

ORIGINAL ARTICLE

Methods for Predicting Diabetes Phase III Efficacy Outcome From Early Data: Superior Performance Obtained Using Longitudinal Approaches

JB Møller¹, NR Kristensen¹, S Klim¹, MO Karlsson², SH Ingwersen¹ and MC Kjellsson²

The link between glucose and HbA1c at steady state has previously been described using steady-state or longitudinal relationships. We evaluated five published methods for prediction of HbA1c after 26/28 weeks using data from four clinical trials. Methods (1) and (2): steady-state regression of HbA1c on fasting plasma glucose and mean plasma glucose, respectively, (3) an indirect response model of fasting plasma glucose effects on HbA1c, (4) model of glycosylation of red blood cells, and (5) coupled indirect response model for mean plasma glucose and HbA1c. Absolute mean prediction errors were 0.61, 0.38, 0.55, 0.37, and 0.15% points, respectively, for Methods 1 through 5. This indicates that predictions improved by using mean plasma glucose instead of fasting plasma glucose, by inclusion of longitudinal glucose data and further by inclusion of longitudinal HbA1c data until 12 weeks. For prediction of trial outcome, the longitudinal models based on mean plasma glucose (Methods 4 and 5) had substantially better performance compared with the other methods.

CPT Pharmacometrics Syst. Pharmacol. (2014) 3, e122; doi:10.1038/psp.2014.20; published online 2 July 2014

In 2012, it was estimated that the total cost of diagnosed diabetes in United States alone was \$245 billion.¹ In order to lower such expenses, there is a clear need for development of cost-effective and safe antidiabetic medicine. At the same time, pharmaceutical companies are struggling to improve R&D productivity² and to fulfill the needs of the growing patient population.³ Key drivers for increased productivity are correct decision making with regard to early selection of candidates and doses—and improved decision making is likely to lead to lower attrition rate in phase II and III.²

In development of antidiabetic medicine, the use of different biomarkers at each development stage⁴ represents one of the main challenges for taking such decisions in drug development. Early clinical studies are typically of short duration, and efficacy is measured using biomarkers that reach steady state within a fairly short timeframe. Dependent on mode of action, efficacy is either assessed using glucose infusion rates from a glucose clamp study, glucose excursions following glucose tolerance tests, or simply by measurements of plasma glucose (typically fasting plasma glucose (FPG)). Clamp tests and glucose tolerance tests often need to be performed in an inpatient setting, whereas plasma glucose measurements can also be performed at home (self-measured plasma glucose (SMPG)), providing the opportunity of frequent assessment of treatment effect. Clearly, both FPG values analyzed in a laboratory, and SMPG values can be biased and hard to interpret due to various factors such as inconsistency in reporting concentration in blood or plasma values and day-to-day variation.^{5,6}

Although debated, the predictive power of HbA1c for development of long-term complications of diabetes is indeed strong.^{7,8} Based on this, steady-state values of HbA1c have been chosen as the main efficacy end point for confirmatory

phase III trials.^{9,10} In contrast, early clinical trials are often of short duration, not allowing HbA1c to reach steady state, and hence, values of glucose at steady state are often used as the main efficacy end point. Thus, in order to predict efficacy outcomes in confirmatory trials (based on early trial data), a translation from values of posttreatment steady-state glucose to posttreatment steady-state HbA1c is needed.

The correlation between HbA1c and FPG has previously been established in several studies.^{11,12} Similar relations have been proposed for the relationship between HbA1c and mean plasma glucose (MPG) calculated from 24-h glucose profiles based on either 7- to 11-point SMPG profiles or continuous measurements.^{13,14} Measuring FPG is less of a burden from a patient's point of view. The drawback of using FPG for efficacy measurement is that the effect of drugs that mainly target the postprandial glucose (PPG) might not be accurately captured. As HbA1c is a function of overall glycemic exposure,^{15,16} one would expect more accurate predictions of HbA1c based on MPG values compared with FPG values.

Although steady-state relations between glucose and HbA1c have been calculated on large populations of subjects, they have generally not been tested and compared for prediction of clinical trial outcome across different types of treatments. Also, due to their inherent structure, they do not take into account the physiological and longitudinal relations between glucose and HbA1c when subjects are under investigation in a clinical trial. Thus, new methods have been proposed that incorporate physiologic knowledge relating glucose measures to HbA1c and enable longitudinal descriptions of HbA1c.^{17–19}

A direct comparison of methods for prediction of confirmatory trial outcome based on early trial data can both provide guidance on which methods to select for prediction of HbA1c

¹Quantitative Clinical Pharmacology, Novo Nordisk A/S, Søborg, Denmark; ²The Pharmacometrics Group, Uppsala University, Uppsala, Sweden. Correspondence: JB Møller (jbem@novonordisk.com)

Received 26 December 2013; accepted 9 April 2014; published online 2 July 2014. doi:10.1038/psp.2014.20

and inform trial design with regard to the choice between FPG and 7-to 11-point SMPG profiles and the choice of sampling schedule. To our knowledge, no extensive comparison of methods for predicting HbA1c across trials and treatments has been published. In this study, we evaluated the possibility of predicting phase III efficacy outcomes using different data sources (FPG, MPG, and early HbA1c).

