The Need for Standardization of Tacrolimus Assays

Daniel M. Levine,1* Gregory T. Maine,2 David A. Armbruster,2 Christopher Mussell,3 Christoph Buchholz,3 Gavin O’Connor,2 Victoria Tuck,4 Atholl Johnston,4 and David W. Holt4

BACKGROUND: Owing to the lack of an internationally recognized tacrolimus reference material and reference method, current LC-MS and immunoassay test methods used to monitor tacrolimus concentrations in whole blood are not standardized. The aim of this study was to assess the need for tacrolimus assay standardization.

METHODS: We sent a blinded 40-member whole-blood tacrolimus proficiency panel (0–30 μg/L) to 22 clinical laboratories in 14 countries to be tested by the following assays: Abbott Architect (n = 17), LC-MS (n = 9), and Siemens Dade Dimension (n = 5). Selected LC-MS laboratories (n = 4) also received a common calibrator set. We compared test results to a validated LC-MS method. Four samples from the proficiency panel were assigned reference values by using exact-matching isotope-dilution mass spectrometry (EM-IDMS) at Laboratory of Government Chemists.

RESULTS: The range of CVs observed with the tacrolimus proficiency panel was as follows: LC-MS 11.4%–18.7%, Architect 3.9%–9.5%, and Siemens Dade 5.0%–48.1%. The range of historical within-site quality control CVs using 3 control levels were as follows: LC-MS low 3.8%–8.9%, medium 2.0%–6.0%, high 2.3%–6.3%; Architect low 2.5%–9.5%, medium 2.5%–8.6%, high 2.9%–18.6%; and Siemens/Dade Dimension low 8.7%–23.0%, medium 7.6%–13.2%, high 4.4%–10.4%. Assay bias observed between the 4 LC-MS sites was not corrected by implementation of a common calibrator set.

CONCLUSIONS: Tacrolimus assay standardization will be necessary to compare patient results between clinical laboratories. Improved assay accuracy is required to provide optimized drug dosing and consistent care across transplant centers globally. © 2011 American Association for Clinical Chemistry

Solid organ transplantation has given people with end-stage organ disease and other diseases the opportunity to resume a normal life through the receipt of a suitable organ from a donor. Once an organ is transplanted, a rather complex therapeutic regimen is begun to prevent rejection of the allograft and allow an otherwise totally foreign body to remain functional for years within the recipient. Optimal immunosuppressive therapy, defined clinically and by therapeutic drug monitoring (TDM),3 is essential to prevent acute rejection and ensure long-term survival of both the patient and the allograft. Tacrolimus is the dominant calcineurin inhibitor used for patients receiving solid organ transplantation (1). The Efficacy Limiting Toxicity Elimination (ELITE)-Symphony study (2, 3) and the European Consensus Conference on Tacrolimus (4) have suggested that tacrolimus minimization strategies be used to prevent rejection of the transplanted organ while reducing long-term nephrotoxicity of the drug. These recommendations require both accurate and precise measurement of the concentration of tacrolimus in whole blood.

Tacrolimus is measured by a variety of LC-MS and immunoassay–based methods that are all independently calibrated without traceability to an accepted reference LC-MS method or standard tacrolimus reference material. Thus, tacrolimus concentration values may not be comparable between methods and/or laboratories, a situation that poses potential risks to patients monitored by TDM. For example, underestimation of blood concentration by some tacrolimus assays may result in unnecessarily high concentrations of immunosuppression, leading to a poor quality of life,

