Received: 24 March 2021 | Revised: 11 June 2021 | Accepted: 27 June 2021

DOI: 10.1111/dom.14483

ORIGINAL ARTICLE

WILEY

Design of clinical trials to assess diabetes treatment: Minimum duration of continuous glucose monitoring data to estimate time-in-ranges with the desired precision

Nunzio Camerlingo MSc[1] | Martina Vettoretti PhD[1] | Giovanni Sparacino PhD[1] | Andrea Facchinetti PhD[1] | Julia K. Mader MD[2] | Pratik Choudhary MD[3,4] | Simone Del Favero PhD[1] | on behalf of the Hypo-RESOLVE Consortium

[1]Department of Information Engineering, University of Padova, Padova, Italy

[2]Division of Endocrinology and Diabetology, Medical University of Graz, Graz, Austria

[3]Department of Diabetes, School of Life Course Sciences, King's College London, London, UK

[4]Department of Diabetes, Leicester Diabetes Centre, University of Leicester, Leicester, UK

Correspondence
Simone Del Favero, PhD, Department of Information Engineering, University of Padova, Via G. Gradenigo, 6B, Padova, 35131, Italy.
Email: simone.delfavero@dei.unipd.it

ABSTRACT

Aim: To compute the uncertainty of time-in-ranges, such as time in range (TIR), time in tight range (TITR), time below range (TBR) and time above range (TAR), to evaluate glucose control and to determine the minimum duration of a trial to achieve the desired precision.

Materials and Methods: Four formulas for the aforementioned time-in-ranges were obtained by estimating the equation's parameters on a training set extracted from study A (226 subjects, ~180 days, 5-minute Dexcom G4 Platinum sensor). The formulas were then validated on the remaining data. We also illustrate how to adjust the parameters for sensors with different sampling rates. Finally, we used study B (45 subjects, ~365 days, 15-minute Abbott Freestyle Libre sensor) to further validate our results.

Results: Our approach was effective in predicting the uncertainty when time-in-ranges are estimated using *n* days of continuous glucose monitoring (CGM), matching the variability observed in the data. As an example, monitoring a population with TIR = 70%, TITR = 50%, TBR = 5% and TAR = 25% for 30 days warrants a precision of ±3.50%, ±3.68%, ±1.33% and ±3.66%, respectively.

Conclusions: The presented approach can be used to both compute the uncertainty of time-in-ranges and determine the minimum duration of a trial to achieve the desired precision. An online tool to facilitate its implementation is made freely available to the clinical investigator.

KEYWORDS
clinical trial, continuous glucose monitoring, cost-effectiveness, type 1 diabetes

1 | INTRODUCTION

Continuous glucose monitoring (CGM) sensors are increasingly used in research and clinical practice.[1] A recent consensus panel[2] identified 'time-in-ranges' as key outcome metrics to assess glycaemic control based on CGM data. The identified time-in-ranges include: time in range (TIR), that is, the percentage of time spent within 70-180 mg/dL (3.9-10 mmol/L); time below range (TBR), that is, the percentage of time

spent with CGM less than 70 mg/dL (<3.9 mmol/L); and time above range (TAR), that is, the percentage of time spent with CGM more than 180 mg/dL (>10 mmol/L).[3-5] Furthermore, the percentage of readings within 70-140 mg/dL (3.9-7.8 mmol/L) is referred to as time in tight range (TITR).[1]

These time-in-ranges were used as final endpoints in several studies evaluating the effects of new treatments and/or drugs on glucose control.[6-9] The duration of these trials, which varied from a single day to several months, strongly impacted the precision of the estimation of time-in-ranges in these studies: the longer the trial duration then the less uncertain/more precise the estimated time-in-ranges,[2,10] and, in turn, the more reliable the clinical conclusions. For example, over a longer monitoring period, confounding factors such as meal times, meal composition and exercise/sports, which cause intra-day variability, but also other factors such as menstrual cycle, shiftwork, vacations, intercurrent illness and weekend lifestyle, which affect inter-day variability, average out. On the other hand, a long trial duration is associated with higher costs,[11,12] increased recruitment difficulties, larger likelihood of withdrawal[13,14] and a greater risk of protocol deviations.[15] Therefore, in the design of a clinical trial, a careful balance between these two opposing needs must be found.[16,17] In particular, for trials with time-in-ranges as final endpoints, understanding the impact of trial duration on the precision of CGM-derived metrics would be particularly useful in power calculations. Similarly, in clinical practice, time-in-ranges are increasingly used to assess current glycaemic status and can influence decisions to change therapy or start new therapies. A 5% change in TIR is believed to be clinically significant,[18] but understanding the effect of assessing TIR over 14 days, 1 month or 3 months can affect clinical decision-making.

The consensus panel of Battelino et al.[2] recommends assessing 14 days of CGM, as the most recent 14 days of CGM data provide a good approximation of time-in-ranges collected over a 3-month period.[10,19] However, this indication is empirical,[20,21] and the literature lacks a description of how the precision in the estimation of time-in-ranges improves as the trial duration increases.

