Journal of Diabetes Science and Technology
2014, Vol. 8(4) 699–708

DOI: 10.1177/1932296814532203
dst.sagepub.com
SAGE

A Comparative Effectiveness Analysis of Three Continuous Glucose Monitors: The Navigator, G4 Platinum, and Enlite

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

Abstract

Background: The effectiveness and safety of continuous glucose monitors (CGMs) is dependent on their accuracy and reliability. The objective of this study was to compare 3 CGMs in adult and pediatric subjects with type 1 diabetes under closed-loop blood-glucose (BG) control.

Methods: Twenty-four subjects (12 adults) with type 1 diabetes each participated in one 48-hour closed-loop BG control experiment. Venous plasma glucose (PG) measurements obtained every 15 minutes (4657 values) were paired in time with corresponding CGM glucose (CGMG) measurements obtained from 3 CGMs (FreeStyle Navigator, Abbott Diabetes Care; G4 Platinum, Dexcom; Enlite, Medtronic) worn simultaneously by each subject.

Results: The Navigator and G4 Platinum (G4) had the best overall accuracy, with an aggregate mean absolute relative difference (MARD) of all paired points of 12.3 ± 12.1% and 10.8 ± 9.9%, respectively. Both had lower MARDs of all paired points than Enlite (17.9 ± 15.8%, $P < .005$). Very large errors (MARD > 50%) were less common with the G4 (0.5%) than with the Enlite (4.3%, $P = .0001$) while the number of very large errors with the Navigator (1.4%) was intermediate between the G4 and Enlite ($P = .1$ and $P = .06$, respectively). The average MARD for experiments in adolescent subjects were lower than in adult subjects for the Navigator and G4, while there was no difference for Enlite. All 3 devices had similar reliability.

Conclusions: A comprehensive head-to-head-to-head comparison of 3 CGMs revealed marked differences in both accuracy and precision. The Navigator and G4 were found to outperform the Enlite in these areas.

Keywords

accuracy, blood glucose, blood glucose meter, CGM, continuous glucose monitoring, reliability

There is an ongoing need to evaluate the relative accuracy and reliability of commercially available continuous glucose monitors (CGMs) over a large range of blood glucose (BG) values and rates of change in BG values. A recent study published by our group[1] evaluated 3 CGMs under these conditions in volunteers with type 1 diabetes who participated in closed-loop BG control experiments. That head-to-head-to-head comparative study revealed that the FreeStyle Navigator (Abbott Diabetes Care, Alameda, CA) outperformed both the Seven Plus (Dexcom, San Diego, CA) and the Guardian with Sof-sensor (Medtronic, Northridge, CA) in accuracy, precision, and reliability.[1] A study testing the accuracy of these sensors over a longer period of wear reached very similar conclusions.[2] All of these devices have now been superseded by next-generation technologies.

The present analysis compares 3 CGM sensors, the FreeStyle Navigator (Abbott Diabetes Care), the G4 Platinum (Dexcom, San Diego, CA), and the Enlite (Medtronic, Northridge, CA). The G4 Platinum (G4) and Enlite were not available at the time of the previous study. The Navigator was included for comparison with the previous study. The present study was conducted in 12 adults (≥ 21 years old) and 12 adolescent subjects (12-20 years old) with type 1 diabetes as part of a closed-loop BG control study.[3] The 3 sensors were worn simultaneously by each subject while reference-quality plasma glucose (PG) levels were measured every 15 minutes continually over 48 hours. Results were analyzed in

[1]Department of Biomedical Engineering, Boston University, Boston, MA, USA
[2]Biostatistics Center, Massachusetts General Hospital, Boston, MA, USA
[3]Diabetes Research Center, Massachusetts General Hospital, Boston, MA, USA

Corresponding Author:
Steven J. Russell, MD, PhD, Diabetes Research Center, Massachusetts General Hospital, 50 Staniford St, Ste 301, Boston, MA 02114, USA.
Email: sjrussell@mgh.harvard.edu

terms of point accuracy, rate-of-change accuracy, and sensor reliability. In addition to using next-generation devices, the present study (1) evaluates the devices in twice as many adult subjects as the previous study, (2) extends the study paradigm to an adolescent population, (3) evaluates the frequency of large errors, (4) compares the performance of each CGM in adolescent versus adult subjects, and (5) compares daytime and nighttime performance.

Methods

Subjects

The clinical protocol was approved by the Massachusetts General Hospital (MGH) and Boston University Human Research Committees. All subjects gave written informed consent. Subjects were required to be 12 years of age or older, to have had type 1 diabetes for at least 1 year, and to have a stimulated C-peptide level ≤ 0.1 nmol/l. Each subject completed one 48-hour experiment.

Experimental Protocol

Subjects were admitted to the MGH Clinical Research Center wearing Navigator, G4, and Enlite transmitters and sensors, which were inserted at approximately 15:00 the previous day. This was done to allow the sensors to equilibrate in vivo and allow any micro-hematoma that developed after insertion to resolve. The transmitters were linked wirelessly to their receivers. Approximately 24 hours after insertion, initial calibrations of the 3 CGMs were performed according to the manufacturers' instructions, except that venous PG measurements (performed with the GlucoScout) rather than capillary BG values, were used for calibration. During each experiment, the Navigator had 1 additional scheduled calibration. The manufacturers' instructions for the G4 and Enlite required 4 additional scheduled calibrations that were performed before breakfast and dinner daily. If the CGM glucose (CGMG) reading of the Navigator did not meet the International Organization for Standardization (ISO) standard for accuracy before breakfast or dinner, then an extra forced calibration was performed (see Supplemental Data). Additional calibrations for the CGMs were performed when requested.

