Clinical Chemistry 54:2
240–248 (2008)

Special Reports

The IFCC Reference Measurement System for HbA_{1c}: A 6-Year Progress Report

Cas Weykamp,[1*] W Garry John,[2] Andrea Mosca,[3] Tadao Hoshino,[4] Randie Little,[5] Jan-Olof Jeppsson,[6] Ian Goodall,[7] Kor Miedema,[8] Gary Myers,[9] Hans Reinauer,[10] David B. Sacks,[11] Robbert Slingerland,[8] and Carla Siebelder[1]

BACKGROUND: The IFCC Reference Measurement System for hemoglobin (Hb)A_{1c} (IFCC-RM) has been developed within the framework of metrologic traceability and is embedded in a network of 14 reference laboratories. This paper describes the outcome of 12 intercomparison studies (periodic evaluations to control essential elements of the IFCC-RM).

METHODS: Each study included: unknown samples (to test individual network laboratories); known samples (controls); recently manufactured calibrators (to check calculated assigned value); stored calibrators (to test stability) and a calibration-set (to calibrate the IFCC-RM). The unknown samples are measured by use of the IFCC-RM and the designated comparison methods [DCMs; the National Glycohemoglobin Standardization Program (NGSP) in the US, Japanese Diabetes Society/Japanese Society for Clinical Chemistry (JDS/JSCC) in Japan, and Mono-S in Sweden] are used to investigate the stability of the Master Equation (ME), the relationship between IFCC-RM and DCMs.

RESULTS: A total of 105 IFCC-RM data sets were evaluated: 95 were approved, 5 were not, and for 5 no data were submitted. Trend analysis of the MEs, expressed as change in percentage HbA_{1c} per year, revealed 0.000% (NGSP, not significant), −0.030%, (JDS/JSCC; significant) and −0.016% (Mono-S; not significant). Evaluation of long-term performance revealed no systematic change over time; 2 laboratories showed significant bias, 1 poor reproducibility. The mean HbA_{1c} determined by laboratories performing mass spectrometry (MS) was the same as the mean determined by laboratories using capillary electrophoresis (CE), but the reproducibility at laboratories using CE was better. One batch of new calibrators was not approved. All stored calibrators were stable.

CONCLUSION: A sound reference system is in place to ensure continuity and stability of the analytical anchor for HbA_{1c}.

Hemoglobin A_{1c} (HbA_{1c})[12] is the most important marker for long-term assessment of the glycemic state in patients with diabetes *(1)*. Goals for therapy are set at specific HbA_{1c} target values *(2)*, and the importance of standardization of HbA_{1c} measurement has been well recognized, as reflected by the establishment of national designated comparison methods (DCMs): the National Glycohemoglobin Standardization Program (NGSP) in the US *(3)*, the Japanese Diabetes Society/Japanese Society for Clinical Chemistry (JDS/JSCC) in Japan *(4)*, and Mono-S in Sweden *(5)*. A disadvantage of DCMs is that they are based on arbitrarily chosen analytical methods, with results in arbitrary units. International support is increasing for standardization of laboratory tests that requires a reference system based on the concept of metrologic traceability: the traceability chain *(6, 7)*. The traceability model is described in an ISO document that forms the basis for the European Directive on In Vitro Diagnostic Devices *(8, 9)*. To begin establishing a system for traceability, the IFCC Working Group on Standardization of HbA_{1c} developed a reference system for HbA_{1c} in which HbA_{1c} is defined as the stable adduct of glucose to the N-terminal valine of the β-chain of hemoglobin *(10)*. Mixtures of

[1] Queen Beatrix Hospital, Winterswijk, The Netherlands (IFCC-network coordinator); [2] Norfolk and Norwich University Hospital, and School of Medicine, Health Policy and Practice, UEA, Norwich, UK; [3] Centro Interdipartimentale per la Riferibilità Metrologica in Medicina di Laboratorio (CIRME), Università degli Studi di Milano, Milano, Italy; [4] Institute of Biopathological Medicine, Kanagawa, Japan (JDS/JSCC network coordinator); [5] University of Missouri School of Medicine, Columbia, MO (NGSP network coordinator); [6] Malmoe University Hospital, Malmoe, Sweden (Coordinator Reference System Sweden); [7] Austin Pathology, Austin Health, Heidelberg, Australia; [8] Isala Klinieken, Zwolle, the Netherlands; [9] Center for Disease Control and Prevention, Atlanta, GA; [10] INSTAND e.V., Duesseldorf, Germany; [11] Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

* Address correspondence to this author at: Queen Beatrix Hospital, Beatrixpark 1, 7101 BN Winterswijk, the Netherlands. Fax +31 543 524265; e-mail c.w.weykamp@skbwinterswijk.nl.

Received September 5, 2007; accepted November 15, 2007.
Previously published online at DOI: 10.1373/clinchem.2007.097402

[12] Nonstandard abbreviations: HbA1c, hemoglobin A1c; DCM, designated comparison methods; NGSP, National Glycohemoglobin Standardization Program; JDS/JSCC, Japanese Diabetes Society/Japanese Society for Clinical Chemistry; IFCC-RM, approved IFCC reference method for HbA_{1c}; ME, master equation.

Table 1. Reproducibility and stability calibrators and controls.

