ORIGINAL ARTICLE

Test–Retest Reliability of a Continuous Glucose Monitoring System in Individuals with Type 2 Diabetes

Tasuku Terada, MSc,^{1,4} Sarah Loehr, MSc,^{2,4} Emmanuel Guigard, BSc,^{2,4}
Linda J. McCargar, PhD, RD,^{2,4} Gordon J. Bell, PhD,^{1,4} Peter Senior, MD, PhD,^{3,4}
and Normand G. Boulé, PhD^{1,4}

Abstract

Aims: This study determined the test–retest reliability of a continuous glucose monitoring system (CGMS) (iPro™2; Medtronic, Northridge, CA) under standardized conditions in individuals with type 2 diabetes (T2D).

Subjects and Methods: Fourteen individuals with T2D spent two nonconsecutive days in a calorimetry unit. On both days, meals, medication, and exercise were standardized. Glucose concentrations were measured continuously by CGMS, from which daily mean glucose concentration (GLU_{mean}), time spent in hyperglycemia ($t_{>10.0\text{ mmol/L}}$), and meal, exercise, and nocturnal mean glucose concentrations, as well as glycemic variability (SD_w , percentage coefficient of variation [$\%cv_w$], mean amplitude of glycemic excursions [$MAGE_c$, $MAGE_{ave}$, and $MAGE_{abs, gos}$], and continuous overlapping net glycemic action [$CONGA_n$]) were estimated. Absolute and relative reliabilities were investigated using coefficient of variation (CV) and intraclass correlation, respectively.

Results: Relative reliability ranged from 0.77 to 0.95 ($P < 0.05$) for GLU_{mean} and meal, exercise, and nocturnal glycemia with CV ranging from 3.9% to 11.7%. Despite significant relative reliability ($R = 0.93$; $P < 0.01$), $t_{>10.0\text{ mmol/L}}$ showed larger CV (54.7%). Among the different glycemic variability measures, a significant between-day difference was observed in $MAGE_c$, $MAGE_{ave}$, $CONGA_6$, and $CONGA_{12}$. The remaining measures (i.e., SD_w , $\%cv_w$, $MAGE_{abs, gos}$, and $CONGA_{1-4}$) indicated no between-day differences and significant relative reliability.

Conclusions: In individuals with T2D, CGMS-estimated glycemic profiles were characterized by high relative and absolute reliability for both daily and shorter-term measurements as represented by GLU_{mean} and meal, exercise, and nocturnal glycemia. Among the different methods to calculate glycemic variability, our results showed SD_w , $\%cv_w$, $MAGE_{abs, gos}$, and $CONGA_n$ with $n \leq 4$ were reliable measures. These results suggest the usefulness of CGMS in clinical trials utilizing repeated measured.

Introduction

THE IMPORTANCE OF CONTROLLING glycemia to avoid microvascular complications has been established.^{1–4} However, the assessment of glycemia under free-living conditions in a minimally invasive manner still remains a challenge. Self-monitored blood glucose and glycosylated hemoglobin (HbA_{1c}) measurements have provided patients with diabetes and their healthcare professionals with the information necessary for planning the most appropriate antidiabetes treatment to optimize blood glucose. However, self-monitored blood

glucose is invasive, cumbersome, and predominantly episodic in nature, making it difficult to capture rapid and overall changes in glucose concentrations in response to various stimuli. Similarly, HbA_{1c} does not reflect short-term glycemic changes, such as postprandial glucose spikes and daily glycemic variability.

Recently, the advent of continuous glucose monitoring systems (CGMSs) has made it possible to measure a continuous temporal line of glucose concentrations. This continuous stream of data captures glucose concentrations in the context of its direction, periodicity, and amplitude in relation

Faculty of ¹Physical Education & Recreation and ³Medicine & Dentistry and ²Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada.

⁴Alberta Diabetes Institute, Li Ka Shing Center for Health Research Innovation, University of Alberta, Edmonton, Alberta, Canada.

to food, exercise, and medication, providing an important overview of glycemic profiles. Furthermore, the continuous glucose data allow the documentation of glycemic variability, a strong stimulus that increases cellular oxidative stress,⁵ which can exacerbate endothelial cellular function.⁶ Detection of abnormalities in these variables can lead to tighter glycemic control and facilitate adjustments in therapy to improve glycemic control.⁷

The iPro™2 CGMS (Medtronic, Northridge, CA) has been approved by the U.S. Food and Drug Administration and Health Canada, and its use has increasingly been investigated in individuals with type 2 diabetes (T2D). Although the accuracy of the Medtronic sensors has been examined under various conditions,^{8–10} the test–retest reliability of the device has not been published. Consequently, to validate the measures of the device, it is important to evaluate its test–retest reliability. The purpose of this study was to determine test–retest reliability of various glucose profiles estimated based on iPro2 CGMS measures in individuals with T2D. In the current study, all CGMS measures were considered equally important, and none was specified as primary or secondary outcomes.

