Methods to Improve Glucose Variability Estimates from Censored Data in Patients with Insulin Dependent T1DM

Nicholas Hein[1], Christopher Wichman[1], Lynette Smith[1]
Jennifer Merickel[2], Andjela Drincic[3], Mathew Rizzo[2], Cyrus Desouza[3]
[1]Department of Biostatistic, University of Nebraska Medical Center, 984375 Nebraska Medical Center, Omaha, NE 68198-4375
[2]Department of Neurological Sciences, University of Nebraska Medical Center, 988440 Nebraska Medical Center, Omaha, NE 68198-8440
[3]Department of Internal Medicine, University of Nebraska Medical Center, 983332 Nebraska Medical Center, Omaha, NE 68198-3332

Abstract
High glucose variability (GV) is associated with risk of serious hypo- and hyperglycemic events in patients with insulin-dependent diabetes. Accurate measurement of GV can help guide the best treatment for a patient. Continuous glucose monitoring systems (CGMS) give clinicians and patients real-time access to glucose levels over extended time frames. However, many patients have glucose levels above or below the CGMS sensor thresholds, producing censored values and challenges for accurate estimates of GV. A statistical technique to impute the censored values to calculate accurate GV measures is proposed.

A simple replacement strategy (replacement method) and a local non-linear least squares regression method (imputation method) were used to impute censored data in a simulation study. Clinically standard methods (standard deviation (SD), mean amplitude of glucose excursion, and coefficient of variation) were used to calculate GV.

Under simulation, the imputation method resulted in a root mean square error (RMSE) of 25.2 and mean bias of 14 when calculating SD. The replacement method had an RMSE 77.3 and mean bias of 65.2. For clinical data, the mean difference in SD was 16.1. We conclude that the imputation method is more accurate than the replacement method.

Key Words: glucose variability, diabetes, continuous glucose monitoring, imputation, censoring

1. Introduction

Diabetes mellitus (DM) affects more than 23 million people in the United States and continues to grow with increases in obesity, aging, and urbanization.[1] This presents a problem of patient safety and public health. DM is associated with a variety of health problems, including heart disease, stroke, blindness, peripheral nerve damage, kidney disease, amputations, and death.[2] Patient outcomes and safety can be improved through treatment and close control of blood glucose.[3,4]

Glucose control quantifies the average level, timing, and amplitude of the patient's glucose levels and provides patients and clinicians valuable information to guide treatment that reduces diabetes-related risk.[5,6] A standard metric of glucose control is hemoglobin A_{1c} (HbA_{1c}), the patient's 3-month average glucose.[7,8] Prospective clinical trials have shown a strong correlation between high HbA_{1c} and serious health problems.[3,7] Shortcomings of HbA_{1c} as a measure of glucose control are well recognized and include factors that affect glycation of hemoglobin and those related to red cell physiology (such as anemia). Further, HbA_{1c} does not account for duration and severity of hypo- and hyperglycemia, pointing to the need for additional indices of diabetes control.[9]

Glucose monitoring technology including continuous glucose monitoring systems (CGMS) has underscored the value of blood glucose variability (GV) as a measure of glucose control in diabetes. GV has been used in clinical and research settings and its relation to diabetes complications is a topic of active study.[10,11]

GV profiles depend on accurate measurement of amplitude and timing of glucose levels which can be obtained with CGMS,[4] which have now become a routine tool for diabetes care. CGMS provide a glucose reading every 1-10 minutes,[12] allowing detailed assessment of blood glucose variability over time. While traditional metrics such as standard deviation (SD), coefficient of variation (CV) and mean amplitude of glucose excursion (MAGE) are well established,[7] the best approach for estimating GV requires clarification and further detail.

The strategies for accurately estimating GV are theoretically and empirically guided by consideration of CGMS sensor-related limitations. Censoring, depending on the device, usually occurs above 400-500 mg/dL and below 40 mg/dL.[13-19] Censored values are typically reported as "HIGH" or "LOW", respectively.[13-19] GV derived from CGMS data in the presence of censoring may be biased, with the degree of bias varying with the degree of censoring. For example, a simple replacement method to account for censored values, e.g., replacing "HIGH" with 401 mg/dL or treating censored values as missing, produces downward biased GV metrics.

While there are multiple metrics that a clinician can use to estimate GV which can capture the amplitude and timing of the blood glucose measurements,[4,5,20] there are no studies describing methods to account for censoring of CGMS data when calculating GV.

2. Methods

We developed a technique that more accurately (less biased) computes GV from CGMS that has censored data. Three traditionally used metrics were chosen to estimate GV: SD, CV, and MAGE measured over a 24-hour period.[5] We chose SD and CV based on the American Association of Clinical Endocrinologists recommendations to report these metrics when publishing CGMS data.[21] We selected MAGE due to its common use in the literature.[22]

We tested the hypothesis that our method would have higher accuracy than a simple replacement strategy when calculating GV using SD, CV, and MAGE over a 24-hour period in patients with DM with censored CGMS data, as measured by root mean squared error (RMSE) and bias. We compared GV measures using a simple replacement and our proposed method using simulation and actual patient CGMS data. All analyses were performed using R[23] v3.4.2 on a PC running Windows 10.

2.1 CGMS data

The CGMS data came from a tertiary methods paper that grew out of a more extensive study.[24] The IRB approved clinical data[24] contains glucose profiles for 18 insulin-dependent individuals with clinically confirmed T1DM, used insulin daily, and had self-reported at-least biweekly episodes of hypoglycemia. Participants were 21-59 years of age and were excluded if major confounding medical conditions were reported during the first study visit.[24] 24-hour glucose profiles were excluded from this analysis if a particular 24-hour period did not contain any censored values or there was a substantial amount of missing data. The CGMS data were collected using a Dexcom G4 professional version over a four week period by the Mind & Brain Health Labs, Department of Neurological Sciences, University of Nebraska Medical Center. Dexcom CGMS censors data over 400 mg/dL and under 40 mg/dL as "HIGH" and "LOW", respectively.[17] Furthermore, the data were previously processed for quality.

We computed SD, CV, and MAGE for each 24-hour glucose profile using our proposed method and simple replacement method. We report the mean difference and minimum/maximum difference between the two methods for each metric. The calculated standard deviation from the two methods was compared using the Brown-Forsythe[25] test. Percentage of standard deviations that were significantly different at alpha level 0.05 are reported.

2.2 Simulated data

Blood-glucose concentration and blood-insulin concentration both vary as a function of time.[26] Furthermore, blood-glucose concentration and blood-insulin concentration are interlocked in a feedback loop, making glucose profiles oscillate.[26] The sine wave is a natural choice to model periodic systems.

