Standardization of Insulin Immunoassays: Report of the American Diabetes Association Workgroup

Santica Marcovina, Ronald R. Bowsher, W. Greg Miller, Myrlene Staten, Gary Myers, Samuel P. Caudill, Scott E. Campbell, and Michael W. Steffes

Background: Circulating insulin concentration in serum or plasma provides important information for the estimation of insulin secretion and insulin resistance. Currently, lack of standardization of insulin assays hinders efforts to achieve consistent measures for treatment guidelines.

Methods: A Workgroup convened by the American Diabetes Association evaluated 12 different commercial insulin methods from 9 manufacturers.

Results: The within-assay CVs ranged from 3.7% to 39.0%, with 7 of 10 assays having a CV ≤10.6%. The among-assay CVs ranged from 12% to 66%, with a median value of 24%. A common insulin reference preparation did not change the among-assay CV and failed to improve harmonization of results among assays. Results from 6 of 10 assays agreed within the total error of 32% that is allowable based on biological variability criteria. Seven of 10 assays recovered insulin added to a serum pool within 15.5% of the expected concentration. In 9 of 10 methods, there was <2% cross-reactivity with intact human proinsulin, and 8 of 10 methods had <3% cross-reactivity with split (32, 33) proinsulin. For 9 of 10 assays, the cross-reactivity of des (64, 65) proinsulin exceeded 40%. Overall, most assays had acceptable imprecision and specificity for insulin.

Conclusion: The discordance in test results for commercial insulin reagent sets is likely multifactorial and will require a continuing effort to understand the differences and achieve the desired consistency and harmonization among commercial immunoassays.

The prevalence of diabetes mellitus has increased markedly during the past decade, with projections for continued increase worldwide. Most type 2, and some type 1, patients may benefit therapeutically from more specific information regarding their insulin secretion, as well as insulin resistance. The current definitions of insulin resistance and hyperinsulinemia are problematic, due largely to the fact that immunoassays for measuring circulating insulin have not been standardized across manufacturers and clinical laboratories. To date, this issue has limited our ability to make data comparisons for different clinical studies and to develop uniform clinical practice guidelines.

Highly standardized and specific assays for measuring circulating insulin are essential to achieve a consistent measure of insulin resistance and insulin secretion among different healthcare facilities worldwide. Despite the awareness of analytical differences (1) and the potential cross-reactivity of proinsulin and/or proinsulin-related peptides in insulin immunoassays, there is no consensus on the analytical requirements (quality specifications) for the assays. To address these issues, the American Diabetes Association (ADA) in conjunction with the National Institute of Diabetes and Digestive and Kidney Diseases

© 2007 American Association for Clinical Chemistry
and the CDC convened an international work group in 2004 to evaluate the specificity of different assays, to establish guidelines for assay acceptability, and to develop a standardization program to achieve uniform, accuracy-based values. In this paper, we report the findings and conclusions from 2 phases of this standardization effort.

Materials and Methods
A complete description of materials and methods, including statistical analysis, is available in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol53/issue4). A brief description is provided here.

The ADA Workgroup agreed on 2 major points: (a) the standardization effort would primarily target commercially available assays that were verified to measure insulin with negligible cross-reactivity with proinsulin and cleaved intermediate products of proinsulin conversion to insulin, and (b) insulin concentrations would be reported in Système International (SI) units (pmol/L) rather than the traditional units based on insulin biologic activity. Highly purified, recombinant human insulin provided by Novo Nordisk and Eli Lilly and Company with a potency of 6.0 nmol/IU or 28.7 IU/mg pure insulin was used as the basis for SI molar units of insulin, assuming a molecular mass of monomeric insulin of 5808 Da. A multiplication factor of 6.0 was used to convert concentrations in mU/L to pmol/L (1, 2).

STUDY DESIGN
The following manufacturers participated in at least 1 of the 2 phases of the investigation: Abbott, Japan; Bayer HealthCare, USA; Beckman Coulter Inc., USA; Dako Cytomation, Ltd., United Kingdom; Diagnostic Products Corporation, USA; Diagnostic Systems Laboratories, USA; Linco Research Inc., USA; Mercodia, Sweden; and Roche, Germany. An assay from Tosoh Bioscience Inc., Japan, was evaluated in-house at the University of Washington by one of the authors. Two manufacturers used more than one assay. The participating manufacturers requested that their respective performance outcomes remain anonymous; consequently, each assay was assigned an arbitrary number for reporting results.