Fully automated closed-loop BG control was initiated at 15:00 and ran continuously for 51 hours; the last 48 hours of each experiment were included in this analysis. Venous PG levels were measured every 15 minutes with the GlucoScout (International Biomedical, Austin, TX) and confirmed hourly with a YSI 2300 STAT Plus Analyzer (YSI Life Sciences, Yellow Springs, OH). Subjects ate 6 high-carbohydrate meals over the course of 2 days and 2 nights, and participated in 1 exercise session involving approximately 30 minutes of moderate intensity cycling on a stationary bicycle with a heart rate goal of 120-140 beats per minute.[3]

Data using the Enlite sensor were collected initially using the Guardian receiver. Medtronic has integrated the Enlite sensor into their Veo and 530G devices, which uses a different calibration algorithm from the Guardian or the Revel pump. We provided to Medtronic the raw Enlite sensor data and calibration PG values that were entered into the Guardian receiver. Medtronic postprocessed the data with the Veo/530G algorithm. If the calibration algorithm requested additional calibrations, we provided to Medtronic the first PG measurement after the calibration request. Therefore, calibrations were provided prospectively to the calibration algorithm just as they would have been if the Veo or 530G systems had been used. Thus, the results presented here are representative of the performance of the Enlite sensor with the Veo or 530G device, not with the Guardian or Revel. Abbott Diabetes Care declined to similarly postprocess the Navigator data with their latest calibration algorithm, which is used in the product currently available outside of the United States.

Accuracy, Precision, and Reliability Metrics

The point accuracy of each CGM is measured in terms of the relative difference (RD), $(CGMG - PG)/PG$, and the absolute relative difference (ARD), $|CGMG - PG|/PG$. Positive RD values correspond to an over-estimation of PG by the CGM. While RD indicates the extent and direction of bias in estimation of PG by a CGM, the ARD is a better measure of the average error across a set of data because of the cancelation that occurs when summing positive and negative RD values.

The mean ARD (MARD) relative to PG and the mean and standard deviation (SD) of the MARDs from all 24 experiments were computed for each CGM. The mean of the 48-hour MARDs characterizes the mean accuracy and the SD of these MARDs characterizes precision, that is, the variation around mean accuracy from 1 sensor session to the next.[1]

Rate-of-change accuracy was also evaluated, where reference data were obtained by taking the difference between 2 consecutive PG values and dividing by the 15-minute PG sampling interval. The CGMG sampling interval was 5 minutes. Device reliability was measured with the percentage of glucose values reported by the CGM relative to the total number possible. Nighttime was defined as 23:00 to 07:00.

Statistical Analyses

For comparisons of ARDs, repeated measurements models were used for within-subject repeated measurements on the differences between the measurements. This accounted for both within-subject correlations and correlations in paired measurements. The repeated measurements models were fitted with the generalized estimating equation (GEE) method and z-statistics and P values were computed. Analysis of variance (ANOVA) was applied to identify differences in performance between sensors. The P values for comparisons were based on F-statistics in the ANOVA. Statistical analyses were performed using SAS version 9.2 (SAS Institute,

Cary, NC). Nominal *P* values without correction for multiple comparisons are presented throughout, for ease of comparison with our previous report.[1] Use of the conservative Bonferroni method of correction would yield a threshold of significance for uncorrected *P* values of .017 instead of .05.

Results

Twelve adult subjects 21 years and older (6 male, 6 female) and 12 adolescent subjects 12-20 years old (3 male, 9 female) each completed a 48-hour experiment. Adult subjects were 45 ± 14 (26-66) years old and weighed 77 ± 11 (64-99) kg. Adolescent subjects were 15 ± 2 (12-18) years old and weighed 60 ± 9 (48-75) kg.

CGM Calibrations

The first 2 calibrations of the Navigator, and the first calibration of the G4 and the Enlite were performed 2-3 hours before the start of the experiment. Per experiment, there were on average 5.2 ± 3.2 (1-13) calibrations for the Navigator, 5.5 ± 1.9 (4-11) for the G4, and 6.4 ± 1.1 (5-9) for the Enlite (see Supplemental Data, available at http://dst.sagepub.com, for further details).

CGM Point Accuracy: MARD

The CGMG was compared with venous PG every 15 minutes. A representative experiment is shown in Figure 1A and data from each experiment are shown in Supplemental Figures 1 to 24 (available at http://dst.sagepub.com). The mean of all 24 individual 48-hour MARDs (Figure 1B) was 12.3 ± 4.7% for the Navigator, 10.8 ± 2.8% for the G4, and 18.3 ± 8.0% for the Enlite ($F = 3.29$, $P = .08$, Navigator versus G4; $F = 12.03$, $P = .002$, Navigator versus Enlite; $F = 19.71$, $P = .002$, G4 versus Enlite).

The point accuracy of each CGM for the entire aggregated data set is reported in Clarke error grids in Figure 2, as MARD ± SD at each PG value and in clinically relevant PG ranges in Figure 3, and in the form of RD and ARD distributions in Figure 4. The aggregate MARD of all paired points was 12.3 ± 12.1% for the Navigator ($N = 4645$), 10.8 ± 9.9% for the G4 ($N = 4634$), and 17.9 ± 15.8% for the Enlite (4521 paired points); $z = 1.92$, $P = .06$, Navigator versus G4; $z = –3.70$, $P = .0002$, Navigator versus Enlite; $z = –4.92$, $P = .0001$, G4 versus Enlite. The upper bound on the MARDs for each CGM (obtained by recalculating the MARD after randomly shuffling the paired CGMG and PG values for each data set) were found to be 44% for the Navigator, 48% for the G4, and 52% for the Enlite (see Supplemental Data for details).

The MARD of the GlucoScout relative to the YSI Stat Plus Glucose monitor when measuring PG of the same blood sample was 6.0% for 1184 paired points.