Using half-cycle sine waves, we simulated glucose variability data. We arbitrarily translated the first half-cycle of the sine wave such that its initial value was 110 mg/dL. The amplitude and period of the half-cycle were allowed to vary randomly. The amplitude and period random variables were constrained based on the classification of glucose level, which was itself random. Probabilities for the classification, i.e., severe hyperglycemia, hyperglycemia, normal, etc. were determined using empirical observations. Glucose profiles from 37 clinical observations[24] (see Section 6) were used to determine the proportion of time spent in each category. The probabilities of the simulation algorithm were adjusted such that the simulated data resembled the empirical[24] results. Additionally, the length of a half-cycle, i.e., the period was established using empirical data.[24] Specifically, the minimum and maximum length of a period for a specific classification was the 5% and 95% quantile of the empirical data[24] for the respective classification.

We alternately generated a relative maximum half-cycle and relative minimum half-cycle of the sine wave. The generation of a half-cycle of the sine wave proceeds as follows.

1. A random uniform(0, 1) is generated and used to classify the glucose level as severe hyperglycemia (varies based on simulation), hyperglycemia (180 – 400 mg/dL), or normal (70 – 180 mg/dL) for half-cycles that are local maximums. The probability of severe hyperglycemia is 0.3, the probability of hyperglycemia is 0.6, and the probability of normal is 0.1. Similarly, for half-cycles that are relative minimums, a random uniform(0, 1) is generated and used to classify the

glucose level as severe hypoglycemia (varies based on simulation – 56 mg/dL), hypoglycemia (56 – 70 mg/dL), or normal/hyperglycemia (70 – 5 less than max of last value mg/dL). The probabilities for each category are 0.1 for severe hypoglycemia and hypoglycemia and 0.8 for normal/hyperglycemia.

2. Based on the classification of the half-cycle, a random uniform(a, b) and random uniform(c, d) were used to generate the amplitude and period, respectively. The amplitude represents the maximum/minimum glucose measured in mg/dL and the period represents the time in minutes for this particular half-cycle. Table 2 lists parameters based on half-cycle classification.

3. The first 70% to 100% of the half-cycle of the sine wave was randomly retained. A random uniform(0.7, 1) was generated and multiplied by the length of the half-cycle of the sine wave to determine the simulated values of the half-cycle of the sine-wave to keep. This was done to mimic a damped sine wave and allows for the simulation function to differ from the function being fit.

4. The average rate of change was calculated. If the magnitude of the average rate of change was greater than 3 mg/dL/min, the procedure restarts at step 2.

5. The last value of the half-cycle of the sine wave is the initial value of the next half-cycle of the sine wave.

6. The procedure was repeated until 24 hours of glucose data were simulated. Data points were simulated every 5 minutes to mimic the clinical data.[24]

7. Random noise was added to 70% of the simulated data. The simulated data that includes noise was chosen randomly using a random uniform(0,1). Random noise is in the form of random normal with mean 0 and standard deviation of 5.

8. If no censoring occurred during the simulated 24 hours or the last value was a censored value, the procedure restarts from the beginning.

9. Glucose levels that were > 400 mg/dL or < 40 mg/dL were censored based on limitations of the CGMS. The data before censoring is referred to as the full data set and after censoring is referred to as the censored data set.

We performed thirty-three simulations with N=500 replicates (see footnote Table 2). Each replicate contained 24-hours of glucose levels that we simulated at 5-minute intervals. The simulated data includes full data and censored data. Figure 1A and Figure 1B are two replicates of the simulation that show the simulated data before censoring and the imputed data. Figure 1A is a simulation where the imputed method performs well. Figure 1B is a simulation where the imputed method would produce biased GV calculations. We calculated SD, CV, and MAGE for the full data for every 24 hours of simulated glucose levels at the individual level. Using the imputation method and simple replacement method, SD, CV, and MAGE was calculated for the censored data for each replicate. The RMSE and mean bias were then calculated for SD, CV, and MAGE over all N=500 simulations. We repeated these calculations for each of the 33 simulations. We calculated RMSE as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_i - \hat{\theta}_i\right)^2}$$

where θ is estimated using the full data, and $\hat{\theta}$ is estimated using data from either the imputation method or simple replacement method. Mean bias is the average difference between the metric using the full data and either statistical technique using the censored data. Additionally, we report minimum/maximum bias. Lastly, three pairwise comparisons of the SDs were made using the Brown-Forsythe[25] test. Comparisons were made on individual 24-hour simulations. P-values were adjusted using a Bonferroni correction. Percentage of simulations that were significantly different at the alpha 0.05 level are reported.

3. GV metrics

SD: The most commonly used measure of GV is SD.[27] GV has a linear relationship to SD and HbA_{1c} especially in T1DM patients who have high HbA_{1c}.[4,27]

CV: CV is the ratio between SD and the arithmetic mean glycemic level.[4,5] CV aims to correct the error between the mean glucose level and SD.[4,5]

MAGE: MAGE is the mean of the daily glucose excursions that exceed the SD as measured over a 24 hour period.[4,5,27] MAGE is calculated from continuously monitored glucose. Recent literature has noted potential weaknesses in using MAGE. Specifically, the area under the curve as measured by the trapezoidal method is a more accurate approach; assessment of MAGE is dependent on the operator; MAGE is highly correlated with SD, and it is unclear whether smaller excursions would still have clinical importance.[4,27]

Other measures of GV exist[4,5,27] that we did not consider here.

4. Statistical methods

4.1 Simple replacement method

With this method, we replaced glucose levels that were censored with a fixed value. We set glucose levels that were censored because the reading was > 400 mg/dL to 401. Similarly, we replaced glucose levels that were censored because they were < 40 mg/dL with 39 mg/dL.

4.2 Imputation method

The imputation method relies on imputing the censored values of the CGMS and measuring variability using the imputed data. Censored values were imputed locally by fitting a sine function proposed by Ackerman and colleagues[26] that minimized the squared error of the values near the censored values. We considered Ackerman's et al.[26] function since a complex set of parameters can be reduced to a few and easily estimated using glucose data and Ackerman's et al.[26] model was developed from a biological perspective. The sine function that was fit is as follows

$$f(t) = Ae^{-\alpha t}sin(\omega t) + D_0$$

Using the data near the censored values, parameters $A, \alpha,$ and ω were estimated using a perturbation search method similar to the method proposed by Shiang and colleagues.[28] D_0 is set to $f(t_0)$. The perturbation search method requires each parameter to be given a set of constraints. We imposed the following constraints when performing the perturbation search method. We constrained Amplitude (A) between 60 and 800 with a perturbation of $\pm$ 10. The damping factor (α) was limited between -0.1 and 0.1 with a perturbation of $\pm$ 0.001. The constraint for the period (ω) varied based on the local region of imputation. We first calculated the period of the values near the censored values (described below). The minimum and maximum value of the period was then set to be 80% or 120%, respectively, of the calculated period near the censored values. The perturbation of the period was $\pm$ 0.0001.