For the first phase of the study, heparin plasma aliquots from 40 donors (verified to have received no exogenous insulin) were measured in duplicate on 3 separate days by each assay using the manufacturer’s product calibrators and a highly purified recombinant insulin preparation as a proposed reference material for insulin (PRMI). To prepare working calibrators the PRMI was diluted in each reagent set manufacturer’s recommended calibrator diluent.

For phase 2 of the study, the recoveries of known concentrations of insulin (PRMI) added to a low-insulin serum pool were measured by each assay. The specificity of each assay in phase 2 was assessed by determining the degree of cross-reactivity with known concentrations of intact human proinsulin (hPI) and the proinsulin intermediate metabolites, split (32, 33) hPI [B–C junctional cleaved metabolite] and des (64, 65) hPI des [A–C junctional cleaved metabolite] added to aliquots of the low-insulin serum pool that had been supplemented with PRMI to a final concentration of 240 pmol/L (40 mIU/L) to ensure that the insulin was at an easily measured concentration. In addition, each participant prepared dilutions of PRMI using the assay’s recommended diluent to evaluate parallelism of insulin recovery from the diluent series to the recovery from the series of serum pools to which insulin had been added.

Results
COMPARABILITY OF RESULTS AMONG ASSAYS
In phase 1, all the manufacturers’ product calibrators were reported to have assigned values traceable to the WHO 1st international reference preparation 66/304. One assay was excluded from the evaluation, because it was not compatible with plasma samples. The remaining methods claimed that plasma was an acceptable sample. Results for 6 and 10 samples were excluded for assays 4 and 8, respectively, because they were below the assay’s lower limits of quantification (12 and 18 pmol/L).

Fig. 1 shows the among-assay CV for each of the 40 patient samples (0.6 to 1008 pmol/L) first using manufacturer’s product calibrator and then PRMI as a common calibrator for 6 assays that were verified in phase 2 to be specific for insulin (see Table 4 below, 1–6). Results were excluded for assay 7 because it was not specific for insulin; for assays 8 and 9 because they did not participate in phase 2, which evaluated specificity for insulin; for assay 10 because it was not able to use PRMI as a calibrator; and for assays 11 and 12 because they did not participate in phase 1. As expected, the imprecision, when expressed as CV, was greatest at low insulin concentra-
tion. Use of PRMI as a common calibrator did not affect the overall imprecision. The volunteers’ test sample results were divided into 2 groups of 20, corresponding to concentrations below and above ~69 pmol/L to isolate the influence of differences in imprecision at lower and higher concentrations on the evaluation of the effect of PRMI as a common calibrator. For the lower concentration group, the among-assay SDs using manufacturer’s calibrators and PRMI as common calibrator were 11.3 and 9.9 pmol/L, respectively (F-test P = 0.270), and for the higher concentration group were 62.2 and 78.7 pmol/L, respectively (F-test P = 0.136).

Harmonization of results among the methods was evaluated by determining the number of results, measured using the manufacturer’s recommended calibration, that were within a total error allowance of ±25.6% from the mean value for all insulin-specific assays. The 25.6% criterion was obtained by adjusting the 32.0% total error criterion for a single reportable result (see Discussion) for the replication used in phase 1 by dividing the imprecision component in the equation by the square root of 3. Six of 10 assays (1, 2, 6, 7, 8, and 10) had >87% of patients’ sample results within the acceptable total error range. All results that were discrepant for this group had mean concentrations ≥27.0 pmol/L, except 1 result for assay 8 with a mean concentration of 68.4 pmol/L. The other 4 assays (5, 3, 4, and 9) had poorer agreement, with 5%, 55%, 62% and 77% of results, respectively, within the acceptable total error range.

### IMPRECISION OF INSULIN ASSAYS

Within-assay CVs for the results for the 40 patient samples when each assay was calibrated with the respective manufacturer’s product calibrators are shown in Table 1. The median value within each group was used as the denominator to calculate CV from the within-assay SD.

