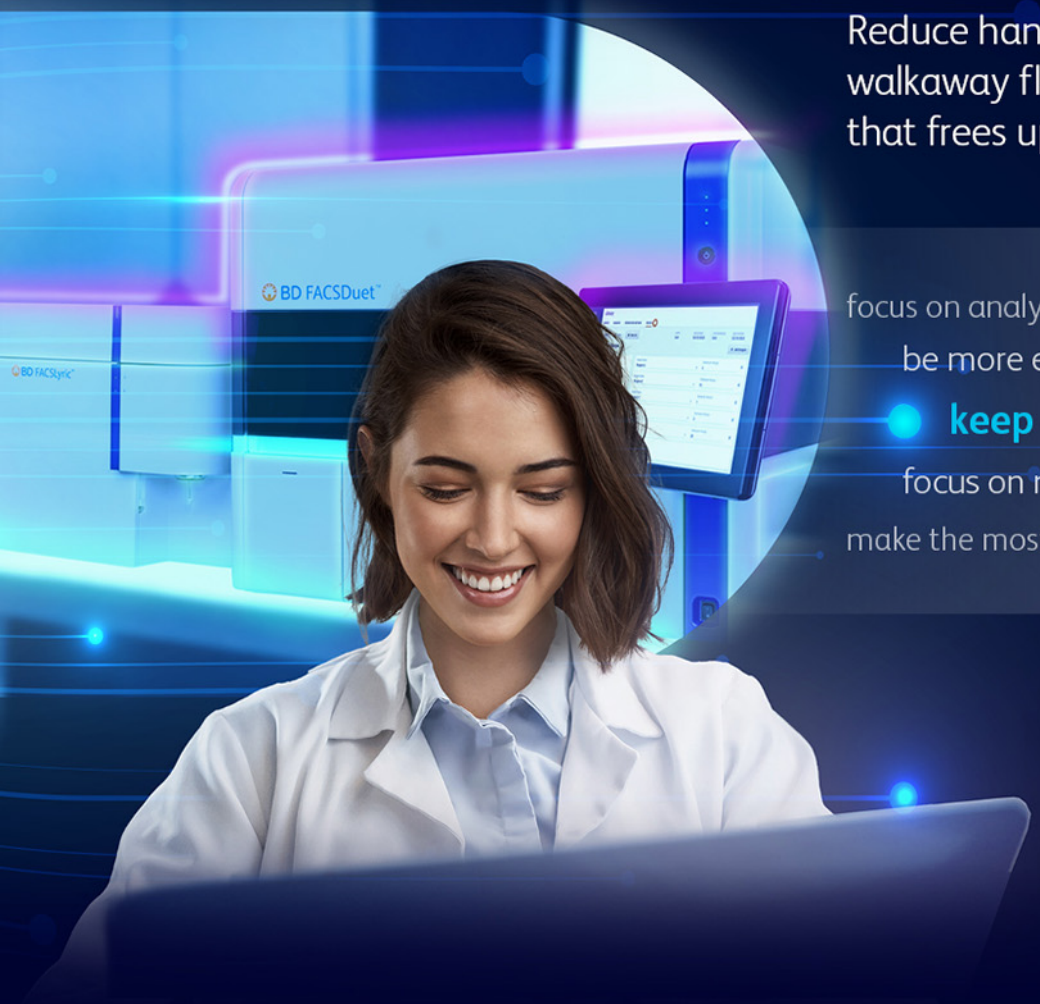
Automate more.
Get more done.



Reduce hands-on time with a true walkaway flow cytometry solution that frees up your team to:

focus on analysis
be more efficient
keep up with testing demands
focus on more critical lab tasks
make the most of the workday

Drive efficiency from sample to answer with the power of the fully automated BD FACSDuet™ Premium Sample Preparation System, offering cocktailing, washing, centrifugation and sample transfer to a physically integrated flow cytometer.

Get more done. Discover how at **bdbiosciences.com/duet**

BD

BD, the BD Logo, BD FACSDuet and BD FACSLyric are trademarks of Becton, Dickinson and Company or its affiliates. © 2023 BD. All rights reserved. BD-88682 (05/23)

Check for updates

Received: 28 October 2022 | Revised: 30 August 2023 | Accepted: 7 September 2023

DOI: 10.1002/cyto.b.22144

ORIGINAL ARTICLE

CLINICAL CYTOMETRY WILEY

Analytical assay validation for acute myeloid leukemia measurable residual disease assessment by multiparametric flow cytometry

Jesse M. Tettero[1,2] | Naveen Dakappagari[3] | Maaike E. Heidinga[1,2] | Yvonne Oussoren-Brockhoff[1,2] | Diana Hanekamp[1,2,4] | Anil Pahuja[3] | Kerri Burns[3] | Pavinder Kaur[3] | Zeni Alfonso[3] | Vincent H. J. van der Velden[5] | Jeroen G. te Marvelde[5] | Willemijn Hobo[6] | Jennichjen Slomp[7] | Costa Bachas[1,2] | Angele Kelder[1,2] | Kevin Nguyen[3] | Jacqueline Cloos[1,2]