Using these constraints, the perturbation search method proceeds as follows:

1. Calculate initial parameter values by randomly generating parameter values from within the parameter constraints.

2. Calculate the residual sum of squares (RSS) using the initial parameter values.

3. Randomly generate parameter values $A, \alpha,$ and ω from within the parameter constraints.

4. Select each parameters' three trial values, e.g., $A + \Delta_A$, A, and $A - \Delta_A$. Where Δ is a predetermined perturbation for the specified parameter. The trial values are set for each parameter.

5. Estimate the function values using the trial values and calculate the RSS. If the RSS of a trial value is less than the current RSS, set the parameter value to the trial value and set the current RSS to the trial RSS. Repeat for each parameter.

6. Repeat steps 3-5 n times.

Using the parameter estimates from the perturbation search method, we estimated the damped sine function used to impute censored values within the local region. The procedure was repeated if there were multiple local regions of censoring.

To determine values near the censored values, we located the relative minimum and maximum values for the censored data. Censored values > 400 are always a relative maximum. To the left and right of these censored values should be a relative minimum. The values between these relative minimums are the values near the censored values. If there were not at least five values (i.e., 25 minutes of data) to the left and five values to the right of the censored values, then imputation did not occur. We fit separate quadratic regressions on the left side and right side of the censored values. Using the fit quadratic regression, we examined the slope of each data point starting with the observations nearest the censored values and moving away. The mean of the first two slopes nearest the censored values was the reference slope. We excluded subsequent observations once a slope differed by at least 1 SD (calculated using the quadratic regression). Due to the small amount of time an individual glucose level is $\leq$ 40 mg/dL, these censored values were not imputed.

5. Simulation Results

The range of the upper bound of the amplitude caused the most variation in the results. There appears to be a very minimal difference in the results with different lower bounds of the amplitude. Therefore, results discussed will be based on the eight simulations where the lower bound of the amplitude was set to 25 mg/dL. We will limit the discussion to the estimation of SD. Results for CV and MAGE are similar since the calculation of CV and MAGE depend on SD (see Table 1). For all results see Table 3, Table 4, and Table 5.

As the upper bound of the amplitude increased, so did the Type I error rate, i.e., percent of the SDs different from the actual SDs as determined by the Brown-Forsythe test, regardless of the method (Figure 2A). The minimum Type I error rate was 0, and the maximum was 0.284 under the imputation method. Compared to a minimum Type I error rate of 0 and a maximum rate of 0.994 under the replacement method. When examining the RMSE of the two methods, a similar trend occurs. Higher amplitudes result in higher RMSE regardless of the method (Figure 2B). The minimum and maximum RMSE under the imputation method was 2.25 and 47.69, respectively. Compared to 3.89 and 109.03, respectively, under the replacement method. The actual mean SD for the respective RMSE were 101.64 and 215.21. The mean bias of the imputation method stayed close to zero with a slight trend towards over-estimation. As amplitude increased, the replacement method further under-estimated the SD (Figure 2C). The mean bias (min, max) at an amplitude of 800 was -1.1 (-137.3, 177.7) and -108.0 (-138.1, -0.04) when calculating SD using the imputation and replacement method, respectively.

Under the simulation where the range of values was 25 to 800 mg/dL inclusive, the imputation method had a Type I error rate of 0.124, and the replacement method had a Type I error rate of 0.626 when calculating SD. The actual mean SD was 147.0. The imputation method had an RMSE of 26.4 and mean bias (min, max) of 0.67 (-102.1, 166.7). While the replacement method had an RMSE of 55.1 and mean bias (min, max) of -45.0 (-129.3, 0).

6. Application

We excluded seven subjects for the following two reasons. The first is if glucose data for a 24-hour period contained five or fewer censored values (i.e., glucose higher than 400 mg/dL) or 24-hour periods where censoring occurred at the start or end of the 24-hour period were excluded. Additionally, we excluded 24-hour periods with substantial amounts of missing data. There were a total of 37 out of 542, 24-hour glucose measurements for the 11 individuals. The median number of 24-hour glucose measurements periods from a participant was 2 with one subject contributing 12 glucose measurements. Computing SD using the imputation and replacement method resulted in 24.3% of the calculations being significantly different. The mean SD for the 37 profiles was 114.5 and 97.8 for the imputation method and replacement method, respectively. The average difference in standard deviation between the two methods was 16.8 with a maximum difference of 109.5. The mean CV and MAGE under the imputation method was 0.45 and 268.3, respectively. While the replacement method produced a mean CV of 0.41 and mean MAGE of 207.1 for the 37 profiles. The mean difference when calculating CV and MAGE using the two methods was 0.04 and 61.5, respectively. The maximum

difference between the two methods when calculating CV was 0.2 and 397.0 when calculating MAGE.

7. Validation

The clinical data[24] with alternative inclusion criteria were used to validate the imputation method. 24-hour observations were included if that observation did not contain censored values (i.e., glucose > 400 mg/dL) and the 24-hour observation contained glucose readings between 350 to 400 mg/dL. We artificially censored the data at 300 mg/dL, i.e., we set glucose readings greater than 300 mg/dL to 301 mg/dL. Furthermore, 24-hour observations were excluded if censoring occurred at the beginning or end of a 24-hour observation, there was substantial missing data, or there were too few observations on either side of the censored data (see Section 4.2). The imputation method and simple replacement method were compared using the methods described in Section 2.2.

There were a total of 38, 24-hour glucose measurements for the 16 individuals. The median number of 24-hour glucose measurements periods from a participant was 2 with a maximum contribution of 5. The actual mean SD for all 38, 24-hour observations was 75.6. The imputation method had an RMSE of 5.5 and mean bias (min, max) of 1.62 (-7.8, 13.9) when calculating SD. 0.0% of SDs calculated using the imputation method were significantly different from the true SD. Whereas the replacement method had an RMSE 10.0 and mean bias (min, max) of -9.4 (-17.4, -3.1), and 21.1% of the sample SDs were significantly different than the actual SDs. Directly comparing the calculation of SD under the imputation and replacement method resulted in 26.3% of the estimates being significantly different.