RESULTS

An overview of trials included in this study is presented in [Table 1](#). With this data, we evaluated five different methods ([Table 2](#)) for prediction of steady-state HbA1c (here defined as HbA1c at 26/28 weeks in a clinical trial) based on data up to 12 weeks of treatment. In order to both evaluate the impact of using FPG vs. MPG and steady-state vs. longitudinal data, two steady-state methods were applied. Method 1 applies the steady-state relation between FPG and HbA1c previously derived by Samtani.¹⁷ Method 2 is based on the steady-state relation between MPG and HbA1c previously published by Nathan *et al.*¹⁴

For incorporation of longitudinal FPG and MPG data, the models proposed by Samtani (Method 3)¹⁷ and Lledó-García *et al.* (Method 4)¹⁸ were applied. Method 3 is an indirect response model linking FPG to HbA1c, whereas Method 4 is based on the relation between MPG and the lifespan and

glycosylation of red blood cells (RBCs). Method 5 consists of two linked indirect response models in which predictions of MPG are related to predictions of HbA1c. Thus, only Method 5 used both MPG and HbA1c as dependent variables. For each longitudinal model, all data up to 12 weeks were included. Thus, Method 3 included FPG and HbA1c data up to 12 weeks, Method 4 used MPG data up to 12 weeks, and Method 5 used MPG and HbA1c data up to 12 weeks. All methods were applied for prediction of all 12 treatment arms evaluated in the four trials as shown in [Table 1](#).^{20–23}

Mean prediction errors at 26/28 weeks are presented for each treatment arm in [Figure 1](#). Individual predictions from each of the five methods were obtained from 12-week data in line with the description for each method (see Methods section). All methods underestimated the observed values to various degrees. In particular, the application of FPG values (Methods 1 and 3) seemed to cause predictions to be much lower than the observed values. Numerical values of mean absolute prediction errors are presented in [Table 3](#) for each treatment arm for the five different methods. For Method 5, the overall mean prediction error was -0.13% points, whereas for the other methods, it ranged from -0.37 to -0.61% points. As seen, the regression-based method using FPG (Method 1) significantly underestimated HbA1c at end of trial for basal insulins (arms 1, 2, and 4). The predictions from the MPG-based methods (Methods 2 and 4) were more consistent

Table 1 Summary of studies applied for comparison of steady-state HbA1c prediction methods

Study	Arm	<i>n</i>	HbA1c at baseline (%)	Fasting plasma glucose (weeks)	24-h glucose profiles (weeks)	HbA1c measures (weeks)	Steady-state HbA1c evaluated at
Rosenstock <i>et al.</i> ²⁰	Insulin glargine (comparator)	211	8.5	0, 1, 2, 3, 4, 6, 8, 10, 12	0, 12	0, 12	28 weeks
	Insulin detemir	198	8.6				
Raskin <i>et al.</i> ²¹	Biphasic insulin aspart	95	9.5	0–12	0, 12	0, 12	28 weeks
	Insulin glargine (comparator)	104	9.7				
Zinman <i>et al.</i> ²²	Liraglutide 1.8 mg + metformin + rosiglitazone	134	8.6	0, 2, 8, 12	0, 8, 12	0, 8, 12	26 weeks
	Metformin + rosiglitazone (comparator)	123	8.3				
	Liraglutide 1.2 mg + metformin + rosiglitazone	145	8.5				
Nauck <i>et al.</i> ²³	Liraglutide 0.6 mg + metformin	182	8.3	0, 2, 4, 8, 12	0, 8, 12	0, 8, 12	26 weeks
	Liraglutide 1.8 mg + metformin	168	8.4				
	Glimeperide + metformin (comparator)	191	8.4				
	Metformin	78	8.3				
	Liraglutide 1.2 mg + metformin	176	8.3				

Table 2 Summary of five methods compared for prediction of HbA1c at steady state

Method number (study)	Independent variables (data source)	Dependent variables (prediction)	Model description
Method 1 (Samtani ¹⁷)	FPG: samples taken at 12 weeks	Steady-state HbA1c	Linear regression model
Method 2 (Nathan <i>et al.</i> ¹⁴)	MPG: 12-week SMPG profile	Steady-state HbA1c	Linear regression model
Method 3 (Samtani ¹⁷)	FPG: baseline to 12-week samples	Longitudinal predictions of HbA1c	Indirect response model for HbA1c
Method 4 (Lledó-García <i>et al.</i> ¹⁸)	MPG: baseline to 12-week SMPG profiles	Longitudinal predictions of HbA1c	Simulation model based on glycosylation of red blood cells
Method 5 (Møller <i>et al.</i> ¹⁹)	MPG: baseline to 12-week SMPG profiles HbA1c: screening to 12-week samples	Longitudinal predictions of MPG and HbA1c	Coupled indirect response model for MPG and HbA1c

See detailed description of each method in Methods section.

FPG, fasting plasma glucose; MPG, mean plasma glucose; SMPG, self-measured plasma glucose.

but still underestimated HbA1c by 0.3–0.4% points. Only Method 5 overestimated HbA1c for some treatments (arms 5, 7, and 11). Precision was substantially higher for the three methods using MPG (Methods 2, 4, and 5) than for the two methods using FPG (Methods 1 and 3). When evaluating the model performance by means of the root mean square error (RMSE), a measure of individual prediction errors (Table 3), the above conclusions were largely confirmed: the steady-state methods (Methods 1 and 2) provided the largest RMSE

with overall larger errors of predictions based on FPG than those based on MPG. Furthermore, the RMSE was considerably lower for Method 5 than for the other four methods.

Figure 2 presents the prediction errors in terms of treatment contrasts between study drugs and comparators ($\Delta\Delta\text{HbA1c}$). The $\Delta\Delta\text{HbA1c}$ in each panel of Figure 2 was calculated by subtracting the baseline-corrected mean of HbA1c at 26/28 weeks of the comparator arms ($\Delta\text{HbA1c}_{\text{comparator}}$) from the baseline-corrected mean of HbA1c at 26/28 weeks of the study drug arm ($\Delta\text{HbA1c}_{\text{study drug}}$) for observed data or for predictions obtained using the five methods. The confidence intervals of $\Delta\Delta\text{HbA1c}$ were obtained by comparing the distribution of baseline-corrected HbA1c of treatment and comparator, using *t*-test. Lines through 0 and -0.3% are inserted to indicate no difference and the noninferiority limits for comparison against a comparator defined in regulatory guidance.^{9,10} In line with the results presented in Figure 1, the steady-state methods more frequently provided an incorrect conclusion compared with the longitudinal models. Additionally, the longitudinal model based on MPG observations (Method 4) and the model based on MPG observations in combination with early HbA1c observations (Method 5) provided the correct conclusion for all comparisons. The longitudinal approach using FPG (Method 3) failed to correctly predict trial outcome in several of the comparisons.

