



Minutes of the NGSP/IFCC Manufacturer Forum

Monday, July 25, 2011 10:00AM-12:00PM
Marriott Marquis, Atlanta, GA

Presenters:

David Sacks —Chair, NGSP Steering Committee
Randie Little—NGSP Network Coordinator
Cas Weykamp—IFCC Working Group Network Coordinator
Carol Benson—U.S. Food and Drug Administration

Present were members of the NGSP Steering Committee and representatives from various manufacturers, laboratories and agencies.

1. Welcome and Introduction— David Sacks, Chair, NGSP Steering Committee

- D. Sacks welcomed those in attendance on behalf of the NGSP and IFCC.

2. NGSP Progress Report—Randie Little, NGSP Network Coordinator

- The NGSP is overseen by a Steering Committee and an administrative core, and a laboratory network consisting of Primary and Secondary Reference Laboratories in the U.S. and Europe.
- The network is tied to the IFCC network via sample comparisons performed 2X/year.
- The NGSP assists manufacturers with calibrating their assays, has a formal certification process and monitors performance of HbA_{1c} testing in the field via the CAP survey.
- Status of HbA_{1c} measurement
 - The numbers of certified methods and laboratories continues to increase.
 - Most certified laboratories are level 1 laboratories, largely due to requirements of pharmaceutical companies performing clinical trials, and are located outside the U.S.
 - There has been much improvement in the comparability of HbA_{1c} results since 1993.
 - CAP GH2 survey 2011A:
 - Three HbA_{1c} levels: low (5.4%), middle (6.4%) and high (8.5%).
 - CAP acceptable limits are now +/-7% of the NGSP target value.
 - The method-specific means were all within 0.35 at all levels. Only one method showed a bias >0.3% HbA_{1c}. 23/30 method groups showed mean biases of <0.3% HbA_{1c} at all three levels.
 - Method-specific, between-laboratory CV's ranged from 1.4% to 7.2%. All but 3 methods had CVs below 5% for all 3 HbA_{1c} levels.
 1. Method-specific, between-lab CVs ranged from 1.4 to 6.0%.
 2. However, ~93% of labs used methods with CVs of <5% at all HbA_{1c} levels.
 - There has been improvement in the all-method CVs over time. In 2000 the CVs were ~5%, on the 2011A survey the CVs were ≤3.5% at all three levels.
 - Method-specific between-lab CVs
 1. CVs were ≤3.5% for the ion-exchange and boronate affinity methods, many of the immunoassay methods also had CVs ≤3.5%
 2. The POC methods on the survey did not necessarily show worse performance than the laboratory methods.
 3. The methods with the worst CVs are used by a very small number of labs.
 - Pass rates by method
 1. Most methods had pass rates ≥90%.
 2. Pass rates for many of the immunoassay methods were comparable to those for the HPLC methods.
 3. Again, POC methods did not necessarily show lower pass rates.
 - Tightening the NGSP criteria

- Current criteria: 95%CI of differences must be within $\pm 0.75\%$ HbA1c
- The new criteria will be based on $\pm 7\%$ (CAP)
 - The details are yet to be finalized.
 - More information will be sent to each manufacturer prior to implementation.

3. CAP Grading, future plans—David Sacks, Chair, NGSP Steering Committee

- In the past CAP used peer group grading for HbA1c.
- In 2007 CAP began using accuracy grading with acceptable limits of $\pm 15\%$ of the NGSP target. This resulted in a 99% pass rate.
- The acceptable limit was reduced to $\pm 12\%$ in 2008, then $\pm 10\%$ in 2009 and $\pm 8\%$ in 2010.
- The plan was to further tighten the limit to $\pm 6\%$ in 2011. However, CAP was concerned that given the current technology this change would fail too many labs and decided to tighten the acceptable limit to $\pm 7\%$ for 2011 and 2012 instead of the originally proposed $\pm 6\%$.
- For the GH2 2010A survey the overall pass rates at the three HbA1c levels were 95.5%, 95.4% and 95.2% at the limit of $\pm 8\%$. Using the projected limit of $\pm 6\%$, the respective pass rates would be 91.0%, 91.6% and 88.6%.
- In the 2011A survey, the pass rates were 95.1%, 92.8% and 95.2% at the three HbA1c levels using the current CAP acceptable limit of $\pm 7\%$. Using a limit of $\pm 6\%$, the pass rates would be 95.1%, 92.8% and 92.4%.
- The grading criteria will remain at $\pm 7\%$ for 2011 and 2012.
- In 2012 the data will be evaluated to determine whether the acceptable limit will be further tightened to $\pm 6\%$ for 2013.

