Mechanistic modeling of hemoglobin glycation and red blood cell kinetics enables personalized diabetes monitoring

Roy Malka,1,2 David M. Nathan,3,4 John M. Higgins1,2*

The amount of glycated hemoglobin (HbA1c) in diabetic patients’ blood provides the best estimate of the average blood glucose concentration over the preceding 2 to 3 months. It is therefore essential for disease management and is the best predictor of disease complications. Nevertheless, substantial unexplained glucose-independent variation in HbA1c makes its reflection of average glucose inaccurate and limits the precision of medical care for diabetics. The true average glucose concentration of a nondiabetic and a poorly controlled diabetic may differ by less than 15 mg/dl, but patients with identical HbA1c values may have true average glucose concentrations that differ by more than 60 mg/dl. We combined a mechanistic mathematical model of hemoglobin glycation and red blood cell kinetics with large sets of within-patient glucose measurements to derive patient-specific estimates of nonglycemic determinants of HbA1c, including mean red blood cell age. We found that between-patient variation in derived mean red blood cell age explains all glucose-independent variation in HbA1c. We then used our model to personalize prospective estimates of average glucose and reduced errors by more than 50% in four independent groups of greater than 200 patients. The current standard of care provided average glucose estimates with errors to personalize prospective estimates of average glucose and reduced errors by more than 50% in four independent groups of greater than 200 patients. The current standard of care provided average glucose estimates with errors >15 mg/dl for one in three patients. Our patient-specific method reduced this error rate to 1 in 10. Our personalized approach should improve medical care for diabetes using existing clinical measurements.

INTRODUCTION

Diabetes mellitus is a growing global health burden affecting about 400 million people worldwide (1). A person’s glycated hemoglobin fraction (HbA1c) reflects the average concentration of glucose in the blood (AG) over the past 2 to 3 months and is the gold standard measure for estimating the risk for diabetes-related complications in patients with type 1 or type 2 diabetes (2–4). An HbA1c greater than or equal to 6.5% is diagnostic for diabetes, and the treatment goal for most people with diabetes is an HbA1c less than 7% (5). HbA1c is used to infer AG because continuous glucose measurements (CGMs) are not routinely available (6).

Glycation of hemoglobin occurs in a two-step process including the condensation of glucose with the N-terminal amino group of the hemoglobin β chain to form a Schiff base and the rearrangement of the aldimine linkage to a stable ketoamine (7). The kinetics of this slow nonenzymatic posttranslational modification are thought to depend largely on the concentration of glucose, with previous studies establishing the first-order kinetics (7–9) and irreversibility of HbA1c formation (10, 11). Other glycated forms are generated, but HbA1c is the clinically relevant glycation product and is therefore the focus of this analysis. Hemoglobin in older red blood cells (RBCs) has had more time to become glycated, and older RBCs therefore have higher glycated fractions. HbA1c is measured as an average over RBCs of all ages in the circulation and therefore depends on both AG and M_{RBC}. Other factors may also be involved, including glucose gradients across the RBC membrane, intracellular pH, and glycation rate constants.

Here, we study the glycemic and nonglycemic determinants of HbA1c. First, we dissected the contributions of glycemic and nonglycemic factors by deriving a mechanistic mathematical model quantifying the dependence of HbA1c on the chemical kinetics of hemoglobin glycation in a population of RBCs in dynamic equilibrium. Second, we personalized the model parameters for individual patients using existing CGM data. Third, we validated the personalized model’s utility in accurately estimating future AG from future HbA1c for each patient, and we compared the accuracy of the patient-specific model estimates of AG with the accuracy of those made using the current standard regression method.

RESULTS

A mathematical model for HbA1c can be derived from hemoglobin glycation and RBC kinetics

The process of HbA1c formation inside a single RBC can be described by the irreversible chemical reaction of hemoglobin with glucose to form glycated hemoglobin (gHb) with rate $k_g$:

$$\text{Hemoglobin} + \text{glucose} \xrightarrow{k_g} g\text{Hb} \tag{1}$$

The rate of change in gHb can be modeled with a differential equation:

$$\frac{d}{dt} \text{gHb}(t) = k_g \cdot \text{AG} \cdot (t\text{Hb} - \text{gHb}(t)) \tag{2}$$

tHb is the concentration of total hemoglobin in the RBC. The variable $t$ is the time for the glycation reaction and is equivalent to the RBC’s age. This model of glycation kinetics in general has been reported previously (7, 9, 12, 13). We use it here to describe glycation in a single RBC. Equation 2 can be solved analytically and scaled by tHb to yield the HbA1c in an RBC of age $t$:

