EARLY ONLINE RELEASE

Note: This article was posted on the Archives Web site as an Early Online Release. Early Online Release articles have been peer reviewed, copyedited, and reviewed by the authors. Additional changes or corrections may appear in these articles when they appear in a future print issue of the Archives. Early Online Release articles are citable by using the Digital Object Identifier (DOI), a unique number given to every article. The DOI will typically appear at the end of the abstract.

The DOI for this manuscript is doi: 10.5858/arpa.2019-0493-CP

The final published version of this manuscript will replace the Early Online Release version at the above DOI once it is available.
Participation in the College of American Pathologists Laboratory Accreditation Program Decreases Variability in B-Lymphoblastic Leukemia and Plasma Cell Myeloma Flow Cytometric Minimal Residual Disease Testing

A Follow-up Survey

Meghan M. Hupp, MD; Christine Bashleben, MT(ASCP); Jolene L. Cardinali, MT(ASCP); David M. Dorfman, MD, PhD; William Karlon, MD, PhD; Michael Keeney, ART; Catherine Leith, MBCh; Thomas Long, MPH; Claire E. Murphy, MD; Vinodh Pillai, MD, PhD; Flavia N. Rosado, MD; Adam C. Seegmiller, MD, PhD; Michael A. Linden, MD, PhD

Context.—Minimal residual disease (MRD) testing by flow cytometry is ubiquitous in hematolymphoid neoplasm monitoring, especially B-lymphoblastic leukemia (B-ALL), for which it provides predictive information and guides management. Major heterogeneity was identified in 2014. Subsequently, new Flow Cytometry Checklist items required documentation of the sensitivity determination method and required lower level of detection (LLOD) inclusion in final reports. This study assesses Laboratory Accreditation Program (LAP) participation and new Checklist items’ impact on flow cytometry MRD testing.

Objectives.—To survey flow cytometry laboratories about MRD testing for B-ALL and plasma cell myeloma. In particular, enumerate the laboratories performing MRD testing, the proportion performing assays with very low LLODs, and implementation of new Checklist items.

Design.—Supplemental questions were distributed in the 2017-A mailing to 548 flow cytometry laboratories subscribed to the College of American Pathologists FL3 Proficiency Testing Survey (Flow Cytometry–Immunophenotypic Characterization of Leukemia/Lymphoma).

Results.—The percentage of laboratories performing MRD studies has significantly decreased since 2014. Wide ranges of LLOD and collection event numbers were reported for B-ALL and plasma cell myeloma. Most laboratories determine LLOD by using dilutional studies and include it in final reports; a higher proportion of LAP subscribers used these practices than nonparticipants.

Conclusions.—Several MRD testing aspects vary among laboratories receiving FL3 Proficiency Testing materials. After the survey in 2014, new Checklist items were implemented. As compared to 2014, fewer laboratories are performing MRD studies. While LLOD remains heterogeneous, a high proportion of LAP subscribers follow the new Checklist requirements and, overall, target LLOD recommendations from disease-specific working groups are met.

(Arch Pathol Lab Med. doi: 10.5858/arpa.2019-0493-CP)
residual disease (MRD) after treatment is an important prognostic marker and guides future therapy. While too rare to be identified by normal light microscopy, MRD can be investigated by using clonality in immunoglobulin or T-cell receptor gene rearrangements, polymerase chain reaction for a certain mutation or translocation, flow cytometry, or any combination of these methods. Flow cytometry uses aberrant and unique immunophenotypes of cancer cells to detect very small populations that remain after therapy and has the advantages of being simpler, quicker, and more cost-effective to perform than other methods.

In response to the relative rise in MRD-testing popularity, the DIFCC developed voluntary supplemental questions with the A-mailing of the FL3 Survey (Flow Cytometry–Immunophenotypic Characterization of Leukemia/Lymphoma) in the spring of 2014. Overall, 32.8% of the 500 responding laboratories were performing MRD for any indication, reflecting the growing use of flow cytometric MRD testing, which has expanded from a handful of European and American studies to worldwide use. MRD testing was performed by 87% of these laboratories for lymphoblastic leukemia, 61% for myeloid leukemia, 52% for chronic lymphocytic leukemia, and 47% for plasma cell myeloma (PCM). This survey also demonstrated major heterogeneity in the reported lower limit of detection (LLOD); 10-fold to 100-fold differences in reported LLOD were common and published elsewhere.