Subjects and Methods

Participants

Fifteen individuals with a clinical diagnosis of T2D treated with or without oral hypoglycemic agents were recruited. Inclusion criteria were as follows: between 40–75 years of age; not on exogenous insulin; nonsmokers; comfortable staying in the isolated room over two 24-h periods; no dietary restrictions (i.e., celiac disease or severe food allergies); not on medication with known effects on energy expenditure; blood pressure <140/90 mm Hg; stable body weight for the previous 6 months (<2.3 kg change); and no history of cardiovascular disease.

Pretests

Participants reported to the laboratory on two separate occasions. During the first visit, the nature of the study was explained, and anthropometric measures were obtained. Weight and height were measured with a stand-on scale (Health o meter®; Pelstar LLC, McCook, IL) and wall-mounted stadiometer. Each participant was also introduced to the calorimetry unit. During the second visit, participants performed a modified Bruce graded exercise test.¹¹ Expiratory gases were analyzed by a calibrated TrueOne® 2400 (ParvoMedics, Sandy, UT) metabolic measurement system, and ventilatory threshold was determined using the V-slope method.¹² An oxygen consumption–carbon dioxide production curve was graphically monitored during the incremental exercise test. Exercise was terminated once a change in the slope of the curve caused by excess carbon dioxide production was visually determined. The ventilatory threshold was used to establish the intensity of exercise to be performed within the calorimetry unit.

Standardized condition

Each participant spent two nonconsecutive days (1 day in between at home) in the whole-body indirect calorimetry unit, a self-contained airtight unit comprising a bed, chair, sink, toilet, television, personal computer, and treadmill. Participants were instructed to avoid intense physical activity

the day before the test and to report to the unit by car or public transportation after at least 10 h of fasting. On the first testing day (Day 1), participants reported to the laboratory at 07:00 h. When the subject arrived, a Sof-sensor® (Medtronic) was subcutaneously inserted to the anterior abdominal area with the help of a designated device (Sen-serter; Medtronic). After enough time (>15 min) was allowed to wet the sensor, the iPro2 CGMS (Medtronic, Northridge, CA) was connected. At 08:00 h, the participant entered the calorimetry unit and remained supine on the bed for 1 h. Meals standardized to meet individual energy requirement and a specific macronutrients distribution (50% carbohydrate, 30% fat, and 20% protein) were provided at 09:05, 12:00, and 18:00 h. Standardized snacks were also provided at 15:00 and 21:00 h. Participants were instructed to eat all the food and beverages provided within 30 min. Participants consumed their prescribed oral glucose-lowering medication at their usual time as instructed by their physicians.

Capillary blood glucose was measured 5 min before each meal and 15 min before bed with a OneTouch® Ultra® 2 (LifeScan, Milpitas, CA) handheld glucose monitor, which has previously been validated.¹³ At 14:00 h, each participant performed a 30-min individualized treadmill walking protocol followed by 5 min of cool-down. The walking intensity was determined as the intensity corresponding to the stage below ventilatory threshold in the graded exercise test. Lights were turned off at 22:30 h, and participants were instructed to sleep until 06:30 h the following day. At 07:15 h, participants left the calorimetry unit. During the hours of no assigned tasks, participants were instructed to stay awake and perform their preferred sedentary activities, such as reading or watching television. On completion of 23 h 15 min in the unit, participants spent 1 day outside of the calorimetry unit and returned on the following day at 07:00 h to repeat the protocol (Day 2). At 07:15 on the final experimental day, participants exited the calorimetry unit, and the CGMS was removed. The temperature and relative humidity of the unit were maintained at approximately 22°C and 55%, respectively, and the unit was continuously monitored by research assistants on both testing days. Energy expenditure was recorded every minute during the participants' stay.

Dietary intake

Meal portion size was individualized to target energy balance based on caloric expenditure measured in the calorimetry unit. ESHA food processing software (ESHA Research, Salem, OR) was used to standardize macronutrients proportions (i.e., 50% carbohydrate, 30% fat, and 20% protein) despite different meal portion sizes among participants. Participants were provided with the same food on Day 2. In rare cases where participants were unable to consume all food provided, caloric contents of leftovers were determined and subtracted from total daily caloric intake.

CGMS measures

Stored CGMS data were exported to an online program (CareLink iPro; Medtronic). The glucose data from a handheld glucose monitor were also entered to convert measured signals into glucose values as per the manufacturer's instruction. Because CGMS glucose values are only available after the entry of the first handheld glucose monitor