	Calibrator range, mmol/mol HbA_{1c} in 10 batches					
Calibrators	**A**	**B**	**C**	**D**	**E**	**F**
Maximum allowable difference	0.0–0.0	28.1–29.6	54.4–59.7	86.1–88.1	115.4–119.6	140.0–147.3
(Measured minus calculated HbA_{1c})	±0.8	±0.9	±1.2	±1.7	±2.3	±2.9
Production control new calibrators (n = 10)						
Range observed differences	-0.1 to 0.0	−0.2 to +0.8	−0.5 to +3.6[a]	−1.6 to +1.4	−1.3 to +2.1	−1.8 to +1.9
Mean difference	−0.1	+0.5	+0.4	+0.3	−0.4	+0.3
Stability control old calibrators (n = 5)						
Range observed differences		+0.1 to +0.6	0.0 to +1.2	−1.0 to +1.6	−0.9 to +1.2	
Mean difference		+0.4	+0.6	−0.1	+0.5	

	Control range, mmol/mol HbA_{1c} in 12 batches			
	Low (30.0–33.4)		**High (82.2–93.9)**	
Controls	**After 1 year**	**After 2 years**	**After 1 year**	**After 2 years**
Maximum allowable difference	±0.7	±0.7	±1.3	±1.3
(Measured minus previous measured HbA_{1c})				
Range observed differences	−0.2 to +0.3	−0.7 to +1.0[b]	−0.7 to +0.3	−0.8 to +0.4
Mean difference	0.0	−0.2	−0.1	−0.2

[a] In 1 of 10 batches the maximum allowable difference was exceeded.
[b] In 1 case the maximum allowable difference was exceeded.

pure HbA_{1c} and pure HbA_0 are prepared as primary calibrators *(11)* for the reference method (RM) *(12)*, which has been approved by member societies of the IFCC *(13)*. Relationships between values derived from the approved IFCC RM for HbA1c (IFCC-RM) and the respective DCMs, termed master equations (ME), have been established *(14)*. A recently signed Consensus Statement by the American Diabetes Organization, European Association for the Study of Diabetes, International Diabetes Federation, and IFCC states that the IFCC-RM represents the only valid anchor to implement standardization and also that HbA1c results be represented worldwide in IFCC Units (mmol/mol) and derived NGSP units (%) using the IFCC-NGSP master equation *(15)*. To maintain continuity the IFCC-RM has been embedded in a network of approved reference laboratories. The network is coordinated by a network coordinator who organizes periodic (twice a year) studies to investigate the validity of the essential elements of the reference system. These periodic evaluations, termed intercomparison studies, are the cornerstone of the work described in this paper. Intercomparison studies are designed to meet 5 essential aims: (*a*) joint approval of network laboratories; (*b*) evaluation of reproducibility in the relation between the IFCC-RM and DCMs; (*c*) value assignment and expanded uncertainty; (*d*) evaluation of long-term trend, bias, and reproducibility of network laboratories; and (*e*) assessment of reproducibility and stability of calibrators and controls. This paper describes the outcome of the 12 intercomparison studies performed between 2001 and 2006.

Materials and Methods

DESIGN

This intercomparison study was performed with calibrators and patient specimens. Once a year a 6-level batch of calibrators is manufactured in the ISO 9001: 2000 certified laboratory of the Network Coordinator by mixing pure HbA_{1c} and pure HbA_0 *(11)*. After manufacture, calibrators are included in the intercomparison study as "new calibrators" and measured as samples. The measured HbA_{1c} is compared with the HbA_{1c} calculated from the mixed amounts of pure HbA_{1c} and HbA_0 and if there is a good match (criteria in Table 1) the batch is approved. In the second year the

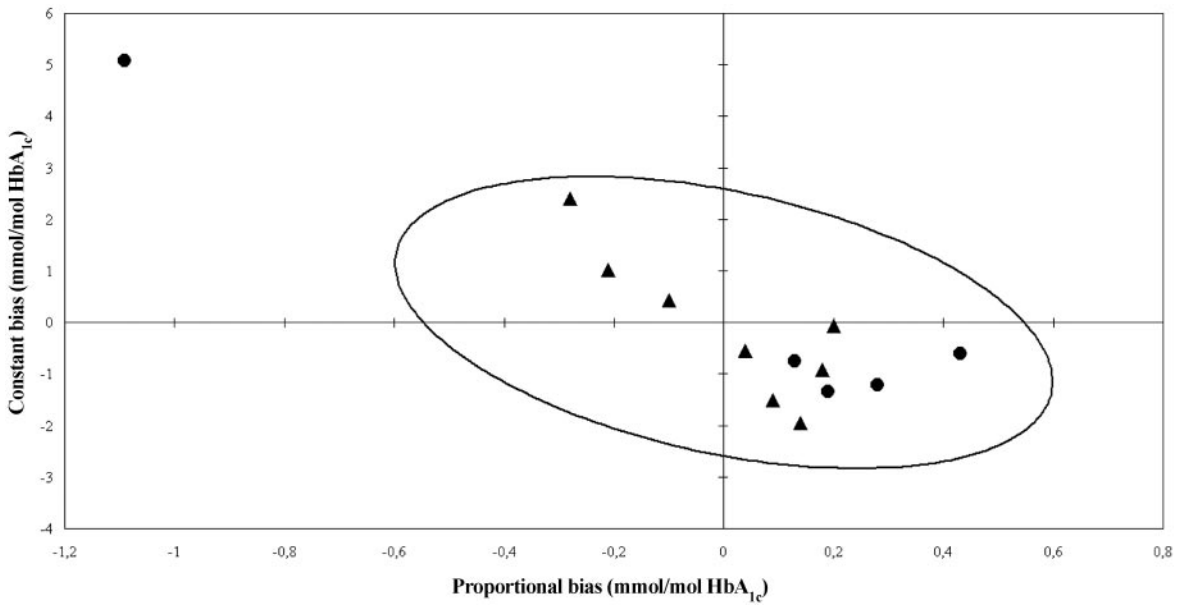

Fig. 1. Approval ellipse intercomparison study.

