Imprecision nutrition? Duplicate meals result in unreliable individual glycemic responses measured by continuous glucose monitors across three dietary patterns in adults without diabetes

Aaron Hengist¹, Juen Guo¹, Kevin D Hall¹*

¹National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD, USA

*Corresponding Author: Kevin D Hall, kevin.hall@nih.gov

Supported by the Intramural Research Program of the National Institutes of Health, National Institute of Diabetes & Digestive & Kidney Diseases.

Abbreviations: CGM, continuous glucose monitor, ICC, intra-class correlation coefficient

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

**Background:** Continuous glucose monitors (CGMs) are being used to characterize postprandial glycemic responses and thereby provide personalized dietary advice to minimize glycemic excursions. However, the efficacy of such advice depends on reliable CGM responses.

**Objective:** To explore within-subject variability of CGM responses to duplicate meals in an inpatient setting.

**Methods:** CGM data were collected in two controlled feeding studies (NCT03407053 and NCT03878108) in 30 participants without diabetes capturing 948 meal responses in duplicate ~1 week apart from three dietary patterns. One study used two different CGMs (Abbott Freestyle Libre Pro and Dexcom G4 Platinum) whereas the other study used only Dexcom. We calculated the incremental area under the curve (iAUC) for each 2-h post-meal period and compared within-subject iAUCs using the same CGM for the duplicate meals using linear correlations, intra-class correlation coefficients (ICC), Bland-Altman analyses, and compared individual variability of glycemic responses to duplicate meals versus different meals using standard deviations (SDs).

**Results:** There were weak to moderate positive linear correlations between within-subject iAUCs for duplicate meals (Abbott r=0.47, \( p<0.0001 \), Dexcom r=0.40, \( p<0.0001 \)), with low within-participant reliability indicated by ICC (Abbott 0.31, Dexcom 0.16). Bland-Altman analyses indicated wide limits of agreement (Abbott -31.4 to 31.5 mg/dL, Dexcom -32.3 to 31.6 mg/dL) but no significant bias of mean iAUCs for duplicate meals (Abbott 0.1 mg/dL, Dexcom -0.3 mg/dL). Individual variability of glycemic responses to duplicate meals was similar to that of different meals evaluated each diet week for both
Abbott (SD_{duplicate} = 10.7 \text{ mg/dL}, SD_{week 1} = 12.4 \text{ mg/dL}, SD_{week 2} = 11.6 \text{ mg/dL}, p=0.38) and Dexcom (SD_{duplicate} = 11.8 \text{ mg/dL}, SD_{week 1} = 12.2 \text{ mg/dL}, SD_{week 2} = 12.4 \text{ mg/dL}, p=0.80).

**Conclusions:** Individual postprandial CGM responses to duplicate meals were unreliable in adults without diabetes. Personalized diet advice based on CGM measurements in adults without diabetes requires more reliable methods involving aggregated repeated measurements.

This secondary analysis contains data from two trials registered at [clinicaltrials.gov](http://clinicaltrials.gov) (NCT03407053 and NCT03878108).

**Keywords:** continuous glucose monitor, CGM, glycemia, glucose variability, personalized nutrition, precision nutrition
Introduction

Postprandial glycemic responses to different foods as measured by continuous glucose monitors (CGM) are highly variable between individuals, with some people exhibiting large glycemic excursions in response to one food compared to another whereas a different person might experience the opposite results (1, 2). Such observations provide the rationale for personalizing diet advice to minimize glycemia excursions by attempting to identify the foods that result in reliably low postprandial glucose in each person (2, 3). The fundamental assumption of such precision dietary advice is that repeated glycemic responses to the same meal within an individual are much less variable than their responses to different meals. However, this assumption has not been rigorously tested.

