620 Final Report: Assessing the Impact of Shortened Follow-up time in Continuous Glucose Monitor Clinical Trials

Walter Williamson, Neo Kok, Will Tackett

Abstract

Continuous glucose monitoring (CGM) has transformed the management of type 1 diabetes by measuring high-resolution glucose data in real time. While CGM-derived metrics such as time in range are increasingly used in clinical trials, the optimal duration of follow-up needed to detect treatment effects remains unclear. This study examined whether shorter follow-up periods can yield results comparable to the standard six-month duration used in a landmark randomized trial of a closed-loop insulin delivery system.

We conducted a secondary analysis of CGM data from Brown et al. (2019), evaluating TIR and other glucose metrics across multiple follow-up windows (7 to 90 days). Linear mixed-effects models were used to estimate treatment effects and assess whether results from shorter durations differed significantly from the full trial period.

Findings showed that treatment effect estimates for TIR, time below range, and hyperglycemic episode rate remained consistent across follow-up windows 14 days. Only the 7-day window for coefficient of variation differed significantly. These results suggest that, in trials of fast-acting interventions like closed-loop systems, shorter follow-up periods may be sufficient to detect treatment effects, potentially reducing trial cost and complexity without compromising analytic validity.

Introduction

Type 1 diabetes is a chronic autoimmune disease in which the body's immune system destroys insulin-producing cells in the pancreas. Thus, individuals with this condition lack sufficient amounts of insulin, a hormone that facilitates the transfer of blood glucose into cells where it can be used for energy. The result is abnormal, often higher levels of glucose in the blood, which, if left unmanaged, can lead to serious complications over time. These complications include cardiovascular disease, nerve damage (neuropathy), kidney failure (nephropathy), vision

loss (retinopathy), and poor wound healing. Effective management of type 1 diabetes requires interventions to maintain glycemic control, as well as careful monitoring of blood glucose levels to reduce the risk of long-term health issues.

Continuous glucose monitoring (CGM) has become a valuable tool for the management of Type I diabetes and diabetes in general, providing continuous, real-time glucose measurements from interstitial fluid. The introduction of CGM devices has advanced diabetes management by facilitating improved glycemic control and reducing the frequency of hypo- and hyper-glycemic events (Klupa et al). While diabetes studies have historically used the Hemoglobin A1c (HbA1c) test as an outcome measure of blood sugar level, the increasing use of CGMs has enabled clinical trials to examine other outcomes such as time in range, the percentage of over-all time in which glucose measurements stayed within a predetermined healthy range of glucose levels. These CGM-derived metrics offer a more granular measure of glucose levels compared to A1c, for example, which represents a 3-month average. These metrics, typically calculated using a minimum of fourteen days of CGM data (Bergenstal 2018), are increasingly widely used in clinical practice and research to assess the effectiveness of diabetes interventions.

However, the appropriate follow-up period for these measures has not been examined. Longer follow-up periods may provide data on the sustained effectiveness and stability of glucose con-trol interventions, but they incur greater costs, greater logistical complexity, and a higher risk of participant attrition. Conversely, shorter follow-up times may reduce participant burden, trial complexity, and overall cost, but the validity of outcomes derived from shorter observa-tion periods is uncertain. There is some evidence that 14 days of CGM data can predict up to 3 months (Bergenstal 2018). Determining whether shorter durations of follow-up are sufficient to yield reliable and sensitive clinical endpoints is potentially important for improving the efficiency of future clinical trials in CGM research.

This project aimed to evaluate whether shorter follow-up periods in CGM-based clinical trials can yield results comparable to those obtained from standard, longer-duration follow-up. To investigate this, we performed a secondary analysis of publicly available data from a land-mark clinical trial conducted by Brown et al. (2019), published in the New England Journal of Medicine. This trial examined the efficacy of a closed-loop insulin delivery system, the Medtronic MiniMed 670G, in improving glucose control among individuals with Type 1 Dia-betes. A closed-loop system is a kind of device that combines CGM with an automated insulin delivery system of some kind, so that real-time glucose measurements inform the automatic delivery of insulin, minimizing user burden.

Specifically, the goal of this project was to compare the sensitivity and consistency of clinical endpoints measured by CGM calculated at shorter follow-up durations (e.g., 2 weeks, 1 month) against the full 6-month period used in the original trial. By addressing the scientific question of whether CGM clinical trial endpoints are sensitive to differing lengths of follow-up after an intervention, this project will provide insight into optimal clinical trial design.