1 The Rogosin Institute, New York, NY; 2 Abbott Diagnostics, Abbott Park, IL; 3 Chemical Measurement and Calibration, LGC Limited, Teddington, UK; 4 Analytical Services International Ltd, St. George’s–University of London, London, UK.
* Address correspondence to this author at: The Rogosin Institute, 310 E. 67th Street, New York, NY 10065. E-mail dml9006@nyp.org.
A preliminary account of this work was presented at the IFCC-WorldLab-EuroMedLab, Berlin, Germany, May 15–19, 2011, and at the Canadian Laboratory Medical Congress, Vancouver, British Columbia, Canada, June 4–8, 2011. Received July 2, 2011; accepted September 26, 2011. Previously published online at DOI: 10.1373/clinchem.2011.172080
5 Nonstandard abbreviations: TDM, therapeutic drug monitoring; ELITE, Efficacy Limiting Toxicity Elimination; EM-IDMS, exact-matching isotope-dilution mass spectrometry; LGC, Laboratory of Government Chemists; LC-MS/MS, liquid chromatography/tandem mass spectrometry; CRM, certified reference material; TE, total error.
whereas overestimation may result in inadequate immunosuppression and subsequent acute rejection followed by possible loss of the allograft. Furthermore, tacrolimus test methods that are not standardized serve as a confounding variable in multicenter clinical studies of tacrolimus. Although other analytes have been standardized in the US and globally \((5, 6)\) to optimize patient care, no attempt has been made to do so for an immunosuppressive therapeutic drug. The aim of this work was to conduct a global proficiency study in 22 laboratories in 14 countries using LC-MS and immunoassay–based methods to assess the current comparability of tacrolimus measurements and the need for standardization of assays that measure this critically important therapeutic drug in allograft recipients.

Materials and Methods

TACROLIMUS PROFICIENCY PANEL

Tacrolimus powder was a gift from Astellas Pharma Europe. Tacrolimus-free EDTA anticoagulated human blood was supplied by Biological Specialty. We produced a 200-mg/L stock solution of tacrolimus by dissolving 10.21 mg tacrolimus (corrected for purity, 97.8%) into 50 mL methanol and a 1-mg/L substock by diluting 0.1 mL of the stock solution into 20 mL tacrolimus-free blood. We made appropriate dilutions of this substock into tacrolimus-free blood to produce 10 spiked samples with nominal tacrolimus concentrations in the range of \(0–30 \mu\text{g/L}\). We used surplus whole blood samples from patients undergoing routine tacrolimus TDM to produce the pooled samples. Patient samples with different tacrolimus concentrations were pooled to achieve a set of 10 samples with nominal tacrolimus concentrations in the range of \(2–30 \mu\text{g/L}\). Tacrolimus concentrations were assigned by the LC-MS/MS method validated to GLP criteria and used by Analytical Services International on the St. George’s campus (validation data on file at the laboratory). In addition, 4 of the pooled samples were also tested by exact-matching isotope-dilution mass spectrometry (EM-IDMS) at Laboratory of Government Chemists (LGC). Each laboratory was sent a total of 40 blinded samples (10 spiked and 10 pooled samples in duplicate) in a randomized order for testing.

PARTICIPANTS AND INSTRUMENTATION

Twenty-two clinical laboratories from the following 14 countries participated: Argentina \((n = 1)\), Australia \((n = 2)\), Austria \((n = 1)\), Belgium \((n = 1)\), Canada \((n = 2)\), Colombia \((n = 1)\), France \((n = 1)\), Germany \((n = 1)\), Japan \((n = 2)\), Korea \((n = 2)\), Norway \((n = 1)\), Singapore \((n = 1)\), UK \((n = 1)\), and US \((n = 5)\). All of these laboratories routinely perform TDM of tacrolimus. Seventeen laboratories reported results using the Architect Tacrolimus assay (Abbott Laboratories); 1 used LC-MS and 8 used liquid chromatography/tandem mass spectrometry (LC-MS/MS) (collectively referred to as LC-MS); and 5 used the Dade Dimension Tacrolimus assay. Each LC-MS laboratory has established their own assay acceptance criteria, which usually involves running a set of controls with limits for assay imprecision \((<10\%–15\% \text{ CV})\) and deviation from the target control values, usually 2SD. The Architect and Dade Dimension assay acceptance criteria were followed per manufacturer’s package insert instructions. Each laboratory also reported historical quality control imprecision data for the tacrolimus test methods they used. A description of the LC-MS and LC-MS/MS methods is summarized in Table 1. The validated LC-MS/MS method used by Analytical Services International (site AU) was selected as the method of comparison for purposes of this study, in the absence of a true reference method that could be applied to all samples in the proficiency panel. Five laboratories running LC-MS/MS (AU, CS18, CS19, CS20, and CS22) were sent a common calibrator set (from the Mass Trak Immunosuppressants Kit, Waters Corp.) and were asked to measure the concentration of tacrolimus in the panel using their routine method and the common calibrator set.