In a recent work,[22] we analytically obtained a mathematical formula that predicts the uncertainty of TBR estimates based on the number of CGM days. Briefly, it can be assumed that an accurate estimate of TBR, $TBR(N)$, can be obtained from a sufficiently long CGM recording of N days (e.g. $N \geq 90$ days). The estimate $TBR(n)$, obtained using only n days of CGM data $(n < N)$, is affected by an estimation error of $e_{TBR}(n) = TBR(n) - TBR(N)$ (see[19,21] for similar definitions). This error is distributed around 0 with a certain standard deviation, $SD[e_{TBR}(n)]$, which describes the uncertainty around the estimate of TBR. We derived an explicit equation describing how fast the uncertainty of the TBR estimate, $SD[e_{TBR}(n)]$, decreases as the length of the trial n increases:

$$SD[e_{TBR}(n)] = \sqrt{\frac{p_r(1-p_r)}{k\,n}\left(1+\frac{2\alpha}{1-\alpha}+\frac{2\alpha}{k\,n}\frac{(\alpha^{k\,n}-1)}{(1-\alpha)^2}\right)}. \quad (1)$$

Equation (1) involves three parameters, k, p_r and α.

1. The first parameter, k, is the number of CGM samples produced in 1 day when no measurement is missed (e.., $k = 288$ for a CGM sensor providing measurements every 5 minutes, $k = 96$ for CGM measurements collected every 15 minutes).
2. The second parameter, p_r, represents the expected average TBR in the population.
3. The third parameter, α, depends on the CGM sensor sampling period (but also on the glycaemic range considered, as will be illustrated in the following).

Moreover, in[22] we provided a single set of parameters able to predict the uncertainty of TBR estimates of a whole population of heterogeneous subjects.

Notably, the mathematical machinery used to derive Equation (1) also holds for any other time-in-range, provided that the parameters (in particular, the parameter α) are changed accordingly.

In the current paper, we first provide suitable values of the parameters k, p_r and α for TIR, TITR, TBR and TAR, thus providing four different formulas. Each of these formulas is then validated by comparing the predicted uncertainty of time-in-ranges with the variability observed in the data. The generalizability of the proposed formulas is also tested for different sensors and different populations. To do that, we illustrate and validate how to adjust the parameters to cope with sensors with different sampling rates.

Finally, we discuss two possible clinical cases: (a) 14-day monitoring of a person with diabetes provides the following results: 5% TBR, 70% TIR, 50% TITR and 25% TAR. How precise are these estimates, that is, what are the confidence intervals around them? (b) An investigator is designing a new clinical trial, requiring a maximum confidence interval of ±1% around the resulting TBR. Which is the minimum monitoring duration that warrants such precision?

2 | METHODS

2.1 | Data: study A and study B

In this work we considered two different studies, labelled A and B.

Study A[23] involves 226 subjects (112 women) with type 1 diabetes followed for 177 ± 20 days (mean ± SD) using an unblinded CGM sensor with or without confirmatory fingerpick (adjunctive and non-adjunctive CGM therapy, respectively). The sensor used was a Dexcom G4 Platinum, with enhanced accuracy through Software 505,[24] providing one sample every 5 minutes. The overall population had an HbA1c of 7.0% ± 0.7% (53.0 ± 7.7 mmol/mol), mean glucose of 160.5 ± 22.1 mg/dL, TIR of 63.3% ± 12.5%, TITR of 39.8% ± 12.0%, TBR of 3.7% ± 2.6%, TAR of 33.0% ± 13.4% and coefficient of variation (CV) of 37.1% ± 4.8%.

Study B was an observational study (not published yet) performed at the Medical University of Graz (IRB approval number: 29/522 ex 16/17). It involved 45 subjects (16 women) with type 1 diabetes monitored for 357 ± 14 days (mean ± SD), using a blinded Abbott Freestyle Libre sensor (Abbott Laboratories), providing one sample every

15 minutes. The overall population had an HbA1c of 7.7% ± 1.2% (60.3 ± 13.2 mmol/mol), mean glucose of 172.8 ± 34.1 mg/dL, TIR of 54.0% ± 17.7%, TITR of 34.5% ± 16.9%, TBR of 5.2% ± 4.3%, TAR of 40.8% ± 19.3% and CV of 48.6% ± 6.7%.

To ensure that an adequate amount of CGM data were available for each subject, only participants with at least 173 days of monitoring were retained for study A and only participants with at least 330 days were retained for study B, resulting in the exclusion of approximately 10% of subjects (22 from study A and four from study B).

2.2 | Proposed formulas for uncertainty in time-in-ranges

As anticipated, the mathematical framework used in Camerlingo et al.[22] also holds for any other time-in-range. The parameters of Equation (1) should be altered to cope with different glycaemic ranges. In particular:

1. The parameter p_r should be set to the average percentage of time (known or expected) spent by the population in the range under analysis: average TIR when TIR is considered, average TAR when TAR is considered, etc.

2. The parameter α, as well as depending on the CGM sampling rate, also depends on the glycaemic range under analysis, that is, a different α should be used for different time-in-ranges. Notably, the parameter α is not very sensitive to the population under analysis (unlike p_r). Therefore, in this work we will provide suitable values for parameters α for TIR, TITR, TBR and TAR (Table 1) that can be used for a generic population. These values were estimated using the procedure proposed in[22] on the population of study A. As such, they should be used with a 5-minute CGM sensor. If a sensor with a different sampling rate is used then these parameters can be easily adjusted. Specifically, the α_T value to be used for a sensor providing one measurement every T minutes can be linked to α_5, the parameter α used for a 5-minute CGM sensor, by

$$\alpha_T = \alpha_5^{\frac{T}{5}} \tag{2}$$

As an example, for 10- and 15-minute CGM sensors, α_5 can be adjusted to $\alpha_{10} = \alpha_5^2$, and $\alpha_{15} = \alpha_5^3$, respectively.

3. Lastly, the parameter k in Equation (1), representing the number of CGM samples provided in one day, remains independent of the glycaemic range.