Methods

We performed a retrospective secondary analysis of a previously published, parallel-group, randomized clinical trial that enrolled 168 adults with type 1 diabetes across seven sites (Brown et al., 2019). The original study evaluated the efficacy of a fully automated closed-loop control (CLC) insulin delivery system compared to a sensor-augmented pump (SAP) system. The detailed trial protocol can be found in Brown et al (2019).

All subjects wore Dexcom G6 continuous glucose monitors (CGMs) that record interstitial glucose every 5 minutes. CGM data were collected for at least 2 weeks before randomization for the baseline phase and up to six months during the treatment phase. All participants from the original trial were also included in this analysis. The data were extracted from the JAEB Center for Health Research at https://public.jaeb.org with pre-processing available through the Awesome-CGM repository (Xu et al., 2024).

The primary outcome metric for this analysis, like in the original, is time in range (TIR) - the proportion of continuous glucose readings between 70 and 180 mg/dL. Secondary endpoints include time-below-range (TBR), defined as the percentage of readings less than 70 mg/dL; the coefficient of variation (CV), calculated as $100 \times$ (glucose standard deviation / mean); and the hyperglycemic episode rate, the average number of daily events in which glucose exceeds 180 mg/dL for at least 15 minutes (Battelino et al., 2023). All metrics are calculated using the iglu R package (Broll et al., 2021). Each metric was calculated for the baseline period and six post-randomization windows: 7, 14, 30, 60, and 90 days, as well as the full treatment duration, resulting in seven values per metric for each person.

To quantify treatment-related changes across follow-up windows, we fitted a separate linear mixed-effects model for each CGM-derived outcome. The response was the within-participant change from baseline in window j:

$$
\begin{aligned}
\Delta \text{Metric}_{ij} = \beta_0 &+ \beta_1 \, \text{Treatment}_i + \beta_2 \, \text{Metric\_Baseline}_i \\
&+ \beta_3 \, \text{Age}_i + \beta_4 \, \text{InsulinModality}_i + \beta_5 \, \text{PriorCGM}_i \\
&+ \beta_6 \left(\text{Treatment}_i \times \text{Window}_j \right) + u_{\text{subject}(i)} + \varepsilon_{ij}
\end{aligned}
$$

Where
$$
\Delta \text{Metric}_{ij} = \texttt{Metric\_Window}_{ij} - \texttt{Metric\_Baseline}_i;
$$

SAP is the reference category for the treatment variable; *Metric_Baseline* denotes each participant's baseline value of the outcome metric; *Age* is the participant's age at enrollment; *InsulinModality* records the pre-trial insulin-delivery method (pump vs. injections); *PriorCGM* is a binary flag for prior CGM use; and *Window* is a categorical factor for post-randomization time windows, with the full-length window as the reference level. The random effect is written $u_{\text{subject}(i)}$ for the participant-specific intercept, and the residual error is ε_{ij}. This mirrors the

original trial model except for two changes: the explicit inclusion of *Window* as a fixed effect and the replacement of site-level random effects with participant-level intercepts, because variance attributable to study site was negligible.

Residual diagnostics were examined for heteroskedasticity. When non-constant variance was detected, a $\log(\cdot + 1)$ transformation was applied to both the outcome (ΔMetric) and its baseline covariate (*Metric_Baseline*) to stabilize variance. All other variables remained untransformed.

$$\Delta\log(\text{Metric})_{ij} = \beta_0 + \beta_1 \text{Treatment}_i + \beta_2 \log(\text{Metric\_Baseline}_i + 1)$$
$$+ \beta_3 \text{Age}_i + \beta_4 \text{InsulinModality}_i + \beta_5 \text{PriorCGM}_i$$
$$+ \beta_6 \left(\text{Treatment}_i \times \text{Window}_j \right) + u_{\text{subject}(i)} + \varepsilon_{ij}$$

All models were fitted using the lme4 R package (Bates et al., 2003).

We performed a likelihood ratio test comparing the full models to a reduced model without the interaction term to evaluate whether including the interaction significantly improves model fit. We followed this by estimating the magnitude and statistical significance of differences in treatment effects across windows, using the lmerTest R package, which applies Satterthwaite's method for computing p-values for fixed effects in mixed effects models (Kuznetsova et al., 2013).

Family-wise type I error for the five post-randomization windows was controlled with a Bonferroni-adjusted alpha = 0.01. We set alpha = 0.05 for all other statistical tests in this analysis.

All analyses were performed in R version 4.4.1 and are publicly available on GitHub.