$$\text{HbA1c}(t) = \frac{\text{gHb}(t)}{t\text{Hb}} = 1 - e^{\left(-k_g \cdot \text{AG} \cdot t\right)} + \frac{\text{gHb}(0)}{t\text{Hb}} e^{\left(-k_g \cdot \text{AG} \cdot t\right)} \tag{3}$$
We use AG instead of a time-varying glucose to simplify this initial derivation, and we account for the effects of time-varying glucose below, beginning in “Measured variation in $M_{\text{RBC}}$ is sufficient to explain all nonglycemic variation in HbA1c.” gHb(0) is the concentration of glycated hemoglobin in the RBC when it is a reticulocyte and has just entered the circulation. Because $e^x = 1 + x$ when $x$ is small, we can approximate Eq. 3 with a linear function. By linearizing the exponential in this way, we can then average over the roughly uniformly distributed ages of RBCs ($t$) in a patient’s circulation (14–16) to provide the clinically measured HbA1c:

$$\text{HbA1c} = \frac{\text{HbA1c}(0)}{\text{Intercept}} + \frac{[1 - \text{HbA1c}(0)]}{\text{Slope}} \cdot gHb(0) \cdot M_{\text{RBC}} \cdot AG$$

This linear relationship between AG and HbA1c has been reported in several studies such as the A1c-Derived Average Glucose (ADAG) study (see Fig. 1) (8). These studies also show that direct estimation of AG based solely on HbA1c may be inaccurate, in part because of the imprecision and inaccuracies of the component measurements. In addition, significant glucose-independent variation has been described (17–20), including a linear relationship between $M_{\text{RBC}}$ and HbA1c (10, 14). Regardless of the cause, an AG of 150 mg/dl may be associated with HbA1c anywhere between 5.5 and 8.0%, and an HbA1c of 6.5% may reflect AG anywhere between 125 and 175 mg/dl. See “Derivation of the AG-HbA1c linear regression from the physiological model of glycation” and “Synopsis of prior models of hemoglobin glycation” in the Supplementary Materials for more details.

**Patient-specific differences in the AG-HbA1c relationship are caused by slope variation and not by intercept variation**

The scatter of data points away from the regression line shown in Fig. 1 represents patient-specific deviation from the regression model (Eq. 4) in terms of intercept or slope. Published estimates of the intercept HbA1c(0) are small (~0.3%). One previous study measured HbA1c in diabetic and nondiabetic subjects. (red dots, $n = 268$) and type 2 (green dots, $n = 159$) diabetes, as well as nondiabetic subjects (red dots, $n = 80$). The lower dashed line is the 6.5% threshold for initial diagnosis of diabetes, and the upper dashed line is the 7% treatment target.

![Fig. 1. Linear relationship between AG and HbA1c in diabetic and nondiabetic subjects.](http://stm.sciencemag.org/)

**Fig. 1.** Linear relationship between AG and HbA1c in diabetic and nondiabetic subjects. Data from the ADAG study (8) are as follows: 507 subjects with type 1 (blue dots, $n = 268$) and type 2 (green dots, $n = 159$) diabetes, as well as nondiabetic subjects (red dots, $n = 80$). The lower dashed line is the 6.5% threshold for initial diagnosis of diabetes, and the upper dashed line is the 7% treatment target.

Transferrin receptor–positive RBCs, which are typically reticulocytes, and found results consistent with glycation in the bone marrow proceeding at a rate similar to that in the peripheral circulation (14). The range of potential interpatient variation in HbA1c(0) (<0.6%) is too limited to cause significant interpatient variation in HbA1c, though it is difficult to measure accurately in vivo.

Analysis of HbA1c variance as a function of AG suggests that interpatient variation in the slope rather than the intercept is a much more significant cause of glucose-independent variation in HbA1c. The amount of variation in HbA1c in Fig. 1 is different at different AG levels, with apparently less variation in HbA1c at lower AG. Interpatient variation in slope will have a different effect from variation in intercept (Fig. 2, A and B). We analyze this relationship in more detail by calculating the HbA1c variance within AG intervals of 10 mg/dl. We first assess the possibility that there is significant interpatient variability in the intercept, as illustrated in Fig. 2A. Figure 2C simulates the effect of variability in the intercept [HbA1c(0) or reticulocyte HbA1c] when

![Fig. 2. The variance of HbA1c increases with AG and suggests that interpatient differences in slope are more important than differences in intercept for determining nonglycemic variation in HbA1c.](http://stm.sciencemag.org/)

**Fig. 2.** The variance of HbA1c increases with AG and suggests that interpatient differences in slope are more important than differences in intercept for determining nonglycemic variation in HbA1c. A deviation from the regression line in Fig. 1 can be explained by a patient-specific line that has a different intercept, or a different slope, or both. (A) Variation in the intercept (reticulocyte glycation fraction) alone will lead to fixed deviations from the regression line that are independent of the AG. (B) In contrast, variation in the slope will lead to increased variance as AG increases. We can use the model to simulate the effect of each type of interpatient difference and compare the simulation results with the actual data. (C) Effect of simulated interpatient differences in the intercept (that is, the reticulocyte glycation fraction). $r^2$ is the rank correlation coefficient for the raw ADAG data (hence the “D” in $r_{ADAG}^2$), shown as red dots in both (C) and (D), and its correlation is the same in each panel because it is independent of the model of the source of variation. $r^2$ (s for simulation) in (C) is the correlation for the blue dots in (C), which show the simulated effect of interpatient differences in intercept in variance in AG. Their correlation (0.13) is different from $r_{ADAG}^2$ (0.8). The simulations show that interpatient differences in intercept would generate data (blue dots) whose trend (black line) is different from what was actually seen in the real data (red dots). (D) Effect of interpatient differences in slope on AG variance. The raw data (red dots) are repeated from (C) for comparison and have the same correlation coefficient ($r_{ADAG}^2$), $r^2$ in (D) is the correlation for the blue dots in (D), which show the simulated effect of interpatient differences in slope on variance in AG. This simulation shows that interpatient differences in slope would generate data (blue dots) whose trend (black line) is consistent with what was actually seen in the real data (red dots), and $r^2$ is similar to $r_{ADAG}^2$. 