Suitable values of the parameters p_r and α for the four considered glycaemic ranges are obtained using study A. Specifically, patients in study A are randomly split into training (70%) and test (30%) sets. The training set is used to estimate the parameters and the test set is then used to validate the resulting formulas.

The estimated values of p_r are: 64.5% for TIR, 40.0% for TITR, 3.10% for TBR and 33.0% for TAR. The values of α are: 0.961 for TIR, 0.958 for TBR, 0.940 for TBR and 0.968 for TAR (also reported in the third column of Table 1).

In Table 1 we summarize the formulas with p_r as the only parameter (α is fixed to the values above) for the four time-in-ranges under analysis and a sensor with a 5-minute sampling rate. In this table, to improve readability, the term α^{kn} of Equation (1) is neglected because it is close to zero.

2.3 | Validation of the formulas

To validate the equations of Table 1, we evaluated their ability to predict the decrease in the standard deviation of the estimation error $e(n)$, computed by CGM data, as the trial duration n increases. This was performed for the four time-in-ranges under study (i.e. TIR, TITR, TBR and TAR), using the test set extracted from study A.

To explain our validation methodology, let us focus on the TBR. For each participant, the most accurate estimate of TBR, that is, $TBR(N)$, was evaluated over the whole trial duration N. Then, to simulate a short-term trial, we extracted several shorter windows of fixed duration, n, ranging from 1 to $n_{max} = 30$ days. For each window, we computed $TBR(n)$ and compared it with $TBR(N)$ by computing the error

$$e_{TBR}(n,j) = TBR(n,j) - TBR(N),\ j \in \{1,2,\ldots,M_p\},$$

where M_p is the number of different windows available for each patient p, obtained by considering different starting points (with a window shift of 1 day), as proposed in.[21] Repeating this procedure for all the patients, we obtained a total of M values of the estimation error committed using windows of duration n in the whole population.

Finally, for each window duration n, the standard deviation of the estimation error was computed, that is:

$$SD_S[e_{TBR}(n)] = \sqrt{\left(\frac{1}{M-1}\sum_{j=1}^{M} e(n,j)^2\right)}. \tag{3}$$

Notation: to distinguish the standard deviation provided by the proposed formula from the standard deviation computed by data, from now on we will denote the first as theoretical standard deviation, $SD_T[e(n)]$, and the second as sample standard deviation, $SD_S[e(n)]$.

Remark: when dealing with extremely long datasets, the ground-truth TIR can vary during the trial. In this case, $TIR(N,p)$ should be computed over a long-term window of an arbitrary duration, $N_l < N$ (e.g. 90 days), extracted around the centre of the short-term window of duration n.

2.4 | Generalizability of the formula: Validation with other sampling rates and populations

To show the effectiveness of the proposed formulas for sensors with different sampling rates, we modified the test set extracted from

TABLE 1 Proposed formulas for the uncertainty around time-in-ranges estimates, for four different time-in-ranges: time in range (TIR), time in tight range (TITR), time below range (TBR) and time above range (TAR), for sensors with 5-min sampling rate

Glycaemic metric	Glycaemic range	Estimated α_5	Uncertainty around the estimate
TIR	70-180 mg/dL (3.9-10 mmol/L)	0.961	$\sqrt{\frac{p_r(1-p_r)}{288\,n}\left(50.28-\frac{4.388}{n}\right)}$
TITR	70-140 mg/dL (3.9-7.8 mmol/L)	0.958	$\sqrt{\frac{p_r(1-p_r)}{288\,n}\left(47.78-\frac{3.962}{n}\right)}$
TBR	<70 mg/dL (<3.9 mmol/L)	0.940	$\sqrt{\frac{p_r(1-p_r)}{288\,n}\left(24.00-\frac{0.998}{n}\right)}$
TAR	>180 mg/dL (>10 mmol/L)	0.968	$\sqrt{\frac{p_r(1-p_r)}{288\,n}\left(61.50-\frac{6.565}{n}\right)}$

study A, to simulate 10- and 15-minute CGM sensors. This operation was performed by averaging the 5-minute CGM values over the previous 10 and 15 minutes, respectively. Then, for each considered glycaemic range, we compared the sample standard deviation, $SD_S[e(n)]$, computed using this modified dataset against the theoretical standard deviation, $SD_T[e(n)]$, predicted by the proposed formulas, adjusting the values of α by means of Equation (2).

In addition, to show that the parameters reported in Table 1 can be used, with limited approximation, for different populations, we also validated the formulas on the whole of study B. Because the CGM used in study B provides one sample every 15 minutes, the values of α_5 in Table 1 were adjusted according to Equation (2).

These analyses investigate the generalizability of the parameters proposed.

All the analyses were performed in Matlab 2020b (MathWorks, Natick, MA, USA). All the scripts for implementing the methodology in Matlab are publicly available at https://github.com/NunzioCamer/AnalyticalTBRestimation.

3 | RESULTS

3.1 | Validation of the formula (using study A data)

In Figure 1A-D, the sample and theoretical standard deviations of the estimation error (SD_S and SD_T, respectively) are compared considering different window durations n, from 1 to $n_{max} = 30$ days, and for the time-in-ranges under study: TIR (A), TITR (B), TBR (C) and TAR (D). In particular, SD_S, reported as a solid red curve, was computed on a test set extracted from study A, as in Equation (3), while SD_T, reported as a dashed blue curve, was obtained by substituting the values of the parameters p_r and α estimated by the training set data of study A into Equation (1).