Results

Summary statistics of demographic distribution reveal a relatively balanced composition of individuals randomized to treatment and control groups, as shown in Table 1.

	CLC (N=112)	SAP (N=56)	Overall (N=168)
Age	32.7 (15.7)	32.9 (16.7)	32.8 (16.0)
Age Group			
<18	31 (27.7%)	17 (30.4%)	48 (28.6%)
>=18	81 (72.3%)	39 (69.6%)	120 (71.4%)
BMI	25.3 (34.0)	28.3 (46.3)	26.3 (38.4)
Gender			
F	54 (48.2%)	30 (53.6%)	84 (50.0%)
M	58 (51.8%)	26 (46.4%)	84 (50.0%)
Race			
Asian	3 (2.7%)	2 (3.6%)	5 (3.0%)
Black/African American	4 (3.6%)	0 (0%)	4 (2.4%)
Other	11 (9.8%)	1 (1.8%)	12 (7.1%)
White	94 (83.9%)	53 (94.6%)	147 (87.5%)
Ethnicity			
Hispanic or Latino	13 (11.6%)	5 (8.9%)	18 (10.7%)
Not Hispanic or Latino	99 (88.4%)	51 (91.1%)	150 (89.3%)
Previous CGM Use			
Current	78 (69.6%)	40 (71.4%)	118 (70.2%)
In past, but not current	26 (23.2%)	10 (17.9%)	36 (21.4%)
Never	8 (7.1%)	6 (10.7%)	14 (8.3%)
Administration Type			
Injections	22 (19.6%)	13 (23.2%)	35 (20.8%)
Pump	90 (80.4%)	43 (76.8%)	133 (79.2%)

Table 1: Demographic characteristics of trial participants, stratified by treatment group, where SAP represents the sensor augmented pump control group, and CLC represents the closed loop system treatment group. Administration type represents the personal treatment method preceding the study.

Visual comparison of baseline metric distributions with those obtained from all shortened follow-up windows and the original follow-up window across endpoints suggest a similar treatment effect estimate across shortened windows, and similar treatment effect compared to the initial full length follow-up period (Figure 1). There appears to be a consistent treatment effect difference across windows and endpoints, and consistent distribution of control group metrics across windows and endpoints. Closed loop system treatment appears to increase time in range, and decrease time below range, hyperglycemic episode rate, and coefficient variation, all of which indicate an improvement in diabetes management.
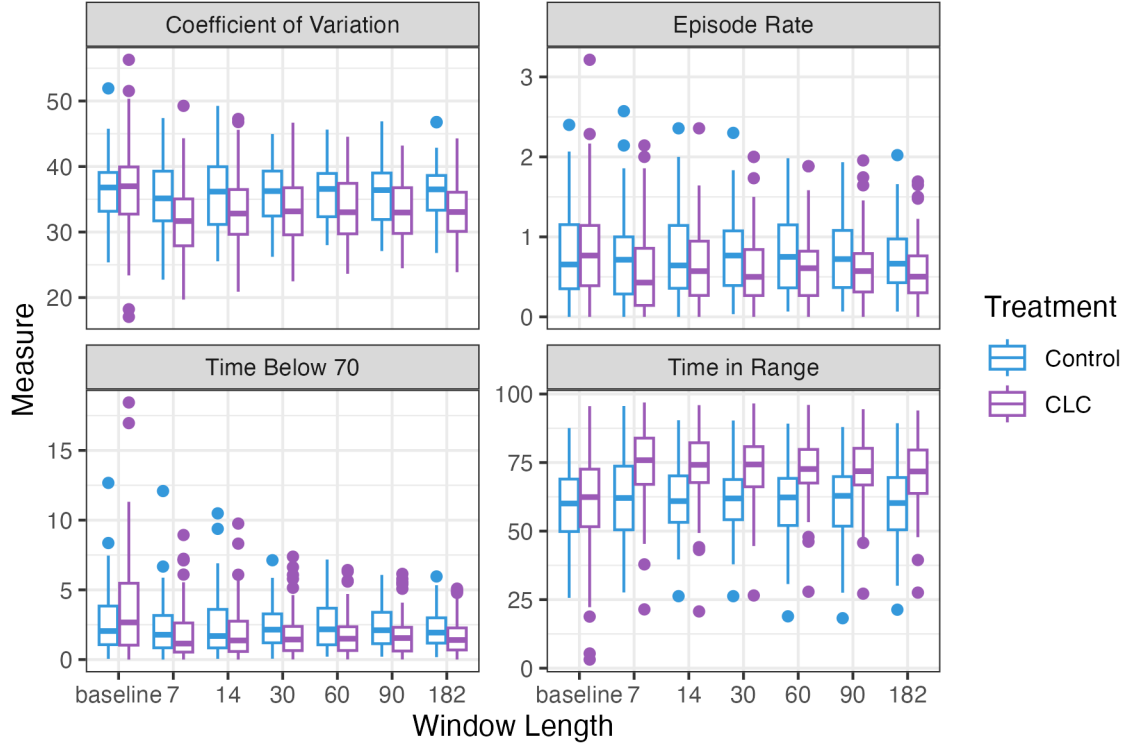
Figure 1: Metric distribution at baseline and with various intervention follow-up windows. Distributions of treatment and control group metrics are represented by purple and blue boxplots, respectively. Window length is in days.