CGM Point Accuracy: Precision

The means and SDs for all PG–ARD pairs associated with PG values within 70-300 mg/dl (Figures 3A-C) were much smaller on average for the Navigator (9.1 ± 3.3%) and G4 (8.7 ± 3.2%) than for the Enlite (14.3 ± 4.5%), indicating higher precision of the Navigator and G4. In terms of means and SDs of ARDs from individual experiments (indicating sensor to sensor variability), the SD for the G4 was lower than for the Navigator, which was lower than for the Enlite (2.8%, 4.7%, and 8.0%, respectively).

In the PG ranges 70-120 and 120-180 mg/dl (Figure 3D) the Navigator and G4 were comparable and outperformed the Enlite, both in terms of MARD and SD. The G4 outperformed the other 2 CGMs in the >250 mg/dl range. The G4 showed the most consistent performance across all 4 PG ranges, both in terms of MARD and SD.

CGM Point Accuracy: Large Errors

To evaluate the frequency of very large errors,[4] we identified all PG–CGM glucose pairs for which the ARD was ≥50% (Figure 3E). There were 64 of 4645 pairs (1.4%) with such errors for the Navigator, 22 of 4634 (0.5%) for the G4, and 197 of 4521 (4.4%) for the Enlite. The rate of very large errors for the G4 was nearly 3-fold lower than for the Navigator, but this difference was not significant ($P = .1$). The rate of very large errors for the Enlite was significantly higher than for the G4 ($P = .0001$) but not the Navigator ($P = .06$).

CGM Point Accuracy: Bias

The RD and ARD distributions of all PG–CGM pairs are shown in Figures 4A and 4C, respectively. The large errors can be seen graphically in Figure 3D as points lying at or above 50% on the ordinate in the ARD versus PG plots of Figure 4C and above or below the ordinate in Figure 4A. Most of the Navigator values with ARD > 50% (61 of 63) had a positive bias (RD 51-167% for PG values 35-139 mg/dl) and approximately half of these corresponded to PG values < 70 mg/dl. Thus, very large errors by the Navigator were uncommon but many of those that did occur reduced the sensitivity for detection of hypoglycemia. Although very large errors were less common for the G4 than the Navigator, the nature of these errors was similar. The majority of G4 ARD > 50% (18 of 21) had positive bias (RD 52-123% for PG values 35-197 mg/dl) and half of these corresponded to PG values < 70 mg/dl. Very large errors were substantially more common for the Enlite (3-fold higher than Navigator and more than 9-fold higher than the G4) and they were more evenly distributed between those with negative and positive bias. Of the 193 values with ARD > 50%, 89 had a negative bias (RD –50% to –70% for PG values 85-284 mg/dl) and 104 had a positive bias (RD 50 to 187% for PG values 35-259