2 of 9

Downloaded from http://stm.sciencemag.org/ by guest on November 9, 2017
the slope is fixed. This hypothesized model of interpatient differences in intercept (black line in Fig. 2C) generates data (blue points in Fig. 2C) that do not agree with the experimental data (red dots in Fig. 2C). Thus, interpatient differences in reticulocyte HbA1c are unlikely to be responsible for the glucose-independent variation seen in HbA1c.

Figure 2B depicts the effect of interpatient differences in the slope of the regression line. In Fig. 2D, the correlation between the simulated and actual data in this case is very high. Overall, it is much more likely that interpatient differences in slope are responsible for nonglycemic variation in HbA1c. See “Conditional variance of HbA1c controlling for AG” in Materials and Methods for more details.

**Measured variation in M\textsubscript{RBC} is sufficient to explain all nonglycemic variation in HbA1c**

We therefore focus on interpatient variation in the slope \( \theta = [1 - \text{HbA1c}(0)] \cdot k_g \cdot M_{\text{RBC}} \). The first component \([1 - \text{HbA1c}(0)]\), as discussed above, varies too little overall (\(-0.994 to 1.00\)) to be a significant cause of glucose-independent variation in HbA1c. The second component \((k_g)\) is not currently possible to measure directly in vivo, but it does not appear to vary between patients (21), and there is no reason to expect that a first-order chemical reaction rate would vary systematically between patients. The third component \((M_{\text{RBC}})\) has been measured with demanding and sophisticated labeling methods and has a mean of about 58 days and an SD of 4.5 to 6.5 days, for a coefficient of variation \(CV(M_{\text{RBC}}) = \frac{SD(M_{\text{RBC}})}{\text{mean}(M_{\text{RBC}})}\) between 7.8 and 11.2% (14, 15).

We can calculate a patient-specific corrected slope \(\hat{\theta} = \frac{\text{HbA1c} - \text{HbA1c}(0)}{\text{AG}}\) at the time of a specific HbA1c measurement using AG determined from large intrapatient CGM data sets. (We use the symbol \(\theta\) to represent an estimate of the true patient-specific slope \(\theta\), which cannot be measured directly.) AG is calculated from CGM data using a weighted average of individual glucose measurements, because glucose levels in the blood immediately before the HbA1c measurement influence the glycation levels in RBCs of all ages, whereas more distant glucose levels influence only those RBCs that are old enough to have been in the circulation at that time (22). See “Calculation of AG and coverage from CGM data” in Materials and Methods for more details.

We calculated \(\hat{\theta}\) for 36 distinct patients at our hospital and found that \(CV(\hat{\theta}) = 10.8\%\), within the range of variation that can be explained entirely by interpatient variation in \(M_{\text{RBC}}\). In three additional independent sets of 339 patients, we found that \(CV(\hat{\theta})\) was equal to 8.8% for 30 patients, 9.4% for 234 patients, and 9.9% for 75 patients. Analysis of all four populations suggests that glucose-independent variation in HbA1c can be explained entirely by the variation in \(M_{\text{RBC}}\). (See “Patient populations” in Materials and Methods for more details.) Figure 3 further shows that if \([1 - \text{HbA1c}(0)]\) and \(k_g\) are constant, all measured glucose-independent variation in HbA1c in the ADAG study data (8) can be accounted for by simulating the variation in \(M_{\text{RBC}}\) with a magnitude equivalent to that previously measured (14, 15). If either or both of the other two slope components \([1 - \text{HbA1c}(0)]\) and \(k_g\) vary significantly, they must be strongly negatively correlated with \(M_{\text{RBC}}\) or else \(CV(\hat{\theta})\) would be much greater than \(CV(M_{\text{RBC}})\).