In an attempt to increase assay disease-specificity and transparency in MRD testing, the committee created 2 new CAP Flow Cytometry Checklist items in the “Rare Event Flow Cytometric Assays” section, as a requirement for Laboratory Accreditation Program (LAP) participants. Items FLO.30800 and FLO.30820 require (1) that the LLOD be validated by performing dilutional studies of known patient samples or reference samples, and (2) that this LLOD be clearly stated in the final diagnostic report. Since their incorporation in 2015, data have not yet been collected about their adoption, both within the LAP and in LAP nonparticipants.

Two disease-specific clinical groups have published criteria requiring high-sensitivity flow cytometric MRD testing. The Children’s Oncology Group (COG) calls for B-lymphoblastic leukemia (B-ALL) MRD testing sensitivity of at least 0.01%, and the International Myeloma Working Group (IMWG) recommends PCM sensitivity of at least 0.001%. Because these criteria are widely accepted, MRD testing for B-ALL and PCM was of interest to the DIFCC.

The primary goal of this study was to examine and report features of MRD testing for B-ALL and PCM among the participants of our leukemia and lymphoma Flow Cytometry Survey. In particular, we sought to enumerate those laboratories performing MRD analysis and for which diseases it was performed, and to compare this to the 2014 survey data. We also investigated changes in the proportion of laboratories performing high-sensitivity MRD testing (based on reported detection thresholds), the method by which LLOD was validated, and LLOD reporting in the final diagnostic report. The latter 2 goals directly assess the adoption of the new Checklist items since the 2015 implementation.

**MATERIALS AND METHODS**

A supplemental questionnaire was sent to 548 flow cytometry laboratories participating in the CAP FL3 (Flow Cytometry) Survey. In these “wet” challenges, laboratories receive tumor cell samples, which must be processed, analyzed, and interpreted. The questionnaire (Figure 1) was developed by the members of the CAP DIFCC and was distributed with the FL3-A (Immunophenotypic Characterization of Leukemia/Lymphoma) Survey mailing in 2017. This questionnaire was similar to the questionnaire that was distributed to 549 laboratories in 2014 with the FL3-A Survey. All P values are calculated by using a 2-tailed χ² test, except for 2 tests where Fisher exact test was used (denoted).

**RESULTS**

Of the 531 laboratories that provided Proficiency Testing data for the FL3 Survey in early 2017, 498 laboratories completed the questionnaire, with an overall response rate of 93.8%. The first question (Figure 1) was intended to enumerate the number of laboratories that perform rare event analysis, or MRD testing, by flow cytometry, most commonly on bone marrow aspirate material. As shown in Table 1, of 498 respondents, 110 (22.1%) currently perform rare event analysis as compared to 164 of 500 (32.8%) who performed this analysis in 2014. This is a significant decrease (P < .001) in laboratories performing MRD testing by flow cytometry. Of laboratories that perform MRD testing, 76.4% (84 of 110) also indicated that they are concurrent LAP participants, whereas 83.5% (324 of 388) of laboratories that do not perform MRD subscribe to LAP (not statistically significant, P = .09).

As childhood B-ALL MRD testing is the standard of care, and high-sensitivity MRD testing for PCM is rapidly emerging, we focused additional questions on these 2 disease types. More laboratories perform MRD testing for B-ALL than PCM: 93 of 110 laboratories (84.5%) and 60 of 110 laboratories (54.5%), respectively. This is not statistically different from the answers provided in 2014 (Table 2).

The survey asked each laboratory a series of questions about its LLOD. Because all flow cytometric MRD assays are laboratory-developed tests (ie, there are none currently approved by the US Food and Drug Administration), some variability was expected. It is understood that there are many variables that contribute to the LLOD of a given assay, including the number of fluorochromes used, the antigens studied, the number of events collected, and the instrumentation used. The method of LLOD determination is another pivotal contributor to the minimal threshold of detection achieved.