Proportional and constant bias are plotted on the x- and *y*-axis, respectively. Individual laboratories are represented by dots (HPLC-MS) or triangles (HPLC-CE). The ellipse is the graphical representation of criteria for approval (laboratories within the ellipse are approved).

(approved) batch is included as the "calibrator-set". In the third year the batch is included as "old calibrators" and again analyzed as sample (stability check). Thus, vials of 3 different batches of calibrators are included in each intercomparison study. Analogous to the president-elect, president, and past-president in organizations, there is a calibrator-elect, a calibrator, and a past-calibrator to ensure continuity and quality. The second type of samples are specimens derived from donated patient samples *(11)*, 5 samples with unknown HbA_{1c} for the purpose of approval of the network laboratories and 4 samples with known HbA_{1c}, 2 of which were manufactured 1 year before and 2 manufactured 2 years before, as controls. The 5 samples with unknown HbA_{1c} are also assayed by the respective DCM networks to evaluate the reproducibility of the ME. Data are presented for 12 consecutive studies. All specimens are shipped on dry ice to the participating laboratories *(14)*. All calibrators and specimens are stored at −84 °C.

MATERIALS AND METHODS AND UNITS

The IFCC Network Laboratories used the IFCC-RM *(12, 13)*. DCMs used the national reference methods: 3 JDS/JSCC network laboratories in Japan *(16)*, 8 NGSP network laboratories in the US *(17)*, and 1 Mono-S reference laboratory in Sweden *(5)*. According to the Consensus Statement *(15)*, HbA_{1c} is expressed as mmol HbA_{1c}/mol Hb for the IFCC-RM and as percentage HbA_{1c} (%HbA_{1c}) for the DCMs.

STATISTICS AND NETWORK RULES

The performances of laboratories in each intercomparison study were evaluated according to statistical rules for networks of reference laboratories *(18)*. In essence these rules are derived from a robust approach with all of the intercomparison samples examined together, allowing proportional and constant bias to be calculated. Mathematically, the differences between the results of an individual laboratory and the overall median are plotted against the overall median of the network laboratories, and the slope (proportional bias) and intercept (constant bias) are calculated from the linear relation. Slope and intercept together determine a criterion that is expressed as a graph (Fig. 1). The criterion for approval is fixed and set empirically: a laboratory is approved when the joint CI of constant and proportional bias is within the 95% CIs as derived from the first 6 intercomparison studies.

The relations between IFCC-RM and DCMs were evaluated using a method based on linear regression

analysis, using the formula: y = ax + b, in which x is the concentration HbA_{1c} of the IFCC-RM and y = the $\%HbA_{1c}$ of the respective DCM. The uncertainty in the relation is determined by the uncertainty in the ME, and the uncertainties in the outcome of the IFCC-network and DCM networks in the individual studies. This protocol implies that the CI is variable from study to study. In addition, the CI varies with the HbA_{1c} concentration. To examine the trend in the relationship between IFCC-RM and DCMs, a Shewhart chart with variable limits can be constructed (unpublished data). This report includes a chart for the defined target for diabetic control (53 mmol/mol).

Assigned values and the expanded uncertainty of specimens are calculated according to *(18)*.

To evaluate long-term performance of individual network laboratories, we included data from the last 8 intercomparison studies (4 laboratories were not approved at the time of the first 4 studies).

The criterion to approve a new batch of calibrators, and stability of calibrators of stored batches is the maximum allowable difference between calculated and measured HbA1c, defined as the combined uncertainty in the calculated and measured HbA1c concentrations at the 95% confidence level *(19)*. The criterion to approve stability of controls is the maximum allowable difference between assigned value and HbA1c measured after 1 and 2 years of storage, defined as the uncertainty in the assigned value at the 95% confidence level *(19)*.

For purpose of the evaluation of the network, specific statistics have been developed (*18*, *19*, and unpublished data). The statistics are implemented in R, a software environment for statistical computing and graphics, and an Excel add-in to import the results submitted by the network laboratories, as well as an R add-on package that contains functions specially defined for the analysis of the network.

It is the policy of the IFCC Working Group to have 8–15 approved network laboratories distributed in a number of countries. Network laboratories participate twice a year in an intercomparison study to demonstrate their competence. A laboratory wishing to be a reference laboratory first becomes a candidate reference laboratory and has to meet the defined analytical criteria in 2 consecutive intercomparison studies to gain the status of approved network laboratory. An approved network laboratory loses the status of approved network laboratory when it fails (or does not submit results) in 2 consecutive studies. A meeting is organized once a year for the network laboratories to exchange knowledge and discuss problems. The educational part of the network statistics helps laboratories that fail these criteria to identify the source of their problems.