**The personalized model increases the accuracy of prospectively estimated AG**

Because \(CV(M_{\text{RBC}}) \approx CV(\hat{\theta})\), we can estimate a patient’s \(M_{\text{RBC}}\) using published estimates of \([1 - \text{HbA1c}(0)]\) and \(k_g\cdot M_{\text{RBC}} = \frac{\theta}{[1 - \text{HbA1c}(0)] \cdot k_g}\). We therefore focus on interpatient variation in the slope \(\theta = [1 - \text{HbA1c}(0)] \cdot k_g \cdot M_{\text{RBC}}\). The first component \([1 - \text{HbA1c}(0)]\), as discussed above, varies too little overall (\(-0.994 to 1.00\)) to be a significant cause of glucose-independent variation in HbA1c. The second component \((k_g)\) is not currently possible to measure directly in vivo, but it does not appear to vary between patients (21), and there is no reason to expect that a first-order chemical reaction rate would vary systematically between patients. The third component \((M_{\text{RBC}})\) has been measured with demanding and sophisticated labeling methods and has a mean of about 58 days and an SD of 4.5 to 6.5 days, for a coefficient of variation \(CV(M_{\text{RBC}}) = \frac{SD(M_{\text{RBC}})}{\text{mean}(M_{\text{RBC}})}\) between 7.8 and 11.2% (14, 15).

We can calculate a patient-specific corrected slope \(\hat{\theta} = \frac{\text{HbA1c} - \text{HbA1c}(0)}{\text{AG}}\) at the time of a specific HbA1c measurement using AG determined from large intrapatient CGM data sets. (We use the symbol \(\theta\) to represent an estimate of the true patient-specific slope \(\theta\), which cannot be measured directly.) AG is calculated from CGM data using a weighted average of individual glucose measurements, because glucose levels in the blood immediately before the HbA1c measurement influence the glycation levels in RBCs of all ages, whereas more distant glucose levels influence only those RBCs that are old enough to have been in the circulation at that time (22). See “Calculation of AG and coverage from CGM data” in Materials and Methods for more details.

We evaluated AG estimates for 16 HbA1c measurements from nine distinct adult patients at Massachusetts General Hospital (MGH). The patient-specific model reduced the median absolute error in estimated AG from more than 15 mg/dl to less than 5 mg/dl, an error reduction of more than 66%. It is most informative to compare AG predictions where the model-based method differs from the current standard regression–based approach. Figure 5 compares errors in predicted AG when the two methods differ by at least 10 mg/dl and confirms the superior accuracy of model-based AG prediction in three additional independent patient populations totaling more than 300 individuals. (See “Patient populations” in Materials and Methods for further details.) Figure 5 shows that in each of these patient populations, the model-based approach reduced the median absolute error in estimated AG by at least 50%. The substantial improvement in accuracy achieved by the model is highlighted by the fact that, for all four independent study groups, the 75th percentile of the model-based estimation error is less than the median error for the current regression-based prediction.

The difference in AG between a nondiabetic (HbA1c < 6.5%) and a diabetic with suboptimal disease control (HbA1c > 7.0%) can be ~15 mg/dl (5). Thus, errors of 15 mg/dl or less in estimated AG could mislead clinicians and patients and compromise patient care and optimal
management of long-term risk of complications. Across our four sets of patients, the current regression method generated AG estimation errors greater than 15 mg/dl for about 1 in 3 patients (31.4%), whereas the patient-specific model produced errors this large for only 1 in 10 patients (9.6%). An error of 28.7 mg/dl in estimated AG is equivalent to an error of ~1% point in HbA1c. The fact that \( M_{\text{RBC}} \) improves the accuracy of HbA1c-derived AG is not entirely unexpected; however, the prospective utility of \( M_{\text{RBC}} \) to improve accuracy suggests that it is consistent in individuals over time. Optimal diagnosis and management of diabetes require an accurate estimate of AG. The improvement in AG calculation afforded by our model presented here has two advantages over these other approaches in that it controls for patient-specific variation in nonglycemic factors influencing HbA1c and it also takes advantage of the vastly richer glucose characterization provided by CGM. It is therefore not surprising that our patient-specific method estimated HbA1c with significantly higher accuracy. We estimated HbA1c for 200 patients in our study populations and found a correlation of 0.90 and an average deviation from measured HbA1c of 0.3%, meaning, for example, that if the measured HbA1c was 7.0%, this method would typically estimate an HbA1c between 6.7 and 7.3%. Given that analytic variation in HbA1c assays would be expected to generate an uncertainty range of at least 6.9 to 7.1% (27), the patient-specific model thus makes a significant advance toward optimal estimation.

**DISCUSSION**

We have developed a model of glycation kinetics and derived a patient-specific correction factor (\( M_{\text{RBC}} \)) to improve the accuracy of AG estimation from HbA1c.