The CAP Flow Cytometry Checklist, a component of the LAP, recommends that the "[analytic] sensitivity of the lower detection limit should be validated by performing

<table>
<thead>
<tr>
<th>Disease (Assay Type)</th>
<th>2014, No. (%)</th>
<th>2017, No. (%)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma cell myeloma</td>
<td>91 (47.4)</td>
<td>60 (54.5)</td>
<td>.23</td>
</tr>
<tr>
<td>B-lymphoblastic leukemia</td>
<td>167 (87.0)</td>
<td>93 (84.5)</td>
<td>.56</td>
</tr>
</tbody>
</table>
dilutional studies” and that the LLOD be reported on the final report (items FLO.30800 and FLO.30820). These items were first incorporated into the checklist in 2015. It was therefore expected that those who subscribe to the LAP would calculate their LLOD by using dilutional studies and report the LLOD on the final diagnostic report. Most respondents did, in fact, perform dilutional studies to find the LLOD for their assay (74.3%, or 81 of 109), while 22 of 109 (20.2%) estimated it from experience (Table 3). Other methods were chosen by 6 of 109 respondents (5.5%). When cross-referenced with LAP subscription, 68 of 84 responding LAP participants (81.0%) performed dilutional studies and 16 of 84 (19.0%) either used experience-based LLOD values or “other” methods. Only 52.0% (13 of 25) of the LAP nonparticipants reported performing dilutional studies, significantly fewer than LAP laboratories (P = .004). The proportion of laboratories reporting or not reporting their LLOD on the final diagnostic report segregated similarly to the determination method: 85 of 107 (79.4%) report their LLOD and 22 of 107 (20.6%) do not. Of the laboratories that participate in LAP, 82.9% (68 of 82) include the LLOD on the final diagnostic report, compared to 17 of 25 nonparticipants (68.0%), a difference that does not reach statistical significance (P = .11, Table 4). For a laboratory to be in compliance with the LAP checklist, the LLOD must be validated with dilutional studies and reported in the final diagnostic report. LAP participation is clearly associated with higher adherence to CAP checklist items.

For both B-ALL and PCM, we asked each laboratory to provide its LLOD: 1%, 0.1%, 0.01%, 0.001%, or “other.” Those laboratories that reported “other” were asked to provide their threshold. The 60 laboratories who perform PCM MRD testing reported LLOD ranging from $7 \times 10^{-7}$ to 1% (n = 60; Figure 2, a). Five laboratories (8.3% of 60) endorsed a threshold of 1%. The most common LLOD was 0.01% (24 of 60, 40.0%), followed by 0.001% (17 of 60, 28.3%), 0.1% (15 of 60, 25.0%), and 1% (5 of 60, 8.3%).

### Table 3. 2017 Method of Lower Limit of Detection Determination, n = 109 (P = .004)

<table>
<thead>
<tr>
<th>Method</th>
<th>LAP Participant, No. (%)</th>
<th>Nonparticipant, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilutional studies</td>
<td>68 (81.0)</td>
<td>13 (52.0)</td>
</tr>
<tr>
<td>Estimated based on experience</td>
<td>13 (15.5)</td>
<td>9 (36.0)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (3.5)</td>
<td>3 (12.0)</td>
</tr>
<tr>
<td>Total</td>
<td>84 (100)</td>
<td>25 (100)</td>
</tr>
</tbody>
</table>

Abbreviation: LAP, Laboratory Accreditation Program.

### Table 4. 2017 Number of Laboratories Reporting Lower Limit of Detection on Final Report, n = 107 (P = .11)

<table>
<thead>
<tr>
<th>Response</th>
<th>LAP Participant, No. (%)</th>
<th>Nonparticipant, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>68 (82.9)</td>
<td>17 (68.0)</td>
</tr>
<tr>
<td>No</td>
<td>14 (17.1)</td>
<td>8 (32.0)</td>
</tr>
<tr>
<td>Total</td>
<td>82 (100)</td>
<td>25 (100)</td>
</tr>
</tbody>
</table>

Abbreviation: LAP, Laboratory Accreditation Program.
The “other” responses ranged from $7 \times 10^{-7}$ to 0.05%. Of the 44 current LAP-participating laboratories, 20 (44.4%) reported an LLOD of 0.01%. Of the 18 who use the EuroFlow method (see Table 5 and discussion below), 11 (61.1%) reported an LLOD of 0.001%. As compared to 2014, a higher percentage of laboratories reported using a threshold of 0.001%, which is recommended by both EuroFlow and the International Myeloma Working Group (IMWG),12,13 with 7 of 91 respondents in 2014 (7.7%) versus 17 of 60 respondents in 2017 (28.3%, $P = .001$). Of those not using the EuroFlow method, most used 0.01% as a cutoff (61.8%, 21 of 34).