[1]Department of Hematology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2]Cancer Center Amsterdam, Imaging and Biomarkers, Amsterdam, The Netherlands

[3]Navigate BioPharma (a Novartis Subsidiary), Carlsbad, California, USA

[4]Department of Hematology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands

[5]Department of Immunology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

[6]Department of Laboratory Medicine—Laboratory of Hematology, Radboud University Medical Center, Nijmegen, The Netherlands

[7]Department of Clinical Chemistry, Medisch Spectrum Twente/Medlon, Enschede, The Netherlands

Correspondence
Jacqueline Cloos, Amsterdam UMC, Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081HV, Amsterdam, The Netherlands.
Email: j.cloos@amsterdamumc.nl

Funding information
KWF Kankerbestrijding, Grant/Award Number: 13650/2021-Infra; Navigate BioPharma

Abstract

Background: Measurable residual disease (MRD) assessed by multiparametric flow cytometry (MFC) has gained importance in clinical decision-making for acute myeloid leukemia (AML) patients. However, complying with the recent In Vitro Diagnostic Regulations (IVDR) in Europe and Food and Drug Administration (FDA) guidance in the United States requires rigorous validation prior to their use in investigational clinical trials and diagnostics. Validating AML MRD-MFC assays poses challenges due to the unique underlying disease biology and paucity of patient specimens. In this study, we describe an experimental framework for validation that meets regulatory expectations.

Methods: Our validation efforts focused on evaluating assay accuracy, analytical specificity, analytical and functional sensitivity (limit of blank (LoB), detection (LLoD) and quantitation (LLoQ)), precision, linearity, sample/reagent stability and establishing the assay background frequencies.

Results: Correlation between different MFC methods was highly significant ($r = 0.99$ for %blasts and $r = 0.93$ for %LAIPs). The analysis of LAIP specificity accurately discriminated from negative control cells. The assay demonstrated a LoB of 0.03, LLoD of 0.04, and LLoQ of 0.1%. Precision experiments yielded highly reproducible results (Coefficient of Variation <20%). Stability experiments demonstrated reliable measurement of samples up to 96 h from collection. Furthermore, the reference range of LAIP frequencies in non-AML patients was below 0.1%, ranging from 0.0% to 0.04%.

Conclusion: In this manuscript, we present the validation of an AML MFC-MRD assay using BM/PB patient specimens, adhering to best practices. Our approach is

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Cytometry Part B: Clinical Cytometry* published by Wiley Periodicals LLC on behalf of International Clinical Cytometry Society.

expected to assist other laboratories in expediting their validation activities to fulfill recent health authority guidelines.

KEYWORDS
assay validation, flow-cytometry, measurable residual disease

1 | INTRODUCTION

Acute myeloid leukemia (AML) is a malignancy of hematopoietic cells that occurs in the bone marrow (BM), characterized by defects in the maturation program with abnormal proliferation and accumulation of myeloid progenitor cells, leading to impaired hematopoiesis and ultimately BM failure (Dohner et al., 2015). The disease is highly heterogeneous concerning morphology, immunophenotype, cytogenetics, molecular aberrations, gene expression signatures, and survival outcomes following current approved treatment modalities (Hou & Tien, 2020). The current management of AML involves induction chemotherapy to achieve complete remission (CR), followed by post-remission treatment, which consists of either several courses of additional chemotherapy or stem cell transplantation (either autologous or allogeneic) (Dohner et al., 2017). Despite high initial remission rates of up to 90% after intensive chemotherapy, the 5-year survival for adults over 60 years of age is only 30%, depending on disease-related characteristics and measurable residual disease (MRD) status after two cycles of chemotherapy (Short et al., 2018; Short et al., 2020). This dismal outcome is predominantly due to the occurrence of relapses that are often resistant to chemotherapy, highlighting the need for early detection of relapse to guide post-remission therapy (Ossenkoppele & Schuurhuis, 2014).