The curves describing SD_S and SD_T overlap well (the relative discrepancy between the two curves is smaller than 10% for all the glycaemic ranges for most of the durations considered). Therefore, we conclude that the proposed formulas are able to effectively describe the uncertainty of all time-in-ranges estimates for the overall population of study A.

In addition, in each part of Figure 1, the boxplot of the estimation error $e(n,j)$ is also reported.

3.2 | Generalization of the results for different sampling rates (using study A data)

In Figure 1E-H, we investigate the adjustment of α proposed in Equation (2). Specifically, we report SD_S computed on the test set extracted from study A, modified to emulate a 15-minute sensor. This SD_S is compared with SD_T, which is obtained by Equation (1) using the values of α corrected as in Equation (2). Also in this case, SD_S and SD_T overlap well (the relative discrepancy between the two curves is smaller than 10% for most of the durations considered), thus proving the efficacy of Equation (1) combined with Equation (2).

Similar results are also obtained when emulating a 10-minute sensor (see Section **S1**).

A comparison between 5-, 10- and 15-minute CGM sensors in terms of the predicted decrease in the uncertainty of time-in-ranges is reported in Figure S2.

3.3 | Generalization of the results for different populations (using study B data)

In Figure 1I-L, we stress the generalizability of the formulas on a different dataset. In this case, for each time-in-range, SD_S was computed using the whole study B, while SD_T was obtained by substituting the values of the parameters p_r and α estimated by the training set data of study A into Equation (1), and adjusting α as in Equation (2) (i.e., $\alpha_{15} = \alpha_5^3$).

Despite the approximations introduced, the SD_S curve overlaps well with the curve of SD_T, for TIR, TITR and TAR. For these ranges, the relative discrepancy between the curves is below 13% for most of the durations considered. As expected, the agreement between the two curves achieved in this case is smaller compared with that observed in Figure 1A-H. The impact of the introduced approximation is larger for TBR, mainly as a result of a large difference in the incidence of hypoglycaemia in the two studies: the average TBR in the training set extracted by study A is $TBR_A = 3.10\%$, while in study B it reaches $TBR_B = 5.15\%$ (~ 40% larger).

In Section S2, we show that a better estimate of the incidence of hypoglycaemia in the population of study B can significantly improve the agreement between the curves for TBR too.