Results

APPROVAL OF NETWORK LABORATORIES

A typical example of an approval ellipse from an intercomparison study is shown in Fig. 1. Thirteen laboratories (5 HPLC-MS and 8 HPLC–capillary electrophoresis) submitted results. Twelve of these results were within the ellipse indicating that these laboratories met the performance criteria and were approved. One laboratory failed. Fig. 1 also illustrates the phenomenon that a positive constant bias is associated with a negative proportional bias and vice versa. Linear regression analysis of data in 12 intercomparison studies revealed that this correlation is statistically significant: $r = 0.765$ ($P < 0.05$) and y = 0.733x + 0.02 (x = the proportional bias and y the constant bias).

REPRODUCIBILITY AND TREND IN THE RELATION BETWEEN IFCC-RM AND DCMS

The relation between IFCC-RM and DCMs is expressed by the equation y = ax + b, in which y is the HbA_{1c} value of a DCM, x is the HbA_{1c} value measured by the IFCC-RM, a is the slope, and b is the intercept. The median r value in the 12 studies was 0.9990 (NGSP), 0.9984 (JDS/JSCC), and 0.9985 (Mono-S). The "Equations and individual studies" section of Table 2 shows slope, intercept, and $\%HbA_{1c}$ (calculated at 53 mmol/mol, the defined target for diabetic control) for the 12 intercomparison studies for the respective DCMs.

The "Master Equations " section (Table 2) shows ME4 and ME12. ME4 is the mean relation between IFCC-RM and DCMs based on results of the first 4 intercomparison studies *(14)*. ME12 is calculated from all 12 intercomparison studies.

The "Trend Analysis " section deals with the trend in the IFCC-RM DCM relation and is expressed as %HbA1c/year over the period 2001–2006.

The Shewhart chart in Fig. 2 allows inspection of whether the relation of IFCC-RM and DCMs in each of the individual intercomparison studies is significantly different from the published ME. Observations within the limits (NGSP) imply compliance of those studies with the ME. Observations just outside or close to the limits (JDS studies 2003–2, 2005–2, 2006–1, 2006–2 and Mono-S 2002–2) suggest non- and borderline compliance, respectively.

VALUE ASSIGNMENT AND EXPANDED UNCERTAINTY

In 8 intercomparison studies, values have been assigned to 40 samples with HbA_{1c} concentrations from 32–121 mmol/mol. At the lower HbA_{1c} concentrations (30–40 mmol/mol HbA_{1c}), the mean (range) expanded uncertainty (k = 2) was 0.6 (0.4–0.8) mmol/

Table 2. Relationship and trend relationship between IFCC-RM and DCMs.

	USA (NGSP)			Japan (JDS/JSCC)			Sweden (Mono-S)		
	8 Network laboratories			3 Network laboratories			1 Network laboratory		
Studies	Slope	Intercept	%HbA_{1c}	Slope	Intercept	%HbA_{1c}	Slope	Intercept	%HbA_{1c}
Equations and individual studies									
Marrakech (2001–1)	0.0926	2.14	7.05	0.0934	1.76	6.71	0.1008	0.90	6.24
Chicago (2001–2)	0.0926	2.05	6.96	0.0926	1.67	6.58	0.0941	1.09	6.08
Kyoto-1 (2002–1)	0.0906	2.21	7.01	0.0920	1.78	6.66	0.1002	0.78	6.09
Kyoto-2 (2002–2)	0.0912	2.17	7.00	0.0943	1.68	6.68	0.0968	1.15	6.28
Barcelona-1 (2003–1)	0.0905	2.23	7.03	0.0912	1.78	6.61	0.0964	0.95	6.06
Barcelona-2 (2003–2)	0.0897	2.21	6.96	0.0916	1.70	6.55	0.0963	0.92	6.02
Los Angeles-1 (2004–1)	0.0901	2.24	7.02	0.0880	1.95	6.61	0.0949	1.10	6.13
Los Angeles-2 (2004–1)	0.0907	2.23	7.04	0.0911	1.73	6.56	0.0997	0.91	6.19
Orlando-1 (2005–1)	0.0913	2.15	6.99	0.0892	1.84	6.57	0.0961	1.01	6.10
Orlando-2 (2005–2)	0.0924	2.07	6.97	0.0928	1.63	6.55	0.0998	0.81	6.10
Amsterdam-1 (2006–1)	0.0890	2.28	7.00	0.0866	1.89	6.48	0.0968	0.89	6.02
Amsterdam-2 (2006–2)	0.0932	2.10	7.04	0.0932	1.59	6.53	0.0989	0.87	6.11
CV 12 Studies	1.4%	3.3%	0.5%	2.5%	6.1%	1.0%	2.3%	12.3%	1.3%
MEs									
ME 4	0.0915	2.15	7.00%	0.0927	1.73	6.64%	0.0989	0.88	6.12%
ME 12	0.0912	2.17	7.00%	0.0913	1.75	6.59%	0.0976	0.95	6.12%
Trend analysis									
Trend %HbA_{1c} /Year	<0.001%			−0.030%[a]			−0.016%		

[a] Statistically significant (linear regression; $P < 0.05$).

mol; in the middle HbA_{1c} concentrations (50–70 mmol/mol) it was 1.0 (0.8–1.2) mmol/mol; and at higher concentrations (80–120 mmol/mol) it was 1.5 (1.0–2.0 mmol/mol).