The EM-IDMS calibration procedure is routinely used at LGC to assign reference values to certified reference materials (CRMs) \((7)\). An EM-IDMS–based method has been developed at LGC for the characterization of a candidate tacrolimus in whole-blood CRM. We used this method to assign reference mass fractions to 4 of the pooled samples. We conducted a density determination on each sample to enable conversion of the mass fraction \((\text{ng/g})\) to concentration \((\mu\text{g/L})\). Natural tacrolimus powder was characterized at LGC \([97.9\% (0.5\%) \text{ m/m purity at the } 95\% \text{ CI}])\). A non–commercially available isotope-labeled tacrolimus (custom synthesis, no detectable naturally abundant tacrolimus present) was used as an internal standard. Solvent standards of tacrolimus were prepared in acetonitrile. We performed initial quantification of the 4 pooled blood samples using linear calibration/IDMS and used the mass fraction calculated from this initial assay to determine the amount of natural and labeled tacrolimus required to prepare calibration blends for subsequent EM-IDMS iterations. Two EM-IDMS iterations were performed, each analyzing 2 aliquots of each sample. Only the EM-IDMS data were used to calculate reference values. The amount of internal standard added to sample and calibration blends was calculated to provide a response/peak area equivalent to that of the tacrolimus present in the sample. The main components to measurement uncertainty were the SD associated with the 5 repeat measurements of a sample blend, the purity of the tacrolimus powder, and the SD of the mean of the mass fraction of the 4 aliquots. Minor components were the
uncertainties associated with the gravimetric preparation of solvent standards and the sample and calibration blends. The standard and combined uncertainties were calculated in accordance with the Eurachem guidelines (8).

STATISTICS
Passing–Bablok correlations and bias analyses were performed on the data collected using Analyse-it Clinical Laboratory software, version 1.73. CIs were calculated using Microsoft Excel.

Results

WITHIN-LABORATORY TEST METHOD ASSAY IMPRECISION
Participating sites routinely used different controls with various target values. Therefore, we generated median values for the control values and median CVs for each test method for controls with low (2–7 μg/L), medium (7–13 μg/L), and high (>13 μg/L) target values. For low target values, control and median control value/median (%CV) were 4.6 μg/L and 5.4% for Architect, 2.8 μg/L and 6.4% for LC-MS, and 3.0 μg/L and 17.5% for Dade Dimension; medium target values, 11.0 μg/L and 4.8% for Architect, 9.0 μg/L and 4.0% for LC-MS, and 7.0 μg/L and 11.5% for Dade Dimension; high target values, 20.1 μg/L and 4.2% for Architect, 19.8 μg/L and 3.4% for LC-MS, and 13.9 μg/L and 7.4% for Dade Dimension. Median CVs for both Architect and LC-MS methods were comparable across the range of controls, whereas the median CVs for the Dade Dimension were higher, especially for the low control.

BETWEEN-LABORATORY TEST METHOD ASSAY BIAS AND IMPRECISION
We determined between-laboratory mean values and assay imprecision for each unique panel member by each test method (Table 2). The mean values obtained with all 3 methods for the spiked samples (C1–C9) were in close agreement with target values, with the following mean ranges of recovery: Architect, 95%–105%; LC-MS, 95%–104%; Dade Dimension, 92%–115%. Analysis of the mean values obtained with all 3 methods for the pooled samples (P-1 to P-10) relative to the validated site AU LC-MS/MS method demonstrated that Architect had a median positive bias of 4.5%, whereas LC-MS and Dade Dimension exhibited median negative biases, −5.2% and −3.6%, respectively.