#### INSULIN RECOVERY AND PARALLELISM

In phase 2 of the evaluation, standard additions of the PRMI were made to a serum pool with low insulin concentration (40.2–65.4 pmol/L by the different assays). Linear regression was performed between the expected and recovered insulin concentrations for the 120 to 600 pmol/L data points for each assay in Fig. 2A to estimate a mean recovery. The 60 pmol/L data were not used in the regression due to the wider dispersion of results for most methods at this concentration. Seven assays recovered insulin within 15.5% of the added amounts, and 3 assays had recoveries of 64%–79% of the added amounts (Table 2).

Phase 2 also included each assay manufacturer making dilutions of PRMI with the assay’s recommended diluent to the same approximate concentrations used for product calibrators in that assay. The diluted PRMI specimens were then measured with that assay calibrated according to the manufacturer’s usual practice. Fig. 2B shows the recovery of the concentrations expected from the dilution ratios for each assay. The slopes for regression of the measured vs the expected insulin concentrations for the PRMI diluted with manufacturer’s recommended diluent (Fig. 2B) were compared with those for the PRMI added to a serum pool (Fig. 2A). For this comparison, only the PRMI dilutions that were in the same approximate concentration range as the samples with PRMI added to serum (60–600 pmol/L) were used. Table 3 shows the results for comparison of slopes. The slopes were significantly different for all assays (P < 0.001 to 0.050). Slopes with an apparently small magnitude difference yet having a strong statistical significance occurred because of very small standard errors for 1 or both of the slopes.

#### INFLUENCE OF PROTEASE INHIBITOR ADDITION ON INSULIN QUANTIFICATION

Because proinsulin and its cleaved metabolites have been reported to have limited stability in human serum (personal communication, Ronald Bowsher), the protease inhibitor Pefabloc SC (AEBSF) was added to the serum aliquots before supplementation to promote stability of the proinsulin-related peptides. One serum aliquot was not supplemented with AEBSF and served as a control for the impact of the protease inhibitor. The addition of AEBSF decreased results of 5 assays by 0%–5%, 3 others by 5%–10%, and 1 assay by 13.6%, but increased results for a final assay by 24.4% (from 218 to 272 pmol/L).

#### CROSS-REACTIVITIES OF INTACT PROINSULIN AND CLEAVED PROINSULIN METABOLITES

The cross-reactivities of intact proinsulin (hPI) and the proinsulin metabolites des (64, 65) hPI [A–C junctional cleaved metabolite] and split (32, 33) hPI [B–C junctional

### Table 1. Within-assay CVs. a

<table>
<thead>
<tr>
<th>Assay</th>
<th>&lt;69 pmol/L</th>
<th>&gt;69 pmol/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.0</td>
<td>9.5</td>
</tr>
<tr>
<td>2</td>
<td>7.4</td>
<td>10.4</td>
</tr>
<tr>
<td>3</td>
<td>4.4</td>
<td>3.7</td>
</tr>
<tr>
<td>4</td>
<td>11.2 b</td>
<td>12.3</td>
</tr>
<tr>
<td>5</td>
<td>27.9 c</td>
<td>16.5</td>
</tr>
<tr>
<td>6</td>
<td>4.2</td>
<td>6.7</td>
</tr>
<tr>
<td>7</td>
<td>21.5</td>
<td>6.2</td>
</tr>
<tr>
<td>8</td>
<td>39.0 d</td>
<td>16.8</td>
</tr>
<tr>
<td>9</td>
<td>5.7</td>
<td>7.3</td>
</tr>
<tr>
<td>10</td>
<td>5.7</td>
<td>8.7</td>
</tr>
</tbody>
</table>

* a Determined from the within assay SD from AVOVA on replicate results for patient samples, and the median value within each group. Two groups of n = 20, corresponding to concentrations below and above ~69 pmol/L, were used because the CVs were higher (as expected) at lower insulin concentrations.

b This assay had a lower measurement limit of 12 pmol/L, and results for 6 samples were not available in the low concentration range.

c This assay had a lower measurement limit of 12 pmol/L, and results for 5 samples had missing values for some replications in the 12–24 pmol/L range.

d This assay had a lower measurement limit of 18 pmol/L, and results for 10 samples were not available in the low concentration range.
cleaved metabolite] in each assay are summarized in Table 4. For 8 of 10 assays, intact hPI produced estimated cross-reactivities of no more than 2.3%. However, for assay 11, the cross-reactivity from intact hPI had an estimated potency \( \frac{1}{27} \) that of insulin. The extent of cross-reactivity could not be estimated exactly, as the assay’s response produced by the added proinsulin exceeded the upper limit of quantification.