LONG-TERM TREND, BIAS, AND REPRODUCIBILITY RESULTS FOR NETWORK LABORATORIES

Table 3 shows the results of long-term evaluation of the network laboratories. Trend is expressed as the HbA_{1c} change (mmol/mol) per year of an individual laboratory. Bias is expressed as the differences between the mean HbA_{1c} value of a laboratory and that of the network mean in 8 intercomparison studies. Reproducibility is expressed as the SD of differences between a laboratory and the network-mean in the 8 intercomparison studies. Statistical significance is evaluated with linear regression analysis (trend), *t*-test (bias), and *F*-test (reproducibility). An evaluation of 8 intercomparison studies conducted between 2003 and 2006 (at an HbA_{1c} concentration of 53 mmol/mol), revealed that the trend over time (Table 3, column 2) ranges from +0.5 to −0.4 mmol/mol HbA_{1c} /year. Regression analysis shows that there is no significant trend for any of the network laboratories. Bias (Table 3, column 3) of network laboratory 2 (+0.8 mmol/mol) is statistically significantly higher than the network-mean, whereas the bias of network laboratory 4 (−0.9 mmol/mol) has a significantly lower outcome. Reproducibility (Table 3, column 4) ranges from 0.1–1.4 mmol/mol. Seven network laboratories have a very low variation (SD <0.4 mmol/mol) in their difference from the network-mean. Laboratory 3 has a statistically significant higher variation (SD 1.4 mmol/mol).

REPRODUCIBILITY AND STABILITY CALIBRATORS AND CONTROLS

The upper part of Table 1 shows essential data describing the reproducibility test of new batches of calibrators and summarizes the stability test of old calibrators. From 2000–2006, 10 batches of 6-concentration sets of calibrators were manufactured. To check the assigned value, new calibrators are measured as samples,

Table 3. Long-term performance of network laboratories in 8 intercomparison studies (2003–2006) at 53 mmol/mol HbA_{1c} concentration.

	Trend	Bias	Reproducibility
Network Laboratory	Shift in HbA_{1c} outcome of a laboratory in respect to the network-mean, mmol/mol HbA_{1c} change/year	Mean difference of a laboratory in respect to network-mean, mmol/mol HbA_{1c}	SD of differences of laboratory and network-mean, mmol/mol HbA_{1c}
	(r^2 in brackets)		0.4
1 (MS)	<0.1 (0.04)	+0.5	0.3
2 (MS)	<0.1 (0.03)	+0.8[a]	1.4[c]
3 (MS)	+0.5 (0.15)	−0.1	1.1
4 (MS)	−0.2 (0.10)	−0.9[b]	0.9
5 (MS)	−0.4 (0.19)	+0.2	0.3
6 (MS)	+0.1 (0.23)	−0.1	0.6
11 (CE)	−0.1 (0.09)	+0.5	0.5
12 (CE)	+0.1 (0.15)	−0.2	0.3
13 (CE)	<0.1 (0.00)	−0.3	0.3
14 (CE)	−0.1 (0.15)	−0.3	0.7
15 (CE)	<0.1 (0.00)	−0.1	0.6
16 (CE)	<0.1 (0.00)	−0.1	0.3
17 (CE)	+0.2 (0.18)	+0.2	0.2
18 (CE)	+0.1 (0.15)	0.0	0.1
19 (CE)	−0.2 (0.23)	+0.4	0.7[d]
Mean MS Laboratories		+0.1	0.4
Mean CE Laboratories		0.0	

[a] Mean HbA_{1c} outcome of this laboratory is significantly higher than network mean (*t*-test; $P < 0.05$).
[b] Mean HbA_{1c} outcome of this laboratory is significantly lower than network mean (*t*-test; $P < 0.05$).
[c] Variation in HbA_{1c} outcome over 8 intercomparison studies of this laboratory is significantly higher than network mean (*F*-test; $P < 0.05$).
[d] Variation in HbA_{1c} outcome of MS laboratories using mass spectrometry is significantly higher than that of CE laboratories (*F*-test; $P < 0.05$).

and outcome is compared with the HbA_{1c} calculated from the weighed amounts pure HbA_{1c} and pure HbA_0. The mean difference between measured and weighed HbA_{1c} ranges from −0.4 to +0.5 mmol/mol in the respective levels. The maximum allowable difference was exceeded only once (3.6 mmol/mol observed vs 1.2 mmol/mol allowed).

To test stability of the calibrators, a limited number of the moderate concentrations (B to E) of stored calibrators were assayed as samples 2–3 years after their manufacture. No significant difference between measured and weighed HbA_{1c} was observed in any of these "old calibrators".

In the lower part of Table 1 the outcomes for 12 batches of control samples, assayed 1 and 2 years after manufacture, are summarized. Control specimens are spare samples of previous intercomparison studies. In each intercomparison study, 4 controls are included: 2 (1 low and 1 high HbA_{1c} level) from an intercomparison study performed the previous year, and 2 from a study conducted 2 years earlier. On 1 occasion, for the low concentrations after 2 years of storage, the difference was 1.0 mmol/mol and exceeded the maximum allowable difference of 0.7 mmol/mol.

Discussion

APPROVAL OF NETWORK LABORATORIES

The model for approval was applied in 8 intercomparison studies in 2003–2006. All 105 datasets have been evaluated. On 95 occasions the laboratories passed, 5 times a laboratory failed, and 5 times a laboratory did not submit results. A network laboratory loses the status of approved laboratory when it fails (or does not submit results for) 2 consecutive studies, a situation that has not occurred to date. A candidate network laboratory can gain the status of approved network laboratory when it meets the criteria in 2 consecutive intercomparison studies. Four candidate laboratories have achieved this goal.