The range of CVs across the entire proficiency panel (2–30 μg/L tacrolimus concentration) were as follows: Architect, 3.9%–9.5%; LC-MS, 11.4%–18.7%; and Dade Dimension, 5.0%–48.1%. It is noteworthy that the CVs for Architect across the entire panel and for Dade Dimension on panel members >10 μg/L were lower than the CVs with LC-MS. On average, the CVs for LC-MS averaged about twice the CVs for Architect.

METHOD COMPARISON TO EM-IDMS
We performed tacrolimus test method comparison with EM-IDMS on 4 of the 10 pooled samples (P-1, P-4, P-6, and P-9). These samples were chosen to span the dynamic range of the assays under evaluation (Table 3). The comparative LC-MS/MS method at site AU

---

Table 1. Comparison of LC-MS tacrolimus measurement methods.

<table>
<thead>
<tr>
<th>Site</th>
<th>Extraction</th>
<th>Calibratorsa</th>
<th>LC manufacturerb</th>
<th>MS manufacturerc</th>
<th>MS model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS8</td>
<td>PPT + SPEd</td>
<td>Chromsystems</td>
<td>Jasco</td>
<td>Thermo Fisher Scientific</td>
<td>AQA MS</td>
</tr>
<tr>
<td>CS10</td>
<td>Turf flow</td>
<td>In-house</td>
<td>HP</td>
<td>Thermo Fisher Scientific</td>
<td>Quantum</td>
</tr>
<tr>
<td>CS12</td>
<td>PPT + SPEd</td>
<td>Chromsystems</td>
<td>Perkin-Elmer</td>
<td>AB Sciex</td>
<td>API 2000</td>
</tr>
<tr>
<td>CS15</td>
<td>L/L</td>
<td>Chromsystems</td>
<td>Shimadzu Corp.</td>
<td>AB Sciex</td>
<td>API 4000</td>
</tr>
<tr>
<td>CS18</td>
<td>L/L</td>
<td>Chromsystems</td>
<td>Waters Corp.</td>
<td>Waters Corp.</td>
<td>Quattro</td>
</tr>
<tr>
<td>CS19</td>
<td>PPT + SPEd</td>
<td>Chromsystems</td>
<td>Thermo Fisher Scientific</td>
<td>Thermo Fisher Scientific</td>
<td>Quantum</td>
</tr>
<tr>
<td>CS20</td>
<td>PPT</td>
<td>Chromsystems</td>
<td>Waters Corp.</td>
<td>Waters Corp.</td>
<td>Quattro</td>
</tr>
<tr>
<td>CS21</td>
<td>L/L</td>
<td>EMIT 2000</td>
<td>Agilent</td>
<td>AB Sciex</td>
<td>4000 Q trap</td>
</tr>
<tr>
<td>CS22</td>
<td>PPT</td>
<td>Chromsystems</td>
<td>Waters Corp.</td>
<td>Waters Corp.</td>
<td>Quattro XE</td>
</tr>
<tr>
<td>AU</td>
<td>PPT + L/L</td>
<td>In-house</td>
<td>HP</td>
<td>AB Sciex</td>
<td>API 4000</td>
</tr>
</tbody>
</table>

a All LC-MS and LC-MS/MS labs used acomycin as an internal standard except sites CS8 (32-desoxy rapamycin) and CS15 (sirolimus).

b All labs used ammonium acetate and/or formate in methanol for LC except for sites CS8 (sodium acetate/methanol), CS10 (formic acid/methanol), and CS19 (ammonium acetate/acetonitrile).

c All labs used electrospray ionization for mass spectrometry except site CS19 (atmospheric pressure chemical ionization) and CS 15 (turbo spray).

d PPT, precipitation with organic solvent mixture and centrifugation; SPE, solid phase extraction; turf flow, alkaline turbulent flow separation with Cyclone P column

(Cohesive Technologies); L/L, liquid/liquid extraction; AU, analytical unit, St. George’s University of London.
had a positive bias with EM-IDMS, ranging from 2.0% to 2.5%. The mean bias observed with Architect, LC-MS, and Dade Dimension relative to EM-IDMS ranged from 2.5% to 9.2%, 4.3% to 2.8%, and 0.3% to 6.4%, respectively. The 95% CI for the SD of the mean for Architect was < LC-MS < Dade. The range of values for each method is annotated in the box-and-whisker plots shown in Fig. 1.