Detecting MRD using various techniques allows for the prediction of impending relapse (Schuurhuis et al., 2018). Multiparametric flow cytometry (MFC) can be used to quantify MRD in more than 90% of AML cases, with a sensitivity that surpasses detecting one tumor cell among 1000 white blood cells (WBC) (Buccisano et al., 2017). MRD status is prognostic for overall survival (OS) and relapse-free survival (RFS) (Schuurhuis et al., 2018; Short et al., 2020). In recent years, MRD has gained importance not only in treatment decision-making but also in patient selection for clinical trials (Hu et al., 2020; Löwenberg et al., 2021; Venditti et al., 2019). More importantly, MRD is expected to become a surrogate endpoint for survival based on ongoing research collaborations such as HARMONY and MPAACT (MRD Partnership and Alliance in AML Clinical Trials), potentially expediting new drug effectivity assessments over the next few years. Consequently, there is a growing emphasis on ensuring the reliability of MRD results and the appropriate validation of underlying assays. Hence, to enable clinical decision-making based on MRD, it is essential to use an assay that adheres to the In Vitro Diagnostic Regulations (IVDR). Similarly, the U.S. Food and Drug Administration (FDA) requires comprehensive analytical validation of MRD assays prior to their use in interpreting drug efficacy in investigational clinical trials (U.S. Department of Health and Human Services OCoE, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, 2020; Wood et al., 2013). Currently, MFC-MRD assays vary among different centers and depend on extensive scientific expertise and knowledge of inter-patient differences for accurate interpretation of data (Grimwade & Freeman, 2014; Keeney et al., 2015). To harmonize interpretation and reporting, recently, the European Leukemia Network (ELN) MRD working party on AML has recommended the utilization of a single consensus panel of markers for MRD monitoring (Heuser et al., 2021).

In this current study, we report the validation of a semi-quantitative MFC-MRD assay using a 4-tube, 8-parameter panel incorporating cell differentiation markers recommended by the ELN AML MRD working party. The panel was originally designed by the AML MRD working group of the Dutch cytometry association and has been employed in multiple large clinical studies conducted by the Hemato-Oncology Foundation for Adults in the Netherlands (HOVON)/Swiss Group for Clinical Cancer Research (SAKK) (Löwenberg et al., 2021; Terwijn et al., 2013; Zeijlemaker, Grob, et al., 2019). While previous trials have provided a clinical validation, establishing a correlation between flow MRD positivity and outcomes, this manuscript solely focuses on the analytical validation. The validation principles adheres to the assay performance criteria proposed by Wood et al. (2013) and CLSI H62 (2021). The validation process of the MRD assay involves assessing assay accuracy using a separate MFC assay, determining analytical specificity by utilizing a predefined set of leukemia associated immuno-phenotype (LAIP) with established limit of blanks in regenerating/within-normal bone marrows. Additionally, it involves evaluating analytical and functional sensitivity (limit of detection and quantitation), examining precision/reproducibility, assessing linearity, evaluating sample/reagent stability, and establishing the assay background frequencies through a reference range. The validation is conducted for the entire assay and, although preferred, not for each LAIP separately as such an exercise would be impractical. Our results demonstrate that a well-validated AML MFC-MRD assay can accurately quantify a LAIP at diagnosis and MRD at follow-up. With the reliability of MFC assay readouts, MRD results could be used in clinical trial and support new drug approvals in the treatment of AML (U.S. Department of Health and Human Services OCoE, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, 2020). Importantly, our results provide a framework for other clinical laboratories to validate their own MFC panel or replicate our validated assay for use in their clinical practice or trials.

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | MATERIALS AND METHODS

The MRD qualification experiments were conducted at the Cancer Center Amsterdam. In addition, accuracy, sensitivity, and linearity experiments were carried out simultaneously at Navigate BioPharma, a subsidiary of Novartis. Furthermore, the inter-laboratory experiments involved the collaboration of the two aforementioned laboratories and three other laboratories affiliated with the AML MRD working group of the Dutch cytometry association, namely Erasmus MC in Rotterdam, Radboud UMC in Nijmegen and Medlon/Medisch Spectrum Twente (MST) in Enschede (Dutch Cytometry Association, 2022). All MRD measurements adhered to the protocol published by Cloos et al. (2018). Flow cytometry data acquisition was performed using FACS Canto-II flow cytometer (BD Biosciences, San Jose, CA, USA) and Navios cytometer (Beckman Coulter, Miami, FL, USA), with daily calibration following the EuroFlow procedures or Harmonemia settings (Kalina et al., 2012; Lacombe et al., 2016). Detailed information on the instruments and software used for the experiments can be found in Table S1.

2.1 | Antibody panel

The assay consists of four eight-color tubes, with each tube containing antibodies for a common set of backbone markers: CD13, CD34, CD45, CD117, HLA-DR, along with three additional antibodies for informative markers to enhance the identification of relevant LAIPs. The specific composition of the four tubes is detailed in Tables S2–S5. The choice of antibodies and their conjugated fluorochromes in the panels is carefully designed to minimize spectral overlap. The panel consists of antibodies against backbone markers to facilitate consistent gating of WBC, lymphocytes, CD45dim, and the primitive marker. The additional lymphocytic and monocytic markers are included in one tube and are used for the identification of all other relevant populations, including the determination of LAIPs.

