Components of Total Measurement Error for Hemoglobin A₁c Determination, George Phillipov* and Patrick J. Phillips (Endocrinology, The Queen Elizabeth Hospital, Woodville, South Australia 5011, Australia; * author for correspondence: fax 61-8-8222-6021, e-mail george.phillipov@nwahs.sa.gov.au

The Diabetes Control and Complications Trial (DCCT) and the UK Prospective Diabetes Study (UKPDS), undertaken in people with type 1 and 2 diabetes, respectively (1, 2), established the significance of glycohemoglobin (HbA₁c), and in particular hemoglobin A₁c (HbA₁c), as a prognostic indicator for long-term micro- and macrovascular complications. However, the HbA₁c measured during the DCCT and UKPDS represents a Hb fraction characterized by its retention time on cation-exchange HPLC rather than its unique chemical structure (deoxyfructosylhemoglobin). Accordingly, the DCCT HbA₁c procedure represents a selective, but not a specific assay method (3).

Because no definitive or reference method exists for quantification of HbA₁c (4), the American Diabetes Association (ADA), in collaboration with the Association of Clinical Chemists, implemented the National Glycohemoglobin Standardization Program (NGSP) (5) to standardize HbA₁c values determined by methods different from that used in the DCCT. The ADA (6) now states that their recommended HbA₁c thresholds, with respect to patient management goals, are valid only for NGSP-certified methods.

The variability of HbA₁c measurements depends on both analytical and biological variation. However, because HbA₁c concentrations are used for individual patient management, only analytical imprecision and within-person biological variation (s1²) are relevant. Whereas the NGSP (5) states that within-person HbA₁c variance is negligible, previous studies (7, 8) have reported s1 estimates of 0.17–0.79, and investigations measuring gHb (9) or HbA₁c (10) have reported values between 0.45 and 1.03. Accordingly, both analytical and within-person variability, but particularly the latter, increase measurement uncertainty and, therefore, the potential for clinical misinterpretation at ADA-specified HbA₁c thresholds.

The following definitions are provided to avoid ambiguity with respect to terminology:

Within-person variance (s1²): the degree of random fluctuation of values around a person’s homeostatic set-point for a particular biological analyte. For people with diabetes, the HbA₁c homeostatic set-point is controlled by dietary and/or pharmacologic treatment, and not by the normal physiologic mechanism. Repeatability (s1): closeness of agreement between successive results obtained with the same method on identical test material and under the same conditions (same operator, same apparatus, same laboratory, and same time).

Reproducibility (s2): closeness of agreement between individual results obtained with the same method on
identical test material but under different conditions (different operator, different apparatus, different laboratory, and/or different time).

Analytical variance ($s_a^2$): comprises both within- ($s_o^2$) and between-assay components of variance.

Total measurement variance ($s_e^2$): comprises both biological and analytical variance.

Serial HbA$_{1c}$ measurements were made in a cohort of 26 diabetic patients, in stable metabolic control, taking part in a 48-week multicenter trial (11). The mean number of specimens per patient was 7.2 (range, 6–9). All gHb determinations were performed with an affinity microcolumn assay (12), and results were converted to HbA$_{1c}$ percent-equivalents based on an algorithm originally derived by comparison (n = 186) with a HPLC method (13). During the trial period (2 years), $s_a$ for the affinity column method was 0.47 at a mean HbA$_{1c}$ concentration of 9.6%. The standard error of the estimate ($s_e$), calculated by nonparametric regression (HbA$_{1c}$ vs time) (14), was used to determine long-term variability associated with each patient’s serial HbA$_{1c}$ measurements. The corresponding mean $s_a$ was determined as the root mean square of the individual estimates (15).

Four different blood samples (~100 μL of each in a sealed ampoule), spanning HbA$_{1c}$ concentrations of ~6–13%, were hand-delivered on the same day to the five pathology laboratories performing physician-referred HbA$_{1c}$ assays in this State. The protocol (16) requires that all samples be analyzed in duplicate within a single analytical run. Three laboratories used Bio-Rad Variant HPLCs (NGSP-certified), and two used Pharmacia Mono S column HPLC systems (13) traceable to the Bio-Rad Diamat HPLC method. The cooperative trial method was defined as ion-exchange HPLC.

Routine data analysis was performed using SPSS for Windows, Release 10.0.7 (SPSS Inc.) and the Cbstat program (http://www.cbstat.com). The procedures described by Steiner (16) were used to calculate the corresponding estimates of repeatability and reproducibility for the interlaboratory study.

The individual estimates of $s_a$ were not statistically related to the respective baseline HbA$_{1c}$ concentrations (Kendall tau-b, 0.039; $P = 0.8$). The overall mean $s_a$ was 0.65, and given $s_e^2 = s_o^2 + s_a^2$, the within-person standard deviation $s_i = \sqrt{(s_o^2 - s_a^2)} = 0.44$.