The pattern and extent of cross-reactivity were similar for intact hPI and the split (32,33) hPI junctional metabolite in all assays but 1 (assay 7). Assay 7 differed from the others in that it displayed a high level of cross-reactivity from split (32, 33) hPI but negligible cross-reactivity with both intact hPI and des (64, 65) hPI. In all cases, except assay 7, the degree of cross-reactivity was greater for the des (64, 65) hPI junctional metabolite than split (32, 33) hPI. This finding occurred although the serum pool was supplemented with des (64, 65) hPI at one tenth the molar concentration.

### Table 2. Cross-reactivities of insulin-related peptides in insulin assays.  

<table>
<thead>
<tr>
<th>Assay</th>
<th>Intact hPI</th>
<th>Split (32, 33) hPI</th>
<th>Des (64, 65) hPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7</td>
<td>1.3</td>
<td>61.2</td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
<td>1.5</td>
<td>67.2</td>
</tr>
<tr>
<td>3</td>
<td>-1.8</td>
<td>-2.2</td>
<td>41.9</td>
</tr>
<tr>
<td>4</td>
<td>0.8</td>
<td>1.0</td>
<td>54.0</td>
</tr>
<tr>
<td>5</td>
<td>1.8</td>
<td>2.1</td>
<td>68.8</td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>0.3</td>
<td>42.2</td>
</tr>
<tr>
<td>7</td>
<td>0.2</td>
<td>98.1</td>
<td>2.5</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>0.0</td>
<td>41.2</td>
</tr>
<tr>
<td>11</td>
<td>&gt;27.0c</td>
<td>&gt;27.0c</td>
<td>20.5</td>
</tr>
<tr>
<td>12</td>
<td>2.3</td>
<td>2.6</td>
<td>39.8</td>
</tr>
</tbody>
</table>

*As described in Materials and Methods, portions of a serum pool were supplemented with 6000 pmol/L intact hPI, 6000 pmol/L split (32, 33) hPI, and 600 pmol/L Des (64, 65) hPI.

a Estimated percent cross-reactivity could not be estimated as the assay’s results after addition of intact hPI and split (32, 33) hPI exceeded the upper limit of quantification.
concentration of either intact hPI or split (32, 33) hPI. For 8 of 10 assays, the approximate extent of cross-reactivity for the des (64, 65) hPI junctional metabolite was ≅40%.

Discussion

QUALITY SPECIFICATIONS, PERFORMANCE CHARACTERISTICS, AND HARMONIZATION OF ASSAYS

An allowable total error for a single reportable insulin measurement can be based on the average biological variability of insulin concentrations and the generally accepted premise that a desirable measurement error should not cause more than ~12% error in the ability to categorize an individual result as belonging to a population for which a distribution-based reference interval has been established (3). The within-individual biological variability for insulin has been reported as 21.1% CV, and the within-group variability (for nondiseased individuals) has been reported as 58.3% CV (3). From these biological variability values for insulin, one can derive a desirable allowable measurement bias of ±15.5%, imprecision of 10.6% CV, and total analytical error of 32.0% for a single reportable result at concentrations within or near the normal reference interval, ~15–120 pmol/L (2.5–20 mIU/L), and presumably for values at other concentrations.

For 7 assays, recovery of added insulin was within 15.5% of the expected values, suggesting that assay calibration was acceptable and reasonably uniform over the concentration range evaluated. Three assays had lower recoveries (64%–77%), suggesting their calibration was not adequate for clinical comparisons of results with those from other methods. The biases of these assays were reasonably proportional over the concentration range and could be corrected by having the calibration traceable to an appropriate reference system (Fig. 2A).