measurement, CGMS data from 09:00 to 07:00 h the following day obtained from Day 1 and Day 2 in the calorimetry unit were used for analysis. Several outcome variables were calculated from CGMS data, including daily mean glucose over the entire 22-h period (GLU_{mean}), glycemic variability, mean glucose concentrations in response to breakfast, lunch, and supper (2-h mean since each meal was provided), mean glucose during exercise and 2-h post-exercise, mean nocturnal glucose (24:00–05:00 h), and time spent in hyperglycemia (glucose concentrations > 10.0 mmol/L [$t_{>10.0}$ mmol/L]) for both testing days. Within-day glycemic variability was assessed by glucose SD_w ,¹⁴ within-day percentage coefficient of variation ($\%cv_w$),¹⁴ mean amplitude of glycemic excursions (MAGE),¹⁵ and continuous overlapping net glycemic action (CONGA_n).¹⁶ SD_w was the SD of all of the measurements over the 22-h period. $\%cv_w$ was calculated as $([SD_w/mean] \times 100)$. MAGE was estimated using three different protocols: the classical protocol as developed by Service et al.,¹⁵ which is the mean of single direction glucose excursions (either nadir to peak or peak to nadir) that exceeded 1 SD (MAGE_c); the average of both upward and downward excursions that exceeded 1 SD (MAGE_{ave})¹⁷; and an absolute group of signs method using both upstroke and downstroke excursions introduced by Zaccardi et al.¹⁸ (MAGE_{abs.gos}). Graphical approaches were used for MAGE_c and MAGE_{ave}. The value of 1 SD calculated for each testing day for each participant was used to calculate MAGE on each testing day. All calculation was completed by a single researcher. CONGA_n was calculated as the SD of the glycemic differences between a specific point on the CGMS measure and another measure “n” hour(s) previous to the observation.¹⁹ CONGA₁, CONGA₂, CONGA₄, CONGA₆, CONGA₈, CONGA₁₀, and CONGA₁₂ were examined for their reliability. Lastly, the mean of daily differences (MODD) over 2 days (MODD₂)¹⁴ was also examined for between-days glycemic variability. MODD₂ was calculated as the mean of absolute differences between glucose values at the same time of the two testing days.

Statistical analysis

Paired *t* test was initially performed for each variable to assess if there was a significant systematic difference between the testing days. Subsequently, both absolute and relative reliabilities were assessed to determine test-retest reliability of CGMS-estimated variables. The relative reliability was assessed using intraclass correlation coefficients (ICCs) based on a two-factor mixed-effects model and type consistency using average measures (ICC_{3,k}): $(MS_S - MS_E)/(MS_S)$, where MS_S and MS_E represented the subject's mean square and an error mean square, respectively.^{20,21} The absolute reliability was assessed by coefficient of variation (CV) $([SD/mean] \times 100)$. All data are presented as mean $\pm$ SD values. The α level was set at 0.05 for statistical significance. Statistical analyses were performed with SPSS statistical software (SPSS Inc., Chicago, IL).

Results

Participants

Fourteen participants completed both testing days. One participant only completed the first day because of a minor

allergic response in the calorimetry unit and was excluded from analysis. Of the 14 participants, one participant did not consume glucose-lowering medication consistently on both days and was excluded. In addition, sporadic data were obtained from one participant on the second day, probably because of a loose sensor. Consequently, the data from 12 participants were included for analyses. No participant reported discomfort or negative effects for CGMS use except for minor bruising and irritation caused by adhesive tapes.

Descriptive characteristics of the 12 participants (six males and six females) are summarized in Table 1. Of the 12 participants, three were not on any oral glucose-lowering medication, three were on metformin alone, and the rest consumed various combinations of oral glucose-lowering medication (e.g., metformin, sitagliptin/metformin, gliclazide, and repaglinide). Glycemic responses over two 22-h periods are presented in Figure 1.

Dietary intake and energy expenditure

The participants' daily average energy expenditures for Day 1 and Day 2 were $2,273 \pm 367$ and $2,224 \pm 353$ kcal, respectively (mean difference of 49 ± 39 kcal; $P = 0.001$). Mean caloric intakes for Day 1 and Day 2 were $2,367 \pm 330$ and $2,337 \pm 348$ kcal, respectively (mean difference of 30 ± 97 kcal; $P = 0.300$). Of the 12 participants, one experienced low glucose concentration (< 4.0 mmol/L) on the first day, and glucose tablets and juice were provided as prescribed by our safety protocol, making energy intake on the first day greater than the second day ($+150$ kcal). In another case, one participant showed markedly higher energy intake than energy expenditure on the first day ($+250$ kcal), and thus the food portion was made smaller on the second day (-230 kcal). Lastly, one participant showed lower caloric intake on the second day after subtracting calories from leftover food (-130 kcal).

Daily glycemic responses

Table 2 presents the reliability of various CGMS-related outcome measures. One participant was excluded from daily glucose analysis because of an unexpected large glucose spike in the morning immediately after the beginning of CGMS measurement on Day 1. Four individuals were not included in the analysis for $t_{>10.0}$ mmol/L as they spent no time in glucose concentration > 10.0 mmol/L. Relative reliability as examined by ICC_{3,k} was significant for GLU_{mean} ($P < 0.001$) and $t_{>10.0}$ mmol/L ($P < 0.01$). Absolute reliabilities for GLU_{mean} and $t_{>10.0}$ mmol/L were 3.9% and 59.4%, respectively. Among the various measures of glycemic variability, paired *t* test showed a systematic difference between the testing days in MAGE_c, MAGE_{ave}, CONGA₆, and CONGA₁₂. SD_w , $\%cv_w$, and MAGE_{abs.gos} showed significant intraclass correlation ($P < 0.05$) and similar absolute reliability with CV ranging from 13.2% to 17.4%. Among different CONGA values, those with smaller “n,” such as CONGA₁, CONGA₂, and CONGA₄, showed significant relative reliability ($R \geq 0.86$; $P < 0.01$). However, both absolute and relative reliabilities deteriorated as “n” or the distance between the two measures increased. For daily glucose profiles, heteroscedasticity was evident in Bland-Altman plots²² (data not shown), indicating the tendency for larger measurement differences with higher mean values. The overall