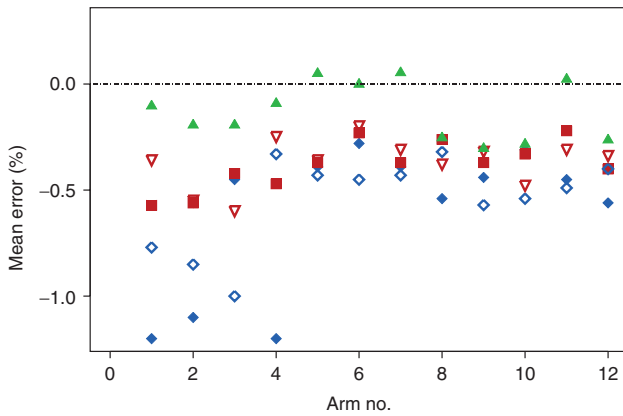


Figure 1 Mean prediction performance per arm of each method for HbA1c prediction at end of trial (26/28 weeks). Predictions were obtained using 12-week data and prediction errors were expressed as the deviation between the mean of individual predictions and the mean of the observations per arm. Arm 1: Insulin glargine, Arm 2: Insulin detemir, Arm 3: Biphasic insulin aspart, Arm 4: Insulin glargine, Arm 5: Liraglutide 1.8 mg + metformin, Arm 6: Metformin + rosiglitazone, Arm 7: Liraglutide 1.2 mg + metformin + rosiglitazone, Arm 8: Liraglutide 0.6 mg + metformin, Arm 9: Liraglutide 1.8 mg + metformin, Arm 10: Glimiperide + metformin, Arm 11: Metformin, Arm 12: Liraglutide 1.2 mg + metformin. Key to symbols: Method 1 (filled blue diamonds),¹⁷ Method 2 (filled red squares),¹⁴ Method 3 (open blue diamonds),¹⁷ Method 4 (open red squares),¹⁸ Method 5 (filled green triangles).¹⁹

DISCUSSION

We compared various methods for predicting steady-state HbA1c outcome (defined as HbA1c at 26/28 weeks in a clinical trial). In order to evaluate both the impact of the type of data used (MPG vs. FPG) and the method (steady-state vs. longitudinal models), classical regression techniques linking glucose to HbA1c at steady state and model-based longitudinal approaches were applied for prediction. The performance of each method was evaluated based on its ability to predict mean HbA1c at 26/28 weeks in each arm

Table 3 Mean prediction errors and RMSE of methods for predicting HbA1c treatment outcome

Arm no.	Arm	Method 1 (Samtani—SS)	Method 2 (Nathan—SS)	Method 3 (Samtani)	Method 4 (RBC)	Method 5 (Møller)
1	Insulin glargine (comparator)	-1.2 (2.4)	-0.4 (1.4)	-0.7 (1.3)	-0.4	-0.1 (0.5)
2	Insulin detemir	-1.1 (2.0)	-0.4 (1.4)	-0.8 (1.6)	-0.4	-0.2 (0.4)
3	Biphasic insulin aspart	-0.5 (1.9)	-0.4 (2.0)	-1.0 (2.1)	-0.6	-0.2 (0.6)
4	Insulin glargine (comparator)	-1.2 (2.6)	-0.5 (1.7)	-0.3 (1.7)	-0.3	-0.1 (0.5)
5	Liraglutide 1.8 mg + metformin	-0.4 (0.8)	-0.4 (0.8)	-0.4 (0.7)	-0.3	0.1 (0.2)
6	Metformin + rosiglitazone (comparator)	-0.3 (0.8)	-0.2 (1.0)	-0.4 (0.7)	-0.3	0.0 (0.3)
7	Liraglutide 1.2 mg + metformin + rosiglitazone	-0.4 (0.7)	-0.4 (0.9)	-0.4 (0.8)	-0.2	0.1 (0.5)
8	Liraglutide 0.6 mg + metformin	-0.5 (1.2)	-0.3 (1.4)	-0.3 (0.8)	-0.4	-0.3 (0.4)
9	Liraglutide 1.8 mg + metformin	-0.4 (1.3)	-0.4 (0.9)	-0.6 (1.1)	-0.4	-0.3 (0.5)
10	Glimiperide + metformin (comparator)	-0.3 (1.0)	-0.3 (1.1)	-0.5 (1.0)	-0.4	-0.3 (0.6)
11	Metformin	-0.5 (1.0)	-0.2 (1.4)	-0.5 (1.0)	-0.3	0.0 (0.5)
12	Liraglutide 1.2 mg + metformin	-0.6 (1.1)	-0.4 (1.0)	-0.5 (1.0)	-0.4	-0.3 (0.5)
All	Overall mean prediction error (mean RMSE)	-0.61 (1.4)	-0.38 (1.3)	-0.55 (1.2)	-0.37 (1.2)	-0.13 (0.4)
All	Overall absolute mean prediction error	0.61	0.38	0.55	0.37	0.15

Mean prediction errors and RMSE of HbA1c (%) predictions calculated as mean of individual predictions subtracted by mean of individual observations. HbA1c predictions at end of trial (26–28 weeks) were based on 12-week data. Steady-state predictions (Methods 1 and 2) were obtained using FPG/MPG values at 12 weeks only. Overall mean and absolute mean prediction errors are shown in bold. FPG, fasting plasma glucose; MPG, mean plasma glucose; RBC, red blood cell; RMSE, root mean square error; SS, steady state.