The personalized model provides more accurate real-time estimates of HbA1c for patients with CGM

A method to estimate HbA1c from CGM in real time would provide useful feedback for patients trying to optimize glucose management between clinic visits. Patients are already accustomed to thinking about the quality of their glucose control in terms of HbA1c. Previous studies have developed sophisticated methods to estimate HbA1c by combining prior HbA1c levels with multipoint profiles of self-monitored glucose (26). These methods have generated impressive results with sparse measurements of glucose, achieving a correlation between estimated and measured HbA1c as high as 0.76, with estimates of HbA1c deviating from measured HbA1c by an average of as little as 0.5%. For example, if the measured HbA1c was 7.0%, this method would typically estimate an HbA1c between 6.5 and 7.5%. The patient-specific model presented here has two advantages over these other approaches in that it controls for patient-specific variation in nonglycemic factors influencing HbA1c and it also takes advantage of the vastly richer glucose characterization provided by CGM. It is therefore not surprising that our patient-specific method estimated HbA1c with significantly higher accuracy. We estimated HbA1c for 200 patients in our study populations and found a correlation of 0.90 and an average deviation from measured HbA1c of 0.3%, meaning, for example, that if the measured HbA1c was 7.0%, this method would typically estimate an HbA1c between 6.7 and 7.3%. Given that analytic variation in HbA1c assays would be expected to generate an uncertainty range of at least 6.9 to 7.1% (27), the patient-specific model thus makes a significant advance toward optimal estimation.

**DISCUSSION**

We have developed a model of glycation kinetics and derived a patient-specific correction factor (\( M_{\text{RBC}} \)) to improve the accuracy of AG estimation from HbA1c.

The fact that \( M_{\text{RBC}} \) improves the accuracy of HbA1c-derived AG is not entirely unexpected; however, the prospective utility of \( M_{\text{RBC}} \) to improve accuracy suggests that it is consistent in individuals over time. Optimal diagnosis and management of diabetes require an accurate estimate of AG. The improvement in AG calculation afforded by our model should improve medical care and provide for a personalized approach to determining AG from HbA1c. The model would require one pair of CGM-measured AG and an HbA1c measurement that would be used to determine the patient’s \( M_{\text{RBC}} \). \( M_{\text{RBC}} \) would then be used going forward to refine the future AG calculated on the basis of HbA1c.

Our study follows a rich history of mathematical modeling in diabetes, which has revealed important pathophysiologic insights with great potential to inform early diagnosis and effective treatment (28–32), as well as more recent studies modeling other aspects of diabetes.
diabetes, including models classifying diabetes subtypes by integrating medical record data (33), predicting near-term glucose based on dietary intake (34, 35), identifying patients at high short-term risk of diabetes (36), controlling for nonbiologic measurement errors (37), and optimizing treatment strategies using fasting plasma glucose measurements (38).

Future work is needed to define the duration of CGM required for sufficient calibration of $M_{RBC}$. Our analysis of these four data sets suggests that no more than 30 days is required, and we find statistically significant improvement in as few as 21 days. If a patient’s monthly glucose averages are stable, then the prior 1 month would be sufficient for calibration, and if the patient’s weekly glucose averages are stable, then even 1 week of CGM might be sufficient. The patients in our four study populations all received regular routine medical care and were generally healthy. Follow-up study is necessary to assess model accuracy in the setting of more acute and serious comorbid disease, including conditions known to affect RBC turnover. We note that our patient-specific model may be particularly helpful in situations where plasma glucose is likely to deviate significantly from the longer-term average reflected in HbA1c, such as optimization of treatment for a patient recently diagnosed with diabetes (38). By controlling for patient-specific nonglycemic factors, the model should improve the clinical utility of HbA1c to provide more information regarding AG levels.

Our patient-specific model provides a substantial improvement in the accuracy of AG estimates, but its estimates are not perfect. When used to estimate AG from HbA1c, the model's sensitivity to variation in true AG will depend on the accuracy of the input HbA1c and CGM. HbA1c is typically rounded to multiples of 0.1%, which means that the model is theoretically sensitive to changes of 2 to 3 mg/dl in AG, and higher-resolution HbA1c measurements would increase the model's sensitivity. Analytic variation in current HbA1c measurements is reported to be ~3% (27), and this variation alone would be expected to generate AG estimation errors of ~7 mg/dl. The median error in the model-based estimate of AG may thus be as low as possible given current HbA1c measurement methods, but errors for some individual patients are higher, and the source of those errors warrants further investigation. Individual CGMs have a reported error of about 10% (39), but because AG is an average over thousands of separate CGMs with frequent calibration, the expected error in AG is about 0.1%. Systematic bias in CGM or calibration

![Fig. 5. Model-based inference of AG from HbA1c reduces estimation errors by about 50%](http://stm.sciencemag.org/)

Fig. 5. Model-based inference of AG from HbA1c reduces estimation errors by about 50%. Top row shows histograms of errors in AG estimation for four different sets of patients using the current standard regression-based formula. Second row shows histograms of errors using model-based estimation of AG. Histograms include predictions where estimation methods differ by at least 10 mg/dl. Errors for model-based predictions are significantly more tightly clustered around zero. The bottom panel compares box plots of median absolute error and shows that the model reduces error by at least 50% in each of the four independent sets of patients. The model-based estimates are superior to the standard method in all four cases with $P < 0.001$. (See “Patient populations” in Materials and Methods and the Supplementary Materials for more details.)
might be involved in the differences between measured and calculated AG is limited, and factors such as glycation rates or intracellular pH would not currently be practical to measure. Future studies that directly measure \( M_{RBC} \) will be required to increase confidence that interindividual variability in \( M_{RBC} \) is the key factor, and improvements in methods that directly measure \( M_{RBC} \) will be important to determine how much nonglycemic variation in HbA1c remains unexplained. In the meantime, the correction factor we have identified appears to be sufficient to improve the accuracy of the AG estimation from HbA1c. More generally, our study demonstrates how clinical accuracy can be enhanced in a patient-specific manner by combining large intra-patient data sets with mechanistic dynamic models of physiology.