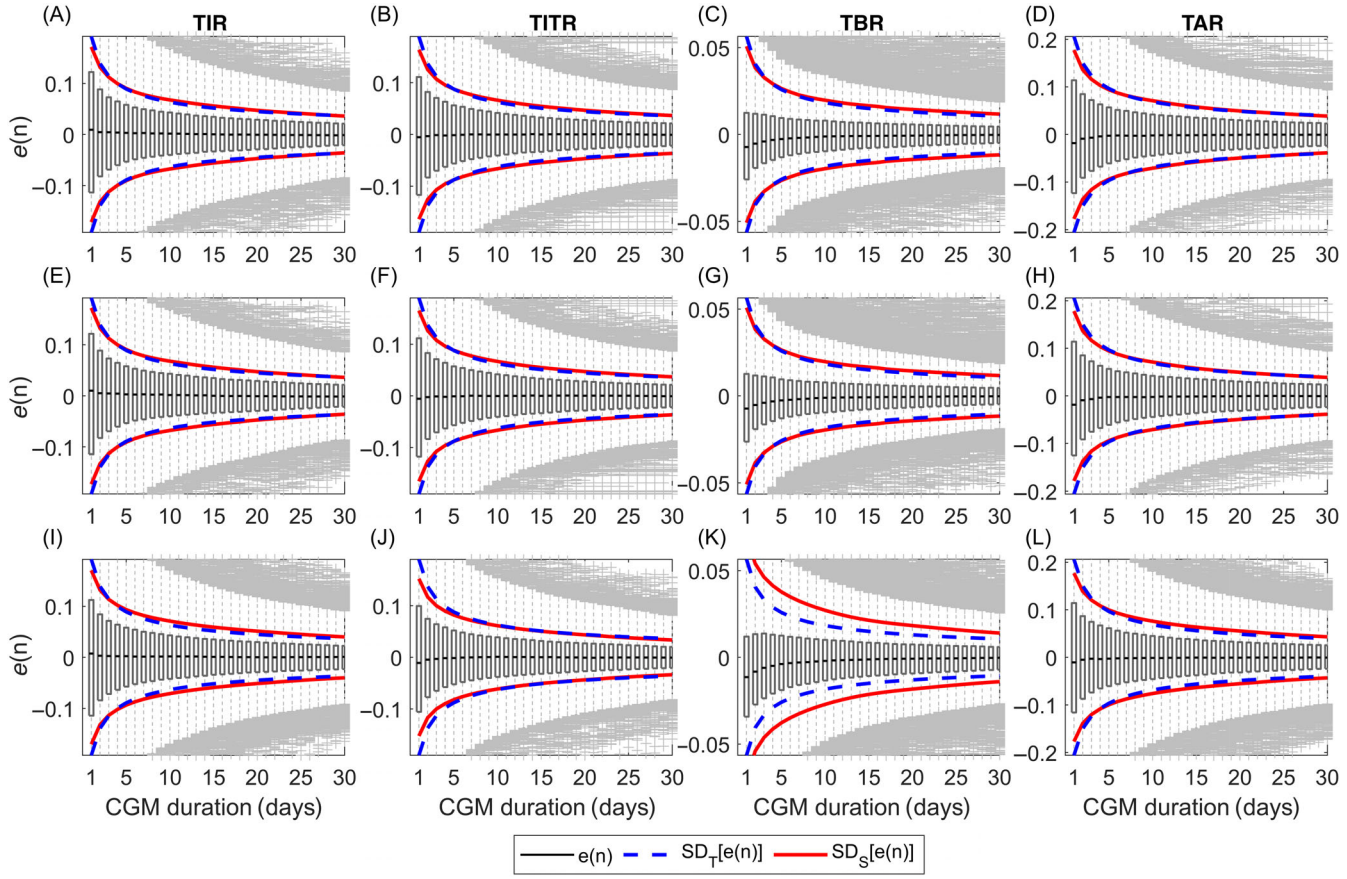

FIGURE 1 Estimation error $e(n)$ for four time-in-ranges: time in range (TIR; A, E and I), time in tight range (TITR; B, F and J), time below range (TBR; C, G and K) and time above range (TAR; D, H and L). The dashed blue curve is the theoretical standard deviation, $SD_T(n)$, returned by the proposed formulas with the parameters estimated using the training set of study A, while the solid red curve is the sample standard deviation, $SD_S(n)$, computed retrospectively by the test set of study A (A-D), the modified test set of study A, simulating a 15-min continuous glucose monitoring (CGM) sensor (E-H) and the whole of study B (I-L). In the proposed boxplot representation, the median, 25th percentile and 75th percentile of the estimation error are shown in black, while the grey stars indicate the outlier values

4 | DISCUSSION AND CLINICAL CASES

We introduced four formulas linking the precision/uncertainty of four time-in-ranges (TIR, TITR, TBR and TAR) to the number of CGM days used to estimate them. We showed that by setting the formulas' parameters to the values suggested in Table 1, one can calculate the uncertainty of time-in-ranges estimates, matching the uncertainty observed on the CGM data collected in two different populations of subjects with heterogeneous characteristics (e.g. age, body mass index and diabetes duration) and wearing different sensors. Moreover, in Section S5, we show that no significant differences in the reliability of the formulas were observed for patients of different ages, body mass index or gender.

Nonetheless, in the case of clinical studies focused on a specific subpopulation (e.g. children or the elderly), one may also consider optimizing the formula's parameters for the specific subpopulation being studied. This can be carried out by re-running the parameters' estimation procedure described in[22] on data collected in populations with similar characteristics.

To facilitate the use of these formulas, we implemented an online, freely accessible calculator, available at http://computecgmduration.dei.unipd.it. The tool requires the user to insert some information about the ongoing experiment then computes the parameters of the formulas. Next, the user can decide either to compute the uncertainty around time-in-ranges estimated in previous clinical trials, or to compute the optimal number of days necessary to achieve the desired uncertainty around the selected time-in-range.

The resulting number of days refers to days with no missing CGM data. To deal with the common issue of gaps in CGM data, a practical approach is to consider the total amount of available data (e.g. a 30-day study with four gaps of 6 hours could be considered as a 29-day study). Therefore, in the design of a clinical trial, if we expect 80% sensor usage (i.e. data loss of 20%), we suggest setting the study duration to 20% longer than the one provided by the proposed formula. In Section S5, we show that this is a conservative approach, as sporadic missing samples have no practical impact on the precision of the estimated time-in-ranges.

The consensus panel of Battelino et al.[2] recommended a minimum CGM duration of 14 days to accurately estimate time-in-ranges.[10,19] However, it was recently shown in[20,21] that this approach[10,19] may produce inconsistent results in different datasets. The approach proposed in this work offers an alternative way to overcome this limitation.

The proposed formulas can be used for multiple applications. In the following, we will focus on two of them and illustrate four clinical cases. Furthermore, we will discuss the analogy of the proposed formula with tools for power calculation.

4.1 | Application 1: Precision of time-in-ranges estimated from previous studies

4.1.1 | Clinical case 1.1

Suppose that an individual was diagnosed with type 1 diabetes and the investigator prescribed 14-day CGM monitoring to evaluate the overall glycaemic control. At the end of the 2-week period, the time-in-ranges observed were 5% TBR, 70% TIR, 50% TITR and 25% TAR. The investigator also needs to know how precise the estimated values are, focusing in particular on TBR. Using the formula for TBR (third row of Table 1) and substituting the values $n = 14$ and $p_r = 0.05$, $SD[e_{TBR}(n)] = 1.95\%$ (the computation can be performed using the online calculator, available at http://computecgmduration.dei.unipd.it). This means that the confidence interval around the estimated TBR is $5\% \pm 1.95\%$. Repeating the computation for the other time-in-ranges, the investigator attains analogous confidence intervals, as summarized in the upper panel of Table 2, second column. The same table also reports the uncertainty obtained with different CGM monitoring durations (i.e. 7, 14, 30, 60 and 90 days). As expected, the time-in-ranges estimated by studies using short monitoring periods (e.g. 7 days) will have wide confidence intervals, while longer studies (e.g. 90 days) provide better precision.

4.1.2 | Clinical case 1.2

The I HART CGM trial[25] compares the effectiveness of real-time CGM (RT-CGM) versus flash CGM on hypoglycaemia for adults with type 1 diabetes. After 8 weeks ($n = 56$ days), the RT-CGM group exhibits an average TBR of 6.20%. Our formula suggests a confidence interval of $\pm 1.08\%$ for all individuals. After 16 weeks of RT-CGM ($n = 112$ days), the average TBR is equal to 5.40%, and the formula indicates a much narrower confidence interval of $\pm 0.72\%$ for all individuals.