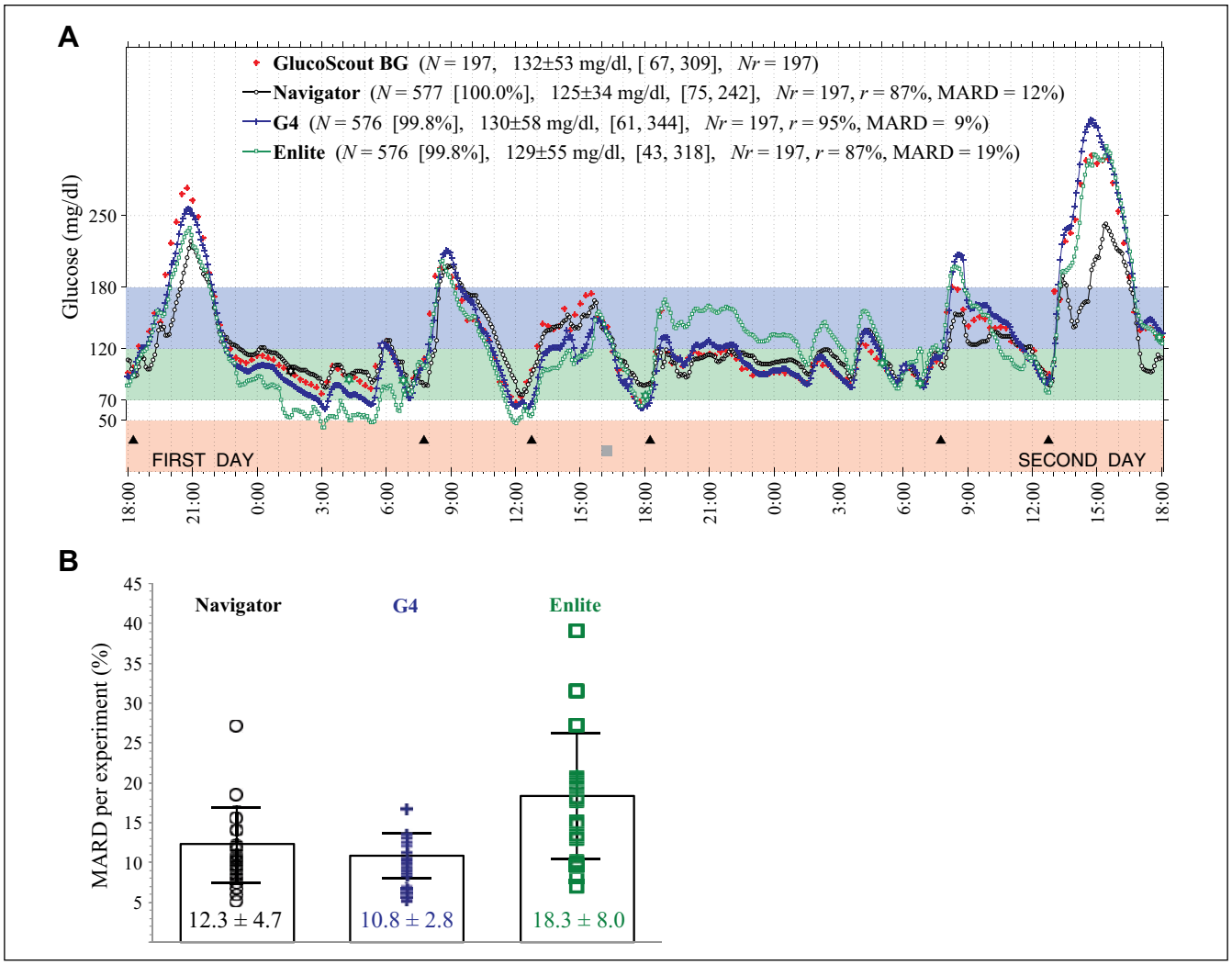

Figure 1. (A) Representative results from 1 of twenty-four 48-hour closed-loop BG control experiments in 1 of 24 subjects showing venous PG concentrations measured every 15 minutes with the GlucoScout (red symbols) and CGMG values measured approximately every 5 minutes with the Navigator (black symbols), G4 (blue symbols), and Enlite (green symbols). The timing of 6 meals is indicated by black triangles. One period of structured exercise at 16:00 (2 hours before the fourth meal) is indicated by a gray square. Listed in the legend for each CGM is the number, *N*, of glucose values measured, the data reporting percentage (in square brackets), and the MARD averaged over the 48-hour period (based on 193 paired PG–CGMG values for each CGM). Results shown in (A) for each experiment are shown in Supplemental Figures 1 to 24. The 48-hour MARDs computed in each of the 24 experiments are shown in (B) for each sensor with the mean and SD of each of those MARDs superimposed on the data for each device.

mg/dl). Within the positive RD group, 8 of 104 corresponded to PG values < 70 mg/dl.

When only errors with a smaller absolute magnitude (<50%) were considered, another pattern emerged. For the Navigator, 93% of points with PG values > 250 mg/dl had negative RD (Figure 4A). The G4 and Enlite showed a smaller negative bias than the Navigator's persistent negative bias in the hyperglycemic range. For PG values > 250 mg/dl, only 67% of G4 and 64% of Enlite values had negative RD. Further evidence for the negative bias of the Navigator at high PG values was the slope of a linear least squares fit to its data (Figure 2B), which was 0.77. The smaller systematic bias in the G4 and Enlite data was evident in the near-unity slopes (0.94 and 0.95, respectively) in linear least squares fits (Figures 2C and 2D).

Bias in the CGMs was also assessed by comparing mean CGMG to mean PG (154 ± 18 mg/dl). The Navigator underestimated the mean PG by 7 mg/dl (147 ± 16 mg/dl, F = 18.54, P = .0003); the G4 by 4 mg/dl (150 ± 17 mg/dl, F = 6.91, P = .02), and the Enlite by 3 mg/dl, (151 ± 24 mg/dl, F = 0.45, P = .5).

With the high data density in the distributions shown in Figures 4A and 4C, we were able to analyze the data in frequency bins (7% by 7 mg/d) in the PG–RD plane of Figure 4A and in the PG–ARD plane of Figure 4C, and generate histograms over the PG–RD plane (Figure 4B) and PG–ARD