TABLE 1. PARTICIPANTS' CHARACTERISTICS

Variable	All (n=12)	Women (n=6)	Men (n=6)
Age (years)	57±10	51±8	62±9
Duration of T2D (years)	11±9	15±12	8±4
Height (cm)	168.4±7.8	165.8±9.4	171.0±5.4
Weight (kg)	80.2±19.3	71.6±21.6	88.9±13.1
BMI (kg/m ²)	28.1±5.5	25.8±6.3	30.4±3.8

Data are mean±SD values.

BMI, body mass index; T2D, type 2 diabetes.

results were unaffected when statistical analyses were repeated after excluding the participants who showed large day-to-day differences in energy intake. Between-day variability as determined by MODD₂ was 0.9 ± 0.2 mmol/L.

Post-meals, exercise, and nocturnal glycemia

Reliability statistics for CGMS-measured glucose responses to meals and exercise and for nocturnal glycemia are also shown in Table 2. All but one participant had a treadmill exercise speed of 1.7 mph and slope of either 5% or 10%; one participant walked with the speed of 2.5 mph and 12% because of better performance during the baseline fitness test. All 12 participants completed the prescribed exercise protocol.

There was no difference in energy expenditure during exercise between Day 1 and Day 2 (5.3 ± 1.2 and 5.2 ± 1.0 kcal/min, $P=0.173$). Paired *t* tests showed no between-day differences in glucose profiles in responses to meal intake or exercise, or in nocturnal glycemia. Relative reliability was significant for mean post-meal glucose concentrations ($0.77 \leq R \leq 0.88$; $P < 0.01$ for post-breakfast and post-lunch, $P < 0.05$ for post-supper), mean glucose concentrations during and post-exercise ($0.91 \leq R \leq 0.95$; both $P < 0.001$), and mean nocturnal glycemia ($R=0.92$; $P < 0.001$). Absolute reliability as determined by CV ranged from 6.5% to 11.7%.

Discussion

To our knowledge, this is the first test-retest reliability study of CGMS. Under standardized 22-h periods, the Medtronic iPro2 system displayed high relative reliability for outcomes in both long-term and short-term measures such as GLU_{mean}, glycemic responses to meals and exercise, and nocturnal glycemia. Furthermore, compared with a previous study showing a test-retest intra-T2D-individual biological CV of 7.0% and 12.7% for fasting blood glucose and 2-h post-oral glucose tolerance test, respectively,²³ we found the absolute reliability of these CGMS variables also high. Among different measures for glycemic variability, our results indicated that SD_w, %cv_w, and MAGE_{abs, gos} had no systematic differences between the testing days and showed significant relative reliability. Absolute reliability was similar among these measures.

FIG. 1. Mean±SD glucose concentrations over two 22-h periods using continuous glucose monitoring system devices. Meals were provided at 09:05, 12:00, and 18:00h. Exercise was performed from 14:00 to 14:35h. Participants were instructed to sleep from 22:30 to 06:30h. Participants performed sedentary activities of their choice between instructed tasks.