We investigated the reliability of within-subject postprandial CGM responses to duplicate ad libitum meals consumed ~1 week apart by 15 participants residing at the NIH Clinical Center Metabolic Clinical Research Unit during two inpatient controlled feeding studies whose primary results have been reported elsewhere (4, 5). Study participants were presented with three daily meals from 7-day rotating menus for two weeks each such that each meal was provided in duplicate. Participants experienced three dietary patterns with one being a minimally processed plant-based, low-fat diet and the other two patterns had moderate macronutrient compositions but either high in ultra-processed or unprocessed foods. We excluded data from a very low carbohydrate ketogenic dietary pattern because CGM responses during this diet were minimal (5).
Methods

We performed an exploratory analysis of data from two clinical research protocols approved by the institutional review board of the National Institute of Diabetes and Digestive and Kidney Diseases and are registered at clinicaltrials.gov (NCT03407053 and NCT03878108). Participants provided written informed consent and eligibility criteria for both studies were (1) ages 18–50 years; (2) body mass index $>$18.5 kg/m$^2$; and (3) weight stable (<5% change in past 6 months). Both studies were within-subject, random-order crossover designs where participants were exposed to two diets for 14 days each on 7-day rotating menus, consuming each meal twice (once during week 1 and once during week 2). After excluding data from the very low carbohydrate ketogenic dietary pattern, this enabled comparison of up to 21 repeated meals within each of the 3 dietary patterns. The daily menus had three meals (breakfast, lunch, and dinner), where the order within day was always fixed and the time of day when each meal was consumed was similar between the first and second week. Furthermore, the sequence of the daily menu was the same on the first and second week except in cases when the respiratory chamber day (whose menu was fixed within each diet pattern) had to be scheduled on a different day due to availability.

Meals were provided to participants alone in their inpatient rooms and photographs of the meals have been published previously alongside the primary outcomes (4, 5). Participants were instructed eat as much or as little food as they wanted and asked to not intentionally change their weight throughout the study. All foods were weighed to the nearest 0.1 g before and after consumption, and energy intake was calculated using
ProNutra software (v.3.4, Viocare). A limitation of these studies with respect to our analyses of post-meal glycemic responses is that they included bottled water and snacks available throughout the day, but the timing of their consumption was not recorded.

Interstitial glucose concentrations were obtained from two brands of monitor: Abbott Freestyle Libre Pro (Abbott) and Dexcom G4 Platinum (Dexcom). In NCT03407053, some participants wore both Abbott and Dexcom, and in NCT03878108 study participants wore Dexcom. The Abbott device records glucose every 15 minutes and the Dexcom every 5 minutes. For accurate postprandial analysis, only duplicate meals with measured start time and sufficient data availability were included. For Abbott the mean (range) of duplicate meals within-participant was 28 (11 to 34) from 14 participants providing 392 total comparisons and for Dexcom the mean (range) of duplicate meals within-participant was 19 (2 to 38) from 30 participants providing 556 total comparisons.

Data were aligned to the nearest 15-minute (Abbott) or 5-minute (Dexcom) CGM reading for calculation of post-meal responses. Baseline was assigned as the first time-point after the meal was provided. The 2-hour postprandial glucose incremental area under the curve (iAUC) was calculated for each meal using the trapezoid method, with dips below baseline assigned a negative value for iAUC (i.e., netAUC from (6)). Values of iAUC were reported as time-averaged glucose concentrations across the 2-h postprandial period.
Statistical analyses and data visualization were performed in R (v4.2.3) and GraphPad Prism (v9.5.0). Major axis regression was used to plot trends between meal 1 and meal 2 using lmodel2 in R. Simple linear correlation was calculated using Pearson’s correlation coefficient ($r$), with $<0.4$ interpreted as weak, $0.4$ to $0.8$ interpreted as moderate, and $>0.8$ interpreted as strong correlation. Repeatability was estimated by calculating the intra-class correlation coefficient (ICC) for glucose iAUCs. ICC was calculated using the following formula: $ICC = \frac{\text{participant variance}}{(\text{participant variance} + \text{residual variance})}$, which was generated from a linear mixed effects model with participant and residual error as random effects and meal and eating occasion as fixed effects (7). ICC values below 0.5 are considered as indicating poor reliability between measures (8).