**MATERIALS AND METHODS**

**Study design**

Our goal was to develop a more accurate method for estimating AG from HbA1c by adjusting for interpatient variation in nonglycemic factors that help determine HbA1c. Our analysis required three steps.

1. We first quantified the factors determining AG-independent variation in HbA1c by developing a mechanistic mathematical model describing how HbA1c depends on the chemical kinetics of hemoglobin glycation in a population of RBCs at dynamic equilibrium.

2. We then combined the model with CGMs to personalize the model for each patient.

3. Using the patient-specific model in combination with one set of CGM and HbA1c, we derived a patient bin glycation in a population of RBCs at dynamic equilibrium, describing how HbA1c depends on the chemical kinetics of hemoglobin glycation in a population of RBCs at dynamic equilibrium.

We demonstrated the reproducibility of our results by analyzing four independent sets of patients and finding consistent improvement in the accuracy of estimated AG using our model. Because we analyzed CGM and HbA1c data retrospectively, both patients and treating physicians were blinded to the future use of our patient-specific model. Enrollment criteria varied for each patient set, as did any policies for blinding patients to CGM readings or for randomizing patients to CGM use. See “Patient populations” for more details.

**Conditional variance of HbA1c controlling for AG**

The amount of variation in HbA1c in Fig. 1 increases at higher AG levels, with apparently less variation in HbA1c at lower AG. We now analyze the relationship between HbA1c variance and AG in more detail by calculating the HbA1c variance in the ADAG data conditioned on AG. This conditional variance calculation is similar to conditional expectation calculations. Both involve averaging over all measurements that have corresponding AG levels within intervals (of 10 mg/dl in this case). Instead of averaging HbA1c itself as in conditional expectation, we now average the squared deviation of each HbA1c measurement from the mean: \( \text{HbA1c} - \text{E} \{ \text{HbA1c} \} \), \( \text{E} \{ \text{HbA1c} \} \).

The scatter of data around the AG-HbA1c linear regression line may reflect interindividual variation in the slope or the intercept of the regression model, or both. Figure 2 illustrates the different effects the intercept and slope variation would be expected to have on the HbA1c conditional variance. We first assess the possibility of significant interindividual variability in the intercept as illustrated in Fig. 2A. Figure 2C shows a simulation of the effect of variability in reticulocyte HbA1c (equivalent to variation in the intercept \( \beta \)) when \( M_{RBC} \) is fixed. This hypothesized model (black line in Fig. 2C) of variation in the intercept \( \beta \) generates data (blue points in Fig. 2C) that do not agree with the experimental data (correlation coefficient of \( r^2 = -0.05 \)). Thus, variation in reticulocyte HbA1c is unlikely to explain the observed scatter of HbA1c around the regression line.

Figure 2B illustrates the expected effect of variation in the slope of the regression line. The correlation between the conditional variance and AG calculated from simulated HbA1c and the ADAG data is \( r^2 = 0.94 \). Similarly, the correlation in the ADAG data is \( r^2 = 0.65 \) (Fig. 2D). Note that in the ADAG data, of 507 samples, there are 2 outliers both with AG in the range of 110 to 120 mg/dl, creating a single bin for the calculation of conditional variance. In Fig. 2D, we remove these two samples, increasing \( r^2 \) to 0.80. Overall, it is much more likely that interindividual variation in the regression slope is responsible for the variation observed in the AG-HbA1c relationship. Because the model is linear, we can analytically calculate the variance around the regression line (Eq. 5; black dotted line in Fig. 2, C and D).

In the following variance calculations, conditional expectation is taken with respect to the RBC age across the population of RBCs in one patient’s circulation, as well as with respect to the \( M_{RBC} \) for an individual patient across the population of individuals. We assume that the initial glycation fraction \( \frac{g_{\text{Hb}}}{g_{\text{Hb}}} \) is a random variable. To simplify the expression, we treat the glycation rate as a constant.

