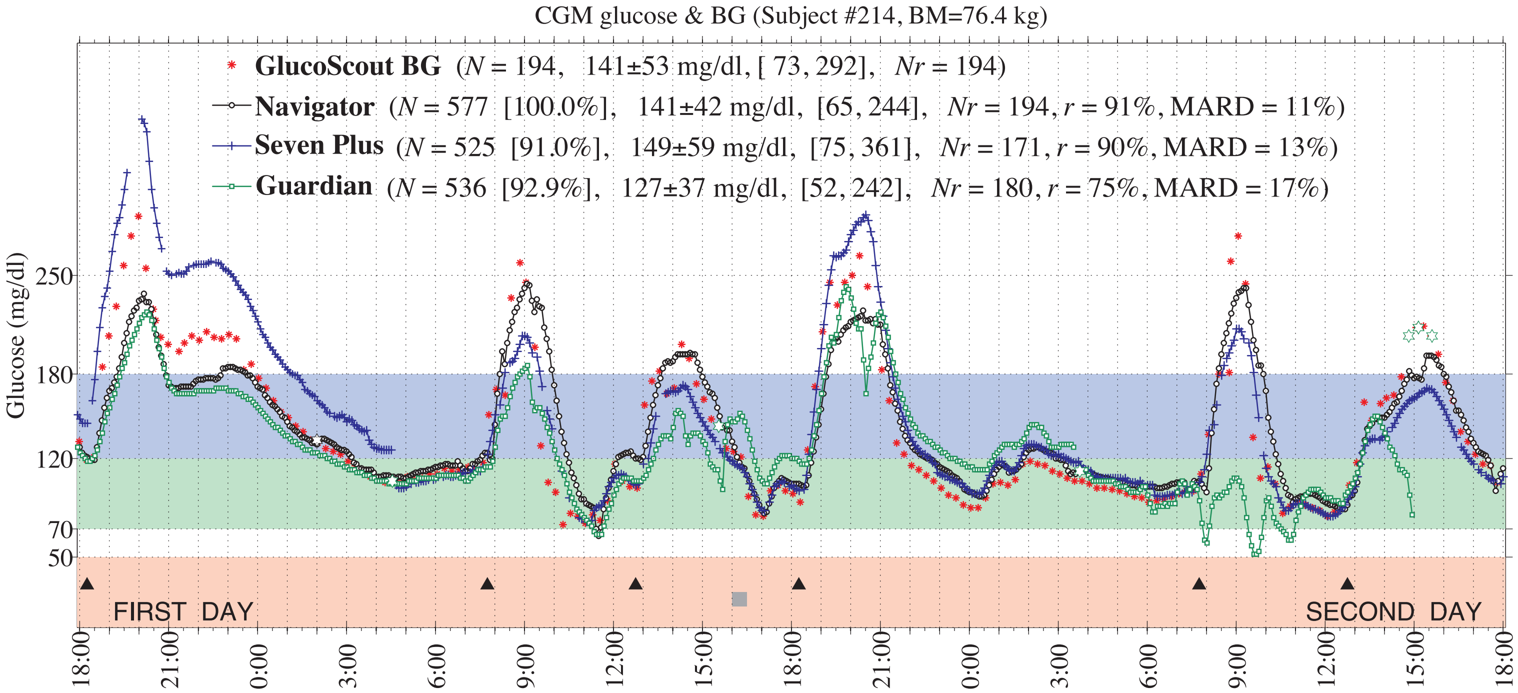
Supplementary Figure 8.

Results obtained during 48 hours of continuous closed-loop control in Subject #211.



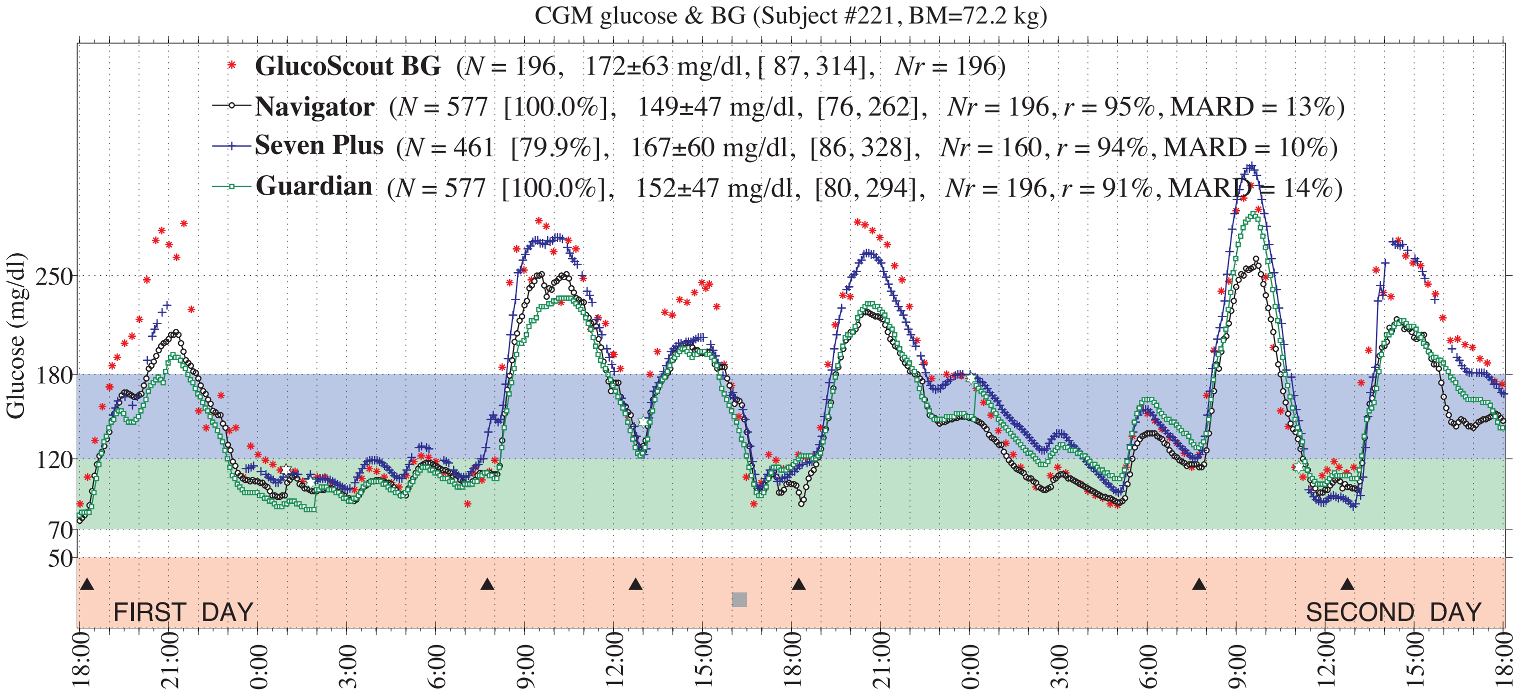
Supplementary Figure 9.