4.2 | Application 2: Minimum trial duration providing the desired precision for estimates of time-in-ranges

4.2.1 | Clinical case 2.1

Suppose that an investigator is designing a clinical trial involving patients with type 1 diabetes to test a new human insulin analogue. The primary outcome is the TBR. Based on a previous pilot study, the investigator expects an average TBR of 4.00% in the population. She/he is interested in setting a suitable trial duration and desires a maximum confidence interval around the final TBR of 1.00%. Using the proposed formula, the investigator can compute the minimum monitoring duration, which warrants the desired uncertainty around the resulting TBR. In particular, by inverting the formula for TBR (third row of Table 1) and setting $p_r = 0.04$ and $SD[e_{TBR}(n)] = 0.01$, the investigator gets that the trial must last at least $n = 44$ days (the computation can be performed using the online calculator available at http://computecgmduration.dei.unipd.it).

4.2.2 | Clinical case 2.2

In the same scenario as the previous clinical case, suppose that the investigator selects the TAR (or similarly, TIR or TITR) as the primary outcome and wants to compute the minimum monitoring duration to provide the desired uncertainty around the final estimate. Based on previous pilot studies, the investigator expects an average TAR of 25%. The investigator prefers to consider the relative uncertainty and deems acceptable a relative uncertainty of 15% around the final TAR. Setting $p_r = 0.25$ and $SD[e_{TAR}(n)] = 0.15 \times p_r = 0.0375$ in the proposed formula, it transpires that the minimum monitoring duration is $n = 29$ days.

In Table 2 we summarize the number of days needed to achieve the desired precision in the estimation for the four main time-in-ranges. Precision is expressed both in absolute terms (2.0%, 1.5%, 1.0% or 0.5%) and in relative terms (20%, 15%, 10% or 5%). The first option is preferable for TBR, while the latter option is preferable for the other ranges. Uncommon options are shaded, as they result in extremely demanding precisions and thus in very long trials. The p_r values used for each time-in-range were extracted from the consensus of Battelino et al.[2] and are also reported in the table.

Table 2 suggests that reaching a tight confidence interval for TBR requires more monitoring days than for the other time-in-ranges. To further assess this consideration, in Section S3 we compared the curves of absolute uncertainty and relative uncertainty for the four time-in-ranges under study.

4.3 | Analogy between power calculation tools and the proposed formula

To better illustrate the main message of the current paper, it is useful to discuss the analogy between our formula and the power calculation tools commonly used by clinical practitioners to answer the question: How many participants should be recruited? These tools are based on a mathematical formula that returns the minimum number of subjects to be monitored, providing the desired study power and thus avoiding a type II error (i.e. the chance of declaring the findings non-significant, while instead the treatment has an effect).[26] The formula requires input of two parameters specific to the population to be monitored: the expected difference in the outcome metric between treatment and control, and the standard deviation of the outcome metric. The values of these parameters are often

TABLE 2 Possible applications of the proposed formulas

Application 1					
Time-in-ranges	**Continuous glucose monitoring duration (d)**				
	7	**14**	**30**	**60**	**90**
Time in range ($p_r = 0.70$)	7.22%	5.12%	3.50%	2.48%	2.03%
Time in tight range ($p_r = 0.50$)	7.59%	5.38%	3.68%	2.60%	2.13%
Time below range ($p_r = 0.04$)	2.47%	1.75%	1.20%	0.85%	0.69%
Time above range ($p_r = 0.25$)	7.53%	5.34%	3.66%	2.59%	2.11%

Application 2								
Time-in-ranges	**Absolute uncertainty**				**Relative uncertainty**			
	2.0%	**1.5%**	**1.0%**	**0.5%**	**20%**	**15%**	**10%**	**5%**
Time in range ($p_r = 0.70$)	93	165	370	1479	2	4	8	31
Time in tight range ($p_r = 0.50$)	102	182	408	1631	4	8	17	66
Time above range ($p_r = 0.25$)	101	179	403	1612	17	29	65	258
Time below range ($p_r = 0.04$)	11	20	44	173	68	120	270	1078

Note: Upper panel (application 1): uncertainty (%) around the estimate of different time-in-ranges, for 7, 14, 30 and 90 days of monitoring. Lower panel (application 2): number of days needed to reach a desired precision of different time-in-ranges. Precision is expressed as absolute uncertainty for time below range, with possible values of 20%, 1.5%, 1.0% and 0.5%. Relative uncertainty is reported for time in range, time in tight range and time above range, with possible values of 20%, 15%, 10% and 5%.

TABLE 3 Analogy between power calculation tool and proposed tool

	Power calculation tool	**Proposed tool**
Question	How many participants should be recruited?	For how long should the participants be monitored to have a realistic evaluation of their time-in-ranges?
Desired result	Sufficient study power (to avoid type II errors)	Sufficient precision in the estimated time-in-ranges
Inputs of the formula	Desired power + expected outcome difference in the two arms, standard deviation	Desired precision + p_r, α
How to set the inputs	Pilot study, educated guess, clinical experience of previously reported data	
Mathematical assumptions	Observations are independent from each other; the outcome parameter is normally distributed	The outcome metric is stationary. Mono-expontential time-invariant autocorreation of the dichotomized trace obtained from continuous glucose monitoring.[22]

unknown when designing the study, and are usually based on previously reported or preclinical studies.

Analogous to power calculation tools, the methodology discussed in the current work can be used by clinical practitioners to answer the question: For how long should the participants be monitored to have a realistic evaluation of their time-in-ranges? In fact, the presented formula returns the minimum number of monitoring days that would provide the desired precision of time-in-ranges indexes selected as endpoints of the study. Besides specifying the desired precision, this formula requires as input two parameters specific to the population to be monitored: p_r (the time-in-ranges in the population) and α (related to the sensor sampling period and the glycaemic range under analysis). Suitable values for α suggested in this work can be used for a generic type 1 diabetes population. As for power calculation tools, the value of p_r should be set to the expected time-in-ranges in the population under analysis, based on pilot studies, clinical experience of previously reported data, or an educated guess.