Table 2. Test method assay imprecision with tacrolimus proficiency panel.

<table>
<thead>
<tr>
<th>Panel</th>
<th>Tacrolimus, μg/L</th>
<th>Architect, 17 sites</th>
<th>LC-MS, 9 sites</th>
<th>Dade Dimension, 5 sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean, g/L</td>
<td>CV, %</td>
<td>Mean, g/L</td>
<td>CV, %</td>
</tr>
<tr>
<td>C-9</td>
<td>2.0</td>
<td>9.3</td>
<td>2.0</td>
<td>18.7</td>
</tr>
<tr>
<td>C-8</td>
<td>4.0</td>
<td>6.0</td>
<td>3.8</td>
<td>15.9</td>
</tr>
<tr>
<td>C-7</td>
<td>6.0</td>
<td>5.1</td>
<td>6.2</td>
<td>14.2</td>
</tr>
<tr>
<td>C-6</td>
<td>8.0</td>
<td>6.3</td>
<td>7.9</td>
<td>13.0</td>
</tr>
<tr>
<td>C-5</td>
<td>10.0</td>
<td>6.5</td>
<td>9.9</td>
<td>14.7</td>
</tr>
<tr>
<td>C-4</td>
<td>15.0</td>
<td>5.8</td>
<td>15.4</td>
<td>12.2</td>
</tr>
<tr>
<td>C-3</td>
<td>20.0</td>
<td>6.9</td>
<td>20.8</td>
<td>14.7</td>
</tr>
<tr>
<td>C-2</td>
<td>25.0</td>
<td>6.7</td>
<td>25.5</td>
<td>16.3</td>
</tr>
<tr>
<td>C-1</td>
<td>30.0</td>
<td>8.6</td>
<td>30.4</td>
<td>11.4</td>
</tr>
<tr>
<td>P-1</td>
<td>2.9</td>
<td>7.9</td>
<td>2.7</td>
<td>14.1</td>
</tr>
<tr>
<td>P-2</td>
<td>4.5</td>
<td>5.8</td>
<td>4.4</td>
<td>16.8</td>
</tr>
<tr>
<td>P-3</td>
<td>6.4</td>
<td>6.2</td>
<td>6.0</td>
<td>14.4</td>
</tr>
<tr>
<td>P-4</td>
<td>8.5</td>
<td>6.1</td>
<td>8.1</td>
<td>15.3</td>
</tr>
<tr>
<td>P-5</td>
<td>10.4</td>
<td>3.9</td>
<td>9.7</td>
<td>12.6</td>
</tr>
<tr>
<td>P-6</td>
<td>12.2</td>
<td>4.9</td>
<td>11.5</td>
<td>12.3</td>
</tr>
<tr>
<td>P-7</td>
<td>13.1</td>
<td>9.5</td>
<td>12.5</td>
<td>15.7</td>
</tr>
<tr>
<td>P-8</td>
<td>16.1</td>
<td>5.8</td>
<td>15.5</td>
<td>14.5</td>
</tr>
<tr>
<td>P-9</td>
<td>20.2</td>
<td>5.9</td>
<td>19.0</td>
<td>13.3</td>
</tr>
<tr>
<td>P-10</td>
<td>26.5</td>
<td>6.9</td>
<td>25.4</td>
<td>13.3</td>
</tr>
</tbody>
</table>

* Panels C-1 through C-9 were spiked samples, and P-1 through P-9 were pooled samples from patients undergoing tacrolimus therapy. The concentration of tacrolimus in Panel C-9 was below the limit of quantitation reported for the Dade Dimension in the package insert. Panel C-10 (not shown) was negative whole blood. Samples were tested in duplicate.

Table 3. Test method comparison to EM-IDMS.