Results obtained during 48 hours of continuous closed-loop control in Subject #212. Note the 5-hour period from 13:00–18:00 hours in which the Guardian essentially missed the hyperglycemic excursion associated with the sixth meal.



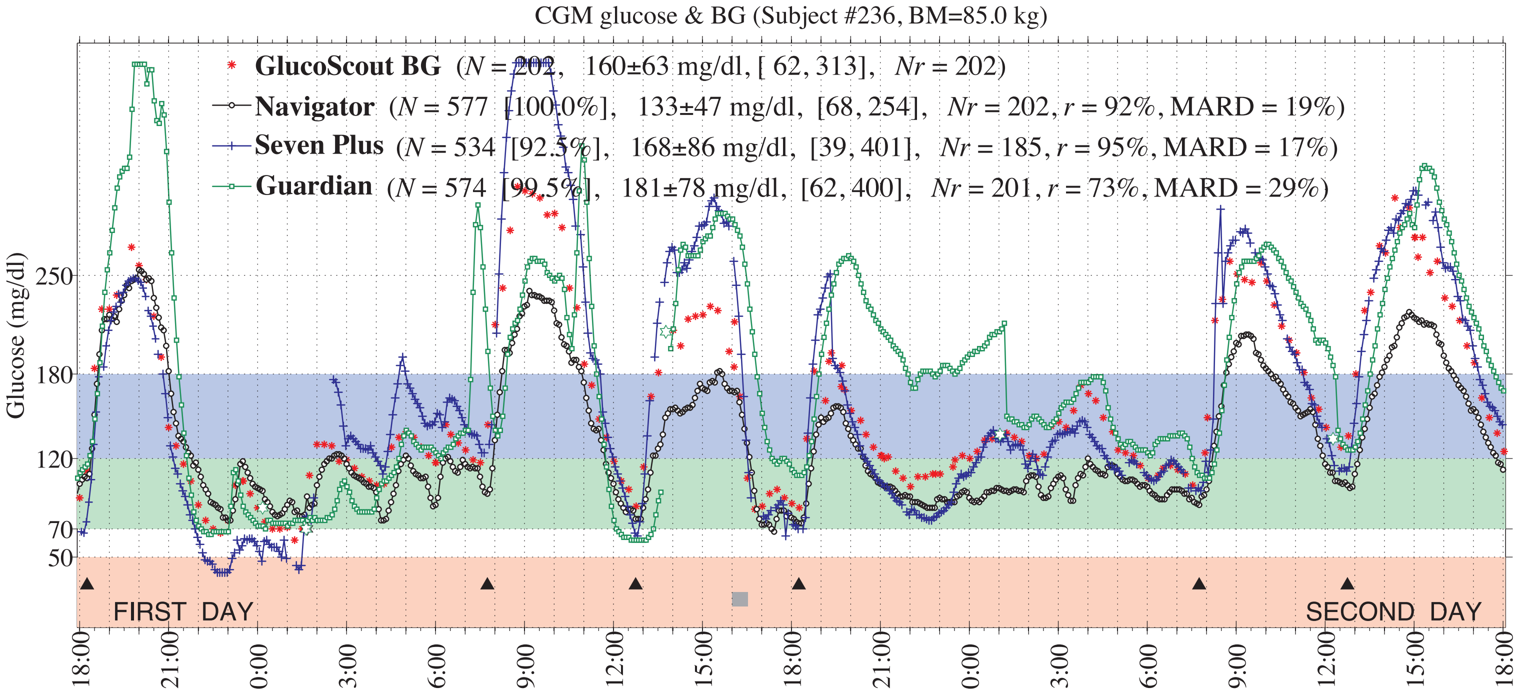
Supplementary Figure 10.

Results obtained during 48 hours of continuous closed-loop control in Subject #214. Note the 2-hour period from 8:00–10:00 hours and the 4-hour period from 14:00–18:00 hours in which the Guardian essentially missed the hyperglycemic excursions associated with the fifth and sixth meals, respectively.



Supplementary Figure 11.

Results obtained during 48 hours of continuous closed-loop control in Subject #221.



Supplementary Figure 12.

Results obtained during 48 hours of continuous closed-loop control in Subject #236. Note the 2-hour period from 8:30–10:30 hours in which the Seven Plus severely over-estimated glucose during the second meal. Note the 2-hour period from 19:00–21:00 hours and the 6-hour period from 19:00–1:00 hours in which the Guardian severely over-estimated glucose during the first and fourth meals, respectively.