To examine individual glycemic variability in response to duplicate meals as compared with different meals, we partitioned the total iAUC variance between diet pattern, meal type (breakfast, lunch, dinner), menu day, and duplicate meals between successive weeks to compute the standard deviation (SD) of duplicate iAUC responses as follows:

$$SD_{\text{duplicate}} = \sqrt{\frac{\sum_{\text{diet}} \sum_{\text{meal}} \sum_{\text{menu}} (\text{iAUC}_{\text{diet,meal,menu}} - \langle \text{iAUC}_{\text{diet,meal,menu}} \rangle)^2}{df}},$$

where $df$ is the degrees of freedom and $\langle \text{iAUC}_{\text{diet,meal,menu}} \rangle$ is the average iAUC over duplicate meals. Similarly, we computed individual SDs of the iAUC responses to different meals on each week of 7-day rotating menus as follows:

$$SD_{\text{week}} = \sqrt{\frac{\sum_{\text{diet}} \sum_{\text{meal}} \sum_{\text{menu}} (\text{iAUC}_{\text{diet,meal,menu}} - \langle \text{iAUC}_{\text{diet,meal}} \rangle)^2}{df}}.$$
where df is the degrees of freedom and \(<iAUC_{\text{diet,meal}}\) is the average iAUC over the 7-day week of menus. One-way ANOVA with Bonferroni correction was used to compare SDs of week 1, week 2, and duplicate meal responses. Energy intake of meals in week 1 and week 2 were compared using a paired t-test. Significance was accepted as \(p \leq 0.05\).

**Results**

**Glycemic responses to the same meals eaten on separate occasions are not well-matched**

We investigated 30 participants whose characteristics are shown in Table 1 who were presented with duplicate meals on two consecutive weeks exactly 7 days apart for 85% of Abbott measurements and 72% of Dexcom measurements. Figure 1A plots the iAUC responses to meals consumed on week 2 versus the duplicate meals on week 1 measured in the same participants using the Abbott device. Figure 1B plots analogous data obtained using the Dexcom device. Regardless of CGM, there were weak to moderate positive linear correlations between the within-subject iAUC responses to duplicate meals across all dietary patterns (Abbott \(r=0.47, p<0.0001\), Dexcom \(r=0.40, p<0.0001\)). Linear correlations were similar when meals were split into breakfast, lunch, and dinner (Abbott breakfast \(r=0.41\), lunch \(r=0.55\), dinner \(r=0.44\); Dexcom breakfast \(r=0.40\), lunch \(r=0.41\), dinner \(r=0.35\), all \(p<0.0001\)).

Intra-class correlation coefficients (ICCs) were 0.31 for Abbott and 0.16 for Dexcom, indicating that there was a low tendency for glucose responses to be similar in duplicate
meals in the same participant. Across all duplicate meals and subjects, Bland-Altman plots are shown in Figures 1C and 1D indicating a low mean bias between iAUC responses to duplicate meals (Abbott 0.1 mg/dL, Dexcom -0.3 mg/dL), but there was a large variability indicated by the wide 95% limits of agreement (LoA) for both CGMs (Abbott -31.4 to 31.5 mg/dL, Dexcom -32.3 to 31.6 mg/dL).

Figures 2A and 2B plot the differences in glycemic responses to duplicate meals for each individual participant using the Abbott and Dexcom devices, respectively, and show highly variable glycemic responses when the same participant consumed duplicate meals on separate weeks, regardless of the CGM device. However, the iAUC bias was relatively low for most participants when averaged across different duplicate meals. Figures 2C and 2D plot the same data separated by individual duplicate meals as measured using the Abbott and Dexcom devices, respectively, and indicate highly variable individual glycemic responses to duplicate meals, regardless of the CGM device. Nevertheless, the iAUC bias was relatively low for most meals when averaged across participants.