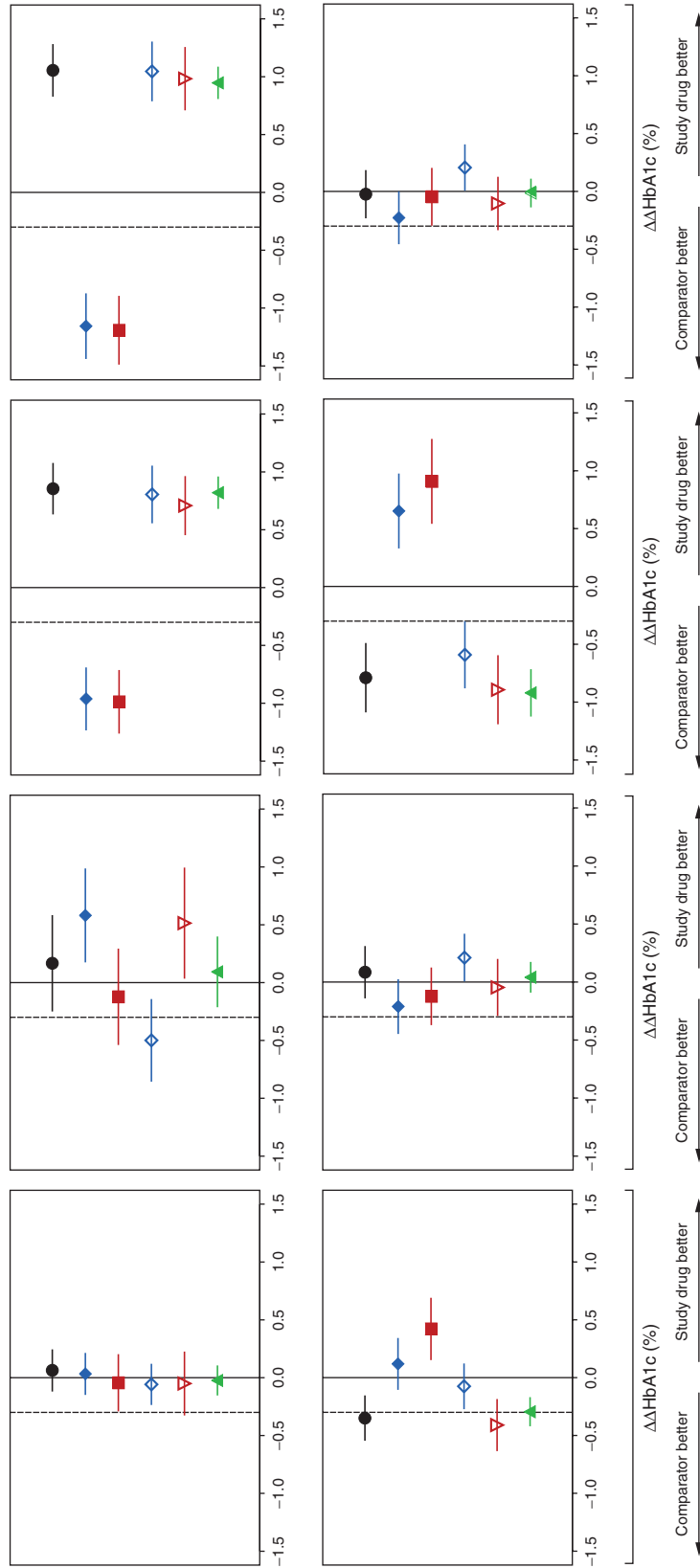


Figure 2 Phase III outcome ($\Delta\Delta\text{HbA1c}$ (%)) observations (black) and predictions across treatment arms. Method 1 (filled blue diamonds),¹⁷ Method 2 (filled red squares),¹⁴ Method 3 (open blue diamonds),¹⁷ Method 4 (open red triangles),¹⁸ Method 5 (filled green triangles).¹⁹ Vertical lines through 0 and -0.3% indicate no difference and noninferiority limits used for treatment comparisons. Left to right-top to bottom: 1) insulin detemir vs. insulin glargine, 2) biphasic insulin aspart vs. insulin glargine, 3) Liraglutide 1.8 mg + metformin + rosiglitazone vs. metformin + rosiglitazone, 4) Liraglutide 1.2 mg + metformin + rosiglitazone vs. metformin + rosiglitazone, 5) Liraglutide 0.6 mg + metformin vs. metformin, 6) Liraglutide 1.8 mg + metformin vs. metformin, 7) Metformin vs. metformin, 8) Liraglutide 1.2 mg + metformin vs. metformin, 9) Liraglutide 1.8 mg + metformin vs. metformin.

(by mean prediction error and RMSE) and on its ability to predict trial outcome expressed as the difference in treatment effect vs. comparator. Using these two evaluation criteria, the methods were in turn compared both with regard to their ability to predict the observed HbA1c value and their ability to predict HbA1c change from baseline vs. comparator. For both criteria, the methods using MPG data provided better predictions than the ones using FPG. This was in accordance with theoretical considerations,²⁴ as the MPG—in contrast to FPG—also contains information of drug effects on PPG.

The best predictive performance was obtained with a dynamic model describing longitudinal progression of both MPG and HbA1c (Method 5). As this model uses all the observations up to 12 weeks collected for both MPG and HbA1c, it is not surprising that this model has the best predictive performance for end-of-trial HbA1c. Moreover, this method accurately predicted the outcome of treatment comparisons obtained for all eight comparisons included. Contrary to this, the two steady-state methods (Methods 1 and 2) were unable to predict the actual outcome for four of eight comparisons.

In the study, we implemented each prediction method as presented in the original papers. Thus, we used fixed parameters and constants as reported in the original publications and explained in Methods section. Based on our results using the published parameter values, all methods significantly underestimated the final HbA1c values.

For the steady-state methods (Methods 1 and 2), we used the regression parameters reported by Samtani¹⁷ and Nathan *et al.*¹⁴ The possibility of reestimating the parameter values using, e.g., week 12 data from each treatment arm would likely have resulted in biased parameter values as HbA1c, contrary to MPG, would not be at steady state at this time point. Moreover, the original parameter values were estimated on larger data sets of a more heterogeneous nature than the clinical trial data investigated in this study and thus we did not find it appropriate to change the regression parameters for the link between FPG/MPG and HbA1c reported in the original papers.