TABLE 2. ABSOLUTE AND RELATIVE RELIABILITIES OF DAILY AND SHORT-TERM GLUCOSE PROFILES

	Day 1	Day 2	T	Average CV (%)	Intraclass correlation	
					R	95% CI
Glucose (mmol/L)						
Mean ^{a,b}	7.0 ± 1.2	7.1 ± 1.1	−0.81	3.9	0.95 [‡]	0.81–0.99
Post-breakfast mean	8.4 ± 2.2	8.6 ± 1.6	−0.59	9.3	0.88 [‡]	0.57–0.97
Post-lunch mean	7.8 ± 1.8	7.8 ± 1.5	−0.05	6.5	0.86 [‡]	0.50–0.96
Post-supper mean	7.2 ± 1.7	7.2 ± 1.8	0.08	11.7	0.77*	0.20–0.93
Exercise mean	7.0 ± 1.9	7.4 ± 1.6	−2.02	7.3	0.95 [‡]	0.83–0.99
Post-exercise mean	7.3 ± 2.0	7.9 ± 1.5	−2.15	8.9	0.91 [‡]	0.70–0.98
Nocturnal mean	6.7 ± 1.5	6.7 ± 1.7	0.03	7.9	0.92 [‡]	0.71–0.98
$t_{>10.0 \text{ mmol/L}}$ (min) ^c	250 ± 311	249 ± 451	0.02	59.4	0.93 [‡]	0.58–0.99
SD _w (mmol/L) ^a	1.16 ± 0.37	1.02 ± 0.35	1.59	16.3	0.80*	0.24–0.95
%cv _w (mmol/L) ^a	16.8 ± 5.2	14.4 ± 4.7	1.80	17.4	0.74*	0.05–0.93
MAGE (mmol/L) ^a						
MAGE _c	2.83 ± 1.17	2.08 ± 0.66	2.38*	20.7	0.56	−0.62 to 0.88
MAGE _{ave}	2.81 ± 0.93	2.31 ± 0.65	2.44*	16.9	0.79*	0.22–0.94
MAGE _{abs.gos}	2.32 ± 0.79	2.11 ± 0.57	1.13	13.2	0.77*	0.14–0.94
CONGA (mmol/L) ^a						
CONGA ₁	0.93 ± 0.30	0.94 ± 0.27	−0.22	7.0	0.96 [‡]	0.84–0.99
CONGA ₂	1.22 ± 0.40	1.22 ± 0.37	−0.06	8.6	0.95 [‡]	0.81–0.99
CONGA ₄	1.56 ± 0.55	1.41 ± 0.51	1.37	15.8	0.86 [‡]	0.48–0.96
CONGA ₆	1.69 ± 0.68	1.32 ± 0.44	2.44*	25.3	0.76*	0.11–0.94
CONGA ₈	1.76 ± 0.68	1.37 ± 0.39	2.18	23.3	0.59	−0.54 to 0.89
CONGA ₁₀	1.75 ± 0.65	1.42 ± 0.64	1.29	29.5	0.26	−1.77 to 0.80
CONGA ₁₂	1.51 ± 0.56	1.05 ± 0.32	2.68*	30.1	0.41	−1.13 to 0.85

^a $n = 11$ after excluding one outlier.^b22-h mean glucose concentration.^cIndividuals who showed no time spent in glycemia > 10.0 mmol/L were excluded ($n = 7$).* $P < 0.05$, [†] $P < 0.01$, [‡] $P < 0.001$.

%cv_w, percentage coefficient of variation; CI, confidence interval; CONGA, continuous overlapping net glycemic action, with “ n ” indicating the number of hours; CV, coefficient of variation ($[\text{SD}/\text{mean}] \times 100$); MAGE, mean amplitude of glycemic excursions, with MAGE_c, MAGE_{ave}, and MAGE_{abs.gos} being MAGE that exceeded 1 SD in a single direction, MAGE calculated from both upward and downward glycemic excursions, and MAGE by the group of signs method, respectively; SD_w, glucose SD within each 22-h CGMS measurement; $t_{>10.0 \text{ mmol/L}}$, time spent in hyperglycemia.

Relative and absolute reliabilities of CONGA _{n} measures were also high when small n was used ($n \leq 4$). These results support the applicability of CGMS to estimate daily glycemic profiles, as well as various short-term treatment effects.

Although Molnar et al.²⁴ measured this value only in one participant, the previously reported MODD under a standardized near-normal conditions in an individual with stable diabetes was reported to be 1.94 mmol/L. Our lower MODD₂ value of 0.9 ± 0.2 showed very small interday variability, which reflected the more standardized and consistent condition we created for both testing days. Under this standardized condition, among measured daily glycemia, the relative reliability of $t_{>10.0 \text{ mmol/L}}$ was comparable to GLU_{mean} and was high. However, the absolute reliability of this variable was poor. This is likely because statistical tests, such as paired t test and ICC, can largely be influenced by sample heterogeneity and random variations between the tests.²⁵ Considering that slight differences in glucose concentrations between days (i.e., 9.9 vs. 10.1 mmol/L) can contribute to large inconsistency, low absolute reliability may not be surprising. Nonetheless, highly heterogeneous results (our mean $t_{>10.0 \text{ mmol/L}}$ of Day 1 and Day 2 ranged from 48 to 1,083 min) seem to have contributed to high relative reliability. As absolute reliability represents the degree to which repeated measures vary for individuals,²⁵ one must be cautious when

investigating glycemia using $t_{>10.0 \text{ mmol/L}}$ at an individual level.

Within-day glycemic variables measured in the present study were similar to previously reported values. SD_w, CONGA₁, and MAGE_c were similar to the previously reported values from individuals with T2D treated with diet only (1.17 ± 0.56 , 1.08 ± 0.48 , and 3.17 ± 1.53 , respectively).²⁶ Although our other CONGA values (CONGA₂–CONGA₆) were lower than previously reported values from a group with T2D,²⁷ this is probably owing to better overall glycemia in our participants (mean glucose concentration, 7.0 ± 1.2 vs. 10.6 ± 3.3 mmol/L). These comparisons showed that our results were comparable to previously reported CGMS variables. Because many variables were estimated for glycemic variability, we also performed correlation analysis among measures (data not shown). The results were also very similar to those of previous reports.^{28,29} There were significant and high linear correlations among most of the measures, indicating these measures are conveying largely the same information. In addition, our ratio of MAGE_c to SD_w was 2.43, which was in close agreement with previously reported values of 2.62²⁶ and 2.54,²⁹ further indicating the relationships among our measured glycemic variability parameters were similar to previously reported ones. Despite the high correlations among measured variables, we observed no between-