Similar individual glycemic response variability to duplicate versus different meals

Surprisingly, we found that everyone’s glycemic response variability to duplicate meals was similar to the variability in their glycemic responses to different meals. Figure 3A plots the SD of the glycemic responses in each individual participant to different meals eaten in either week 1 or week 2 along with the SD of their glycemic responses to
duplicate meals for the Abbott device. **Figure 3B** plots analogous data from the Dexcom device. Regardless of device, the variability in the glycemic response to duplicate meals was similar to each participant’s glycemic response variability to duplicate meals (Abbott: $SD_{week\ 1} = 12.4 \text{ mg/dL}$, $SD_{week\ 2} = 11.6 \text{ mg/dL}$, $SD_{duplicate} = 10.7 \text{ mg/dL}$, $p=0.38$; Dexcom: $SD_{week\ 1} = 12.2 \text{ mg/dL}$, $SD_{week\ 2} = 12.4 \text{ mg/dL}$, $SD_{duplicate} = 11.8 \text{ mg/dL}$, $p=0.80$).

**Potential factors affecting variability in glycemic response to duplicate meals**

We explored numerous factors relating to the study design and behavior that might contribute to explaining the glycemic response variability to duplicate meals. Firstly, there was a moderate positive linear correlation between differences in baseline glucose and differences in iAUC (Abbott $r=0.42 \ p<0.0001$, Dexcom $r=0.50 \ p<0.0001$), suggesting baseline glucose concentrations may have contributed to the low repeatability of glucose iAUC.

Because food intake was ad libitum in our studies, we investigated whether differences in meal energy intake between duplicate meal weeks affected our findings. For Abbott, mean (95% CI) meal energy intake was 777 (746 to 808) kcal in week 1 and 744 (712 to 777) kcal in week 2 ($p=0.0007$). For Dexcom, mean (95% CI) meal energy intake was 784 (752 to 816) kcal in week 1 and 763 (733 to 793) kcal in week 2 ($p=0.02$). Energy intake between duplicate meals was strongly positively correlated (Abbott $r=0.83 \ p<0.0001$, Dexcom $r=0.85 \ p<0.0001$) and there was a weak positive correlation between differences in energy intake and differences in glucose iAUC (Abbott $r=0.24 \ p<0.0001$, Dexcom $r=0.13 \ p=0.002$). However, repeating our analyses using only
duplicate meals where energy intake was within 100 kcal between meals did not materially affect our results regarding the iAUC correlations (Abbott n=201, r=0.43 \( p<0.0001 \), Dexcom n=277, r=0.44 \( p<0.0001 \)) or ICC (Abbott 0.29, Dexcom 0.17) or Bland-Altman analyses (Abbott bias -0.5 mg/dL, LoA -32.0 to 31.0 mg/dL; Dexcom bias -0.8 mg/dL, LoA -32.4 to 30.8 mg/dL).

In addition to the three daily meals provided, participants were also given snacks that could be consumed at any time of day. To examine whether our results may have been affected by differences in snack intake between days with duplicate meals, we filtered the data such that snack intake was <200 kcal on both duplicate meal days resulting in 136 duplicates meals available for Abbott and 207 for Dexcom. For Abbott, mean (95% CI) meal energy intake was 791 (738 to 845) kcal in week 1 and 748 (694 to 802) kcal in week 2 (\( p=0.008 \)) and mean (95%) snack intake was 39 (28 to 50) kcal/d in week 1 and 32 (23 to 40) kcal/d in week 2 (\( p=0.31 \)). For Dexcom, mean (95% CI) meal energy intake was 707 (658 to 757) kcal in week 1 and 703 (657 to 749) kcal in week 2 (\( p=0.72 \)) and mean (95%) snack intake was 29 (22 to 37) kcal/d in week 1 and 20 (15 to 26) kcal/d in week 2 (\( p=0.05 \)). Repeating our analyses using only meals where snack intake was less than 200 kcal did not materially affect our results regarding the iAUC correlations (Abbott r=0.51 \( p<0.0001 \), Dexcom r=0.44 \( p<0.0001 \)), ICC (Abbott 0.34, Dexcom 0.22), or Bland-Altman analyses (Abbott bias 0.4 mg/dL, LoA -27.4 to 28.2 mg/dL; Dexcom bias 1.7 mg/dL, LoA -29.2 to 33.4 mg/dL).