Similar results presented under simulation when calculating CV and MAGE under the two methods. The average actual CV under all observations was 0.38, and the average true MAGE was 153.2. Under the imputation method, the mean bias (min, max) was 0.01 (-0.04, 0.05) and 43.7 (-74.5, 161.0) for CV and MAGE, respectively. The mean bias (min, max) using the replacement method was -0.04 (-0.07, -0.01) and -25.2 (-95.5, 21.8) for CV and MAGE, respectively.

8. Discussion

Through simulation and validation, we showed that the imputation method had a lower RMSE when calculating GV as estimated by SD, CV, and MAGE. Additionally, the mean bias of the imputation method was consistently smaller and nearer to zero than the replacement method. The replacement method under-estimated GV using the three metrics. The under-estimation became more pronounced as the amplitude increased and the replacement method never over-estimated GV using SD or CV as a metric. The imputation method also had better control over the Type I error rate compared to the replacement method. However, the imputation method is not without its limitations. The Type I error rate did not stay below 0.05 which tends to be the acceptable Type I error rate in the medical community; the imputation method would over-estimate GV on occasion. A possible explanation for the increased Type I error rate and over-estimation is the “behavior” of the glucose in the censored region. The imputation method expects a smooth hill (see Figure 1A). However, this does not always occur. The glucose could plateau or be bimodal, trimodal, etc. in the censored region which the imputation method is unable to detect (see Figure 1B). This would cause an over-estimation of the GV,

thereby increasing the RMSE and Type I error rate. This problem becomes exacerbated as the amplitude increases.

Overall, the imputation method was shown to be a more accurate estimate of GV as measured using SD, CV, and MAGE. We recommend plotting the original censored data and imputed data as a visual check. Additionally, when reporting GV using SD, CV, or MAGE we recommend specifying both the replacement method and imputation method.

9. Acknowledgments

We gratefully acknowledge support by the Chancellor Robert D. Sparks, M.D., Public Health Research Student Award, Toyota Collaborative Safety Research Center, and the Great Plains IDEA-CTR Network (1U54GM115458-01).

References

1. Centers for Disease Control and Prevention. National diabetes statistics report. . 2017.
2. Office of the Associate Director for Communication, Digital Media Branch, Division of Public Affairs. Putting the brakes on diabetes complications. https://www.cdc.gov/features/preventing-diabetes-complications/index.html. Updated 2017. Accessed May 08, 2018.
3. Ceriello A, Kilpatrick ES. Glycemic variability: Both sides of the story. *Diabetes Care*. 2013;36 Suppl 2:S272-5.
4. Frontoni S, Di Bartolo P, Avogaro A, Bosi E, Paolisso G, Ceriello A. Glucose variability: An emerging target for the treatment of diabetes mellitus. *Diabetes Res Clin Pract*. 2013;102(2):86-95.
5. Kovatchev B, Cobelli C. Glucose variability: Timing, risk analysis, and relationship to hypoglycemia in diabetes. *Diabetes Care*. 2016;39(4):502-510.
6. Rodbard D. Clinical interpretation of indices of quality of glycemic control and glycemic variability. *Postgrad Med*. 2011;123(4):107-118.
7. Kovatchev BP. Metrics for glycaemic control—from HbA 1c to continuous glucose monitoring. *Nature Reviews Endocrinology*. 2017;13(7):425.
8. Rohlfing CL, Wiedmeyer HM, Little RR, England JD, Tennill A, Goldstein DE. Defining the relationship between plasma glucose and HbA(1c): Analysis of glucose profiles and HbA(1c) in the diabetes control and complications trial. *Diabetes Care*. 2002;25(2):275-278.
9. Vigersky RA. Going beyond HbA1C to understand the benefits of advanced diabetes therapies. *Journal of diabetes*. 2018.
10. Hirsch IB. Glycemic variability and diabetes complications: Does it matter? of course it does! *Diabetes Care*. 2015;38(8):1610-1614.
11. Bergenstal RM. Glycemic variability and diabetes complications: Does it matter? simply put, there are better glycemic markers! *Diabetes Care*. 2015;38(8):1615-1621.
12. Klonoff DC. Continuous glucose monitoring: Roadmap for 21st century diabetes therapy. *Diabetes Care*. 2005;28(5):1231-1239.
13. MiniMed 530G system user guide. https://www.medtronicdiabetes.com/sites/default/files/library/download-library/user-guides/z10-mp6025813-014-000-a--mp6025813-014_a.pdf. Updated 2012. Accessed Apr. 18, 2018.
14. Getting started with minimed 670G continuous glucose monitoring. https://www.medtronicdiabetes.com/sites/default/files/library/download-

library/workbooks/Getting%20Started%20with%20MiniMed%20670G%20Continuous%20Glucose%20Monitoring.pdf. Updated 2016. Accessed Apr. 18, 2018.
15. iPro2 user guide. https://s3.amazonaws.com/medtronic-hcp/iPro2%20with%20Enlite%20Sensor%20User%20Guide.pdf. Updated 2016. Accessed Apr. 18, 2018.
16. MiniMed 630G system user guide. https://www.medtronicdiabetes.com/sites/default/files/library/download-library/user-guides/MP6026157-011_b.pdf. Updated 2016. Accessed Apr. 18, 2018.
17. Dexcom G4 platinum user's guide. https://s3-us-west-2.amazonaws.com/dexcompdf/LBL012528+Rev+004+User's+Guide%2C+G4+PLATINUM+with+Share+US+Web+with+cover.pdf. Updated 2017. Accessed Apr. 18, 2018.
18. FreeStyle libre pro flash glucose monitoring system operator's manual. https://freestyleserver.com/Payloads/IFU/2017_sep/ART26944-002_rev-A_WEB.pdf. Updated 2017. Accessed Apr. 18, 2018.
19. Dexcom G5 mobile system user guide. https://s3-us-west-2.amazonaws.com/dexcompdf/G5-Mobile-Users-Guide.pdf. Updated 2018. Accessed Apr. 18, 2018.
20. Service FJ. Glucose variability. *Diabetes*. 2013;62(5):1398-1404.
21. Klonoff DC, Ahn D, Drincic A. Continuous glucose monitoring: A review of the technology and clinical use. *Diabetes Res Clin Pract*. 2017;133:178-192.
22. Chon S. How can we easily measure glycemic variability in diabetes mellitus? *Diabetes & metabolism journal*. 2015;39(2):114-116.
23. R Core Team. R: A language and environment for statistical computing. . 2017. https://www.R-project.org/.
24. Merickel J, Robin H, Smith L, et al. Driving safety and real-time glucose monitoring in insulin-dependent diabetes. *Paper submitted to the International Symposium on Future Active Safety Technology, Nara, Japan*. 2017.
25. Brown MB, Forsythe AB. Robust tests for the equality of variances. *Journal of the American Statistical Association*. 1974;69(346):364-367.
26. Ackerman E, Rosevear JW, McGuckin WF. A mathematical model of the glucose-tolerance test. *Physics in Medicine & Biology*. 1964;9(2):203.
27. DeVries JH. Glucose variability: Where it is important and how to measure it. *Diabetes*. 2013;62(5):1405-1408.
28. Shiang K, Kandeel F. A computational model of the human glucose-insulin regulatory system. *Journal of biomedical research*. 2010;24(5):347-364.