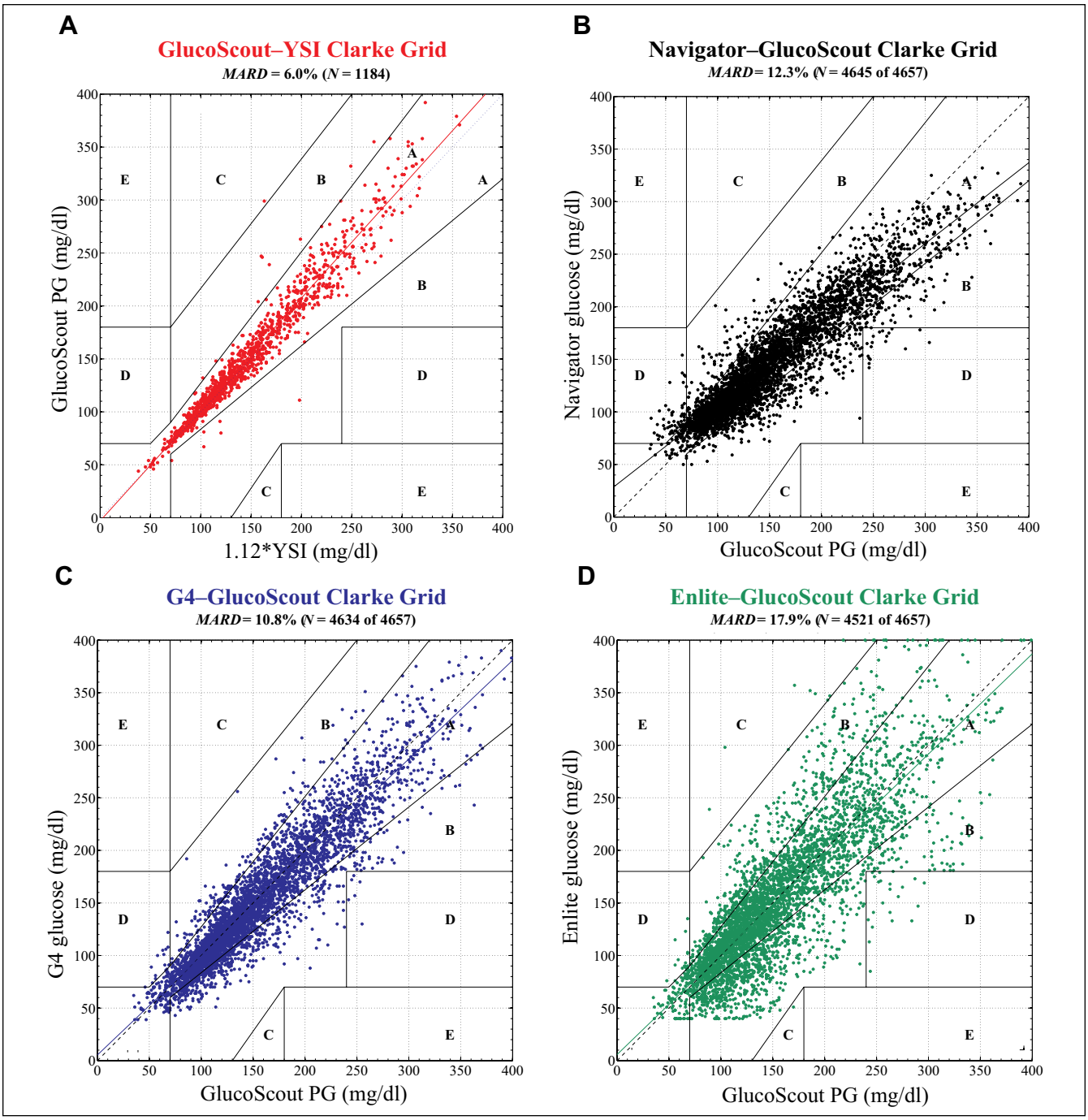

Figure 2. Clarke error grid analyses of (A) venous plasma glucose (PG) measured by the GlucoScout with venous blood glucose (BG) measured by the YSI designated as the reference, and CGMG measured by (B) the Navigator, (C) the G4, and (D) the Enlite with venous PG measured by the GlucoScout designated as the reference. (A) Based on a total of *N* = 1184 GlucoScout–YSI glucose pairs, 98.6% of points fell in zone A, 1.3% in zone B, 0.1% in zone C, and 0% in zone D. The slope and intercept of the linear least squares fit to these data (solid red line) were 1.05 and –3 mg/dl, respectively. The MARD was found to be 6.0% 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 *N* = 4645 Navigator–GlucoScout pairs, the Navigator achieved 84.2% of points in zone A, 14.2% in zone B, 0% in zone C, and 1.6% in zone D. The slope and intercept of the linear least squares fit to these data (solid black line) were 0.77 and 29 mg/dl, respectively. The Navigator achieved an overall data reporting percentage of 99.7% and a MARD of 12.3 ± 12.1%. (C) Based on a total of *N* = 4634 G4–GlucoScout pairs, the G4 achieved 84.5% of points in zone A, 15.1% in zone B, 0% in zone C, and 0.5% in zone D. The slope and intercept of the linear least squares fit to these data (solid blue line) were 0.94 and 5 mg/dl, respectively. The G4 achieved an overall data reporting percentage of 99.5% and a MARD of 10.8 ± 9.9%. (D) Based on a total of *N* = 4521 Enlite–GlucoScout pairs, the Enlite achieved 69.1% of points in zone A, 29.8% in zone B, 0.3% in zone C, and 0.8% in zone D. The slope and intercept of the linear least squares fit to these data (solid green line) were 0.95 and 6 mg/dl, respectively. The Enlite achieved an overall data reporting percentage of 97.1% and a MARD of 17.9 ± 15.8%.