Finally, both approaches are based on simplified mathematical assumptions. For example, in power calculation tools, the study outcome is assumed to be normally distributed, while for the tool proposed in this work, we hypothesize mono-exponential autocorrelation in CGM-based outcome metric samples.[22]

A schematic comparison between power calculation tools and the proposed formula is reported in Table 3.

5 | CONCLUSIONS

With increased use of CGM data and, in particular, time-in-ranges metrics for making therapeutic decisions and assessing differences

between therapies in the clinical and research setting, it is important to understand the reliability of those measurements. While life with diabetes can vary from day to day and month to month, the longer we aggregate data over, the less those data will be subject to the vagaries of individual days.

The formula we have created, which links the number of days of CGM data to the precision/uncertainty around given time-in-ranges values, should help give people a better understanding of the numbers that are so important in current diabetes management.

The formula involves three parameters: k, a constant depending on the sensor sampling rate, α, linked to both the sensor sampling rate and the glycaemic range under analysis, and p_r, related to the characteristics of the population (to be) monitored. Although the values of α can be estimated for different populations, those obtained in this work can be considered general and can be used to apply the proposed formula to other populations, with limited error. For what concerns the values of p_r, we suggest adjusting them to the time-in-ranges expected in the population, based on a pilot study, educated guess, clinical experience or previously reported data (e.g. those considered in this paper, or those identified in the consensus of Battelino et al.[2] as targets for different diabetic populations).

We tested the validity of the formula on patient data and showed its ability to generalize to different CGM sampling rates and populations.

We believe this formula can help the diabetes community in a number of different scenarios. First, when assessing time-in-ranges obtained in published studies, we can determine the precision around the estimated values, based on how long the CGM values were collected over. Some studies using 3 or 7 days of CGM data will have wide confidence intervals, while others using longer durations offer better precision. We believe that this analysis will also be valuable in clinical consultations. Clinicians receive reports on time-in-ranges through software linked to CGM devices, but the proposed formula can help provide an estimate of the confidence intervals (i.e. the precision) of those values.

Second, we believe this formula can be used by academic or industry clinical trial teams, to help them determine a suitable duration of the study. Furthermore, the value of the standard deviation provided by the formula could also be used in power calculations, for example, to determine the number of participants needed in a trial. This will help with designing better studies, providing more accurate results. To support academic colleagues around the world, we have developed an online calculator that we have made freely available and which facilitates the use of the formulas in performing the tasks mentioned above.

In conclusion, together with standard power calculation tools, this formula allows optimization of the cost-benefit ratio of a clinical trial: trials with too many subjects monitored for too long expose the subjects to unnecessary risks. Trials with too few subjects monitored for too short a duration do not permit collection of conclusive scientific evidence and thus waste patients' time and risk exposure. Statistical tools for an effective cost-benefit balance in clinical trials are important for all subjects, but they become essential when recruiting a minority of underprivileged patients.

Future developments include exploring relaxed mathematical assumptions under which the formula is derived, as well as further validation on datasets collected in wider and more heterogeneous populations. We also plan to explore the possibility of optimizing the formula's parameters for specific populations (e.g. type 2 diabetes, pregnant and paediatric) and for different co-variates (e.g. HbA1c, diabetes duration, body weight, CGM sensor model and insulin therapy).

ACKNOWLEDGEMENTS

This study is part of the Hypo-RESOLVE project. The project has received funding from the Innovative Medicines Initiative 2 (https://www.imi.europa.eu/) Joint Undertaking (JU) under grant agreement no. 777460. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and T1D Exchange, JDRF, International Diabetes Federation (IDF), and The Leona M. and Harry B. Helmsley Charitable Trust.

CONFLICT OF INTEREST

NC, MV, AF, GS and SDF declare no conflicts of interest. PC has received personal fees from Medtronic, Abbott, Dexcom, Insulet, Novo Nordisk, Lilly and Sanofi. JKM is a member of the advisory board of Becton-Dickinson, Boehringer Ingelheim, Eli Lilly, Medtronic, Prediktor SA and Sanofi, and has received speaker honoraria from Abbott Diabetes Care, Astra Zeneca, Eli Lilly, Dexcom, Medtronic, Novo Nordisk, Roche Diabetes Care, Sanofi, Servier and Takeda.

AUTHOR CONTRIBUTIONS

NC and SDF contributed to the design of the research and drafted the manuscript. PC, MV, AF, GS and JKM contributed to the interpretation of the results. JKM provided part of the data. The final version was read, reviewed and approved by all authors.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/dom.14483.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Nunzio Camerlingo https://orcid.org/0000-0003-3222-2479
Julia K. Mader https://orcid.org/0000-0001-7854-4233
Simone Del Favero https://orcid.org/0000-0002-8214-2752

REFERENCES

1. Danne T, Nimri R, Battelino T, et al. International consensus on use of continuous glucose monitoring. *Diabetes Care*. 2017;40(12):1631-1640.
2. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care*. 2019;42(8):1593-1603.