When the 94 laboratories reporting B-ALL MRD LLOD data were asked the same question, responses ranged from $6 \times 10^{-6}$ to 1% (Figure 2, b). Again, 5 of 94 laboratories (5.3%) used a threshold of 1%. The most commonly reported LLOD was 0.01% (69 of 94, 73.4%), followed by 0.001% (9 of 94, 9.6%). The write-in responses for the “other” category ranged from $6 \times 10^{-6}$ to 0.02%. As compared to 2014, a higher proportion of respondents reported using an LLOD of 0.01% (as recommended by

---

### Table 5. Plasma Cell Myeloma Sensitivity Cutoff, Separated by Laboratory Accreditation Program (LAP) Participation and Method, 2017

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>LAP Participant, No. (%)</th>
<th>Nonparticipant, No. (%)</th>
<th>EuroFlow Method, No. (%)</th>
<th>Other Method, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>5 (11.1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>0.1%</td>
<td>5 (11.1)</td>
<td>1 (6.7)</td>
<td>1 (5.6)</td>
<td>4 (11.8)</td>
</tr>
<tr>
<td>0.01%</td>
<td>20 (45.4)</td>
<td>4 (26.7)</td>
<td>2 (11.1)</td>
<td>21 (61.8)</td>
</tr>
<tr>
<td>0.001%</td>
<td>9 (20.0)</td>
<td>8 (53.3)</td>
<td>11 (61.1)</td>
<td>6 (17.6)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (13.3)</td>
<td>2 (13.3)</td>
<td>4 (22.2)</td>
<td>3 (8.8)</td>
</tr>
<tr>
<td>Total</td>
<td>45 (100)</td>
<td>15 (100)</td>
<td>18 (100)</td>
<td>34 (100)</td>
</tr>
</tbody>
</table>
COG in 2017: 87 of 167 respondents in 2014 (52.1%) versus 69 of 94 respondents in 2017 (73.4%, P = .001). Of the 74 current LAP participants, 56 (75.7%) reported an LLOD of 0.01% (Table 6). A higher proportion of laboratories following a COG-approved method use an LLOD of 0.01% than do laboratories that use a different method (19 of 21 [90.5%] and 45 of 56 [80.4%], respectively, although this was not a statistically significant difference; Fisher exact test: \( P = .50 \)).

To have useful MRD results, flow cytometry assays must standardize a minimum number of events to collect. Our survey asked for the number of events that laboratories collect for the PCM and B-ALL MRD testing. For myeloma MRD, the number of events ranged from 1000 to 11 000 000 (Table 7) with a median of 1 000 000 events. The middle 50% of respondents (from lower quartile to upper quartile) reported 500 000 to 5 500 000 events. For B-ALL, the events ranged from 20 000 to 20 000 000 with the median of 750 000 events. The lower quartile was 500 000 and the upper quartile was 2 000 000 events. The number of events collected directly correlates with the stated LLODs (Figure 3, a and b)—those that collected more events were able to achieve better analytic sensitivity. The authors note, however, that the wording of the supplemental questions (Figure 1) is not specific, and laboratories were not directed to provide their minimum number of events to achieve a desired LLOD versus the typical number of events collected.

Several guidelines have been published with recommended fluorochrome panels for MRD studies. For PCM MRD studies, the panel with the most clinical data (and IMWG endorsement) is the EuroFlow 2-tube panel (tube 1: CD45, CD138, CD38, CD56, \( \beta_2 \)-microglobulin, CD19, cytoplasmic \( \kappa \) light chain, and cytoplasmic \( \lambda \) light chain; tube 2: CD45, CD138, CD38, CD28, CD27, CD19, and CD117). Other well-known methods include a 10-color, 1-tube assay. When asked if their PCM MRD assay used the EuroFlow method or another method, 20 of 54 laboratories (37.0%) indicated EuroFlow and 34 of 54 (63.0%) indicated “other” methods (Table 8). Among the latter group, 22 entered text indicating they use an “in house” or laboratory-developed test, and 2 laboratories use International Clinical Cytometry Society guidelines.

### DISCUSSION

Previous studies have identified significant heterogeneity in MRD testing, including a study published by the CAP DIFCC in 2015. As no flow cytometry MRD method is US Food and Drug Administration approved, these assays are laboratory-developed tests. There are, however, published guidelines and recommendations that may reduce LLOD heterogeneity between laboratories. After our 2014 survey findings revealed major differences in MRD assays between laboratories, we implemented 2 new CAP Flow Cytometry Checklist items. The “Rare Event Flow Cytometric Assays” portion of the Checklist indicates LLOD must be calculated by dilutional studies and the LLOD must be included on the final diagnostic report.