Discussion:

Regarding the plan to further tighten the criteria, are we approaching the limit of our abilities in terms of random error?

D. Sacks said we will eventually reach a limit where the criteria will not be tightened further. The question is how accurate do we need to be? Talking to clinicians, most would say that a change of 0.5% HbA1c suggests a change in therapy; some of the guidelines specify this as well. The goal would be to get all assays to discriminate between two values 0.5% HbA1c apart. These numbers are based on management of patients; now that HbA1c has been recommended for diagnosis one could argue that 0.5% HbA1c might be a bit wide. Realistically there are limits as to how accurately HbA1c can be measured, but we need to continue efforts to make it more accurate.

Will there ever be a difference between a monitoring claim and a diagnostic claim?

D. Sacks said CB will discuss the issue of diagnostic claims later. The accuracy required for monitoring is probably not substantially lower than for diagnosis, clinically there may not be much distinction but from a regulatory standpoint it is an issue.

Are there issues with pre-analytical error that should be of concern?

D. Sacks said that HbA1c is much more stable than analytes such as glucose. Most assays have eliminated significant pre-analytical issues; of course there can be issues with patients that have variant hemoglobins, renal failure, etc. RL said that sample stability has improved overall with the newer assays.

Looking at 0.5% HbA1c as a clinically significant change, this difference at the diagnostic cutoff of 6.5% HbA1c represents a difference of 7.7%. The CAP limit is already $\pm 7\%$, are we getting beyond medical necessity?

D. Sacks said there was a paper published in Clinical Chemistry earlier this year which looked at whether methods can currently differentiate differences of 0.5% HbA1c, most can but not all. In any case the 0.5% is not evidence based, and if we can be more accurate we should.

The CAP data showed by RL seems to put some of the POC methods on a level playing field with lab methods, how might this affect inclusion or exclusion of methods in terms of use for diagnosis?

DS said that one of the big problems with POC testing is that the federal government decided that these methods could be waived and thereby do not have to participate in proficiency testing (PT). While the devices may perform well in the hands of the end users that do participate in the survey (5% or less of users), we have no idea how they are performing in the hands of users that do not participate. D. Simmons (Bayer Diabetes Care) said that there are two barriers: the lack of PT and the barrier of approval by the regulatory authorities who could see if the method has passed PT and use this as a criteria for use in diagnosis. Right now a POC method could pass PT and demonstrate excellent performance but the recommendations are such that it cannot be used for diagnosis regardless, where is the equality? DS said he cannot comment on regulatory issues, this is a matter for the FDA. If POC methods participate in PT and demonstrate adequate performance perhaps this should be reconsidered, this is just a personal opinion. The caveat is that everyone planning to use the device for diagnosis should be required to participate in PT.