Table 1: Formulas for SD, MAGE, and CV

Standard deviation	$$sd = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$$ where n is number of observations, y_i is an observed value, and μ is the mean of the observed values
CV	$$CV = sd/\mu$$ where sd is the standard deviation and μ is the mean
Mage	$$mage = \frac{1}{m}\sum_{p_i>1\ sd} p_i$$ where $p_i = peak - nadir$ and m is the number of $p_i's$ that exceed 1 sd

Table 2: Parameter values of half-cycle sine waves

	Amplitude		Period	
Classification	a[1]	b[1]	c	d
Severe Hyperglycemia	varies[2,4]	varies[2,4]	83	674
Hyperglycemia	180	400	25	323
Normal	70	180	25	195
Severe Hypoglycemia	varies[3,4]	56	59	533
Hypoglycemia	56	70	33	436
Normal/ Hyperglycemia	70	5 less than previous max	20	259

1. a represents the lower bound of the maximum/minimum of the half-cycle and b represents the upper bound of the maximum/minimum half-cycle.

2. 8 simulations of N=500 replicates each were performed with $(a, b) = (401, 450), (451, 500), \dots, (751, 800)$.

3. $a = 40, 35, 30, 25$

4. 32 simulations of N=500 replicates each were performed using all possible combinations of (1) and (2). 1 simulation was performed were serve high was allowed to vary between 401 and 800 and serve low was allowed to vary between 25 and 56.

Table 3: Simulation results for SD

Lower bound; upper bound (mg/dL)	Actual mean SD	Impute mean SD	Replace-ment mean SD	RMSE Impute	RMSE Replace-ment	Imputation bias (min, max)	Replacement bias (min, max)	Actual v. Imputate[1]	Actual v. Replace-ment[1]	Impute v. Replace-ment[1]
40; 401-450	100.45	100.29	97.31	2.26	3.85	-0.16 (-8.41, 17.38)	-3.15 (-11.51, 0)	0.002	0.004	0.004
40; 451-500	114.97	115.31	101.62	6.85	14.11	0.34 (-20.89, 61.74)	-13.35 (-25.98, 0)	0.034	0.246	0.242
40; 501-550	128.66	129.62	102.76	17.8	27.02	0.96 (-33.71, 227.72)	-25.91 (-46.33, 0)	0.086	0.574	0.548
40; 551-600	142.79	145.13	102.57	26.33	41.38	2.34 (-59.96, 211.1)	-40.22 (-73.29, -0.02)	0.164	0.880	0.780
40; 601-650	160.61	164.05	104.14	30.22	57.81	3.43 (-73.75, 316.6)	-56.47 (-83.37, -0.01)	0.170	0.980	0.912
40; 651-700	177.18	181.24	104.48	39.79	73.95	4.06 (-96.1, 243.23)	-72.71 (-100.78, -0.01)	0.246	0.992	0.92
40; 701-750	194.05	192.44	105.53	42.5	89.75	-1.6 (-113.53, 223.13)	-88.52 (-120.7, -0.01)	0.248	0.992	0.906
40; 751-800	211.11	212.99	104.17	47.62	108.57	1.88 (-132.47, 211.99)	-106.95 (-137.38, -0.01)	0.274	0.984	0.918
35; 401-450	100.90	100.81	97.79	2.24	3.81	-0.09 (-7.23, 15.01)	-3.12 (-9.62, 0)	0.000	0.000	0.004
35; 451-500	115.16	116.43	101.67	7.65	14.35	1.26 (-27.83, 51.64)	-13.49 (-28.13, -0.01)	0.030	0.248	0.246
35; 501-550	128.29	128.37	102.09	15.47	27.4	0.09 (-42.32, 202.02)	-26.19 (-42.65, 0)	0.104	0.596	0.548
35; 551-600	144.46	147.79	103.26	27.31	42.52	3.34 (-54.96, 206.59)	-41.19 (-64.16, -0.01)	0.176	0.888	0.798
35; 601-650	160.14	164.99	104.40	32.57	57.21	4.84 (-74.33, 240.49)	-55.74 (-82.97, -0.01)	0.210	0.982	0.910
35; 651-700	176.98	179.91	104.39	35.51	74.02	2.94 (-94.8, 178.32)	-72.59 (-101.18, -0.01)	0.234	0.988	0.926
35; 701-750	191.87	192.92	103.30	40.97	90.05	1.04 (-117.69, 169.62)	-88.58 (-117.69, 0)	0.248	0.986	0.908
35; 751-800	210.36	211.44	104.65	47.01	107.2	1.09 (-128.13, 257.16)	-105.71 (-138.23, 0)	0.276	0.986	0.918
30; 401-450	99.98	99.89	96.57	2.74	4.11	-0.08 (-9.02, 21.91)	-3.41 (-10.88, -0.01)	0.002	0.002	0.010
30; 451-500	114.72	115.70	101.30	7.29	14.24	0.98 (-21.08, 74.81)	-13.42 (-25.42, -0.01)	0.034	0.256	0.260
30; 501-550	129.40	130.47	103.43	17.65	27.02	1.08 (-43.51, 227.04)	-25.97 (-44.59, 0)	0.082	0.622	0.538
30; 551-600	145.06	147.49	104.15	26.07	42.13	2.42 (-59.18, 267.29)	-40.91 (-63.88, 0)	0.158	0.872	0.796
30; 601-650	160.47	165.23	104.07	31.57	57.62	4.75 (-78.95, 293.05)	-56.4 (-82.84, 0)	0.194	0.990	0.930
30; 651-700	176.55	178.06	104.78	33.8	73.03	1.5 (-90.68, 199.13)	-71.77 (-100.27, 0)	0.200	0.990	0.920
30; 701-750	192.60	192.23	105.30	41.66	89.05	-0.37 (-117.02, 224.13)	-87.3 (-120.42, -0.01)	0.256	0.980	0.894
30; 751-800	210.59	207.83	105.49	45.84	106.97	-2.76 (-140.02, 220.75)	-105.11 (-142.45, -0.01)	0.284	0.978	0.902
25; 401-450	101.64	101.61	98.45	2.25	3.89	-0.03 (-7, 15.5)	-3.19 (-8.96, 0)	0.000	0.000	0.002
25; 451-500	114.27	115.46	100.76	10.11	14.38	1.19 (-19, 172.15)	-13.51 (-24.76, 0)	0.022	0.250	0.264
25; 501-550	128.82	130.69	102.24	16.66	27.43	1.87 (-37.97, 217.28)	-26.58 (-44.63, 0)	0.084	0.634	0.594
25; 551-600	144.16	145.52	103.68	20.99	41.89	1.36 (-61.74, 195.03)	-40.48 (-61.74, 0)	0.160	0.894	0.804
25; 601-650	161.61	163.93	105.77	31.54	57.24	2.31 (-77.73, 213.97)	-55.85 (-80.59, 0)	0.216	0.976	0.886
25; 651-700	177.75	181.13	105.13	36.5	73.85	3.39 (-95.81, 207.44)	-72.61 (-101.59, -0.01)	0.228	0.992	0.914
25; 701-750	194.98	195.83	106.38	43.44	89.92	0.85 (-106.97, 201.01)	-88.6 (-119.56, -0.01)	0.266	0.990	0.896