2.2 | Reagents

In addition to the antibodies, we utilized Phosphate Buffered Saline (PBS) solution containing 0.05% sodium azide and 0.1% Serum Albumin, along with an NH4Cl-based lysing solution. Monoclonal antibody (MoAb) cocktails were prepared regularly (approximately on a weekly basis) using the reagents to make liquid cocktail with volumes sufficient for 60 MRD tests. For a comprehensive overview of the reagents used, see Table S6, while the stability results of the antibody panel used can be found in Section 3.6.

2.3 | Patient specimens and LAIP interpretation

All patient specimens were obtained after informed consent according to institutional guidelines. These patients participated in either HOVON/SAKK or Novartis sponsored CPKC412E2301 clinical trial or received treatment at the Amsterdam University Medical Center (location VUmc). For the validation experiments, fresh frozen samples of BM and peripheral blood (PB) were used. An overview of samples used for each experiment can be found in Table S7. The analysis of samples was performed using Infinicyt™ software version 1.8 (Cytognos, Salamanca, Spain), FlowJo version 10.5.3 (BD Biosciences) or KALUZA™ analysis software version 2.1.3 (Beckman Coulter). Detailed procedures for identifying LAIP at diagnosis and assessing MRD at follow-up are described in Cloos et al. (2018) and Zeijlemaker, Kelder, et al. (2019). The gating strategy for each LAIP is visualized in Figure S1, although adjustments of gates may be necessary for each individual specimen. A table of all immature LAIPs can be found in Table S8. Cell viability of fresh specimens were analyzed using a cellometer (Nexcelom, Lawrence, MA, USA). In addition, samples were assessed to ensure the absence of clumps, clotting, hemolysis, contamination, and debris (as observed in the forward scatter (FSC) versus side scatter (SSC) plot) prior to their use.

2.4 | Statistics

Statistical analyses were performed with IBM SPSS Statistics 26 (Armonk, NY, USA) and the results were visualized with GraphPad prism 9 (San Diego, CA, USA). The Mann–Whitney *U*-test was used to compare non-parametric unpaired variables. Correlation coefficients were calculated using the Pearson Correlation test, with a significance level set at $p < 0.05$ (two-sided).

3 | VALIDATION EXPERIMENTS AND RESULTS

The design and results of each validation experiment are presented along with the specific acceptance criteria for testing and the corresponding outcomes. A summary of all the results can be found in Table 1.

3.1 | Accuracy

Although it would be preferable to determine accuracy of a test by an orthogonal method (i.e., molecular qPCR or NGS), we considered these methods to be not suitable for determining closeness of agreement based on previous observations of the discrepancies and because the origin of these discrepancies are still unknown (Jongen-Lavrencic et al., 2018; Morita et al., 2018; Venditti et al., 2019). Instead, we determined accuracy by comparing our experimental results to an alternative flow cytometry assay routinely used for diagnosing hematological malignancies at VUmc (Zeijlemaker, Kelder, et al., 2019). The comparison was based on the percentage LAIPs and blasts, using Pearson correlation coefficient (PCC), with a threshold of ≥0.9 considered acceptable (Narayanan & Weinberg, 2020). Eleven

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Summary of quality assessment, acceptance criteria and results per specific elements of the MRD assay.

Validation parameter	Experimental design	Acceptance criteria	Results	Pass/fail
Accuracy	%Blasts and %LAIP of 11 AML diagnosis with 21 LAIPS compared in two flow panels: MRD and immunophenotyping	Pearson correlation coefficient (PCC) ≥0.90	PCC %blasts 0.99, %LAIP 0.93	Pass
Specificity	MFI's of positive and negative populations were assessed on NBM, FU and diagnosis samples	Expression consistency for backbone markers with CV ≤20% for 80% of the samples	Mean %CV is 9.9, and in 82.1% of the samples the %CV is <20%	Pass
Limit of blank (LoB)	BM from four healthy donors with four LAIPS analyzed with five replicates	No specific acceptance criteria	LoB = 0.024	Not applicable
Lower limit of detection (LLoD)	Bone marrow mononuclear cells (BMMC) from four AML were stained and spiked at low levels (<0.1% of total WBC) into four unique healthy donor BM which were pre-stained with the total panel to detect four LAIPs and acquired in triplicate	<0.1%	LLoD = 0.036	Pass
Lower limit of quantification (LLoQ)	BMMC of four AML patients with five unique LAIPs acquired in triplicate	The triplicates should have %CV ≤20 and be within 30% of the expected value	LLoQ = 0.1	Not applicable
Inter-operator precision	Three samples acquired by two different operators	≤20 %CV	%CV = 0.1%–14.0%	Pass
Inter-gating precision	Two samples gated by four technicians	≤20 %CV or identical MRD% for low values	%CV = 0%–66.33% with low SD and similar MRD%	Pass
Inter-instrument precision	Three samples measured on three different FACS Canto II	≤20 %CV	%CV = 0.9%–8.2%	Pass
Intra-assay precision	13 sample measured in three parts by the same operator on the same instrument	≤20 %CV	%CV = 0.1%–10.0%	Pass
Inter-laboratory precision	Qualitative: Samples acquired at two different laboratories	>90% similar results in two centers	93.9% similar results	Pass
	Quantitative: Samples processed at two different laboratories	≤20 %CV for at least 80% of LAIPs	%CV ≤20 for 94.4% of LAIPs (0.39%–24.01%)	Pass
Linearity	1. Expected amount of MRD compared with measured MRD for 12 samples	Correlation ≥0.90	Correlation = 0.99	Pass
	2. Spiking BMMCs from four AML patients, into four healthy donors to achieve an expected tumor target range of 20% to 0.1% of total WBC	Correlation ≥0.90	Correlation =0.995	Pass
Specimen stability De Novo BM	Five samples with 11 LAIPs were measured on consecutive days during 1 week, %LAIP were compared to day 1	LAIPs should be present, populations stable and ≤20 %CV	Stable until day 5	LAIPs can be determined reliably in BM until day 5