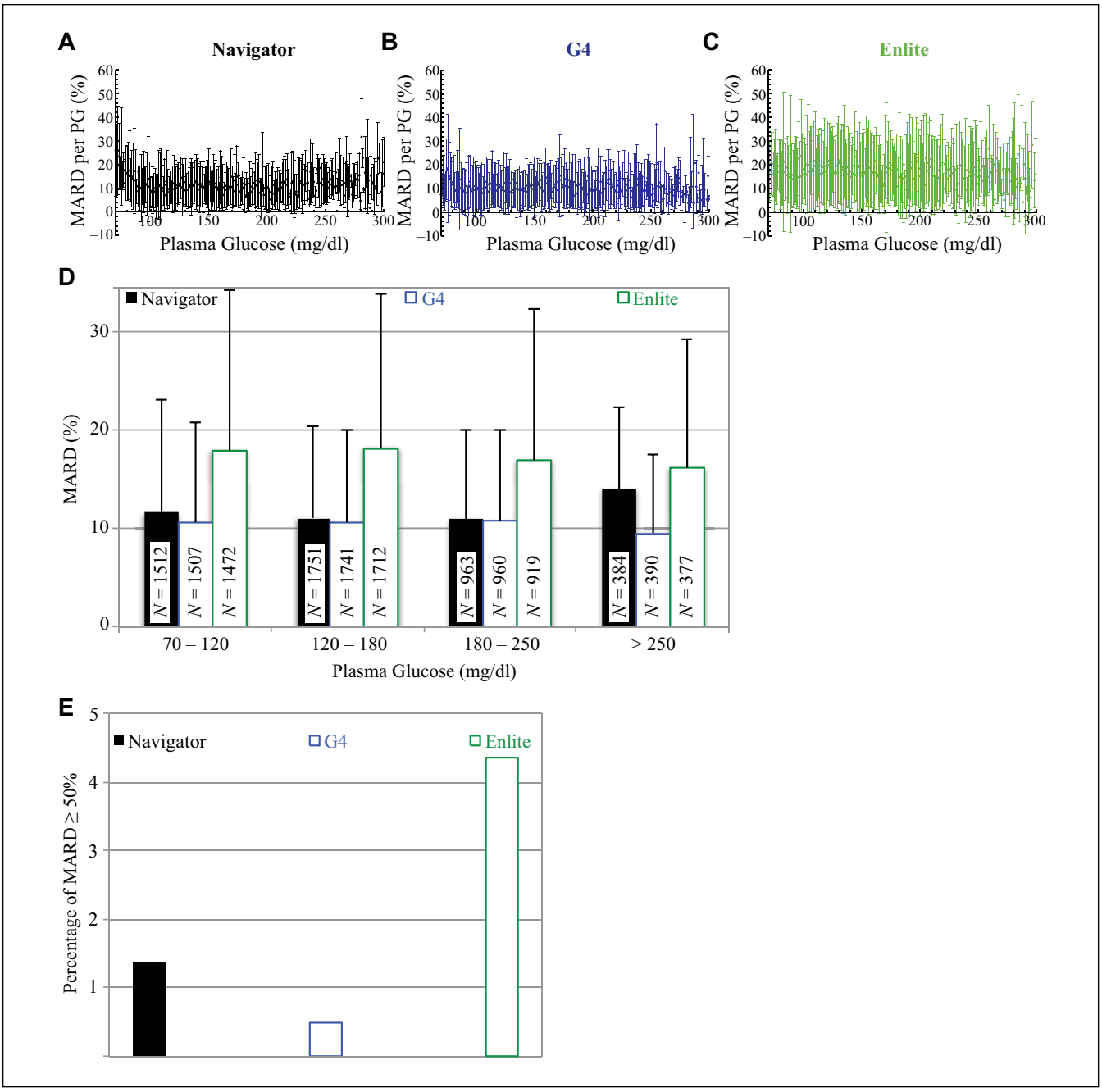

Figure 3. (A)-(C) The MARD and SD in the MARD corresponding to each PG value from 70 to 320 mg/dl for the Navigator, G4, and Enlite, respectively. Data points without error bars represent sole values for that particular PG value. (D) The MARD and SD in the MARD corresponding to the clinically relevant PG ranges from 70 to 120, 120 to 180, 180 to 250, and ≥250 mg/dl for the Navigator, G4, and Enlite. 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 found to be 11.7 ± 11.4% (N = 1512), 10.7 ± 10.2% (N = 1507), and 17.9 ± 16.9% (N = 1472), for the Navigator, G4, and Enlite, respectively. Somewhat less reliable, because of the relatively smaller 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 found to be 36 ± 27% (N = 101), 19 ± 17% (N = 102), and 23 ± 19% (N = 102), for the Navigator, G4, and Enlite, respectively. (E) Occurrence of MARD values ≥ 50% for each of the 3 CGMs. The G4 had the lowest occurrence of 22 events (0.5%) and was followed by the Navigator, with 64 events (1.4%), and the Enlite, with 197 events (4.4%).

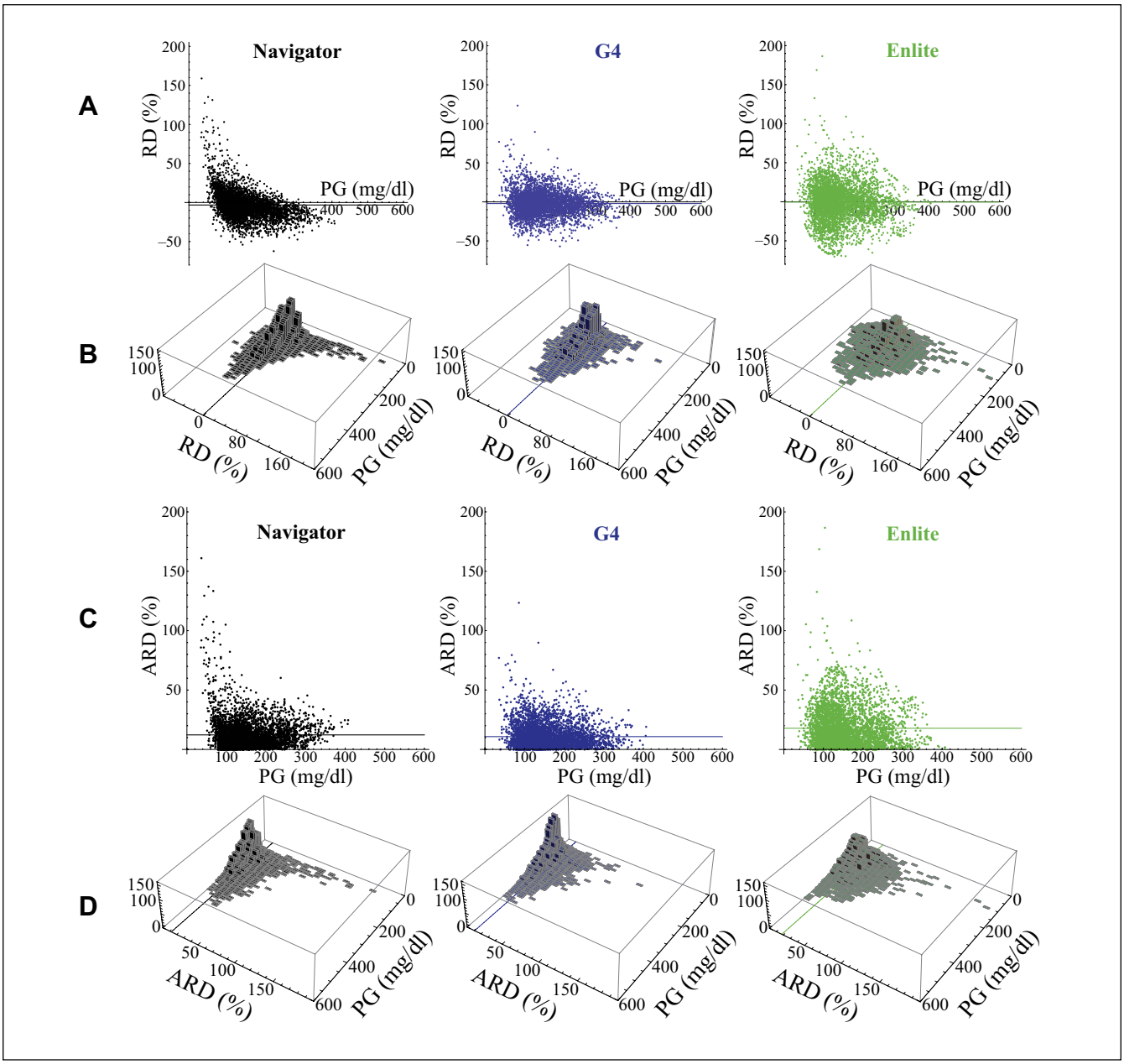

Figure 4. (A) Distribution, as a function of PG, of the relative difference (RD) between each CGMG measurement and its corresponding PG value (measured with the GlucoScout) for the Navigator (black), G4 (blue), and Enlite (green). (B) Histograms in the PG–RD plane for each of the data sets shown above 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 3 data sets. (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), G4 (blue), and Enlite (green). (D) Histograms in the PG–ARD plane for each of the data sets shown above 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 3 data sets. 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.

plane (Figure 4D), respectively.[1] Relative to the Enlite, the data obtained from the Navigator and the G4 were much more concentrated in the 0-7% error bins and showed much less dispersion over the PG–RD and PG–ARD planes (Figures 4B and 4D), demonstrating their greater accuracy and precision.