25; 751-800	215.21	214.12	107.25	47.69	109.03	-1.1 (-137.3, 177.65)	-107.96 (-138.06, -0.04)	0.284	0.994	0.914
25; 401-800	147.01	147.69	102.07	26.43	55.14	0.67 (-102.09, 166.69)	-44.95 (-129.31, 0)	0.124	0.682	0.626

1. Percent different using Brown-Forsythe test and a Bonferroni correction.

Table 4: Simulation results for CV

Lower bound; upper bound (mg/dL)	Actual mean CV	Impute mean CV	Replacement mean CV	RMSE Impute	RMSE Replacement	Impute bias (min, max)	Replacement bias (min, max)
40; 401-450	0.41	0.41	0.41	0.01	0.01	0 (-0.02, 0.04)	-0.01 (-0.02, 0)
40; 451-500	0.45	0.45	0.41	0.02	0.04	0 (-0.05, 0.12)	-0.04 (-0.07, 0)
40; 501-550	0.47	0.48	0.41	0.03	0.07	0 (-0.09, 0.33)	-0.07 (-0.11, 0)
40; 551-600	0.50	0.50	0.40	0.04	0.1	0 (-0.14, 0.26)	-0.1 (-0.14, 0)
40; 601-650	0.52	0.52	0.39	0.05	0.13	0 (-0.17, 0.22)	-0.13 (-0.18, 0)
40; 651-700	0.54	0.55	0.39	0.06	0.16	0 (-0.21, 0.23)	-0.16 (-0.21, 0)
40; 701-750	0.57	0.56	0.39	0.07	0.18	-0.01 (-0.23, 0.18)	-0.18 (-0.25, 0)
40; 751-800	0.58	0.58	0.38	0.07	0.21	0 (-0.25, 0.17)	-0.21 (-0.28, 0)
35; 401-450	0.42	0.42	0.41	0.01	0.01	0 (-0.02, 0.04)	-0.01 (-0.03, 0)
35; 451-500	0.45	0.45	0.41	0.02	0.04	0 (-0.06, 0.1)	-0.04 (-0.07, 0)
35; 501-550	0.47	0.47	0.40	0.03	0.07	0 (-0.08, 0.2)	-0.07 (-0.1, 0)
35; 551-600	0.49	0.50	0.39	0.04	0.1	0 (-0.13, 0.24)	-0.1 (-0.14, 0)
35; 601-650	0.52	0.53	0.39	0.05	0.13	0.01 (-0.16, 0.2)	-0.13 (-0.19, 0)
35; 651-700	0.54	0.54	0.38	0.05	0.16	0 (-0.18, 0.2)	-0.15 (-0.21, 0)
35; 701-750	0.56	0.56	0.38	0.06	0.19	0 (-0.22, 0.17)	-0.18 (-0.24, 0)
35; 751-800	0.58	0.58	0.38	0.07	0.21	0 (-0.24, 0.23)	-0.21 (-0.28, 0)
30; 401-450	0.41	0.41	0.40	0.01	0.01	0 (-0.02, 0.05)	-0.01 (-0.03, 0)
30; 451-500	0.45	0.45	0.41	0.02	0.04	0 (-0.05, 0.11)	-0.04 (-0.07, 0)
30; 501-550	0.47	0.48	0.41	0.03	0.07	0 (-0.08, 0.28)	-0.07 (-0.11, 0)
30; 551-600	0.50	0.50	0.40	0.04	0.1	0 (-0.12, 0.23)	-0.1 (-0.15, 0)
30; 601-650	0.52	0.53	0.39	0.05	0.13	0.01 (-0.16, 0.21)	-0.13 (-0.18, 0)
30; 651-700	0.55	0.54	0.39	0.06	0.16	0 (-0.19, 0.22)	-0.16 (-0.22, 0)
30; 701-750	0.57	0.56	0.39	0.07	0.19	-0.01 (-0.26, 0.17)	-0.18 (-0.26, 0)
30; 751-800	0.59	0.58	0.38	0.07	0.21	-0.01 (-0.26, 0.17)	-0.2 (-0.28, 0)
25; 401-450	0.43	0.43	0.42	0.01	0.01	0 (-0.02, 0.05)	-0.01 (-0.03, 0)
25; 451-500	0.45	0.45	0.41	0.02	0.04	0 (-0.06, 0.23)	-0.04 (-0.07, 0)
25; 501-550	0.47	0.47	0.40	0.03	0.07	0 (-0.11, 0.22)	-0.07 (-0.11, 0)
25; 551-600	0.49	0.50	0.40	0.04	0.1	0 (-0.11, 0.2)	-0.1 (-0.15, 0)
25; 601-650	0.53	0.53	0.40	0.05	0.13	0 (-0.18, 0.21)	-0.13 (-0.18, 0)
25; 651-700	0.55	0.55	0.39	0.06	0.16	0 (-0.18, 0.2)	-0.16 (-0.22, 0)
25; 701-750	0.57	0.57	0.39	0.07	0.18	-0.01 (-0.22, 0.21)	-0.18 (-0.25, 0)
25; 751-800	0.60	0.59	0.39	0.07	0.21	-0.01 (-0.25, 0.18)	-0.21 (-0.28, 0)

25; 401-800	0.50	0.50	0.40	0.05	0.12	0 (-0.2, 0.19)	-0.1 (-0.28, 0)