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Validation parameter	Experimental design	Acceptance criteria	Results	Pass/fail
Specimen stability follow up BM	Seven samples with 15 LAIPs were measured on consecutive days during 1 week, %MRD was compared to day 1	MRD result should be the same as day 1 and populations should be stable (%CV ≤20)	Stable until day 7, %CV of populations were < 20 until day 4	MRD results in BM can be reliably obtained until day 4
Stability of MRD at low WBC	Three AML specimen spiked into healthy donor PB at a concentration of 0.1% of total WBC	Frequency of LAIP at %CV ≤20 with >50 events	Event count became below 50 when 100,000 WBC were acquired	A minimum of 200,000 WBC is required for MRD-positive result
Specimen stability De Novo PB	Five samples with 12 LAIPs were measured on consecutive days during 1 week, %LAIP were compared to day 1	LAIPs should be present, populations stable and ≤ 20 %CV	Stable until day 4	LAIPs can be determined reliably in PB until day 4
Reagent stability	Cryopreserved samples were measured at day 1, 7, 14, 28 and 56 upon staining with the same MRD panel	Panel is stable when all %CV compared to day 1 are ≤20%	At day 28, all markers had <20% CV. At day 56, CD15 had a CV of 20.66%	Panel is stable until day 28
Reference range	LAIPs were measured from 10 NBM and 300 AML samples	Background LAIP should be below 0.1% MRD.	All background %LAIP on NBM were <0.1% (range: 0%–0.04%)	Not applicable

Abbreviations: % CV, coefficient of variation expressed as a percentage; LAIP, Leukemia associated phenotypes; NBM, normal bone marrow; SD, standard deviation.

patients were included in the comparison, resulting in a total of 21 LAIPs across both assays (Cremers et al., 2017; Ossenkoppele et al., 2011). These LAIPs consisted of CD45 + CD34 + CD13 + CD7+ ($n = 3$), CD45 + CD34 + CD13 + HLA-DR− ($n = 4$), CD45 + CD34 + CD13 + CD2+ ($n = 1$), CD45 + CD34 + CD13 + CD22+ ($n = 1$), CD45 + CD34 + CD13 + CD33− ($n = 4$), CD45 + CD13 + CD133 + CD34− ($n = 1$), CD45 + CD34 + CD33+/CD13− ($n = 1$), CD45 + CD34 + CD13 + CD15+ ($n = 2$), CD45 + CD34 + CD13 + CD11b+ ($n = 1$) and CD45 + CD34 + CD13 + CD56+ ($n = 3$). The concordance between the two assays was 0.99 for blast percentage and 0.93 for LAIP percentage (see Figure 1). The largest discrepancy was observed in the LAIP expressing CD15, although the same clone was used. The MRD assay showed relatively lower CD15 expression compared to the immunophenotyping assay, but these differences did not influence result interpretation as the frequency of these cells was above 10% (or 100 fold above the MRD cut-off) in both assays routinely used for LAIP identification at diagnosis. Furthermore, consistent results were obtained when the same sample was measured in different laboratories (see Section 3.4.4), further supporting the accuracy of the assay.