The criteria for approval of network laboratories are empirically based rather than determined on the basis of predefined performance goals *(20)*. One reason for this protocol is the lack of consensus on approval criteria: in a recent review Goodall *(21)* refers to 7 published statements, with a proposed CV of 2% to 5% for routine methods. The other reason is that the development of the RM started with the qualitative aim to be "as precise as possible." With 12 intercomparison studies completed, we now can quantify the performance of the network. According to criteria for approval of reference laboratories, the maximum CV of assigned values is <0.9% (assigned by the network) or <3% (assigned by an individual network laboratory). In light of the most stringent performance goal of 2% for routine laboratories *(21)*, the uncertainty of value assignment by one single reference laboratory (in general) is too high to be acceptable, but suitable when performed by the whole network. However these are maximum CVs. The CV of 0.5% seen over 6 years in relation to the NGSP (Table 2) suggests that the actual CV of the network is substantially lower than 0.9%. From the performance data of 12 intercomparison studies, it can be concluded that the IFCC-RM is suitable for the intended purpose in the top of the traceability chain of HbA1c. The data also suggest that, to limit uncertainty, it is preferable for values to be assigned by the network rather than by individual network laboratories.

STABILITY OF THE RELATIONSHIP BETWEEN IFCC-RM AND DCMS

The stability of the relationship between IFCC-RM and DCMs is of the utmost importance for clinical studies. In 2004 the relationship was calculated on the basis of the 4 completed intercomparison studies and published as the ME *(14)*. To date, the outcome of 12 intercomparison studies is known and allows evaluation of compliance with the published ME for each of the intercomparison studies and trend over time in the relationship of IFCC-RM and DCMs. As demonstrated by the r values, the relationship between IFCC-RM and DCMs is very consistent. Slope and intercept are not independently related, as demonstrated by the CVs (CV %HbA1c <CV slope and intercept). Therefore the % HbA1c is the best parameter to evaluate the stability of the MEs. For the NGSP this relationship is very stable: at the 53 mmol/mol level, the calculated %HbA1c is 7.00% whether ME4 or ME12 is used. There is also no trend (<0.001%/year), and all 12 studies are in compliance with the published ME. The same is true for the Mono-S relationship with the IFCC-RM. For the JDS/JSCC a borderline trend (Fig. 2) is seen that is also reflected by a difference of 0.05% between ME4 and ME12, a finding that will be investigated in future intercomparison studies.

VALUE ASSIGNMENT AND EXPANDED UNCERTAINTY

The assigned values and uncertainty of specimens derives from (i) the uncertainty in the calibrator sets used to calibrate the IFCC-RM, (ii) the uncertainty due to the measurement error of the reference method, (iii) the number of network laboratories involved in the value assignment, and (iv) the number of assays performed by each network laboratory. When values are assigned with the whole network, an expanded uncertainty of 1.0 mmol/mol (0.9% CV) can be achieved in the middle range of HbA_{1c} concentrations, which is acceptable in view of the performance goals for routine laboratories as discussed above.

LONG-TERM TREND, BIAS, AND REPRODUCIBILITY RESULTS FOR NETWORK LABORATORIES

A single intercomparison study is the forum for approval of network laboratories at a given point of time. Evaluation of multiple intercomparison studies discloses small phenomena and trends over time. From Table 3 it can be seen that none of the laboratories has shown a trend over time, that 2 laboratories have a consistent low or high bias, and that 1 laboratory has high variation, indicating lack of reproducibility. From the bottom lines of Table 3, it can be seen that there is no difference in HbA_{1c} outcome between laboratories that use MS vs CE methods (+0.1 mmol/mol vs 0.0 mmol/mol) but that the MS-group has a significantly higher variation (0.7 mmol/mol vs 0.4 mmol/mol). The difference in performance between MS and CE might be explained by nonoptimal HPLC circumstances for the MS method, and a modification leading to improvement *(22)* is under investigation, to be implemented in the IFCC-RM.

REPRODUCIBILITY AND STABILITY OF CALIBRATORS AND CONTROLS

Long-term reproducibility is the cornerstone of the management of the network. Calibrators and controls play key-roles, and their reproducibility and stability are systematically monitored.

After manufacture, a batch of calibrators is measured as a sample by the whole network and the mean measured concentration HbA_{1c} is compared with the HbA_{1c} concentration calculated from the weighed amounts of pure HbA_{1c} and HbA_0. If the difference exceeds the criterion, the calibrator set is rejected. This situation occurred once in our studies. When a batch of calibrators meets the criterion it is approved and used as the calibrator set in the next intercomparison study (1 year after manufacture). Two or 3 years after manufacture, spares are included in an intercomparison study as "old calibrators" to check their long-term stability. Throughout the 12 studies we report, calibrators always proved to be stable.