<table>
<thead>
<tr>
<th>Panel</th>
<th>AU&lt;sup&gt;a&lt;/sup&gt; Mean, μg/L</th>
<th>Reference value, μg/L</th>
<th>Expanded uncertainty, μg/L&lt;sup&gt;b&lt;/sup&gt;</th>
<th>LGC EM-IDMS</th>
<th>Architect, 17 sites</th>
<th>LC-MS, 9 sites</th>
<th>Dade Dimension, 5 sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean, μg/L</td>
<td>SD</td>
<td>Mean, μg/L</td>
</tr>
<tr>
<td>P-1</td>
<td>2.9</td>
<td>2.82</td>
<td>0.09</td>
<td></td>
<td>2.9</td>
<td>0.10</td>
<td>2.7</td>
</tr>
<tr>
<td>P-4</td>
<td>8.5</td>
<td>8.33</td>
<td>0.17</td>
<td></td>
<td>9.1</td>
<td>0.24</td>
<td>8.1</td>
</tr>
<tr>
<td>P-6</td>
<td>12.2</td>
<td>11.91</td>
<td>0.18</td>
<td></td>
<td>12.8</td>
<td>0.32</td>
<td>11.5</td>
</tr>
<tr>
<td>P-9</td>
<td>20.2</td>
<td>19.75</td>
<td>0.38</td>
<td></td>
<td>21.5</td>
<td>0.54</td>
<td>19.0</td>
</tr>
</tbody>
</table>

<sup>a</sup> St. George’s University of London, concentration of tacrolimus determined by the validated LC-MS/MS method.

<sup>b</sup> Expanded uncertainty encompasses all facets of measurement procedure, including the uncertainty associated with the purity of the standard used, enabling traceability to the SI.
We calculated total error (TE) for each assay for each pooled sample tested by EM-IDMS using the imprecision data in Table 2 and the EM-IDMS reference values in Table 3 as follows (9):

$$TE = \text{bias} + 2(CV)$$

where bias = $100 \times |[\text{test method (}\mu g/L) - \text{EM-IDMS (}\mu g/L)]/[\text{EM-IDMS (}\mu g/L)].$

Across the 4 pooled samples, TE ranged from 17.6% to 21.4% for Architect, 28.0%–33.4% for LC-MS, and 17.6%–54.0% for Dade Dimension.

**METHOD COMPARISON TO THE VALIDATED SITE**

**AU LC-MS/MS METHOD**

Because it was impractical to test all 20 unique panel members by the EM-IDMS method, we compared the results of the 3 assays to results obtained by site AU, which were in close agreement to the EM-IDMS method for 4 of the samples (Fig. 2). The mean bias for Architect (n = 17) was positive relative to the site AU method (0.54 µg/L), whereas the mean bias was negative for both LC-MS (n = 9) and Dade Dimension (n = 5) (−0.20 µg/L and −0.75 µg/L, respectively). Assay bias with the Dade Dimension assay tended to be more negative at higher concentrations of tacrolimus (Fig. 2C). The range of average assay bias observed at each site was −0.47 to 1.28 µg/L for Architect, −2.77 to 2.42 mg/L for LC-MS, and −1.53 to 0.08 µg/L for Dade Dimension. The slope for Architect was 1.06, whereas the slopes for LC-MS and Dade Dimension were 0.99 and 0.92, respectively. The range of slopes observed at each site was 0.95–1.16 for Architect, 0.85–1.22 for LC-MS, and 0.87–0.99 for Dade Dimension.

**COMMON CALIBRATION OF LC-MS METHODS**

To investigate the effect of the use of a common set of calibrators by different LC-MS laboratories, 4 LC-MS sites (CS18, CS19, CS20, and CS22) were selected at random to test the blinded proficiency panel using both their normal site procedure for as-

---

**Fig. 1.** Box-and-whisker plots of Architect, LC-MS, and Dade Dimension test values for samples P-1 (A), P-4 (B), P-6 (C), and P-9 (D).