3.2 | Analytical specificity

The assay's analytical specificity was determined by the capability of each marker to identify the desired cell population while accurately discriminating between multiple other cell types found in the sample. To demonstrate specificity of the antibodies used for AML MRD detection, five healthy donor specimens representing bone marrow and peripheral blood were stained using each of the four LAIP test tubes representing all the CD markers used for identification of AML MRD. The specificity of each antibody was analyzed by gating internal positive and negative control populations. Each constituent antibody clearly identified internal positive and negative control populations and normal BM and PB samples showed <0.1% of AML MRD burden of total WBC, a cut point utilized for reporting MRD status (see Figure 2). To further verify antibody specificity, we compared mean fluorescence intensity (MFI) of the test antibodies on the population of interest (blast cells) to a control population (lymphocytes), although other factors such as gating strategy and interactions with other reagents are also important. Consistency in MFI should be consistent for the backbone CD markers in all four MRD test tubes, with a Coefficient of Variation (CV) of ≤20% for 80% of the measurements. The MFI ratio between internal positive and negative control cell populations should be >2. To evaluate these criteria, we assessed the MFIs produced by all test antibodies on BM obtained from three healthy donors and two AML patients (diagnostic and follow-up specimens). A table of the inter positive and negative controls used in this validation are stated in Table S9. The %CV for all backbone markers was below 20% (median %CV of 8.3% for CD117, 7.2% for CD34, 10.6% for HLA-DR and 13.6% for CD13), indicating highly reproducible results (Table S10a). The ratio of internal positive and negative control populations varied between 2.7 and 4127.9, clearly distinguishing the

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

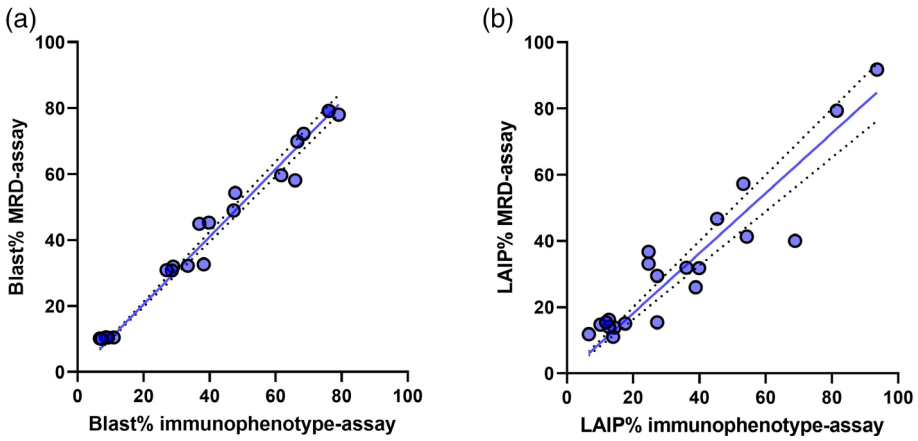


FIGURE 1 Correlation between diagnostic immunophenotypic-assay and MRD assay. (a) Scatter plot demonstrating the correlation between the percentage of blasts detected using the MRD-assay (*y*-axis) and the immunophenotype-assay (*x*-axis) in 11 different bone marrow (BM) diagnosis samples. The analysis includes 21 leukemic-associated immunophenotype (LAIP) markers, yielding a Pearson correlation coefficient (PCC) of 0.99. (b) Scatter plot illustrating the correlation between the percentage of LAIP cells in the same samples using the MRD-assay (*y*-axis) and the immunophenotype-assay (*x*-axis), resulting in a PCC of 0.93. [Color figure can be viewed at wileyonlinelibrary.com]