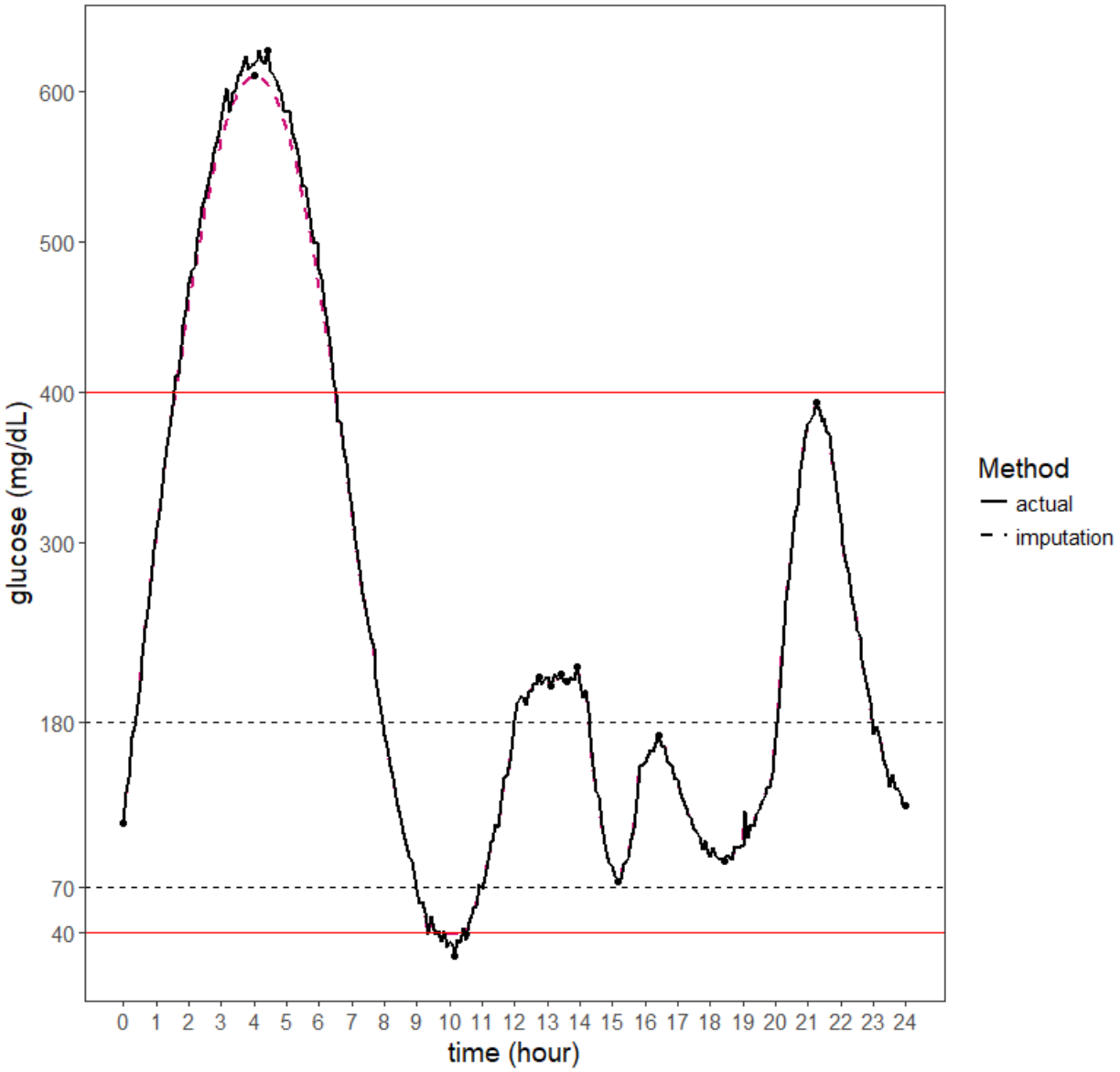
Table 5: Simulation results for MAGE

Lower bound; upper bound (mg/dL)	Actual mean MAGE	Impute mean MAGE	Replacement mean MAGE	RMSE Impute	RMSE Replacement	Impute bias (min, max)	Replacement bias (min, max)
40; 401-450	243.03	241.19	229.35	11.74	20.94	-1.84 (-123.44, 89.38)	-13.68 (-152.16, 56.53)
40; 451-500	273.78	272.17	230.96	23.35	51.83	-1.61 (-151.65, 156.06)	-42.82 (-201.22, 43.83)
40; 501-550	307.12	309.57	230.17	57.95	92.48	2.45 (-218.95, 722.46)	-76.95 (-304.36, 25.5)
40; 551-600	351.08	358.78	231.59	83.89	138.22	7.7 (-272.9, 719.84)	-119.49 (-355.93, 16.88)
40; 601-650	392.24	405.08	234.37	101.55	177.76	12.84 (-315.63, 1011.31)	-157.88 (-417.54, -0.11)
40; 651-700	442.28	457.90	233.59	129.41	230.45	15.63 (-360.26, 901.67)	-208.68 (-466.08, -0.18)
40; 701-750	481.08	482.40	237.35	144.39	267.56	1.32 (-452.63, 773.52)	-243.72 (-503.06, -0.28)
40; 751-800	553.73	562.03	235.47	156.8	343.2	8.3 (-513.9, 688.01)	-318.25 (-530.25, -0.17)
35; 401-450	242.29	240.59	228.78	8.83	19.24	-1.7 (-49.39, 68.72)	-13.51 (-146.09, 44.2)
35; 451-500	276.18	278.16	231.92	24.23	54.27	1.98 (-195.26, 165.67)	-44.26 (-223.93, 39.64)
35; 501-550	311.64	310.80	234.50	46.43	91.56	-0.84 (-234.18, 529.98)	-77.14 (-276.89, 1.52)
35; 551-600	349.86	360.90	231.98	89.98	136.65	11.04 (-326.58, 709.72)	-117.88 (-326.58, 0)
35; 601-650	396.15	411.04	233.98	113.98	184	14.89 (-366.4, 714.91)	-162.16 (-384.84, 11.05)
35; 651-700	443.66	458.23	234.95	118.96	230.29	14.57 (-380.49, 637.59)	-208.71 (-441.54, 0)
35; 701-750	495.65	500.32	234.66	144.59	286.2	4.67 (-476.77, 710.36)	-261 (-484.33, -0.15)
35; 751-800	548.06	557.61	236.32	163.11	338.67	9.56 (-494.71, 889.68)	-311.74 (-545.88, -0.03)
30; 401-450	243.28	241.31	228.79	9.49	19.77	-1.97 (-50.34, 58.67)	-14.49 (-99.96, 46.99)
30; 451-500	272.49	274.84	229.87	27.97	51.93	2.36 (-219.05, 238.49)	-42.61 (-219.05, 44.2)
30; 501-550	311.85	312.71	233.59	60.37	92.91	0.86 (-279.97, 732.51)	-78.26 (-279.97, 13.64)
30; 551-600	351.94	361.29	232.59	85.62	137.12	9.35 (-305.17, 752.24)	-119.35 (-348.55, 36.36)
30; 601-650	389.79	409.86	232.97	115.13	176.08	20.07 (-370.47, 973.18)	-156.82 (-399.9, -0.02)
30; 651-700	441.31	451.33	233.36	115.8	230.1	10.02 (-418.32, 712.07)	-207.95 (-426.12, 0)
30; 701-750	494.58	496.98	235.25	145.9	284.57	2.4 (-497.72, 552.79)	-259.33 (-497.72, -0.36)
30; 751-800	549.56	539.00	237.90	153.16	337.35	-10.55 (-485.84, 771.39)	-311.66 (-540.66, -0.17)
25; 401-450	244.24	243.24	230.10	9.32	19.16	-1 (-26.52, 74.2)	-14.14 (-113.07, 14.22)
25; 451-500	270.74	273.97	229.20	34.35	50.21	3.23 (-133.09, 521.79)	-41.54 (-201.82, 30.5)
25; 501-550	308.56	312.65	228.92	52.94	92.43	4.09 (-199.69, 591.12)	-79.63 (-263.01, 30.17)
25; 551-600	349.92	355.20	236.00	69.84	130.89	5.28 (-247.23, 723.61)	-113.92 (-336.37, -0.1)
25; 601-650	389.11	399.29	237.40	103.41	171.27	10.19 (-309.13, 742.04)	-151.7 (-392.54, -0.26)
25; 651-700	446.38	460.20	235.14	125.59	233.58	13.83 (-370.02, 723.13)	-211.23 (-450.11, -0.19)
25; 701-750	497.28	501.70	240.21	153.4	280.92	4.42 (-490.07, 758.69)	-257.07 (-490.07, 0)
25; 751-800	560.87	556.25	240.35	165.54	343.4	-4.62 (-513.65, 679.79)	-320.52 (-539.42, -1.12)