CGM Rate-of-Change Accuracy

Reference rate-of-change PG data yielded a total of 4437 slopes from the 24 experiments. Time-rate-of-change data corresponding to these reference values were extracted from the CGMG data and the absolute value of the difference

between the PG slopes and each of the corresponding CGMG slopes was computed and averaged for each of the 3 CGMs. The mean time-rate-of-change error was 0.65 ± 0.77 mg/dl/min for the Navigator compared with 0.67 ± 0.89 mg/dl/min for the G4 and 0.76 ± 0.97 mg/dl/min for the Enlite (z = –0.38, P = .7, Navigator versus G4; z = –3.16, P = .002, Navigator versus Enlite; z = –2.09, P = .04, G4 versus Enlite). Therefore, the rate accuracies for the Navigator and the G4 were not significantly different, while both had significantly better rate accuracy than the Enlite. Rate accuracy was best at low rates of change for all CGMs, and worsened with increasing rates of change in PG (Supplemental Figure 25). The largest rates of rise and fall in PG over a 15-minute interval were 8.5 and 9.4 mg/dl/min, respectively. Comparison of CGMG and PG in individual experiments (Supplemental Figures 1-24) revealed that rapidly falling PG accounted for many of the very large errors with positive biases in the low PG range. Much of this error was likely due to the physiologic lag between PG and interstitial-fluid glucose, and was probably not an inherent error in the ability of the CGM to accurately measure interstitial-fluid glucose.

Comparison of Performance Characteristics in Subpopulations and Day Versus Night

We observed a statistically significant difference in several performance characteristics between the adult and adolescent cohorts for the Navigator and G4, but not for the Enlite. The aggregate MARD for the adult cohort was higher than for the adolescent cohort for all 3 CGMs. These differences between the adult and adolescent cohorts were found to be statistically significant for the Navigator (14.8 ± 14.8 versus 9.7 ± 7.8, P = .001, z = 3.23) and the G4 (11.9 ± 11.0 versus 9.6 ± 8.4, P = .02, z = 2.31), but not for the Enlite (20.1 ± 16.9 versus 15.6 ± 14.2, P = .08, z = 1.74). The average 48-hour MARD was also found to be higher for the adult cohort than for the adolescent cohort for all 3 CGMs tested and was statistically significant for the Navigator (14.8 ± 5.5 versus 9.7 ± 1.8, P = .006, F = 3.20) and the G4 (12.0 ± 2.9 versus 9.6 ± 2.3, P = .04, F = 4.96), but not for the Enlite (20.4 ± 6.5 versus 16.3 ± 9.0, P = .2, F = 1.64). No statistically significant differences were found in the mean PG or rate-of-change accuracy between the adult and adolescent cohorts. There were no significant differences in the performance in males and females in the adult cohort. The G4 had a lower MARD in adolescent females than in males (8.4 ± 7.4 versus 11.3 ± 9.4, P = .002), but there were no gender differences in adolescents for the Navigator or the Enlite. The G4 had a lower MARD during the night than during the day for the adult cohort only (10.1 ± 3.9 versus 12.9 ± 3.6, P = .04, z = 2.03) and the Navigator had a lower MARD during the night than during the day in the adolescent cohort only (8.9 ± 2.5 versus 10.1 ± 1.8, P = .05, z = 1.97). There were no other significant differences in performance between day and night in either the adult or adolescent cohorts.

CGM Reliability

The data reporting percentages were 99.7 ± 0.6% for the Navigator, 99.5 ± 2.1% for the G4, and 97.0 ± 8.2% for the Enlite over all experiments. The Enlite transmitter contains a buffer for up to 45 minutes of data, which allows data not transmitted in real time to be back filled to the receiver after transient disconnections. We evaluated reliability based on data collected at the end of each experiment. Therefore, the data reporting percentage for the Enlite represents an upper limit on the completeness of data available in real time.

Discussion

Results from this head-to-head-to-head comparison of 3 CGMs reveal that most measures of accuracy and precision (including aggregate MARD, MARD per experiment, precision measures, distribution of relative errors in the PG–RD plane, rate-of-change errors, and data reporting frequency) are comparable between the Navigator and the G4, with a slight advantage with the G4, while both of these CGMs outperform the Enlite. The G4 significantly outperformed the Navigator in SD of the 48-hour MARD and had fewer very large errors, although the latter difference was not statistically significant. Both the Navigator and the G4 outperformed the Enlite for both of these measures. Finally, for MARDs corresponding to PG values > 250 mg/dl, the G4 outperformed both the Navigator and the Enlite, which had comparable performance in this range due to a strong low bias by the Navigator and wide variability on the part of the Enlite.

One unexpected result was the lower 48-hour MARD and aggregate MARD observed between the adolescent and adult cohorts for the Navigator and the G4. There were minimal differences in accuracy during the night versus daytime hours. In our study, all sensors were placed on the abdomen, which may have reduced the risk of compression artifacts at night,[5] so this finding may not translate to other sensor sites.