Replication of original trial results using mixed effects modeling with the full, 6-month follow-up period resulted in similar treatment effect estimates across endpoints. Accounting for repeated measures from the same individual and interaction between shortened follow-up window and treatment group, and adjusting for the same covariates as the original analysis, resulted in consistent findings across metrics (Figure 2).

For our primary endpoint, time in range, we found there was not a significant difference in estimated treatment effect between shorter follow-up windows and the full length follow-up period, across window lengths 7 days to 90 days, indicating a similar treatment effect is captured within 7 days and 6 months. The same pattern was observed for both time below range and episode rate, which were log transformed and represent a multiplicative treatment effect. However, for coefficient of variation, we observed a significantly different treatment effect in the 7-day follow-up window compared to the full length period, suggesting the treatment effect is exaggerated in the shortest follow-up period. For all other follow-up periods greater than or equal to 14 days in length, there was no significant difference in treatment effect from the full length follow-up period.
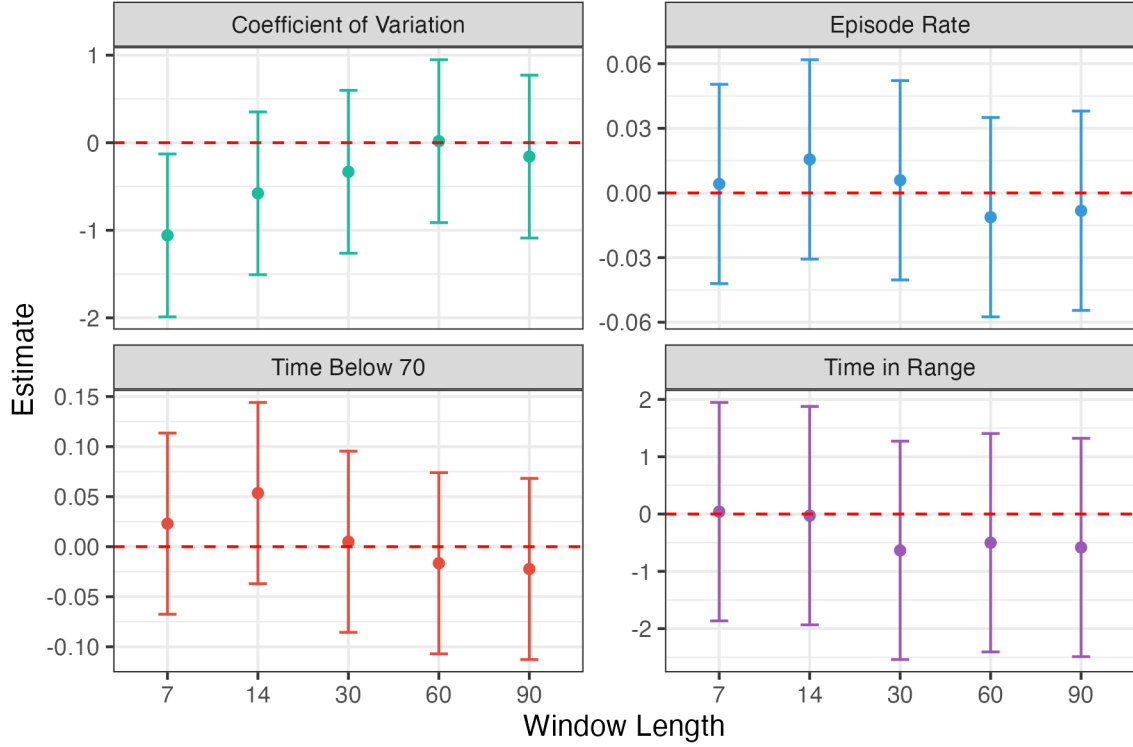
Figure 2: Estimated difference in treatment effect between full length follow-up period and a series of shortened follow-up periods, across CGM metric endpoints. 95% confidence intervals containing the red dashed line indicated a lack of significant difference. Estimates derived from linear mixed effects model.