Vertical solid line across the plots, reference value determined by EM-IDMS; dashed lines, associated expanded uncertainty; values in parentheses, range of tacrolimus concentration values obtained by each tacrolimus test method.
say calibration and a set of common calibrators (see "Materials and Methods") sent to the site. Each of the 4 sites resided in distinct geographic locations in Asia-Pacific, Europe, and the US. The type of calibration and instrumentation used at the 4 sites selected was unknown at the time the sites were chosen. After the data were reported, all LC-MS laboratories that participated in the study were sent a questionnaire (Supplemental Fig. 1, which accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue12) requesting detailed information regarding their LC-MS method and calibration (Table 1). The 4 sites selected used the same commercially available calibration set for their normal calibration that was distinct from the calibration set sent along with the profi-
ciency panel. As an internal control, site AU also calibrated their method with the common calibrator set. The linear regression plots are shown in Fig. 3. The average bias between the 4 sites using the same calibrators ranged from $-2.77$ to $0.62 \mu g/L$.

**Discussion**

We conducted a global tacrolimus proficiency study to assess the comparability of LC-MS and immunoassay measurements and the need for assay standardization.
The methods we chose, Abbott Architect, LC-MS, and Siemens Dade Dimension, account for approximately 70%–85% of all methods used for the measurement of tacrolimus globally (10, 11).

Test results with the Architect assay were the most comparable between laboratories among the methods evaluated with assay imprecision CVs between 3.9% and 9.5%, whereas the CVs for LC-MS and Dade Dimension were higher and ranged from 11.4% to 18.7% and 5.0% to 48.1%, respectively (Table 2). The increased mean CVs with the LC-MS methods were driven primarily by between-laboratory assay imprecision, as within-laboratory median CVs for controls ranged from 3.4% to 6.4%. The Dade Dimension had higher assay imprecision at the lower concentrations tested due, in part, to the assay’s higher functional sensitivity relative to Architect and LC-MS/MS (3.1 μg/L vs 0.5–0.8 and 0.4 μg/L for Architect and LC-MS, respectively) (12–14). Above 10 μg/L, however, the assay imprecision of the Dade Dimension assay was comparable to that of Architect and even lower than that of LC-MS. Interference by heterologous antibodies in some samples may have also contributed to higher assay imprecision with the Dade Dimension assay (15–19). One of the advantages the commercial immunoassay methods share over laboratory-developed LC-MS methods is that each of these immunoassays has the same extraction procedure, instrumentation, detection, and assay calibration, whereas for the laboratory-developed LC-MS methods, these parameters vary and contribute to between-laboratory imprecision (Tables 1 and 2).

To determine if LC-MS methods could be harmonized by common calibration, 4 LC-MS sites were sent the same calibrator set along with the proficiency panel and were asked to report results using both sets of calibrators. Six of the 9 laboratories used calibrators from the same manufacturer to calibrate their assays. Four of these 6 laboratories reported test results using both their normal site calibration and a common calibration. The use of a common calibration or normal site calibration with calibrators from the same manufacturer to calibrate their assays to be standardized to a reference method. Four of the 9 laboratories used calibrators from the same manufacturer to calibrate their assays. The 6 laboratories reported test results using both their normal site calibration and a common calibration. The use of a common calibration or normal site calibration with calibrators from the same manufacturer to calibrate their assays suggests a calibration issue or issues involving the site AU LC-MS/MS method. An increase of positive bias was observed with the common calibrator set, as shown in Fig. 3E. This is clearly a shift in the wrong direction, as the site AU LC-MS/MS method was already shown to have a slight positive bias with respect to EM-IDMS (Table 3). Thus, harmonization among all 5 LC-MS methods using the common calibration was not achieved, suggesting that simply using the same commercially available calibrator set will not harmonize LC-MS results and is not a substitute for addressing other analytical variables followed by proper tacrolimus assay standardization. Matrix effects may be involved to some extent, but the fact that calibrators vary between methods is a major issue. Establishing traceability to a reference standard will be a major improvement toward harmonization of methods.