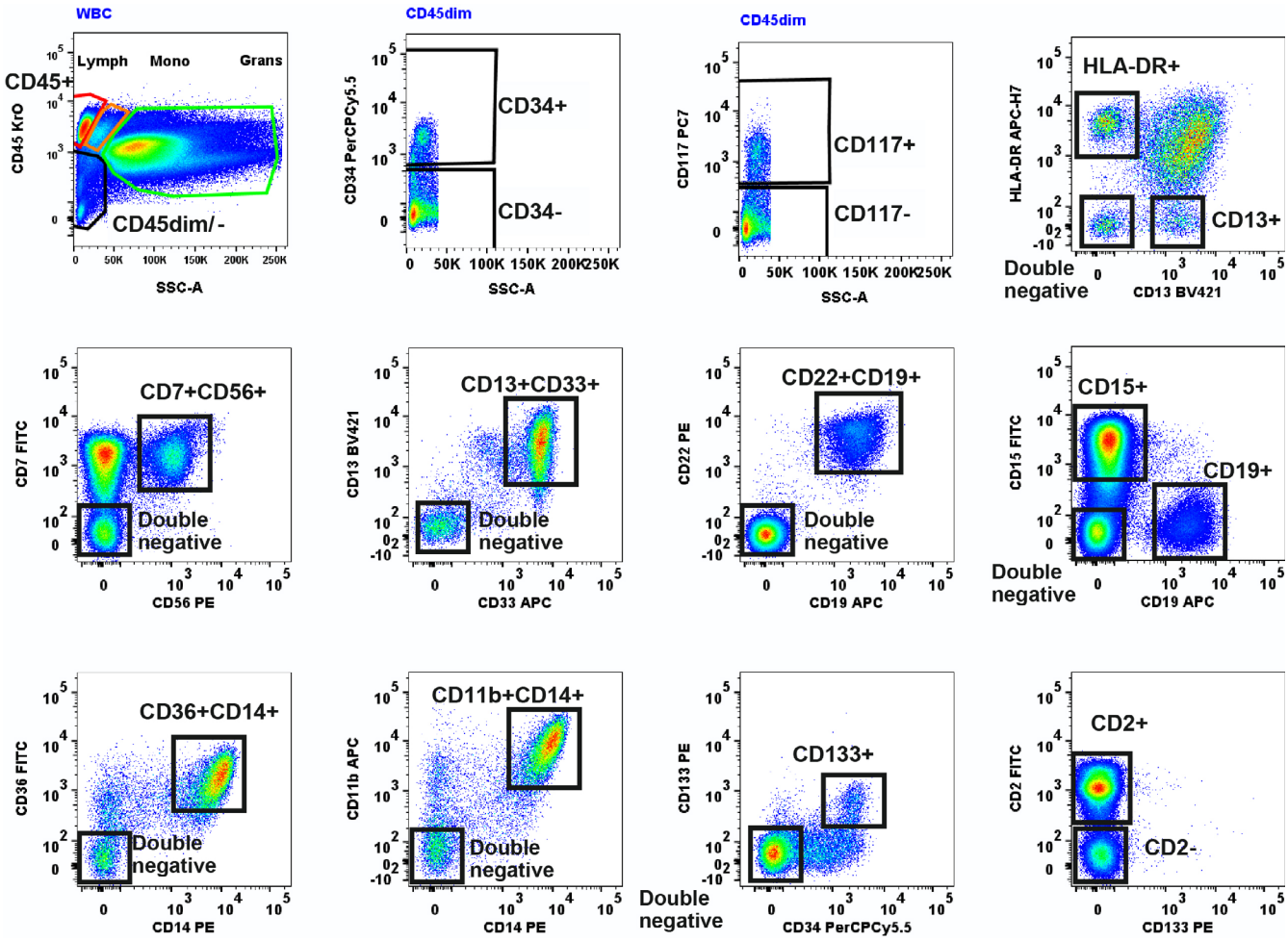


FIGURE 2 Assay specificity. The flow cytometry plots depict the specificity of the assay, demonstrating the recognition of target-specific populations by each antibody used for detection of AML MRD. [Color figure can be viewed at wileyonlinelibrary.com]

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

populations in scatter plots (Table S10b). The MFIs produced by all other test antibodies also showed clear distinction between the internal positive population (blast cells) and negative control (lymphocytes). In addition, specific markers were measured with blast cells as the negative control and another cell type as internal positive population (e.g., CD14 with monocytes as the target population). In summary, internal positive and negative control populations were clearly distinguishable for all markers, and the MFI was highly reproducible across backbone markers (Mean %CV 9.9% and CV <20% for 82.1% of the observations).

3.3 | Detection capability: Limits of blank, detection and quantitation (sensitivity)

Analytical and functional sensitivity of the assay was estimated by measuring either healthy donor samples without LAIP or AML specimens with varying frequencies of LAIP at which MRD could still be assessed accurately. Limit of blank (LoB) (maximum number of LAIP cells acquired in samples lacking leukemia) was calculated using the formula (Mean + 1.645 standard deviation (SD)) and Lower limit of detection (LLoD) (minimal number of LAIP cells that can accurately be distinguished above background) calculated using the formula (LoB + 1.645 SD of low LAIP specimens) in specimen with a spiked MRD level of 0.01%. Lower limit of quantitation (LLoQ) was defined as the value where triplicates exhibited a %CV ≤20 and the MRD value fell within 30% of the expected value for the lowest dilution of LAIP specimens (Donnenberg & Donnenberg, 2007). The acceptance criteria for all LAIPs regarding LLoD are set to be below 0.1%. However, it is desirable for the LLoD to ideally be 0.01%.

Bone marrow mononuclear cells (BMMC) from five AML patients were spiked into three BM samples from three healthy donors to achieve a low level tumor specimens (e.g., 0.01% of total WBC). All serial dilutions generated were stained with the complete AML MRD panel representing real clinical trial/diagnostic specimen testing. Each healthy donor BM samples lacking spiked AML BMMCs cells served as a LoB control. LoB was assessed in BM of four healthy donors with four LAIPS analyzed with five replicates, generating a total of 20 measurements. Mean of blank specimens was 0.024 with a SD of 0.005 (Table S11). To establish LLoD, cryopreserved BMMCs from four AML patients were stained and spiked at low levels (<0.1% of total WBC) into five unique healthy donor BM which were pre-stained with the total panel. The LAIPs used can be found in Table S9. This resulted in the detection of four LAIPs in triplicates. The SD for 12 low positive samples was 0.0025%. LLoD was determined to be 0.036% of total WBC (see Table S12). The lowest quantifiable levels of five unique LAIPs among four AML patients evaluated ranged from 0.1% to 0.01% (Table S13) with acceptable imprecision in replicate assessments and within ±20% CV of expected spike value across 15 measurements (% CV range = 0.0–15.8).