Discussion
CGM devices are becoming widely used in people without diabetes as part of commercial precision nutrition programs that provide personalized diet advice (9), however CGM responses need to be precise and accurate to be useful (10). We recently demonstrated that postprandial glycemic responses using two different brands of CGMs simultaneously worn on different anatomical locations resulted in only moderate correlations of within-subject postprandial responses to simultaneously measured multicomponent meals ($r=0.68$) and modest concordance of the meal rankings by iAUC (Kendall rank correlation $= 0.43$) (11). A subsequent study using simple test meals (i.e., muffins, milkshakes, and energy bars) formulated to have substantial differences in glycemic load confirmed that simultaneous within-subject postprandial iAUCs measured using different CGM devices were only moderately correlated ($r=0.61$) but the rank order of these simple meals according to iAUC was more concordant (Kendall rank correlation $= 0.68$) than the rankings of multicomponent meals in our previous study (12). Interestingly, using identical CGMs resulted in much better agreement ($r=0.97$; Kendall rank correlation $= 0.87$) suggesting that a given CGM device provides valid measures of postprandial glycemic responses to simple test meals on a single occasion (12). However, this does not address the reliability of within-subject responses to repeated multicomponent meals.

The fundamental assumption of personalized or “precision” nutrition is that an individual’s responses to repeated meals are less variable than their responses to different meals. Otherwise, it would be impossible to provide reliable advice to avoid meals that result in poor responses. Previous work found relatively reliable postprandial
CGM responses to a small number of duplicate simple meals like bread (2) or muffins (1), but such meals are not representative of multicomponent meals that are the focus of personalized dietary advice in the real-world. Surprisingly, our study found that the reliability of postprandial CGM responses to many duplicate multicomponent meals was poor and that the within-subject variability to duplicate meals was similar to the variability across different meals. Perhaps this is why recent randomized trials comparing personalized nutrition interventions focused on glycemic responses observed small effects for mean glucose (within 7 mg/dL, 0.39 mmol/L) and HbA1c (within 0.14%) (13), or no differences in glycemic variability and HbA1c (14) as compared to general diet advice.

The low within-subject reliability of postprandial CGM responses to duplicate multicomponent meals in our study occurred under highly controlled metabolic ward conditions where meal order within each day was standardized and was typically preceded by a previous standardized day. Whilst less reflective of free-living conditions, such inpatient controlled feeding studies reduce the amount of variability explained by behavioral factors, enabling better understanding of the amount of glycemic variability that can be explained by ingestion of meals, providing a better indication of measurement error (15, 16). However, despite the strengths of our inpatient controlled feeding design, our study had several limitations. First, the primary aims of the original studies were to measure ad libitum energy intake differences between dietary patterns and therefore duplicate meals were not necessarily consumed in identical amounts, although energy intake of duplicate meals was highly correlated and repeating the
analyses using only meals within 100 kcal of each other did not change interpretation. Furthermore, despite the regimented meal order and timing achieved with implementation of the 7-day rotating menus, snacks were available for consumption at any time of day which may have differentially affected meal responses. Re-analysis using only duplicate meals on days when snack intake was within 200 kcal or when the energy intake difference between duplicate meals was <100 kcal did not materially affect our results.