The clinical utility of CGMs, 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 automated closed-loop BG control would take the closed-loop system offline. All 3 CGMs studied here showed very high data reporting, which would provide sufficient data density and throughput for closed-loop control.[1,3] Another metric of reliability is precision, quantified here in terms of the SD around the aggregate mean of all ARD values and around the mean of ARD values from each individual experiment. The latter confers information about the variation in CGM performance from 1 sensor session to another, and may be a more clinically useful concept than the former. The G4 showed the least variability followed by the Navigator and then the Enlite, for both metrics.

As with our previous study, the data analyzed here were collected as part of a closed-loop study, and therefore contained relatively few points < 70 and > 250 mg/dl. Glucose

values were 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 reliably the accuracy of the 3 sensors in the hypoglycemic range (PG < 70 mg/dl). However, the accuracy in the hypoglycemic range of PG values 50-70 mg/dl was clearly worse than in the other ranges for all 3 sensors (see Figure 3 caption).

Another limitation of this study is that although the timing of calibrations strictly followed manufacturers' specifications, they were carried out using reference-quality PG rather than capillary BG measurements. These factors could have led us to overestimate the accuracy of the CGMs when used as part of current standard-of-care therapy. This did not seem to play a role in the case of the Navigator because its aggregate MARD and SD found here were in excellent agreement with its product labeling by the manufacturer (12.3 ± 12.1% in this study versus 12.8 ± 13.6% in the product labeling). However, the quality of calibrations may have played a role in the discrepancy between the aggregate MARD obtained in this study for the G4 (10.8 ± 9.9%) and that reported in its product labeling (14.1%). On the other hand, since the manufacturer's study in the case of the G4 was limited to subjects > 18 years of age, the comparison is perhaps better made with the aggregate MARD of 11.9% obtained for the G4 in our adult cohort. The MARD found here for the Enlite is higher than the 13.6% claimed for 530 G with Enlite system.

Two of the sensors studied (G4 and Enlite) represent the manufacturers' next-generation devices relative to the devices tested in our previous study while 1 sensor (Navigator) was identical in the 2 trials.[1] The aggregate MARD for the 2356 points obtained with the Navigator in our previous study (which had a very similar design, but was limited to 12 experiments in 6 adult subjects 33-72 years old) was not significantly different from the aggregate MARD for the 2325 points obtained with the Navigator in the 12 experiments of the adult cohort of this study (11.8 ± 11.1% versus 14.8 ± 14.8%, $P = .2$). This continuity allows us to make inferences about the relative accuracies of the G4 versus its predecessor, the Seven Plus, and the Enlite versus its predecessor, the Sof-sensor. The aggregate MARD for the 1799 points obtained with the Seven Plus in our previous study was significantly higher than the aggregate MARD for the 2309 points obtained with the G4 in the adult cohort of this study (16.5 ± 17.8% versus 11.9 ± 11.0%, $P = .008$), demonstrating a clear improvement in performance from 1 generation of Dexcom CGM to the next. A study evaluating an earlier version of the Dexcom G4 sensor, the "version A," versus the Navigator[6] found that the version A was significantly less accurate than the Navigator in an inpatient study (although it performed similarly to the Navigator in the outpatient arm of the study). In combination with the results reported here, these data suggest that there has been significant improvement from the version A to the Platinum version of the G4. On the other hand, the aggregate MARD for the 2328 points obtained with the Guardian in our previous study was not significantly different from the aggregate MARD for the 2263 points obtained with the Enlite in the adult cohort of this study (20.3 ± 18.0% versus 20.1 ± 16.9%, $P = 1.0$).

Conclusions

This head-to-head-to-head comparative effectiveness study reveals the G4 Platinum as the most accurate and precise of the current generation of CGMs, followed closely by the Navigator, with both devices markedly more accurate and precise than the Enlite sensor with the Veo/530G algorithm.

Abbreviations

ANOVA, analysis of variance; ARD, absolute relative difference; BG, blood glucose; CGM, continuous glucose monitor; CGMG, CGM glucose; ISO, International Organization for Standardization; MARD, mean absolute relative difference; MGH, Massachusetts General Hospital; NIH, National Institutes of Health; PG, plasma glucose; RD, relative difference; SD, standard deviation.

Acknowledgments

We 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, Laurel Macey, and Manasi Sinha 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; 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.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Abbott Diabetes Care, Dexcom, and International Biomedical loaned equipment for this study and provided technical support for its use. Abbott, Dexcom, and Medtronic provided sensors for this study. Medtronic processed data from the Enlite sensor with the Veo/530G algorithm. SJR received travel expenses and an honoraium from Abbott Diabetes Care for a lecture, grant support for an investigator initiated study from Abbott Diabetes Care, and consulting fees from Medtronic.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by National Institutes of Health (NIH) grant R01-DK085633 to ERD, a grant to ERD from the Leona M. and

Harry B. Helmsley Charitable Trust, 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, and a grant to DMN from the Charlton Fund for Innovative Research in Diabetes.

References

1. Damiano ER, El-Khatib FH, Zheng H, Nathan DM, Russell SJ. A comparative effectiveness analysis of three continuous glucose monitors. *Diabetes Care*. 2013;36(2):251-259.
2. Freckmann G, Pleus S, Link M, Zschornack E, Klotzer HM, Huag C. Performance evaluation of three continuous glucose monitoring systems: comparison of six sensors per subject in parallel. *J Diabetes Sci Tech*. 2013;7(4):842-853.
3. 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(1):2148-2155.
4. Castle JR, Pitts A, Zheng H, et al. The accuracy benefit of multiple amperometric glucose sensors in people with type 1 diabetes. *Diabetes Care*. 2012;35(4):706-710.
5. Helton KL, Ratner BD, Wisniewski NA. Biomechanics of the sensor–tissue interface—effects of motion, pressure, and design on sensor performance and foreign body response—part II: examples and application. *J Diabetes Sci Tech*. 2011;5(3):647-656.
6. Luijf YM, Mader JK, Doll W, et al. Accuracy and reliability of continuous glucose monitoring systems: a head-to-head comparison. *Diabetes Technol Ther*. 2013;15(8):721-726.