The current survey results show that there are fewer laboratories performing MRD testing, possibly as a result of the Checklist items causing poor-performing or noncompliant laboratories to discontinue MRD testing. A very high majority of laboratories perform dilutional studies to validate their LLOD and report this LLOD on the final diagnostic report, arguably representing better practices in response to the Checklist items. Cross-referencing adherence to these 2 Checklist items to LAP participation,
we found that LAP participants are more likely to be in agreement with the items than LAP nonparticipants. The findings in 2014 showed a large range of LLODs, which was seen again in the responses to this survey. Reported LLODs ranged incredibly widely: more than 100 000-fold differences in B-ALL and almost 1.5 million-fold differences in PCM were reported. Despite the wide LLOD ranges, we observed an increased proportion of laboratories reporting an LLOD that matches the recommended sensitivity cutoffs from prominent clinical groups, such as COG and IMWG. When segregating the PCM data by LAP participation, more LAP participants use cutoffs of 0.01% than LAP nonparticipants (44.4% versus 26.7%, respectively), although this difference did not reach statistical significance (Fisher exact test, $P = .36$).

Several laboratories endorsed LLODs of 1%, which arguably may not qualify as minimal residual disease and is not useful for clinical management. We acknowledge that not all “MRD” testing requires low LLODs; a sensitivity cutoff of 1% may suffice for some patients and in some clinical situations. However, the utility of having uniform and scientifically sound MRD testing based on prospective clinical trials is best illustrated by patients following childhood B-ALL study protocols. At day 29, a bone marrow aspirate is obtained for MRD testing. If MRD values are less than 0.01%, the patient receives standard-risk therapy. If MRD is greater than 0.01%, the patient is

<table>
<thead>
<tr>
<th>Table 8. Method of Detection, 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
</tr>
<tr>
<td>Plasma cell myeloma (n = 54)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>B-lymphoblastic leukemia (n = 78)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: COG, Children’s Oncology Group.
Importantly, the difference between LLOD and lower limit of quantification (LLOQ) must be addressed. LLOD is the lowest analyte (myeloma or leukemia cells, in this study) concentration that can be reliably identified from background “noise.” Analytic sensitivity and LLOD are occasionally used interchangeably, which this article has continued. LLOQ is the lowest concentration at which the analyte can be reliably detected and quantified with a satisfactorily high level of precision. LLOQ is generally a higher concentration than LLOD, though they may be identical. This article has used only one of these terms (LLOD) to minimize confusion, although LLOQ is arguably a more appropriate measurement of assay sensitivity. Furthermore, Borowitz et al reported that patients with B-ALL MRD levels at or above the LLOD but below the LLOQ have inferior event-free survival as compared to truly “MRD-negative” patients.

There continues to be high LLOD variability between laboratories performing MRD testing, ranging from panel of antibodies used to threshold of detection. The decentralization of COG-approved B-ALL MRD testing and the 2016 IMWG response criteria and guidelines for MRD assessment have brought an already existing problem into sharper focus. Even among experienced laboratories that perform B-ALL MRD testing, there was discordance particularly when differentiating between arrested hematogones and lymphoblasts. COG method utilization and educational material significantly improved the concordance, reemphasizing the benefit of method standardization. However, interpretation errors still occurred, revealing the next obstacle to overcome in flow cytometry MRD testing after laboratories have adopted standardized methodologies.

The DIFCC of the CAP introduced rare event analysis Checklist items in 2015 to better ensure transparency and reliability of MRD testing results. We have recently introduced a new Proficiency Testing product for B-ALL MRD testing to evaluate a laboratory’s ability to detect rare events (dry interpretive challenge). A “wet challenge,” which tests both the laboratory’s ability to perform B-ALL MRD as well as the pathologic interpretation of the challenge, was introduced in 2020. Upon comparing survey responses with LAP enrollment, we demonstrated that LAP participants have higher rates of performing dilutional studies to determine LLOD and reporting LLOD on the final report. Future studies to assess how subscription to MRD Proficiency Testing products affects MRD testing heterogeneity are needed. Greater consensus on MRD methodology and sensitivity will help produce results in which providers, clinical trials, and patients can have confidence.

References