day systematic differences and relatively high reliability for SD_w , $\%CV_w$, $MAGE_{abs, gos}$, and $CONGA_n$ with $n \leq 4$, versus either significant systematic difference or nonsignificant ICC for $MAGE_c$, $MAGE_{ave}$, and $CONGA$ with $n \geq 6$. Our results showing that only the $MAGE_{abs, gos}$ computation method was reproducible (significant ICC with no systematic difference) among different MAGE methods are interesting given that $MAGE_c$ has previously been reported as the gold standard metric for measuring glycemic variability.⁶ Although $MAGE_c$ has frequently been used in literature, its ambiguity has been indicated.³⁰ Because $MAGE_c$ only includes the glycemic excursions that correspond to the direction of the first major glucose swing, direction of the glycemic excursions was expected to play a major role in determining $MAGE_c$. However, 10 out of 11 participants in our study consistently showed an upward glycemic swing for their first major glycemic excursion on both testing days. Also, excluding the one participant who indicated inconsistent directions in glycemic excursion did not have a major impact on the results. Therefore, the direction of the excursions cannot explain the low reliability. Instead, it is speculated that less arbitrary determination of peaks and nadirs in $MAGE_{abs, gos}$ ¹⁸ resulted in higher reliability.

For SD_w , its application to CGMS-measured glucose values has been criticized for the lack of normal distribution in glucose concentrations (glucose concentrations of individuals with T2D are often skewed towards hyperglycemia), a mathematical condition for the use of SD.³¹ Consequently, we repeated SD_w calculation using log-transformed glucose values.¹⁴ The results remained robust ($R=0.74$; $P<0.05$). As previously indicated, given the high correlation between $MAGE_c$ and SD_w ($r=0.85$ and 0.87 in the present and previous study,¹⁴ respectively), it is probable that SD_w and $MAGE_c$ convey primarily the same information. Because SD_w is easy to calculate and interpret,³⁰ with its high reliability, SD_w may be a practical method that could be appreciated by clinicians and care practitioners in determination of glycemic variability. A recent study has also shown that SD is a significant predictor of coronary artery calcification in men with type 1 diabetes,³² suggesting its application for diabetes complications research. The $\%CV_w$ also indicated absolute and relative reliabilities to similar to those of SD_w , probably because mean glucose concentrations from Day 1 and Day 2 were very similar (note that $\%CV_w$ was calculated as $[SD_w/\text{mean}] \times 100$).

Lastly, among the different “ n ” for $CONGA$ calculations, $CONGA_1$ and $CONGA_2$ showed particularly high absolute and relative reliability. Nonetheless, reliability of $CONGA_n$ was lowered as n increased. Because $CONGA_n$ is calculated as the SD of the difference between an observation and observation n hour(s) before, our results suggest prolonging the duration between measurements increases test–retest variability. The lower reliability in proportion to an increase in n observed for different $CONGA$ values is likely due to increased variability. As described by Rodbard,¹⁴ glycemic variability increases as the duration of the time frame increases. Thus, increased variability with the use of larger n may have contributed to lower reliability.

It is of importance to note that this also suggests that lower reliability of MAGE and SD_w compared with $CONGA_1$ and $CONGA_2$ may be due to the difference that MAGE and SD_w are based on the whole 22-h period, whereas $CONGA_1$ and

$CONGA_2$ are based on the glycemic variability only 1 or 2 h apart. Because the period of 22 h involves more factors that can affect variability than the shorter time frames, the longer time frame used for MAGE and SD_w calculation may have resulted in lower ICC and higher CV than the short-term criteria such as $CONGA_1$ and $CONGA_2$. Additionally, whether $CONGA_n$ measures with small n provide better and sensitive measure of changes or are equally informative in predicting clinical events or complications has yet to be investigated. In fact, although the $MAGE_c$ predicts total free radical production⁵ and oxidative stress,²⁷ in our study, the correlation between $MAGE_c$ and $CONGA_1$ was relatively small ($r=0.46$, $P=0.155$ for Day 1). It is also of importance to note that the intent of our study was to report the test–retest reliability of variables obtained from CGMS, not to suggest that certain measures are better than the others. Reliability of a measure is one aspect to predict clinical outcome, and high reliability does not necessarily translate into clinical or epidemiological importance. Given that $CONGA_n$ with small n measures high-frequency variability (hour to hour), whereas that with large n measures medium-frequency variability (the circadian changes during the course of the day),³¹ from reliability perspective, our results simply indicated that $CONGA_n$ may be useful if high-frequency glycemic variance is of interest but not for lower-frequency measures. Other methods, such as SD_w , $\%CV_w$, or $MAGE_{abs, gos}$, are more reliable for the longer-term estimation of glycemic variability.