The longitudinal model describing the relation between FPG and HbA1c (Method 3) was implemented according to the original published version of the model. In three of four trials investigated in this work, at least one of the treatment arms involved drugs targeting PPG. As the longitudinal model presented by Samtani¹⁷ was originally estimated on trials with drug classes not targeting PPG, we also investigated an implementation with rate constants estimated for each arm. Estimating the glycosylation constants per arm, as opposed to per trial as recommended in the original publication, allows different glycosylation rates for drugs not targeting PPG and those who do. This did not improve the accuracy of the predictions (results not shown). Estimating the parameter for the lower limit of HbA1c ($HbA1c_{lower\ limit}$) in each trial improved the performance of the model significantly. For the same reason as for the regression methods, we did not find estimating the lower limit of HbA1c appropriate, as this parameter was implemented as being fixed across trials in the original paper, and the data investigated in this study were not as extensive as in the original paper by Samtani¹⁷ where 12 trials were used for estimation.

The second best method (Method 4) was the longitudinal model published by Lledó-García *et al.*¹⁸ This model predicted the $\Delta HbA1c$ just as well as Method 5, but predictive

performance was slightly worse for absolute HbA1c. This method also uses MPG to drive the glycosylation of RBCs but does not require estimation of any parameters. In this study, Method 4 was implemented as a pure simulation model, in the sense that all parameters except for residual variability were fixed using previously published data. Alternatively, the published parameters with their uncertainties could be used as priors for reestimation of parameters based on the study data. Method 4, as it was published by Lledó-García *et al.*, has high parameter certainty and was developed using data much more informative regarding the underlying mechanism of HbA1c formation compared with those included in this analysis. This led us to conclude that reestimation of parameters with priors would not change the parameters significantly. The only parameter reestimated in this setting was the residual error for HbA1c, which was also the most uncertain parameter according to Lledó-García and colleagues.

In a previous paper,¹⁹ we used both observed MPG values and FPG-derived MPG values as glucose input. Here, the implementation of Method 5 only included observed MPG values. This was done to be able to compare models across using only MPG values and models using only FPG values and to show that the use of MPG values alone also provides accurate predictions.

Despite the superior performance of the longitudinal methods based on MPG, it is worth mentioning that each method presented in this paper has its own merit. The regression-based methods can be used to obtain HbA1c predictions based on a single value of either FPG (Method 1) or MPG (Method 2) at steady state. These methods can thus be used to obtain predictions from early phase studies where longitudinal values of glucose are not being sampled. Furthermore, these methods have their merit for diagnostic purpose, by translating single glucose values into expected HbA1c. Clearly, using FPG in a longitudinal approach, as compared with MPG, also needs significantly fewer blood samples, which is one of the advantages of the approach presented by Samtani (Method 3).¹⁷

Various factors could account for the underestimation of HbA1c for each of the methods. One could be that disease progression is not explicitly accounted for in any of the models. The UK Prospective Diabetes Study reports a progression of HbA1c of 0.2% per year, which would thus cause predictions at 26 weeks to be 0.1% higher compared with baseline.²⁵ Such a progression should however be implemented on the glucose component (and not the HbA1c component) in order to align with the physiological principle that the glucose change drives the HbA1c change.

In summary, we have outlined and evaluated different methods for HbA1c predictions based on early trial data. In line with results from previous studies, our results show that using MPG provides more accurate predictions of HbA1c compared with using FPG, independently of treatment. Although the amount of samples needed to obtain estimates of MPG are substantially higher than for FPG, these samples can be taken at home using self-measured glucose values as applied in the studies included in this paper. Furthermore, in case longitudinal HbA1c measurements are available, we suggest including such measurements in the prediction of long-term HbA1c outcome.

METHODS

Trial design and patients

The trials used for evaluation of predictive performance consisted of three phase III trials and one phase IV trial of previously approved pharmaceuticals.^{20–23} The data from these trials were previously used to validate one of the models applied in this paper (Method 5).¹⁹ Trials were required to include glucose profiles and HbA1c sampling at least three times up to end of trial, which was at either 26 or 28 weeks. Furthermore, in order to use similar data for all methods, each subject was required to have an FPG and an MPG value at 12 weeks, and an HbA1c sample at end of trial (26–28 weeks). MPG values were obtained from glucose profiles sampled 7–11 times during the day, calculated as $AUC_{0-24\text{ h}}/24\text{ h}$. FPG was calculated as the mean of all available samples taken before breakfast at each visit. Typically, one FPG sample analyzed in the laboratory, and three prebreakfast SMPG values were available at a given visit, in which case the final FPG value was calculated as the mean of these four values.

Methods used for prediction of HbA1c

Five methods were used to perform predictions based on data up to 12 weeks. The two methods relating single measurements of steady-state glucose to steady-state HbA1c were evaluated assuming the glucose measurement at 12 weeks to be representative of the glucose at steady state (Methods 1 and 2).^{14,17} In contrast, the longitudinal versions of the FPG/MPG models used all FPG/MPG data from baseline up to 12 weeks (Methods 3 and 4).^{17,18} The longitudinal glucose and HbA1c model (Method 5)¹⁹ used both MPG and HbA1c values up to 12 weeks. A detailed description of each model is provided below. Model codes appear in the Supplementary Appendix online.

Method 1 (FPG–HbA1c steady-state solution)

In order to compare longitudinal methods with steady-state solutions,¹⁶ we applied the original steady-state relationship between FPG and HbA1c derived by Samtani.¹⁷ This relation is based on a model built on an extensive data set with many different oral antidiabetic drugs administered to type 2 diabetes patients. The following relationship between glucose and HbA1c was derived in the original publication and applied here for prediction:

$$\text{HbA1c (\%)} = 0.5 \cdot \text{FPG} \left(\frac{\text{mmol}}{\text{l}} \right) + 2.84$$

The original parameter values were used for method comparison.