25; 401-800	358.86	360.90	233.17	90.31	165.79	2.04 (-399.14, 606.4)	-125.7 (-507.25, 93.88)

Figure 1: Sample simulated data

A. Simulation of "smooth" hill



B. Simulation of plateau

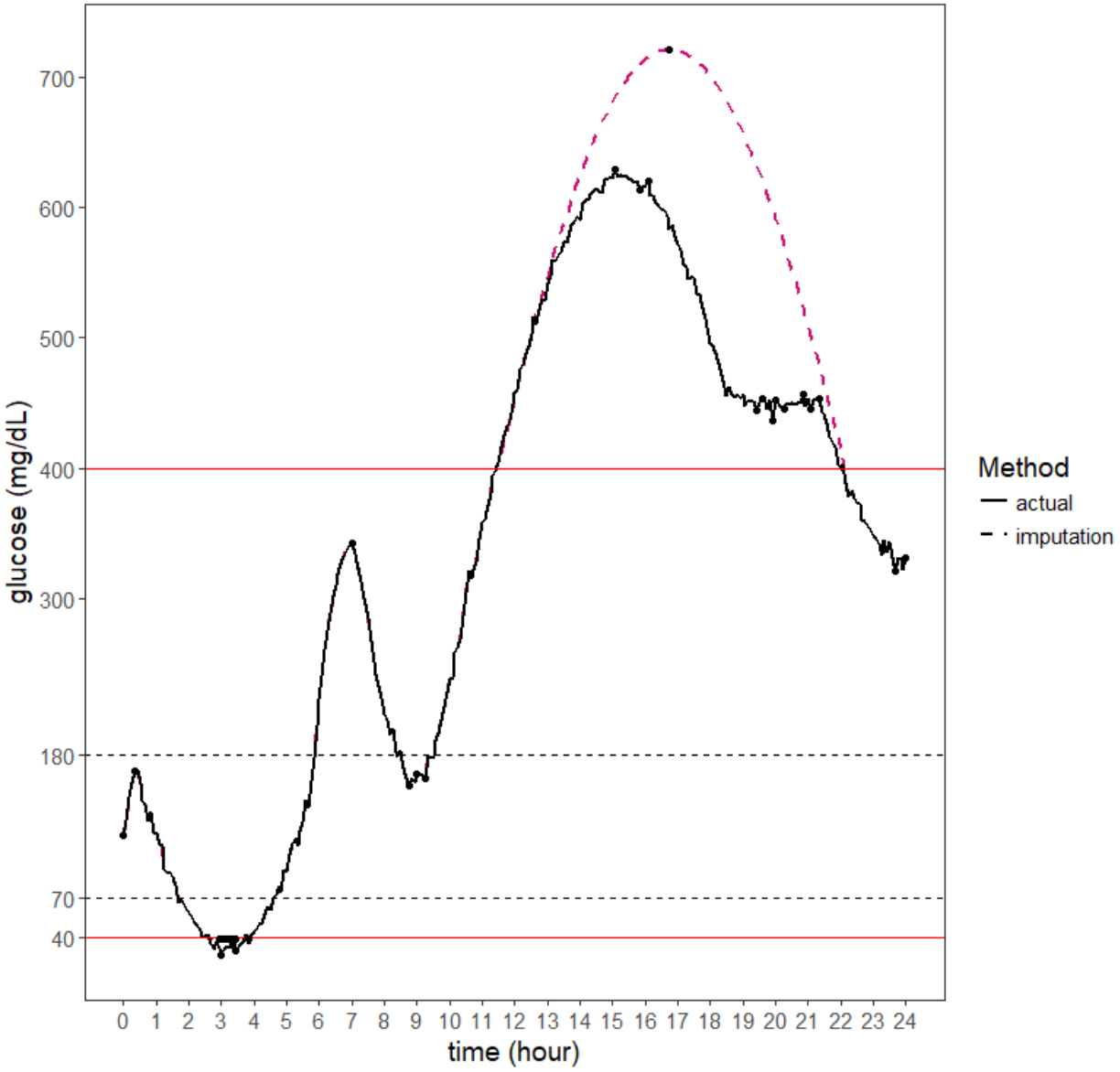


Figure 2: Type I error rate, RMSE, and mean bias of Imputation and Replacement method.