Harmonization of results among assays can be evaluated from the patient samples measured in phase 1. Because no reference measurement procedure exists for insulin, we used the overall mean value from all insulin-specific assays to assess agreement among the commercial methods. The allowable total error derived from biological variability was used as the criterion for agreement of results with the mean value for each sample. Harmonization assessment identified 6 of 10 assays with acceptable agreement among results. Of the 8 assays that were used in both phases 1 and 2, 5 had concordant conclusions regarding harmonization and consistency of results, with 4 having acceptable performance and 1 having unacceptable performance. Two assays had acceptable bias based on recovery of insulin from a serum pool but had only 55% and 5% of the patient samples, respectively, that met the harmonization criterion. One assay had acceptable harmonization results but a 78% recovery of insulin added to a serum pool. The discrepant conclusions from the harmonization and recovery evaluations may be related to the inadequacy of a mean of assays as an appropriate target value for assay agreement. Investigation of a reference measurement procedure for insulin should be a priority to provide a metrologically appropriate basis to evaluate the accuracy of routine methods.

The among-assay CVs for the patient samples were the same when the assays were calibrated either with the manufacturer’s product calibrator or with PRMI diluted in the manufacturer’s recommended diluent. Because a common calibrator is expected to improve the agreement among results, this observation suggests the PRMI diluted in this manner was not commutable with native patient samples.

The suitability of diluting PRMI in the manufacturer’s recommended diluent was evaluated for parallelism of results compared with those for PRMI added to a serum pool. The slopes for recovered vs expected insulin for each type of sample were statistically different for all of the assays examined. For some assays, the manufacturer’s recommended diluent to prepare working concentrations of the PRMI caused a response that was clearly not parallel to that when the diluent was a serum pool. For other assays, the differences between slopes were small enough that the impact on clinical interpretation of results might be acceptable in relation to the desirable bias goal of ±15.5% described earlier. A potential limitation to this evaluation was that a single serum pool was used for the assessment, and an interfering substance could have been present that affected the methods differently. Another potential limitation was that for some methods, the concentration range of samples from dilution of the PRMI matched the concentrations of product calibrators rather than those of the PRMI added to the serum pool, producing fewer points to construct the slopes for comparison in the concentration range 60–600 pmol/L. However, the results are persuasive that PRMI of the formulation used in this study will not be effective to improve the standardization of insulin results. The finding regarding the lack of benefit from using a common reference standard agrees with the finding from a previous study that found little improvement in variation among research-grade insulin assays when a common standard was used for calibration (1), suggesting that commutability with clinical samples must be carefully validated for any proposed reference material.

SPECIFICITY OF METHODS FOR INSULIN

Differences in specificity for proinsulin-related peptides may represent an important source of intermethod variability for insulin assays. Nine of 10 assays had low cross-reactivity with intact proinsulin, and 8 of 10 had low cross-reactivity with split (32, 33) hPI. The assays with unacceptably high cross-reactivity for these peptides are not suitable for use when a specific insulin measurement is needed, because the circulating concentrations of these molecules may be sufficiently high to cause misinterpretation of the result, a situation that is likely to be problematic for samples from patients with insulin resistance who are known to have increased circulating concentrations of proinsulin and split (32, 33) hPI cleaved forms (4).
All assays, except 1, had high cross-reactivity with des (64, 65) hPI. This cross-reactivity is less important for clinical interpretation, however, because the circulating concentrations of this metabolite are usually low relative to intact and split (32, 33) hPI (5–7). It is clear from this study that a common strategy for commercial vendors to produce insulin-specific immunoassays would utilize antisera that recognize the free N-terminal region of the A-chain of insulin. This epitope is absent in proinsulin peptides that possess an intact A-C junctional region, such as intact proinsulin and split (32, 33) hPI.

In conclusion, we found that several commercial assays, but not all, measured insulin with acceptable imprecision and cross-reactivities. In agreement with the conclusions of the earlier ADA task force (1), our findings suggest that the source of discrepancies in results among commercial methods is likely multifactorial and not explainable by a single analytical performance characteristic. Furthermore, not all insulin assays have acceptable performance characteristics at concentrations as low as 12 pmol/L, as is needed for clinical use.

Improvement in harmonization of results will require an ongoing effort to achieve traceability to a reference system for measuring insulin. The work group is investigating alternative preparations for insulin reference materials. An isotope dilution mass spectrometry high-level reference measurement procedure for insulin is also being investigated to establish calibration traceability for routine methods (8). Finalization of this reference system will enable all assays specific for insulin to establish traceability to a common reference system to ensure that results can be compared among different assays and over extended time periods.

We thank Eli Lilly Co. and Novo Nordisk for the generous provision of pure peptides to complete the work.

References