### Calculation of AG and coverage from CGM data

The AG that determines HbA1c is a weighted average of glucose levels before HbA1c measurement (22). As discussed above, the clinically measured HbA1c is an average of single-RBC HbA1c over the ages of RBCs in a patient’s blood sample. The RBC ages are assumed to be uniformly distributed between 0 and 2 \( \cdot M_{RBC} \) days. The blood glucose level on the day before the HbA1c measurement affects the HbA1c of almost every RBC in the blood sample. The blood glucose levels measured much earlier and closer to 2 \( \cdot M_{RBC} \) days before the HbA1c measurement will affect only the small fraction of the oldest RBCs still in circulation. The AG from CGM for the linearized model (Eq. 4) can be calculated using the following equation:

\[
AG = \frac{1}{2 \cdot M_{RBC}} \int_0^{2 \cdot M_{RBC}} \left( t \cdot \int_0^t \text{glucose} (\tau) \, d\tau \right) \, dt
\]

When full CGM data are not available, it is therefore more valuable to have recent CGMs. Defining \( f(t) \) as 1 if there are CGM data within 5 min before \( t \) and as 0 if there are none, we can calculate the fractional...
coverage of CGM data during the desired time period using a related equation:

\[ \text{Coverage} = \frac{1}{2 \cdot M_{\text{RBC}}} \int_{0}^{2 \cdot M_{\text{RBC}}} \left( \frac{1}{t} \int_{-t}^{0} I(\tau) \, d\tau \right) \, dt \]  

(7)

See below for numerical calculation of AG from CGM without assuming the linear approximation.

**Model simulation**

For the simulation in Fig. 3, we assume that the RBC life span is normally distributed among different individuals, with the mean and variance estimated from prior publications. Note that a specific assumption on the parametric distribution of \( M_{\text{RBC}} \) among individuals is necessary only in the simulation and is not required for the analytic calculations. We take \( M_{\text{RBC}} \) to be normally distributed across individuals but find that a gamma distribution yields similar results. The age distribution of RBCs within an individual is assumed to be uniform, with cell ages between 0 and \( 2 \cdot M_{\text{RBC}} \) days. We also assume that the glycation rate is essentially constant as has been demonstrated previously (14, 21). See “Glycation rate and \( M_{\text{RBC}} \)” in the Supplementary Materials for more details.

We then use the fitted average parameter values (slope and intercept) obtained from the corresponding linear regression line, and the model reconstructs the scatter of data points around the regression line, adding variability in \( M_{\text{RBC}} \) equivalent to that previously measured (14, 15). In the simulations, we use a value of 0.001 for the SD of reticulocyte HbA1c for the ADAG data. These values were adopted from the measurements of variation of HbA1c in reticulocytes (14). For \( k_p \), we allow a CV of 1%, although a CV of 5% with constant \( M_{\text{RBC}} \) will reconstruct the variation around the regression line, as expected from the functional form of the model.

We first assume that AG is estimated with high accuracy as a result of the large number of measurements included in the average. In the ADAG study (8), each AG value is calculated using more than 250 samples over the course of 3 months. The SE is determined as follows: \( \text{SE} < \frac{\text{SD}}{\sqrt{250}} \approx \frac{\text{SD}}{15} \). Thus, even if the level of variability in a single glucose measurement is extremely high, for example, \( \text{SD} = 30 \, \text{mg/dl} \), the resulting CV will be less than 3% for all AG values in the ADAG data. The SD for the full ADAG data set is 39 (mg/dl), and 8 (mg/dl) when restricting to the non-diabetic patients, and thus, the uncertainty in AG is expected to be less than 1 (mg/dl).

**Numerical solution**

The physiologic model for glycation can be solved numerically without making a linear approximation. We start with the differential equation model including a time-varying glucose concentration \( G(t) \):

\[ \frac{d}{dt} \text{HbA1c}(t) = k_p \cdot G(t) \cdot (\text{HbA1c} - \text{HbA1c}(t)) \]  

(8)

This equation can be integrated numerically to provide the HbA1c in an RBC of age \( t \):

\[ \text{HbA1c}(t) = \text{HbA1c}(0) + \frac{k_p}{\text{tHb}} \int_{0}^{t} G(\tau) \cdot (\text{HbA1c} - \text{HbA1c}(\tau)) \, d\tau \]  

(9)

The clinical HbA1c measurement over a uniform distribution of RBC ages ranging between 0 and \( 2 \cdot M_{\text{RBC}} \) days:

\[ \text{HbA1c} = \frac{1}{2 \cdot M_{\text{RBC}}} \int_{0}^{2 \cdot M_{\text{RBC}}} \left[ \text{HbA1c}(0) + \frac{k_p}{\text{tHb}} \int_{0}^{t} G(\tau) \cdot (\text{HbA1c} - \text{HbA1c}(\tau)) \, d\tau \right] dt \]  

(10)

Given sufficient CGM data to define \( G(t) \) and a concurrent HbA1c measurement, the above equation can be solved numerically for \( M_{\text{RBC}} \) to provide a patient-specific \( M_{\text{RBC}} \). For the model-based prediction of AG from HbA1c, the patient’s \( M_{\text{RBC}} \) is used, and the following equation is solved numerically for AG:

\[ \text{HbA1c} = \frac{1}{2 \cdot M_{\text{RBC}}} \int_{0}^{2 \cdot M_{\text{RBC}}} \left[ \text{HbA1c}(0) + AG \cdot \frac{k_p}{\text{tHb}} \int_{0}^{t} (\text{HbA1c} - \text{HbA1c}(\tau)) \, d\tau \right] dt \]  

(11)

Results shown in this study use the linear approximation. We replicated all analysis with numerical solutions and reach very similar conclusions.