4. IFCC Reference System—Cas Weykamp, IFCC Network Coordinator

- The mission of the IFCC Network
 - Warrant Continuity of the IFCC Reference Measurement Procedure (IFCC-RMP)
 - Make HbA1c assays worldwide traceable to the IFCC-RMP
- The IFCC Network: Internal Checks
 - Approval of Network laboratories
 - Twice yearly a panel of blind samples that are assayed by the network and candidate laboratories.
 - Statistical analyses are performed to determine if laboratories fall within an acceptance ellipse based on systematic and proportional bias, any labs that do not are not approved.
 - Primary calibrators for the Network
 - Prepared from pure HbA1c and HbA0.
 - They are checked over time by including them periodically in subsequent network studies.
 - Monitoring of the master equations
 - The ME between the IFCC and NGSP networks has remained stable over 10 years.
 - The ME between the IFCC and JDS/JSCC has shown a slight deviation over time in the upper end of the HbA1c range.
 1. The likely cause appears to be double glycation (HbA1c plus an additional labile glucose attached) which is more prevalent at higher HbA1c levels.
 2. The JDS/JSCC method samples must be washed and dialyzed prior to analysis to remove the labile.
 3. We plan to re-establish the ME between the IFCC and JDS/JSCC networks using 30 samples to be analyzed by the JDS/JSCC system and five IFCC network laboratories.
- The IFCC Network in Asia, Europe, U.S.
 - There are currently 3 approved labs in the U.S., 5 in Europe and 3 in Japan. Recently an additional laboratory in France has been approved.
 - There are three candidate labs, one each in Korea, China and India.
- “Apples and Pears”: CV calculations for IFCC vs. NGSP
 - Hypothesis: “The variation in the Temperature of the Human Body is much higher in ordinary Americans than in Scientists”
 - CV in ordinary Americans: 1.8%. CV in scientists: 0.3%
 - It turns out the actual variation (SD) is the same, the difference is that the scientists measure temperature using the Kelvin scale while ordinary Americans use Fahrenheit, resulting in very different CVs even though the actual variation is the same.
 - This also applies to HbA1c when using different number scales.
 - For example, for a SD of 0.4% NGSP which corresponds to 4 mmol/mol on the IFCC scale, the CVs for the NGSP and IFCC numbers at the ADA diagnostic cutoff of 6.5% NGSP (48 mmol/mol IFCC) are 6.2% and 8.3%, respectively.
 - CVs can be converted between the systems using the formula: $CV_{\text{NGSP}} = CV_{\text{IFCC}}(\text{HbA1c}_{\text{IFCC}}/\text{HbA1c}_{\text{IFCC}} + 23.5)$.
 - This difference has implications in terms of goals and quality specifications for assay performance as well as calculations involving biological variation.

- For more information: Weykamp C, Mosca A, Gillery P, Panteghini M. The analytical Goals for Hemoglobin A1c Measurement in IFCC Units and NGSP Units are Different. Clin Chem 2011 May 13 (Epub ahead of print).
- Services for Manufacturers
 - Calibrators to achieve traceability
 - Provided with HbA1c results in IFCC (mmol/mol Hb) and DCCT (%) units, also mmol/L and g/dL.
 - Provided with total hemoglobin in mmol/L and g/dL.
 - All are provided with expanded uncertainties.
 - Controls to check traceability
 - Low, middle and high levels
 - Middle level is provided with low, normal and elevated total hemoglobin
 - Units provided
 1. HbA1c and Total Hb
 2. IFCC- NGSP Units
 3. mmol/mol, %, mmol/L, g/dL HbA1c
 4. mmol and g/dL Total Hb
 - All are provided with expanded uncertainties
 - Monitoring to prove traceability
 - 24 frozen whole blood samples per year, one sample is to be analyzed every-other week
 - Once a year mean deviations from the targets, imprecision and linearity are calculated
 - Certificates of traceability are provided
 - Many manufacturers have been participating in the program for many years.
- IFCC Units in Patient Reports
 - The U.S. has chosen to stay with NGSP units, many countries in Europe are switching to IFCC units; the situation is unclear in some Asian countries and South America.
 - Garry John is the chair of the IFCC integrated project and can answer any questions regarding this.

Discussion:

Is the instability of the ME at the high end with the JDS/JSCC something we should be concerned about in other samples?

CW said this is strictly an issue with the KO500 reference method in Japan, as far as we know the double glycation is not an issue in routine assay methods. The main effect is that we have some uncertainty in the relationship between the IFCC and JDS/JSCC networks at high HbA1c levels; this is why we need to recalculate the equation.

Regarding the use of IFCC units, we should target clinicians if we want them to adopt IFCC numbers. Manufacturers can provide either % or mmol/mol; they will simply provide what is demanded.