One of the limitations to the present study was the significantly different energy expenditure regardless of the sedentary environment we provided. Nonetheless, given the small difference in energy expenditure between the days (<50 kcal), the significant difference was likely attributable to a small random error and consistently lower energy expenditure on Day 2, possibly because of accustomization to the environment (10 out of 12 participants had smaller energy expenditure on the second day). Of the 10 participants, half of them showed a higher mean glucose concentration on Day 1, whereas the other half showed higher values on Day 2, suggesting a relatively small impact of energy expenditure despite the significant difference.

Another limitation was that we were unable to identify to what degree random errors were stemming from sensor errors or from biological fluctuations. The variability estimates reported here consists of the sum of biological fluctuation and possible within-sensor variability. However, a previous report showed that sensor performance was not affected by sensor life.³³ Moreover, based on the fact that the CGMS devices were calibrated on both testing days using the same handheld glucose monitor, we speculate that the CGMS performance was similar on both days. Consequently, the values reported here are expected to be more closely associated with day-to-day biological fluctuations under minimal external stimuli. A less standardized environment, as well as allowing more time between testing sessions, will increase the biological fluctuations.

Lastly, the relatively small sample size used in the present study, especially in examining the reliability of $t_{>10.0 \text{ mmol/L}}$, was another limitation. Because four of our 12 participants had glucose concentrations that remained below 10.0 mmol/L on both testing days and one showed an unexplainable large glucose spike in the morning, the result was based on seven participants. We cannot exclude the possibility that

insufficient sample size is contributing to the large absolute variability for $t_{>10.0\text{ mmol/L}}$. A study with a larger sample size is warranted to elucidate the absolute reliability of $t_{>10.0\text{ mmol/L}}$.

In conclusion, under the condition where glycemia was measured repeatedly within individuals with T2D who are not treated with exogenous insulin, we demonstrated that CGMS measurements are reliable when estimating GLU_{mean} , short-term glycemic responses to meals and exercise, and nocturnal glycemia. Caution is needed when using $t_{>10.0\text{ mmol/L}}$ as it indicated high relative but poor absolute reliability. With regard to glycemic variability measurements, we demonstrated that $CONGA_n$ measurements with $n \leq 4$ are reliable in measuring high-frequency variability, whereas SD_w , $\%CV_w$, and $MAGE_{\text{abs, gos}}$ are reliable methods for longer-term glycemic variability. Although it remains to be seen which of the different glycemic parameters will be embraced by the investigators and clinical practitioners, our study examining the test-retest reliability of various CGMS measures provided one important aspect to consider.

Acknowledgments

The continuous glucose monitors and sensors were graciously provided by Medtronic of Canada. Funding for the Calorimetric Assessment in Diabetes study was provided by the Alberta Diabetes Institute. T.T. was supported by a studentship from the Alberta Diabetes Institute. The authors would like to thank the study participants for their time and efforts.

Author Disclosure Statement

CGMS devices were provided by Medtronic of Canada. P.S. has received honoraria from Medtronic for delivering continuing medical education and fees for consulting work. Medtronic had no role in the study design, analyses of the data, or the writing of the manuscript. T.T., S.L., E.G., L.J.M., G.J.B., and N.G.B. declare no competing financial interests exist.