Method 2 (MPG–HbA1c steady-state solution)

Nathan *et al.*¹⁴ previously established the relation between average glucose and HbA1c data from 507 subjects. The following linear relation was derived in the original paper (revised to relate HbA1c to glucose and not vice versa as presented in the original paper) and applied here to perform predictions of HbA1c:

$$\text{HbA1c (\%)} = 0.62 \cdot \text{MPG} \left(\frac{\text{mmol}}{\text{l}} \right) + 1.62$$

These parameter values were used for method comparison.

Method 3 (FPG–HbA1c model)

The longitudinal model by Samtani¹⁷ is derived based on mean data from various different phase II trials. In order to align with the original publication, we implemented the model using a fixed effects approach, thus keeping all interindividual variability parameters fixed to 0. The following equations describe the relation between glucose and HbA1c:

$$\frac{d\text{HbA1c}}{dt} = k_g \cdot \text{FPG} \left(\frac{\text{mmol}}{\text{l}} \right) - k_d \cdot \text{HbA1c} \cdot \left(1 - \frac{\text{HbA1c}_{\text{lower limit}}}{\text{HbA1c}} \right)$$

Here, k_g and k_d are the two rate constants for HbA1c production and degradation, respectively, which account for the time delay between FPG and HbA1c. The system is initialized at the baseline value of HbA1c. $\text{HbA1c}_{\text{lower limit}}$ is a physiological limit parameter—which enables integration across trials with different baseline HbA1c levels. For our method comparison, k_g and k_d were estimated per trial using FPG and HbA1c values week 12, whereas $\text{HbA1c}_{\text{lower limit}}$ was fixed to 2.84% as reported.¹⁷

Method 4 (MPG–HbA1c model)

In order to incorporate the possibility of glucose levels affecting the life span of RBCs, Lledó-García *et al.*¹⁸ formulated a transit compartment model for description of aging of RBCs. The HbA1c predictions from the model are the estimated fraction of glycosylated to nonglycosylated RBCs—both of which have to be determined. The below equations describe the nonglycosylated part:

$$\text{LS}_i = \text{TVLS} \cdot \left(\frac{\text{MPG}}{149} \right)^\gamma \cdot e^{\eta_i} \quad k_{tr} = \frac{\text{NC}}{\text{LS}_i}$$

$$\text{PRE} = e^{-k_g \cdot \text{MPG} \cdot \text{LSP}_i}$$

$$\frac{d\text{RBC}_1}{dt} = k_{in} \cdot \text{PRE} - \text{RBC}_1 \cdot k_{tr} - \text{RBC}_1 \cdot k_g \cdot \text{MPG}$$

$$\text{RBC}_1(0) = \text{PRE} \cdot \frac{k_{in}}{k_{tr} + k_g \cdot \text{MPG}}$$

$$\frac{d\text{RBC}_2}{dt} = (\text{RBC}_1 - \text{RBC}_2) \cdot k_{tr} - \text{RBC}_2 \cdot k_g \cdot \text{MPG}$$

$$\text{RBC}_2(0) = \text{RBC}_1(0) \cdot \frac{k_{in}}{k_{tr} + k_g \cdot \text{MPG}}$$

$$\frac{d\text{RBC}_3}{dt} = (\text{RBC}_2 - \text{RBC}_3) \cdot k_{tr} - \text{RBC}_3 \cdot k_g \cdot \text{MPG}$$

$$\text{RBC}_3(0) = \text{RBC}_2(0) \cdot \frac{k_{in}}{k_{tr} + k_g \cdot \text{MPG}}$$

...

$$\frac{d\text{RBC}_{12}}{dt} = (\text{RBC}_{11} - \text{RBC}_{12}) \cdot k_{tr} - \text{RBC}_{12} \cdot k_g \cdot \text{MPG}$$

$$\text{RBC}_{12}(0) = \text{RBC}_{11}(0) \cdot \frac{k_{in}}{k_{tr} + k_g \cdot \text{MPG}}$$

Here LS_i represents the mean value life span for the i th individual, η_i is the individual deviation from the population, normally distributed with a mean of zero and a standard deviation of 0.29 (as previously reported).¹⁸ TVLS is the typical value for life span, which in our case was fixed to 91.7 days as previously reported.¹⁸ The parameter γ described the shape of the effect of MPG on the life span of RBCs. k_g is the rate of glycosylation of RBCs, fixed to $8.37 \cdot 10^{-6}$ as previously reported. NC is the number of transit

compartments (i.e., 12) from which the transit rate per day can be calculated (k_{tr}). PRE is the contribution coming from the precursor pool, which is calculated using the glycosylation and the life span of precursor, fixed to the reported value of 8.2 days with a standard deviation between subjects of 0.34.¹⁸ $k_{in} = 1$ is the normalized production rate of RBCs.

The structure of the coupled differential equations system for the glycosylated RBCs follows the one written above, but with the first equation in this system written as

$$PRE_g = 1 - PRE$$

$$\frac{dRBC_{13}}{dt} = k_{in} \cdot PRE_g - RBC_{13} \cdot k_{tr} + RBC_1 \cdot k_g \cdot MPG$$

$$RBC_{13}(0) = \frac{k_{in} \cdot PRE_g + RBC_1(0) \cdot k_g \cdot MPG}{k_{tr}}$$

$$\frac{dRBC_{14}}{dt} = (RBC_{13} - RBC_{14}) \cdot k_{tr} + RBC_2 \cdot k_g \cdot MPG$$

$$RBC_{14}(0) = \frac{RBC_{13}(0) \cdot k_{tr} + RBC_2(0) \cdot k_g \cdot MPG}{k_{tr}}$$

$$\frac{dRBC_{15}}{dt} = (RBC_{14} - RBC_{15}) \cdot k_{tr} + RBC_3 \cdot k_g \cdot MPG$$

$$RBC_{15}(0) = \frac{RBC_{14}(0) \cdot k_{tr} + RBC_3(0) \cdot k_g \cdot MPG}{k_{tr}}$$