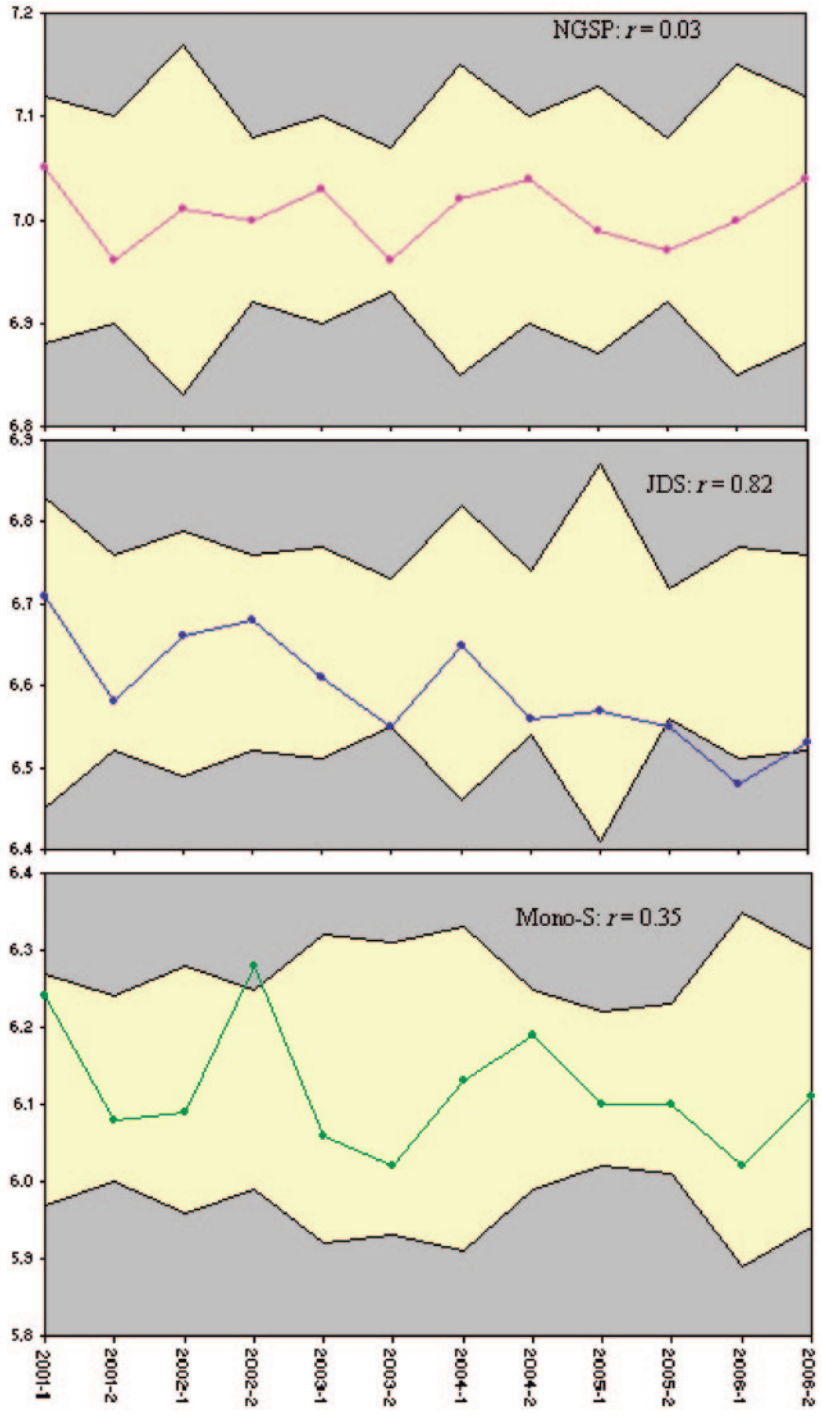

Fig. 2. Trend relationship between IFCC-RM and DCM at 53 mmol/mol HbA_{1c} concentration HbA_{1c} values as calculated from the relation between IFCC-RM and DCM in each of the respective 12 intercomparison studies are plotted with individual studies on the *x* axis and the HbA_{1c} value (expressed as the percentage HbA_{1c} calculated for the respective DCMs at the 53 mmol/mol IFCC-RM HbA_{1c} concentration) on the *y* axis.

The middle of the window of the Shewhart chart is the percentage HbA_{1c} calculated from the ME *(14)*, and the limits are variable ($P < 0.05$; for explanation see the "Statistics" section in "Materials and Methods").

Spare samples of intercomparison studies are stored at −84 °C and after 1 and 2 years of storage are systematically included as controls in intercomparison studies. This protocol allows comparison 0, 1, and 2 years after manufacture and is a parameter for stability. As can be seen from Table 1, changes in the 12 batches of controls have been negligible. The maximum allowable difference was exceeded in only 1 of 48 studies (1.0 mmol/mol vs 0.7 mmol/mol). The data indicate that this type of material can be (and is) used as long-term quality control. A possible objection to this conclusion is that the trend in laboratories may have been compensated by the instability in the controls, but this theory does not hold given that 12 batches showed the same performance.

In conclusion, the results of the 12 intercomparison studies performed during 2001–2006 reported here confirm the robustness of this system to guarantee stability and continuity of the analytical reference method for HbA_{1c}. The results also demonstrate that the concept of a network of reference laboratories as the foundation to develop, implement, and maintain a reference system is very effective and efficient.