References

1. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *N Engl J Med* 1993;329:977–986.
2. Shichiri M, Kishikawa H, Ohkubo Y, Wake N: Long-term results of the Kumamoto Study on optimal diabetes control in type 2 diabetic patients. *Diabetes Care* 2000;23:B21–B29.
3. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* 1998;352:837–853.
4. Ohkubo Y, Kishikawa H, Araki E, Miyata T, Isami S, Motoyoshi S, Kojima Y, Furuyoshi N, Shichiri M: Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes-mellitus—a randomized prospective 6-year study. *Diabetes Res Clin Pract* 1995;28:103–117.
5. Brownlee M, Hirsch IB: Glycemic variability: a hemoglobin A1c-independent risk factor for diabetic complications. *JAMA* 2006;295:1707–1708.
6. Monnier L, Colette C, Boegner C, Pham TC, Lapinski H, Boniface H: Continuous glucose monitoring in patients with type 2 diabetes: Why? When? Whom? *Diabetes Metab* 2007;33:247–252.
7. Klonoff DC: Continuous glucose monitoring—roadmap for 21st century diabetes therapy. *Diabetes Care* 2005;28:1231–1239.
8. Diabetes Research in Children Network (DirecNet) Study Group: The accuracy of the CGMS in children with type 1 diabetes: results of the Diabetes Research in Children Network (DirecNet) accuracy study. *Diabetes Technol Ther* 2003;5:781–789.
9. Bay C, Kristensen PL, Pedersen-Bjergaard U, Tarnow L, Thorsteinsson B: Nocturnal continuous glucose monitoring: accuracy and reliability of hypoglycemia detection in patients with type 1 diabetes at high risk of severe hypoglycemia. *Diabetes Technol Ther* 2013;15:371–377.
10. Clarke WL, Anderson S, Farhy L, Breton M, Gonder-Frederick L, Cox D, Kovatchev B: Evaluating the clinical accuracy of two continuous glucose sensors using continuous glucose-error grid analysis. *Diabetes Care* 2005;28:2412–2417.
11. Lerman J, Bruce RA, Sivarajan E, Pettet GEM, Trimble S: Low-level dynamic exercises for earlier cardiac rehabilitation—aerobic and hemodynamic responses. *Arch Phys Med Rehabil* 1976;57:355–360.
12. Beaver WL, Wasserman K, Whipp BJ: A new method for detecting anaerobic threshold by gas exchange. *J Appl Physiol* 1986;60:2020–2027.
13. Brunner GA, Ellmerer M, Sendlhofer G, Wutte A, Trajanoski Z, Schaupp L, Quehenberger F, Wach P, Krejs GJ, Pieber TR: Validation of home blood glucose meters with respect to clinical and analytical approaches. *Diabetes Care* 1998;21:585–590.
14. Rodbard D: New and improved methods to characterize glycemic variability using continuous glucose monitoring. *Diabetes Technol Ther* 2009;11:551–565.
15. Service FJ, Molnar GD, Rosevear JW, Ackerman E, Gatewood LC, Tayloe WF: Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes* 1970;19:644–655.
16. McDonnell CM, Donath SM, Vidmar SI, Werther GA, Cameron FJ: A novel approach to continuous glucose analysis utilizing glycemic variation. *Diabetes Technol Ther* 2005;7:253–263.
17. Baghurst PA: Calculating the mean amplitude of glycemic excursion from continuous glucose monitoring data: an automated algorithm. *Diabetes Technol Ther* 2011;13:296–302.
18. Zaccardi F, Stefano PD, Busetto E, Federici MO, Manto A, Infusino F, Lanza GA, Pitocco D, Ghirlanda G: Group of signs: a new method to evaluate glycemic variability. *J Diabetes Sci Technol* 2008;2:1061–1065.
19. Standl E, Schnell O, Ceriello A: Postprandial hyperglycemia and glycemic variability: should we care? *Diabetes Care* 2011;34(Suppl 2):S120–S127.
20. Shrout PE, Fleiss JL: Intraclass correlations—uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428.
21. Weir JP: Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19:231–240.

22. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
23. Mooy JM, Grootenhuys PA, deVries H, Kostense PJ, Popp-Snijders C, Bouter LM, Heine RJ: Intra-individual variation of glucose, specific insulin and proinsulin concentrations measured by two oral glucose tolerance tests in a general Caucasian population: The Hoom Study. *Diabetologia* 1996; 39:298–305.
24. Molnar GD, Taylor WF, Langworthy A: Measuring adequacy of diabetes regulations—comparison of continuously monitored blood-glucose patterns with values at selected time points. *Diabetologia* 1974;10:139–143.
25. Atkinson G, Nevill AM: Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;26:217–238.
26. Kohnert K, Heinke P, Fritzsche G, Vogt L, Augstein P, Salzsieder E: Evaluation of the mean absolute glucose change as a measure of glycemic variability using continuous glucose monitoring data. *Diabetes Technol Ther* 2013; 15:448–454.
27. Penckofer S, Quinn L, Byrn M, Ferrans C, Miller M, Strange P: Does glycemic variability impact mood and quality of life? *Diabetes Technol Ther* 2012;14:303–310.
28. Rodbard D: Interpretation of continuous glucose monitoring data: glycemic variability and quality of glycemic control. *Diabetes Technol Ther* 2009;11(Suppl 1):S-55–S-67.
29. Fritzsche G, Kohnert K, Heinke P, Vogt L, Salzsieder E: The use of a computer program to calculate the mean amplitude of glycemic excursions. *Diabetes Technol Ther* 2011;13:319–325.
30. Cameron FJ, Baghurst PA, Rodbard D: Assessing glycemic variation: why, when and how? *Pediatr Endocrinol Rev* 2010;7(Suppl 3):432–444.
31. Siegelaar SE, Holleman F, Hoekstra JB, DeVries JH: Glucose variability; does it matter? *Endocr Rev* 2010;31: 171–182.
32. Snell-Bergeon JK, Roman R, Rodbard D, Garg S, Maahs DM, Schauer IE, Bergman BC, Kinney GL, Rewers M: Glycaemic variability is associated with coronary artery calcification in men with Type 1 diabetes: the Coronary Artery Calcification in Type 1 Diabetes study. *Diabet Med* 2010; 27:1436–1442.
33. Iscoe KE, Davey RJ, Fournier PA: Is the response of continuous glucose monitors to physiological changes in blood glucose levels affected by sensor life? *Diabetes Technol Ther* 2012;14:135–142.

Address correspondence to:

Normand G. Boulé, PhD

Faculty of Physical Education & Recreation

Li Ka Shing Center for Health Research Innovation

University of Alberta

Edmonton, AB, Canada T6G 2H9

E-mail: nboule@ualberta.ca