3. Beck RW, Bergenstal RM, Riddlesworth TD, et al. Validation of time in range as an outcome measure for diabetes clinical trials. *Diabetes Care*. 2019;42(3):400-405.
4. Agiostratidou G, Anhalt H, Ball D, et al. Standardizing clinically meaningful outcome measures beyond HbA1c for type 1 diabetes: a consensus report of the American Association of Clinical Endocrinologists, the American Association of Diabetes Educators, the American Diabetes Association, the Endocrine Society, JDRF International, The Leona M. and Harry B. Helmsley Charitable Trust, the Pediatric Endocrine Society, and the T1D Exchange. *Diabetes Care*. 2017;40(12):1622-1630.
5. Gabbay MAL, Rodacki M, Calliari LE, et al. Time in range: a new parameter to evaluate blood glucose control in patients with diabetes. *Diabetol Metab Syndr*. 2020;12(1):22.
6. Haak T, Hanaire H, Ajjan R, Hermanns N, Riveline JP, Rayman G. Use of flash glucose-sensing technology for 12 months as a replacement for blood glucose monitoring in insulin-treated type 2 diabetes. *Diabetes Ther*. 2017;8(3):573-586.
7. Tauschmann M, Thabit H, Bally L, et al. Closed-loop insulin delivery in suboptimally controlled type 1 diabetes: a multicentre, 12-week randomised trial. *Lancet*. 2018;392(10155):1321-1329.
8. Bolinder J, Antuna R, Geelhoed-Duijvestijn P, Kröger J, Weitgasser R. Novel glucose-sensing technology and hypoglycaemia in type 1 diabetes: a multicentre, non-masked, randomised controlled trial. *Lancet*. 2016;388(10057):2254-2263.
9. Malecki MT, Cao D, Liu R, et al. Ultra-rapid lispro improves postprandial glucose control and time in range in type 1 diabetes compared to lispro: PRONTO-T1D continuous glucose monitoring substudy. *Diabetes Technol Ther*. 2020;22(10):1-8.
10. Xing D, Kollman C, Beck RW, et al. Optimal sampling intervals to assess long-term glycemic control using continuous glucose monitoring. *Diabetes Technol Ther*. 2011;13(3):351-358.
11. Huang ES, O'Grady M, Basu A, et al. The cost-effectiveness of continuous glucose monitoring in type 1 diabetes. *Diabetes Care*. 2010;33(6):1269-1274.
12. Wan W, Skandari MR, Minc A, et al. Cost-effectiveness of continuous glucose monitoring for adults with type 1 diabetes compared with self-monitoring of blood glucose: the DIAMOND randomized trial. *Diabetes Care*. 2018;41(6):1227-1234.
13. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemp Clin Trials Commun*. 2018;11(8):156-164.
14. Desai M. Recruitment and retention of participants in clinical studies: critical issues and challenges. *Perspect Clin Res*. 2020;11(2):51-53.
15. Getz KA, Zuckerman R, Cropp AB, Hindle AL, Krauss R, Kaitin KI. Measuring the incidence, causes, and repercussions of protocol amendments. *Drug Inf J*. 2011;45(3):265-275.
16. Gilron I. Methodological issues associated with clinical trials in neuropathic pain. *Expert Rev Clin Pharmacol*. 2016;9(11):1399-1402.
17. Kilburn LS, Banerji J, Bliss JM. The challenges of long-term follow-up data collection in non-commercial, academically-led breast cancer clinical trials: the UK perspective. *Trials*. 2014;15(1):1-10.
18. Wilmot EG, Lumb A, Hammond P, et al. Time in range: a best practice guide for UK diabetes healthcare professionals in the context of the COVID-19 global pandemic. *Diabet Med*. 2021;38(1):e14433.
19. Riddlesworth TD, Beck RW, Gal RL, et al. Optimal sampling duration for continuous glucose monitoring to determine long-term glycemic control. *Diabetes Technol Ther*. 2018;20(4):314-316.
20. Camerlingo N, Vettoretti M, Cicler M, et al. 877-P: limits of correlation coefficient analysis in determining the minimal duration of CGM data needed to estimate time below range. *Diabetes*. 2020;69(Suppl 1):877.
21. Herrero P, Alalitei A, Reddy M, Georgiou P, Oliver N. Robust determination of the optimal continuous glucose monitoring length of intervention to evaluate long-term glycaemic control. *Diabetes Technol Ther*. 2021;23:314-319.
22. Camerlingo N, Vettoretti M, Facchinetti A, et al. An analytical approach to determine the optimal duration of continuous glucose monitoring data required to reliably estimate time in hypoglycemia. *Sci Rep*. 2020;10(1):1-13.
23. Aleppo G, Ruedy KJ, Riddlesworth TD, et al. REPLACE-BG: a randomized trial comparing continuous glucose monitoring with and without routine blood glucose monitoring in adults with well-controlled type 1 diabetes. *Diabetes Care*. 2017;40(4):538-545.
24. Bailey TS, Chang A, Christiansen M. Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm. *J Diabetes Sci Technol*. 2015;9(2):209-214.
25. Reddy M, Jugnee N, Anantharaja S, Oliver N. Switching from flash glucose monitoring to continuous glucose monitoring on hypoglycemia in adults with type 1 diabetes at high hypoglycemia risk: the extension phase of the i HART CGM study. *Diabetes Technol Ther*. 2018;20(11):751-757.
26. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *BMJ Emerg Med J*. 2003;20(5):453-458.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Camerlingo N, Vettoretti M, Sparacino G, et al. Design of clinical trials to assess diabetes treatment: Minimum duration of continuous glucose monitoring data to estimate time-in-ranges with the desired precision. *Diabetes Obes Metab*. 2021;1-9. https://doi.org/10.1111/dom.14483