All 3 methods were compared to the site AU LC-MS/MS method across the 40-member proficiency panel (Fig. 2A–2C), with a positive bias observed with Architect and negative bias observed with LC-MS and Dade Dimension. A positive bias with the Architect method was expected, due to its detection of tacrolimus metabolites not detected by LC-MS methods. However, the negative bias observed with the Dade Dimension assay suggests a calibration issue or issues with the extraction of tacrolimus from whole blood, since all immunoassay-based test methods cross-react with metabolites of tacrolimus and so would be expected to report higher concentrations for patient samples compared to LC-MS methods.

To further assess the need for tacrolimus assay standardization, we selected 4 pooled samples to be tested by the candidate reference method for tacrolimus developed at LGC, EM-IDMS (Table 3). A positive bias was observed with the Architect assay (range 2.8%–9.2%). The Architect assay had a tighter distribution around the mean value than LC-MS and Dade (Fig. 1A–1D and Table 3). The Architect assay also had the lowest TE of all 3 methods evaluated. Nevertheless, comparison of the range of slopes and bias observed with the methods evaluated in this study suggest that assay standardization will be needed to compare results between these methods. Furthermore, because the within-laboratory assay imprecision for LC-MS is within acceptable limits, the only way laboratory-developed LC-MS test results can be comparable is for those assays to be standardized to a reference method. Hence, caution must be observed when comparing results between laboratories using laboratory-developed LC-MS methods.

The clinical impact of the laboratory variability in tacrolimus concentrations on patient care can be illustrated by comparing the slopes obtained by linear regression from sites that demonstrated the most negative and most positive bias. In this study, the largest
differences we observed in slopes were between LC-MS methods, 0.85 vs 1.22. This is roughly a 20%–30% margin relative to the site AU LC-MS/MS and EM-IDMS methods. Although the reference value for sample P-4 was 8.33 μg/L, the concentrations generated by the lowest- and highest-bias sites for this pooled patient sample were 6.0 and 10.2 μg/L. Hence, if a clinician were targeting a value of 8 μg/L, patients monitored by the site with the most negative bias would need to be dosed higher and patients monitored by the site with the most positive bias would need to be dosed lower to achieve this target concentration. This example also illustrates the confounding variable of tacrolimus test methods in multicenter drug regimen studies such as the ELiTE-Symphony study (2, 3). The quality of life and long-term survival of transplant patients can be only positively impacted by improved accuracy of tacrolimus assays. Although there is never a perfect correlation between pharmacokinetics and pharmacodynamics, and many other factors may play a role in dose response, there is a clear consequence of method bias, plus or minus, from the true value.

Harmonization of tacrolimus assay results for patient samples is a worthwhile aim in clinical medicine and an important step toward providing uniform global care for transplant patients. Tacrolimus assay standardization is needed and must encompass a range of analytical variables, not only issues related to calibration. A standardization exercise does not improve the performance characteristics of an assay, but rather centers the results as close to the true value as possible given the error of each method. Professional associations and academics should take a leadership role and work collaboratively with pharmaceutical and diagnostic companies to fund and promote assay standardization.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: D.A. Armbruster, Global Scientific Affairs Manager, Abbott Diagnostics; C. Buchholz, Trainee Researcher, Mass Spectrometry, LGC, and Analytical Chemiker LC-MS/MS, Metabolomic Discoveries GmbH.

Consultant or Advisory Role: D. Holt, Consultant on TDM to Abbott Diagnostics.


Research Funding: The development of a tacrolimus reference method procedure at LGC was funded by the Chemical and Biology Metrology Program, UK National Measurement System. Research funding was provided by Abbott.

Expert Testimony: None declared.

Other Remuneration: D.M. Levine, travel to give talks, Abbott Diagnostics.

Role of Sponsor: The funding organizations played a direct role in design of study, review and interpretation of data, and preparation and final approval of manuscript.

Acknowledgments: The authors express their thanks to the institutions, laboratory directors, and personnel that participated in the global tacrolimus proficiency study.

References

10. AACCCAP Immunosuppressive Drugs CSM-A 4-22-2010.