Likelihood ratio tests to assess the addition of a treatment group and window length interaction revealed a non-significant finding for TIR (25 = 1.09, p = 0.95), CV (25 = 7.60, p = 0.18), Time below range (25 = 3.68, p = 0.60), and hyperglycemic episode rate (25 = 1.75, p = 0.88)

Discussion

In this study, we investigated the potential for shortening the intervention follow-up period in continuous glucose monitor trials, using the experiment from Brown et al. 2019 as an example. We discovered that across CGM metric endpoints, shortened follow-up intervals resulted in similar estimated treatment effects with linear mixed effects modeling. This may indicate that, for the sake of treatment estimation, a full 6-month follow-up period may not be necessary to capture the treatment effect and diabetes management improvement associated with using a closed-loop system.

The exception to this finding was the 7-day follow-up period for coefficient of variation, which

we found to have a significantly lower CV compared to the full length, 6-month, study period treatment effect. This may be due to the small sample size associated with 7 days of data, as CV is a function of the mean and will thus provide a better estimate of diabetes management with more observations. Notably, this finding does not remain statistically after multiple comparison adjustments.

While treatment effects may be captured in as few as 7 to 14 days in this study, a shorter follow-up period may not be sufficient to identify or observe adverse events or safety concerns associated with treatment. In addition, this finding is unlikely to hold in settings where the treatment effect is delayed or takes time to present itself. In the case of a closed-loop system, it makes sense to see an immediate effect, as glucose management decision-making is modified immediately after the beginning of the trial. However, in the hypothetical case of a prescription medicine or exercise regimen intervention, it may be reasonable to expect a delayed treatment effect.

In light of these considerations, we recommend only shortening clinical trial length in specific cases, where treatment is fast acting and treatment safety concerns are minimal. This limits the generalizability of our result. One potential adaptation to address these concerns could be a flexible study design using intermediate endpoints, where trial length is adapted on an individual basis according to observed treatment effect or adverse events, allowing for early stopping. Based on our results, it may be wise to explore additional short follow-up periods that align with practical considerations, such as the 10-14-day lifetime of commonly used CGM sensors. This area represents a promising direction for study, which may allow CGM trials to reduce per-person cost and increase scale, while ensuring quality findings.

Acknowledgement

References

Bergenstal RM. Understanding Continuous Glucose Monitoring Data. In: Role of Continuous Glucose Monitoring in Diabetes Treatment. ADA Clinical Compendia Series. American Diabetes Association; 2018. Accessed April 24, 2025. http://www.ncbi.nlm.nih.gov/books/NBK538967/

Klupa T, Czupryniak L, Dzida G, et al. Expanding the Role of Continuous Glucose Monitoring in Modern Diabetes Care Beyond Type 1 Disease. Diabetes Ther. 2023;14(8):1241-1266. doi:10.1007/s13300-023-01431-3

Brown SA, Kovatchev BP, Raghinaru D, et al. Six-month randomized, Multicenter Trial of closed-loop control in type 1 diabetes. *New England Journal of Medicine.* 2019;381(18):1707-1717. doi:10.1056/nejmoa1907863

Xinran Xu, Neo Kok, Junyan Tan, Mary Martin, David Buchanan, Elizabeth Chun, et al. IrinaStatsLab/Awesome-CGM: Updated release with additional public CGM dataset and enhanced processing. Zenodo; 2024.

Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement [published correction appears in Lancet Diabetes Endocrinol. 2024 Feb;12(2):e12. doi: 10.1016/S2213-8587(24)00001-9.]. *Lancet Diabetes Endocrinol.* 2023;11(1):42-57. doi:10.1016/S2213-8587(22)00319-9

Broll S, Urbanek J, Buchanan D, et al. Interpreting blood glucose data with R Package Iglu. *PLOS ONE.* 2021;16(4). doi:10.1371/journal.pone.0248560

Bates D, Maechler M, Bolker B, Walker S. LME4: Linear mixed-effects models using "Eigen" and S4. *CRAN: Contributed Packages.* Published online June 25, 2003. doi:10.32614/cran.package.lme4

Kuznetsova A, Bruun Brockhoff P, Haubo Bojesen Christensen R. LmerTest: Tests in linear mixed effects models. *CRAN: Contributed Packages.* Published online January 26, 2013. doi:10.32614/cran.package.lmertest