CW said that it has proven impossible to get clinicians in the U.S. to use IFCC units.

What about the use of eAG, what are your thoughts and will it be used outside of the U.S.

CW said that his personal opinion it is a different thing that should not be reported along with or instead of HbA1c. However, for educational purposes it might have some use. In the U.S. the idea was popular a few years ago but is less so now, outside of the U.S. it will not be on reports but some may use it for educational purposes. DS said some clinicians in the U.S. find it useful while others do not, the important thing is to use the correct equation. For the last 3 years the CAP has sent out a questionnaire along with the survey asking whether the lab is reporting eAG and which equation is being used. The first year ~20-25% of labs reporting eAG were using the correct equation, on the last survey it was ~50%.

Are there any ongoing efforts to go back and repeat the average glucose study looking at ethnic groups, children, etc. that were not represented in the original study?

CW said that from the perspective of the IFCC network, for any new clinical studies HbA1c assays should be anchored to the IFCC network. GJ said that in Europe it was decided that the study did not sufficiently

define the HbA1c/glucose link in any case. It may be useful as an educational tool with patients but there were groups that were not represented, also there is much uncertainty in the relationship. As to whether the study will be repeated it is doubtful, this would be very costly. D. Simmons noted that the original study was funded by the ADA; they used the results of this study to define the HbA1c/glucose relationship and no longer use the relationship from the DCCT study. In terms of the ADA there is currently not as much momentum for the use of eAG as there was several years ago. GJ said that the scientific basis of demonstrating the HbA1c/glucose relationship was scientifically valid; it was the translation to the clinical use of this that was questionable.

In terms of globalization of the IFCC network there is now a candidate lab in India, what efforts are being made to draw in more labs to participate?

CW said that since we now have a lab in India, we will try to work through them to develop contacts and communicate through diabetes congresses, etc. what the IFCC has done and what the IFCC Integrated Project wants to achieve. That is, to achieve global standardization using the IFCC Reference System as the anchor. We will make similar efforts in China and other areas. The issue is finding the proper contacts in these regions. GJ added that there is a big difference between globalization of HbA1c standardization and globalization of laboratories in the network; we have to balance between globalization of the network with how many labs are actually needed. In terms of reporting, the main issue is that in countries as big as India and China we try to avoid the use of different units within a country.

What is happening with the IFCC Integrated Project?

GJ said that it has been slow to start for different reasons, the first meeting was a few months ago. We need to look at what is happening around the world, until we know this we cannot move forward. We will be sending a survey out to international societies to find out what is being reported, what kind of quality assessment is being done and whether laboratories are even standardized in different countries. We will also be publishing a response to a recent article in which it was asserted that DCCT numbers are not traceable to the IFCC. This is not correct, they are linked by the networks, there is some confusion regarding this. Over the next 12-18 months there will be a lot more activity on the part of the Integrated Project.

5. HbA1c for Diagnostic Claim: FDA Update—Carol Benson, FDA

- The ADA recommendation to use HbA1c for diagnosis added HbA1c as a means to diagnose diabetes using assay performed in a laboratory and certified by NGSP and standardized to the DCCT assay.
- Intended use
 - Current assays: For the quantitative measurement of HbA1c in whole blood. The measurements are used to monitor the long term glycemic control of individuals with diabetes.
 - Diagnosis
 - Represents an intended use change
 - Manufacturers cannot promote use of their device for diagnosis until they receive clearance for the new intended use.
- FDA Outreach to Manufacturers
 - Sent out letter to manufacturers with HbA1c assays listed with the FDA in 2010 advising them to contact the FDA.
 - We wish to have communication with the manufacturer regarding protocols and study plans using the “pre-ide” process.
- Why use the pre-ide process?
 - Allows open communication between FDA and manufacturers
 - Different approaches for study design and data analysis
 - Percent or absolute value?
 - Sample in duplicate or singlicate?
 - How many samples to test? CI?
- How is FDA guidance developed?
 - FDA uses good guidance practices
 - Draft open for public comment