...

$$\frac{dRBC_{24}}{dt} = (RBC_{23} - RBC_{24}) \cdot k_{tr} + RBC_{12} \cdot k_g \cdot MPG$$

$$RBC_{24}(0) = \frac{RBC_{23}(0) \cdot k_{tr} + RBC_{12}(0) \cdot k_g \cdot MPG}{k_{tr}}$$

In order to obtain a predicted measure of HbA1c (%), the relation between glycosylated RBCs and the total population of RBCs is calculated according to:

$$HbA1c (\%) = 100 \cdot RBC_{gly} / RBC_{tot}$$

Here, RBC_{gly} represents the amount of glycosylated RBCs obtained as the sum of compartments RBC_{13} to RBC_{24} , whereas RBC_{total} is the total sum of RBCs in compartments RBC_1 to RBC_{24} .

For our method comparison, we used the published parameter values.¹⁸ The only reestimated parameter was the residual error for HbA1c, which was also the most uncertain parameter according to Lledó-Garcia and colleagues.

Method 5 (MPG–HbA1c model)

This model consists of two linked indirect response models, linking MPG dynamics to HbA1c dynamics.¹⁹ The model has both MPG and HbA1c as dependent variables and thus requires longitudinal data of both MPG and HbA1c. The structural model is described by

$$\frac{dMPG}{dt} = -k_{out_MPG} \cdot (MPG - MPG_{target})$$

$$MPG(t = \text{screening}) = MPG_{ss}$$

$$\frac{dHbA1c}{dt} = k_{in_HbA1c} \cdot (MPG + \beta) - k_{out_HbA1c} \cdot HbA1c$$

$$HbA1c(t = \text{screening}) = \frac{k_{in_HbA1c} \cdot (MPG_{ss} + \beta)}{k_{out_HbA1c}}$$

where k_{out_MPG} , k_{in_HbA1c} , and k_{out_HbA1c} are parameters characterizing the first-order turnover of MPG and HbA1c, respectively. MPG_{ss} is a parameter representing the estimated pretreatment steady-state level, whereas MPG_{target} is assigned different values during the course of the trial: during the run-in period, MPG_{target} is assigned the value of MPG_{base} (MPG at randomization), and during the treatment period, it is assigned the value of $MPG_{posttreatment}$ (MPG level at full treatment effect). The time between screening and baseline can vary between trials. As indicated, both MPG and HbA1c are assumed to be at steady state at the time of screening, but not necessarily at randomization. The parameter β relates MPG to HbA1c at steady state and is estimated per subject using a population model.

For method comparison, values of MPG_{ss} , MPG_{target} , k_{out_MPG} , and β were estimated per subject using a population approach. k_{in_HbA1c} and k_{out_HbA1c} were fixed to 0.081% per mmol/L per week, and 0.226 per week, respectively.¹⁹

Evaluation of prediction performance

Two different metrics were used for evaluating prediction performance. Prediction performance on the absolute HbA1c scale was evaluated by comparing mean predictions and mean observations for each method and for each treatment arm at 26–28 weeks, depending on study. The prediction error $\epsilon(HbA1c)_j$ for each arm (j) was calculated according to

$$\epsilon(HbA1c)_j = \overline{pHbA1c}_j - \overline{HbA1c}_j$$

where $\overline{pHbA1c}_j$ represents the mean of the individual predictions for arm j and $\overline{HbA1c}_j$ represents the mean of the observations for arm j . Thus, a negative value indicates underprediction of HbA1c. In order to assess the potential bias as well as the mean deviation, summary statistics were derived using absolute prediction errors as well as the actual numerical values. In case all values have the same sign, the mean deviation is thus equal to the bias.

The RMSE was estimated for each treatment arm j as a summary statistic over all subjects i as:

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{n_j} (pHbA1c_i - HbA1c_i)^2}{n_j}}$$

where $pHbA1c_i$ and $HbA1c_i$ represent the individual predictions and observations for subject i , respectively, and n_j is the number of subjects in treatment arm j .

Furthermore, all methods were evaluated with respect to their ability to predict phase III efficacy outcome with respect to performance against comparator. The difference in the ability to lower HbA1c between trial drug and comparator was calculated according to the following formula:

$$\Delta\Delta HbA1c = \Delta HbA1c(\text{trial drug arm}) - \Delta HbA1c(\text{comparator arm})$$

$$\Delta HbA1c = HbA1c(\text{at 26–28 weeks}) - HbA1c(\text{at baseline})$$

Software

S-PLUS, version 8.0 (TIBCO, Palo Alto, CA) was used for data file processing, explorative data analysis, and plotting.

NONMEM version 7.1.2 (ICON Development Solutions, Ellicott City, MD) was used for model estimation and simulation.

Acknowledgments. The research leading to these results was conducted within the Drug Development Modelling Resources (DDMoRe) consortium and has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115156. Resources for this consortium are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and European Federation of Pharmaceutical Industries and Associations companies' in-kind contribution. The DDMoRe project is also financially supported by contributions from Academic and small and medium-sized enterprises partners. This work was further funded by Corporate Research Affairs, Novo Nordisk A/S.

Author Contributions. J.B.M., N.R.K., S.K., and M.C.K. wrote the manuscript. J.B.M., N.R.K., S.K., M.O.K., S.H.I., and M.C.K. designed the research. J.B.M., N.R.K., S.K., M.O.K., S.H.I., and M.C.K. performed the research. J.B.M. analyzed the research.

Conflict of Interest. J.B.M., N.R.K., S.K., and S.H.I. are all employees of Novo Nordisk A/S and own stocks in the company. The other authors declared no conflict of interest.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

- ✓ Several approaches for predicting steady-state HbA1c based on either FPG or MPG have been proposed. These methods span from simple regression techniques to more mechanistic methods including turnover of glucose and HbA1c.