Due to the inpatient setting, our study has limited generalizability to free-living people. However, free-living behaviors will likely further increase the within-subject variability of CGM responses to similar meals. A plethora of modifiable behavioral factors can also influence postprandial glycemic responses to the same meal within an individual and the reasons for the variable responses to repeated meals in our study are presently unknown. In our study, meals were ad libitum, but participants tended to eat similar amounts of the repeated meals and differences in energy intake did not seem to account for the differences in glycemic response, as the variability was similar when only analyzing meals with similar energy intake within 100 kcal (data not shown). However, variations in the sequence of foods consumed within the repeated multicomponent ad libitum meals may have contributed to the variability because food sequence has been previously shown to result in varying glycemic responses in people with and without type 2 diabetes (17-20). Physical activity differences may have also played a role, as previous studies have shown that breaking up prolonged sitting with small amounts of physical activity during the postprandial period reduces postprandial...
glycemia (21-23), and even leg fidgeting may have an effect (24). Sleep quality and bedtime has recently been associated with changes in CGM-derived measures of postprandial glucose (25), so variations in sleep quality may have contributed to differences in the studies presented. Importantly, if such behavioral factors are indeed important contributors to meal glycemic responses, then an enormous amount of data may be required to capture these behavioral determinants and reliably predict an individual’s glucose excursions and thereby provide personalized “precision” diet advice.

Intriguingly, we found a low mean bias of within-subject iAUCs in response to multiple duplicate meals suggesting that it may be possible to reliably estimate within-subject postprandial responses to the same meals provided that enough repeated measurements are made. How many repeated postprandial CGM measurements in response to the same meals are required within each individual to provide reliable personalized estimates is a critical question for future research, but our results suggest that two measurements are too few even under highly standardized metabolic ward conditions.

In conclusion, our data suggest that personalized diet advice is unlikely to be reliable if it is based primarily on postprandial CGM measurements obtained using very few repeated measurements in adults without diabetes. Instead, precision nutrition requires more reliable methods involving aggregated repeated measurements.
Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Institute of Diabetes & Digestive & Kidney Diseases under award number 1ZIADK013037. We thank the nursing and nutrition staff at the NIH Metabolic Clinical Research Unit for their invaluable assistance with this study. We thank the study participants for their invaluable contribution.

Author contributions

JG and KDH conceptualized and designed the research, AH, JG, and KDH analyzed and interpreted the data, and critically reviewed, drafted, and approved the final manuscript.

Conflict of interest

The authors report no conflicts of interest.
## Tables

**Table 1.** Baseline characteristics of participants included in analyses. Data mean ±SD.

<table>
<thead>
<tr>
<th></th>
<th>Abbott (n=14)</th>
<th>Dexcom (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (F,M)</td>
<td>F=8, M=6</td>
<td>F=15, M=15</td>
</tr>
<tr>
<td>Age (y)</td>
<td>31 ±8</td>
<td>30 ±7</td>
</tr>
<tr>
<td>Body mass (kg)</td>
<td>73.0 ±13.3</td>
<td>80.6 ±19.4</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>25.5 ±5.2</td>
<td>27.6 ±6.3</td>
</tr>
<tr>
<td>Body fat (%)</td>
<td>28.9 ±11.1</td>
<td>31.9 ±9.9</td>
</tr>
</tbody>
</table>
Figures

Figure 1. Comparison of incremental area under the curve (iAUC) of postprandial glucose responses to duplicate meals using A) Abbott and B) Dexcom continuous glucose monitors. Trendline is major axis regression. Bland-Altman plots of the iAUC differences between duplicate meals versus the average of both measurements using C) Abbott and D) Dexcom devices. Solid line indicates mean bias and dashed lines indicate 95% limits of agreement.
Figure 2. Mean ±SD difference and individual comparisons of duplicate meals organized by participant using A) Abbott and B) Dexcom devices. Each data point is the within-participant iAUC difference between duplicate meals. Mean ±SD and individual comparisons of duplicate meals ordered by meal pairing (across all participants) using C) Abbott and D) Dexcom CGMs. Each data point is a duplicate meal eaten in week 2 minus the same meal eaten in week 1 with data from all participants who consumed that meal (abbott has 42 total meals for comparison across the 14 days of rotating menu, 14 days x 3 meals; dexcom has 63 total meals for comparison across 21 days of rotating menu, 21 days x 3 meals).
Figure 3. Mean and individual participant standard deviations (SD) of postprandial glucose responses between different meals across week 1, different meals across week 2, and duplicate meals between weeks using A) Abbott and B) Dexcom devices.
References