- Address comments - finalize
- Guidance is not the law
- How much accuracy is needed?
 - Needs to meet the clinical needs of the test for the intended use
 - Criteria for accuracy is not well defined
 - FDA does not have specific criteria for accuracy
 - Should be tighter than the current NGSP criteria
 - Bias at 6.0%, 6.5% and 7.0% should be low
 - Imprecision should be low
 - Interferences – little to none especially with hemoglobinopathies
- Laboratory and POC devices
 - POC and lab devices have NGSP certification at time of marketing
 - Some POC devices have similar performance to laboratory devices
 - NGSP certification renewed yearly? This is not something FDA can enforce.
 - Some POC devices are waived (no PT) - some are not waived
- FDA actions
 - Continuing to work with manufacturers to have the claim for diagnosis.
 - We believe that obtaining the claim is possible with additional analytical data showing assays are accurate and reliable for this new claim
 - Regulatory processes to clear new intended use – One approach is the de novo which we have experience with and used successfully
 - Want manufacturers to talk to us about the new intended use

Discussion:

How many manufacturers have contacted the FDA regarding the claim?

CB responded that she cannot divulge this information.

Does this process make a difference in clinical practice? Physicians have been using HbA1c to diagnose diabetes for a long time despite there being no diagnostic claims on the package inserts.

CB said having a claim on the package insert does matter to manufacturers; they want to promote the device for this use. C. Harper (FDA) added that some labs are knowledgeable about how HbA1c assays work in terms of performance but others are not, for them the additional information provided by a claim for use in diagnosis may be useful. Providing additional information regarding the performance around the diagnostic cutoff can help these laboratories to choose the appropriate platform for their needs and understand the limitations of their assay methods. We know that physicians are already using the test for diagnosis, from FDA's perspective it has more to do with providing additional information to help in selecting labs/methods that are appropriate for the intended use. It was noted that the physicians do not pay attention to package inserts, they use the test however they wish to, and now they have a guideline that tells them they can use it for diagnosis. However, a diagnostic claim provides an advantage to the physician and the patient in terms of manufacturer support when the test is used for diagnosis. Right now we are very limited if there are questions/problems, we cannot provide educational materials, etc.

Is there openness to discussing the possibility of including POC systems that demonstrate good performance?

CB responded that FDA is open to all of the possibilities that exist. We still have the ADA recommendation that POC not be used; this would need to be addressed.

The reality is that in a large hospital setting, the decision regarding what platform to use for HbA1c is not dictated by information on package inserts; it is mostly dictated by whatever platform is being used for other testing. Also, what will the FDA use as the gold standard to determine if the test can be used for diagnosis? For example, if you use fasting glucose only about 80% of individuals that are above normal will have a HbA1c level $\geq 6.5\%$.

CH said that unfortunately there are some bad HbA1c assays still out there, that is part of the issue. The major platform manufacturers may not have trouble with the diagnostic claim with their current methods; the issue with distinguishing methods is that we have laws that limit our ability to remove methods from

the market once they have been cleared. Thus, some methods that were cleared years ago when overall performance was not as good are still out there. The additional information on the label may help some physicians and others to make the choice as to whether a method meets their needs. Under current law a manufacturer might claim substantial equivalence to an old assay method from the 1990s and there would be little the FDA can do to keep it off the market. However, we may be able set up a system that allows distinguishing the methods that perform the best. We are not planning to redo the DCCT trial; we will simply be looking at CV and bias around the diagnostic cutoff. Right now we are not doing that, we want to differentiate how the tests look around 6.5% in addition to the entire range.

Is the FDA changing the way they are evaluating new methods for the monitoring claim? It seems that it is more difficult to obtain a monitoring claim than in the past.

CB said that for the monitoring claim they do not look specifically at levels around 6 or 7%, they look at overall performance in terms of precision and bias. If manufacturers have issues with obtaining a monitoring claim they need to discuss it with us.

Although you have issued an open invitation to all manufacturers, you have referred to the issue of the best performers getting the diagnosis claim. How do you see then fitting into studies relative to the de novo process in terms of how you see the diagnosis claim appearing for the first time?