**Patient populations**

**Patient set #1.**

We analyzed existing CGM data from 36 adult patients at MGH under a research protocol approved by the Partners HealthCare Institutional Review Board. CGMs were made with Dexcom G4 continuous glucose monitors (Dexcom Inc.). HbA1c was measured either on a Roche COBAS instrument (Roche Diagnostics) or Bio-Rad VARIANT II TURBO (Bio-Rad). Thirty-six patients had at least one HbA1c measurement with concurrent CGM covering a period of time equivalent to the most recent 30 days before HbA1c. See “Calculation of AG and coverage from CGM data” for more details. Of those 36 individuals, 9 had a total of 16 additional future HbA1c measurements with concurrent CGM covering a period of time equivalent to the most recent 30 days before HbA1c. Those 16 future HbA1c measurements were used to validate the accuracy of the model-based AG estimation.

**Patient set #2.**

Data for the second, third, and fourth patient populations were made available by the Jaeb Center for Health Research, a coordinating center for multicenter clinical trials and epidemiologic research. Their studies of diabetic control reported CGM and HbA1c measurements in patients and generously included raw data, enabling us to test our model and hypothesis in three additional independent data sets. Our “patient set #2” comes from a study entitled “Effect of metabolic control at onset of diabetes on progression of type 1 diabetes” (http://direcnet.jaeb.org/Studies.aspx?RecID=165). The original purpose of this study was to investigate the impact of intensive metabolic control from the onset of diabetes on preservation of C-peptide secretion. This study, conducted between November 2008 and October 2013, included patients aged 6 to 46 years. Thirty patients had at least one HbA1c measurement with concurrent CGM covering a period of time equivalent to the most recent 45 days before HbA1c. Of those 30 individuals, 23 had a total of 79 additional future HbA1c measurements with concurrent CGM covering a period of time equivalent to the most recent 45 days before HbA1c. Those 79 future HbA1c measurements and
corresponding CGM were used to validate the accuracy of the model-based AG estimation. The source of the data is the Diabetes Research in Children Network (DirecNet), but the analyses, content, and conclusions presented herein are solely the responsibility of the authors and have not been reviewed or approved by DirecNet.

**Patient set #3.**

The data for this third patient population came from a study entitled “A randomized clinical trial to assess the efficacy of real-time continuous glucose monitoring in the management of type 1 diabetes” (http://diabetes.jaeb.org/RT_CGMRCTProtocol.aspx). This study was designed to compare continuous glucose monitoring versus standard intensive glucose monitoring in three age groups (>25, 15 to 24, and 8 to 14 years), of intensively treated type 1 diabetics having high HbA1c values of 7.0 to 10.0%. A total of 234 patients had at least one HbA1c measurement with concurrent CGM covering a period of time equivalent to the most recent 45 days before HbA1c. Of those 234 individuals, 155 had a total of 276 additional future HbA1c measurements with concurrent CGM covering a period of time equivalent to the most recent 45 days before HbA1c. Those 276 future HbA1c measurements and corresponding CGM were used to validate the accuracy of the model-based AG estimation.

**Patient set #4.**

The data for this fourth patient population came from a study entitled “A randomized clinical trial to assess the efficacy and safety of real-time continuous glucose monitoring in the management of type 1 diabetes in young children (4 to <10 year olds)” (http://direcnet.jaeb.org/Studies.aspx?RecID=162). This study was designed to assess the efficiency of CGM in young children (4 to 10 years old) in terms of tolerability, safety, and effect on quality of life with type 1 diabetes. Thirty-seven patients had at least one HbA1c measurement with concurrent CGM covering a period of time equivalent to the most recent 45 days before HbA1c. See “Calculation of AG and coverage from CGM data” for more details. Of those 37 individuals, 31 had a total of 69 additional future HbA1c measurements with concurrent CGM covering a period of time equivalent to the most recent 45 days before HbA1c. Those 69 future HbA1c measurements and corresponding CGM were used to validate the accuracy of the model-based AG estimation.

### Computational modeling and statistical analysis

Modeling and statistical analyses were performed in MATLAB (MathWorks Inc.).

### REFERENCES AND NOTES

Mechanistic modeling of hemoglobin glycation and red blood cell kinetics enables personalized diabetes monitoring
Roy Malka, David M. Nathan and John M. Higgins

Sci Transl Med 8, 359ra130-359ra130.
DOI: 10.1126/scitranslmed.aaf9304

A better estimate of blood glucose
For optimal medical care, diabetics and their doctors need to know exactly the patient's recent average blood glucose. Malka and colleagues have developed a mathematical model to this end by integrating the mechanisms of hemoglobin glycation (an indication of blood glucose concentrations) and red blood cell kinetics. Combining the modeling with routine clinical measurements yielded personalized estimates of a patient's average blood glucose that reduced diagnostic errors by more than 50% compared to the current method.