For the interlaboratory study, the respective mean HbA$_{1c}$ values for the four samples were 5.8%, 7.8%, 10.4%, and 12.8%, and no laboratory showed consistently high or low HbA$_{1c}$ values, based on a method that ranks the sum of replicates. Similarly, no abnormal data were identified within the four samples when we used the Dixon test, and experimental variation between laboratories and between replicates was homogeneous. ANOVA established a significant variance ratio between laboratories ($F_{4,12} = 6.5; P < 0.05$) and for laboratory-sample interaction ($F_{12,20} = 30.7; P < 0.01$). Solving standard ANOVA equations (16), we calculated the between-laboratory ($s_{b2}^2$) and laboratory-sample interaction ($s_{LS}^2$) variances as 0.145$^2$ and 0.12$^2$, respectively. The reproducibility ($s_e$), or variation arising from different operators, instruments, and laboratories is then given by:

$$s_e^2 = s_{L2}^2 + s_{LS}^2 + s_o^2$$

and calculated as 0.19. Accordingly, among the five laboratories, 57% of total variance is between-laboratory, 40% is attributable to laboratory-sample interaction, and 3% is attributable to repeatability. The total error variance associated with the five HbA$_{1c}$ assays is therefore given by:

$$s_e^2 = s_a^2 + s_i^2$$

and is calculated as 0.48$^2$, of which $s_a^2$ contributes 84%. The design of the present cooperative trial, however, did not allow an estimate of long-term repeatability.

Knowledge of $s_a$ allows estimation of the range within which the true value lies at a reported HbA$_{1c}$ value, assuming that biological variability is that of a typical patient. Moreover, because $s_a^2 < s_e^2$, total error can be decreased more by analyzing additional specimens on the same patient than by performing more assays on the same specimen. This is highlighted in Table 1, which summarizes confidence ranges at different probabilities for analysis of one and two specimens. Accordingly, from Table 1, to be 80% confident that the ADA goal of <7.0% has been achieved (single specimen), the measured HbA$_{1c}$ concentration should be <6.4%. A 95% confidence for the same goal requires a mean HbA$_{1c}$ concentration (two specimens) <6.3%. Alternatively, to be 90% confident (single specimen) that the ADA >8.0% intervention threshold has been exceeded, a measured HbA$_{1c}$ concentration ≥8.7% is necessary.

Although the goal of the NGSP is to minimize bias between the DCCT and other HbA$_{1c}$ methods, and thereby allow uniform application of DCCT-derived HbA$_{1c}$ results, measurement uncertainty at ADA clinical decision-making thresholds has not been thoroughly addressed. In particular, failure to acknowledge the magnitude of within-person variation produces a significant underestimation of total measurement error. Our estimate of $s_i$ (0.44) is remarkably similar to the value of 0.41 reported previously by Hytloft Petersen et al. (7), although both the experimental design and HbA$_{1c}$ methodology were different. In contrast, Kolatkar et al. (8)

### Table 1. Confidence ranges, at different probabilities, around a reported HbA$_{1c}$ concentration.

<table>
<thead>
<tr>
<th>Probability, %</th>
<th>One specimen</th>
<th>Two specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>±0.94$^a$</td>
<td>±0.71</td>
</tr>
<tr>
<td>90</td>
<td>±0.79</td>
<td>±0.60</td>
</tr>
<tr>
<td>80</td>
<td>±0.61</td>
<td>±0.47</td>
</tr>
<tr>
<td>60</td>
<td>±0.40</td>
<td>±0.31</td>
</tr>
</tbody>
</table>

$^a$ Implies that 95% of values should be within 0.94 above or below a reported HbA$_{1c}$ value.
reported lower values of 0.17 and 0.29 for 3- and 12-month study periods, respectively, where all patients were intensively treated and had HbA1c concentrations maintained at <7.0%.

Our findings for the five HPLC methods indicated that reproducibility was much less than within-person biological variation. However, the confidence intervals around a measured HbA1c concentration were still wide (Table 1) in comparison with the small difference (1%) between the ADA HbA1c management thresholds. The potential impact of measurement uncertainty on HbA1c thresholds has been discussed previously by Lytken Larsen et al. (17).

The most recently posted results for the College of American Pathologists HbA1c survey (18) show reproducibility values for certified methods between 0.19 and 0.85; the most common method (n = 335), the Abbott IMx (uncertified), had s = 0.54. Only HPLC methods had s <0.25, whereas all multianalyte methods had s >0.43. Minimal analytical performance has been proposed as 0.75s (19), which based on our results is 0.33. However, the actual DCCT HbA1c procedure had a repeatability of 0.15 on masked split-duplicate specimens, whereas the long-term internal quality control showed a s of ∼0.40 (20).

In summary, the degree of within-person biological variation associated with HbA1c determinations significantly increases the total measurement error. If a HbA1c assay of high reproducibility is not used, the dispersion range of true HbA1c values around the mean true biological set-point will be so wide that the ADA management thresholds may become unworkable. Although the NGSP has significantly reduced intermethod bias, only some HPLC methods currently meet the required analytical performance (19).

We thank P. Charles, M. Haywood, Dr. M. Whiting, Dr. S. Sykes, and D. Moore for taking part in the HbA1c inter-laboratory study.

References