CB said the FDA is open to all manufacturers, that is why we use the pre-ide process. We do not pick and choose which manufacturers can talk to us. The FDA evaluates what is sent to them. There are many mechanisms in terms of how we evaluate the information. We can do so internally, or if we think we do not know the answers we have advisors that are special government employees, we can also have an open panel meeting to have public discussion for comment as to whether performance is adequate.

Do you have any additional guidance regarding the study design, etc. since last year? Also, what is the incentive for manufacturers if the process is perceived to be long and difficult?

CB responded that the FDA believes in general that the criteria need to be tighter than NGSP. We want to see the bias around cutpoints and if there is interference from hemoglobinopathies, etc. We need to look at the consequences if there is a false negative or false positive. CH added that there have been no decisions made as to what the bar for acceptable performance is. The FDA would like to talk to manufacturers regarding what might be needed to show adequate performance, we feel that many platforms can demonstrate adequate performance. We sense frustration on the part of manufacturers, but not many have actually come to talk to us. We would encourage manufacturers to come and talk to us regarding how performance for diagnosis can be established. There is not a cutoff for how many methods can obtain the claim; we are not restricting it to a certain percentage of methods. There are a lot of rumors out there, it is best if you talk to us directly. There is no requirement that manufacturers seek a diagnostic claim if they are not wishing to promote it for that use. However, we would like to see manufacturers of robust methods seek the claim.

Last year there was an offer from opinion leaders in the field to work with the FDA in developing a protocol that makes sense. This would make for a level playing field for all manufacturers. It seems that this would be a better approach as opposed to requiring each manufacture to go to the FDA and "re-invent the wheel".

CB said we discussed a workshop; it must be open to the public so that any interested parties can attend. The FDA does not have the resources to do this, maybe we could get the AACC or some other party to sponsor it. CH said that the FDA is open to proposals, regardless of whether we speak to manufacturers as a group or separate, it will be a level playing field for manufacturers. We are open to suggestions. CB said a proposal can be in the form of a white paper. DS asked if the FDA would be open to a workshop sponsored by the manufacturers, CB said this would be fine as long as it is open to everyone, not just the manufacturers that agree to support the workshop. It must be open to the public. RL asked how we go about this. CB responded that the FDA has a list of all the people that have devices registered with them, also we can have an announcement in the Federal Register notice or announce it at professional conferences, we can also post an announcement letter on the FDA web site. DS asked if FDA would be willing to do this if manufacturers supplied the funding for the workshop, CH said they would be willing to do this, they would just need to check on conflict of interest issues regarding the funding.

Currently, monitoring is cleared but it is in the indications section. How did it get there, what scrutiny was used, and what is the difference between intended use and indications for use? Why can't a manufacturer include diagnosis in the summary of explanations, which many indications have gotten into without scrutiny?

CB said that going back to the regulation of medical devices in 1976, there is a regulation for HbA1c that specifies a claim for monitoring. For 510K the requirement is showing substantial equivalence to a predicate that has the same intended use with similar performance. So, all of the decisions now for obtaining the monitoring claim are based upon obtaining a 510K showing substantial equivalence to a predicate device of their choice. Diagnosis is part of the general intended use, you cannot state that the intended use is for monitoring then say it can be used for diagnosis in the summary, that is an implied claim to us. If you are changing the intended use you cannot just submit a 510K, we must have a process to classify the new intended use.

When you refer to study design, this does not involve additional clinical studies?

CB said that is correct, the FDA is not asking for a re-evaluation of the cutpoint. We are accepting that the cutpoint that has been accepted is correct. We are asking for more robust studies to show performance around the cutpoint and also show the influence of interferences at that level.

DS said that the NGSP is willing to facilitate an open workshop with the FDA if manufacturers are interested. The manufacturers would need to support this; you can contact either DS or RL regarding this. He thanked everyone for their attendance and efforts to improve HbA1c testing. The meeting was adjourned at 12:00PM.

Minutes prepared by Curt Rohlfing 8/16/11. Modified by Randie Little 8/18/11.