Grant/funding Support: None declared.
Financial Disclosures: None declared.
Acknowledgments: We thank the members of the IFCC Scientific Division for reviewing the manuscript before submission. We also thank the following persons of the network laboratories: Masao Umemoto (Institute of Biopathological Medicine, Kanagawa, Japan), Izumi Takei (School of Medicine Keio University, Tokyo, Japan), Maria Ospina (CDC, Atlanta, USA), Yuanfang Deng (Siemens Medical Solution Diagnostics, Norwood, USA), Veronica Luzzi (Washington University, St. Louis, USA), Jean Bucksa (Fairview University, Minneapolis, USA), Uwe Kobold (Roche Diagnostics, Penzberg, Germany), Renata Paleari and Donatella Caruso (Universita degli Studi di Milano, Milano, Italy), Erna Lenters and Janine Slootstra (Isala Klinieken, Zwolle, The Netherlands), Willeke Dekkers and Jeffrey Sigger (Queen Beatrix Hospital, Winterswijk, The Netherlands), Franciscus Susanto (Deutsches Diabetes Zentrum, Duesseldorf, Germany), Patricia Kaiser (INSTAND, Duesseldorf, Germany).

References

1. The Diabetes Control and Complications Trial 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 1993;329:977–86.
2. ADA. Clinical practice guidelines 2007. Diabetes Care 2007;30(Suppl 1):S3.
3. Little RR, Rohlfing CL, Wiedmeyer HM, Myers GL, Sacks DB, Goldstein DE. The national Glycohemoglobin Standardization Program: a five-year progress report. Clin Chem 2001;47:1985–92.
4. Shima K, Endo J, Oimomi M, Oshima I, Omori Y, Katayama Y, et al. Inter-laboratory difference in HbA1c measurement in Japan: a report of the Committee on an Inter-laboratory Standardization of HbA1c Determination, the Japan Diabetes Society. J JPN Diabetes Soc 1994;37:855–64.
5. Arnquist H, Wallensteen M, Jeppsson JO. Standardization of long-term glucose measurements established. Lakartidningen 1997;50:4789–90.
6. Mueller MM. Implementation of reference systems in laboratory medicine. Clin Chem 2000;46: 1907–9.
7. Panteghini M, Forest JC. Standardization in laboratory medicine: new challenges [review]. Clin Chim Acta 2005;355:1–12.
8. International Organization for Standardization. In vitro diagnostic medical devices-measurement of quantities in samples of biological origin-metrological traceability of values assigned to calibrators and control material. ISO 17511. Geneva, Switzerland: ISO, 2003.
9. Directive 98/79/EC of the Eurean Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. Off J L 1998;331: 1–37.
10. Hoelzel W, Miedema K. Development of a reference system for the international standardisation of HbA1c/glycohemoglobin determinations. J Int Fed Clin Chem 1996;9:62–7.
11. Finke A, Kobold U, Hoelzel W, Weykamp C, Jeppsson JO, Miedema K. Preparation of a candidate primary reference material for the international standardisation of HbA1c determinations. Clin Chem Lab Med 1998;36:299–308.
12. Kobold U, Jeppsson JO, Duelffer T, Finke A, Hoelzel W, Miedema K. Candidate reference methods for HbA1c based on peptide mapping. Clin Chem 1997;43:1944–51.
13. Jeppsson JO, Kobold U, Barr J, Finke A, Hoelzel W, Hoshino T, et al. Approved IFCC Reference Method for the measurement of HbA1c in human blood. Clin Chem Lab Med 2002;40:78–89.
14. Hoelzel W, Weykamp C, Jeppsson JO, Miedema K, Barr J, Goodall I, et al. IFCC Reference System for measurement of hemoglobin A1c in human blood and the National Standardization Schemes in the United States, Japan, and Sweden: a method-comparison study. Clin Chem 2004;50:166–74.
15. Consensus Statement on the Worldwide Standardization of the Hemoglobin A1c Measurement. American Diabetes Association, Eurean Association for the Study of Diabetes, International Federation of Clinical Chemistry and Laboratory Medicine and International Diabetes Federation Consensus Committee. Diabetes Care 2007;30: 2399–400.
16. Hoshino T, Nakayama T, Kuwa K, Nakanishi T, Okahashi M, Tominaga M, et al. Reference method for St-GHbA1c determination: standard operating procedure, Ver. 1.4. Tokyo, Japan: Japanese Society of Clinical Chemistry Working Group on SOP for St-GHbA1c Determination, 2000.
17. Goldstein DE, Little RR, England JD, Wiedmeyer HM, McKenzie EM. Methods for quantitating glycosylated hemoglobins: high performance liquid chromatography and thiobarbituric acid colorimetry: In: Clarke WL, Larner, Pohl S, eds. Methods in diabetic research. Vol 2. Clinical methods. New York: John Wiley, 1986:475–504.
18. Konnert A, Arends S, Schubert S, Berding C, Weykamp C, Siebelder C. Uncertainty calculation for calibrators of the IFCC HbA1c Standardization Network. Accred Qual Assur 2006;11:319–28.
19. Konnert A, Berding C, Arends S, Parvin C, Rohlfing C, Wiedmeyer H, Little R, Siebelder C, Weykamp C. Statistical rules for laboratory networks. J Test Eval 2006;34:128–34.
20. Stöckl D, Reinauer H. Development of criteria for the evaluation of reference method values. Scand J Clin Invest 1993;53 Suppl 212:16–8.
21. Goodall I, Colman P, Schneider H, McLean M, Barber G. Desirable performance standards for HbA1c analysis: precision, accuracy and standardisation. Clin Chem Lab Med 2007;45(8): 1083–97.
22. Kaiser P, Akerboom T, Dux L, Reinauer H. Modification of the IFCC reference measurement procedure for determination of HbA1c by HPLC-ESI-MS. GMS Ger Med Sci 2006;4:Doc06.