WHAT QUESTION DID THIS STUDY ADDRESS?

- ✓ We investigated the accuracy of published regression and model-based approaches built on FPG or MPG for predicting HbA1c clinical trial outcome from early data.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

- ✓ In the forecast of HbA1c clinical trial outcome, the use of MPG values is essential for obtaining accurate predictions. Based on our results, we further propose to use a longitudinal model for the predictions—preferably including early HbA1c measurements and integrating both glucose and HbA1c measurements in a pharmacometric framework.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS

- ✓ Our results provide evidence for a significant added value by application of longitudinal pharmacometric approaches compared with steady-state regression techniques. This was shown to provide accurate prediction of late-stage clinical trial efficacy outcome based on early trial data.

- American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. *Diabetes Care* **36**, 1033–1046 (2013).
- Paul, S.M. et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).
- Danaei, G. et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet* **378**, 31–40 (2011).
- Roden, M. *Clinical Diabetes Research: Methods and Techniques* (John Wiley & Sons, Chichester, 2007), 387–392.
- Schrot, R.J., Patel, K.T. & Foulis, P. Evaluation of inaccuracies in the measurement of glycemia in the laboratory, by glucose meters, and through measurement of hemoglobin A1c. *Clinical Diabetes* **25**, 43–49 (2007).
- Frank, E.A., Shubha, M.C. & D'Souza, C.J.M. Blood glucose determination: plasma or serum? *J. Clin. Lab. Anal.* **26**, 317–320 (2012).
- The DCCT Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986 (1993).
- Lind, M., Odén, A., Fahlén, M. & Eliasson, B. The true value of HbA1c as a predictor of diabetic complications: simulations of HbA1c values. *PLoS One* **4**, e4412 (2009).
- Guidance for Industry, *Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention* (U.S. Department of Health and Human Services, FDA, Center for Drug Evaluation and Research (CDER), Rockville p. 12, (February 2008).
- Guideline on Clinical Investigation of Medicinal Products in the Treatment of Diabetes Mellitus (Committee for Medicinal Products for Human Use (CHMP), CPMP/EWP/1080/00 Rev. 1, EMA, London, p. 7, (14 May 2012).
- Aleyassine, H., Gardiner, R.J., Tonks, D.B. & Koch, P. Glycosylated hemoglobin in diabetes mellitus: correlations with fasting plasma glucose, serum lipids, and glycosuria. *Diabetes Care* **3**, 508–514 (1980).
- Karl, D., Zhou, R., Vlainic, A. & Riddle, M. Fasting plasma glucose 6–12 weeks after starting insulin glargine predicts likelihood of treatment success: a pooled analysis. *Diabet. Med.* **29**, 933–936 (2012).
- Rohlfing, C.L., Wiedmeyer, H.M., Little, R.R., England, J.D., Tennill, A. & Goldstein, D.E. 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* **25**, 275–278 (2002).
- Nathan, D.M., Kuenen, J., Borg, R., Zheng, H., Schoenfeld, D. & Heine, R.J. Translating the A1C assay into estimated average glucose values. *Diabetes Care* **31**, 1473–1478 (2008).
- Monnier, L., Lapinski, H. & Colette, C. Contributions of fasting and postprandial plasma glucose increments to the overall diurnal hyperglycemia of type 2 diabetic patients: variations with increasing levels of HbA(1c). *Diabetes Care* **26**, 881–885 (2003).
- Monnier, L., Colette, C. & Owens, D. Postprandial and basal glucose in type 2 diabetes: assessment and respective impacts. *Diabetes Technol. Ther.* **13** (suppl. 1), S25–S32 (2011).
- Samtani, M.N. Simple pharmacometric tools for oral anti-diabetic drug development: competitive landscape for oral non-insulin therapies in type 2 diabetes. *Biopharm. Drug Dispos.* **31**, 162–177 (2010).
- Lledó-García, R., Mazer, N.A. & Karlsson, M.O. A semi-mechanistic model of the relationship between average glucose and HbA1c in healthy and diabetic subjects. *J. Pharmacokin. Pharmacodyn.* **40**, 129–142 (2013).
- Møller, J.B. et al. Longitudinal modeling of the relationship between mean plasma glucose and HbA1c following antidiabetic treatments. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e82 (2013).
- Rosenstock, J., Davies, M., Home, P.D., Larsen, J., Koenen, C. & Schernthaner, G. A randomised, 52-week, treat-to-target trial comparing insulin detemir with insulin glargine when administered as add-on to glucose-lowering drugs in insulin-naive people with type 2 diabetes. *Diabetologia* **51**, 408–416 (2008).
- Raskin, P. et al. Initiating insulin therapy in type 2 diabetes: a comparison of biphasic and basal insulin analogs. *Diabetes Care* **28**, 260–265 (2005).
- Zinman, B. et al.; LEAD-4 Study Group. Efficacy and safety of the human glucagon-like peptide-1 analog liraglutide in combination with metformin and thiazolidinedione in patients with type 2 diabetes (LEAD-4 Met+TZD). *Diabetes Care* **32**, 1224–1230 (2009).
- Nauck, M. et al.; LEAD-2 Study Group. Efficacy and safety comparison of liraglutide, glimepiride, and placebo, all in combination with metformin, in type 2 diabetes: the LEAD (liraglutide effect and action in diabetes)-2 study. *Diabetes Care* **32**, 84–90 (2009).
- Riddle, M., Umpierrez, G., DiGenio, A., Zhou, R. & Rosenstock, J. Contributions of basal and postprandial hyperglycemia over a wide range of A1C levels before and after treatment intensification in type 2 diabetes. *Diabetes Care* **34**, 2508–2514.
- UK Prospective Diabetes Study Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type II diabetes (UKPDS 34). *Lancet* **352**, 854–865 (1998).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (<http://www.nature.com/psp>)