3.4 | Precision

To determine the precision of the assay, multiple aspects were evaluated. Figure 3 provides a visual representation of the precision experiment conducted in this study. Inter-operator variability was assessed by involving two operators who processed and measured three AML samples. The inter-gating was assessed by having two gating experts independently perform gating on two additional samples. Inter-

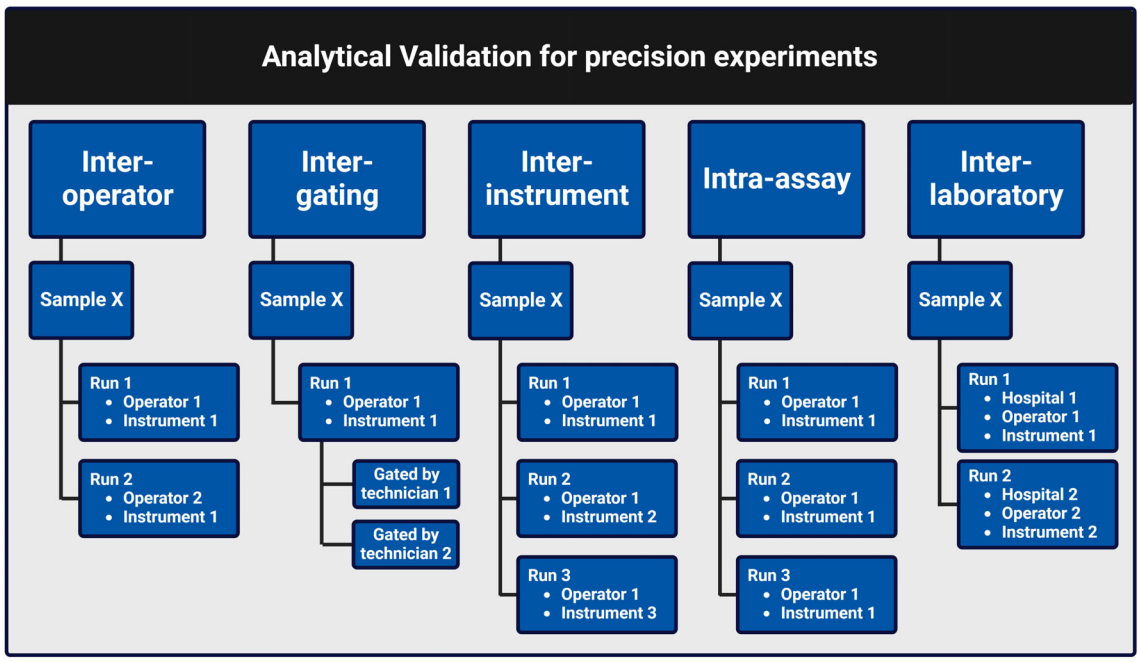


FIGURE 3 Design of the precision experiments. Visual overview of various precision experiments, including inter-operator, inter-gating, inter-instrument, intra-assay and inter-laboratory comparisons. [Color figure can be viewed at wileyonlinelibrary.com]

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

instrument variability was examined by measuring a sample across three different instruments, while intra-run variability was evaluated by dividing one sample through three sub-runs. Additionally, inter-laboratory variability was assessed among five different laboratories by measuring a sample and a sister-sample in two different laboratories. In precision studies, a target % CV of ≤20% was set as the acceptance criteria. However, for samples containing very low frequencies of tumor cells, the absolute frequencies of events were also compared. Results of precision studies are shown in Figure 4.

3.4.1 | Inter-operator precision

To assess the variability between operators, we selected three AML patient samples and processed them by two different operators. The samples were acquired on the same day and measured on the same flow cytometer. Among the three samples, one sample was determined to be MRD-negative (LAIP percentage below 0.1%). For all parameters (% WBC, %blasts, %primitive marker and %LAIPs) measured in these samples, we observed a low SD of 1.0 and a CV of less than 20% (ranging from 0.1% to 14.0%) (see Figure 4a and Table S14 for detailed results).

3.4.2 | Inter-gating precision

To ensure the robustness of the LAIP interpretation, two additional technicians independently analyzed the raw data from the two MRD+ patients to evaluate the inter-gating variability. The interpretation of 11 out of the 14 gates showed high reproducibility (<20% CV). The variability in the frequency of the LAIPs was also relatively low (SD ranging from 0.0 to 0.39) (see Figure 4d and Table S15 for detailed analyses). Two out of the fourteen gates exhibited higher variability

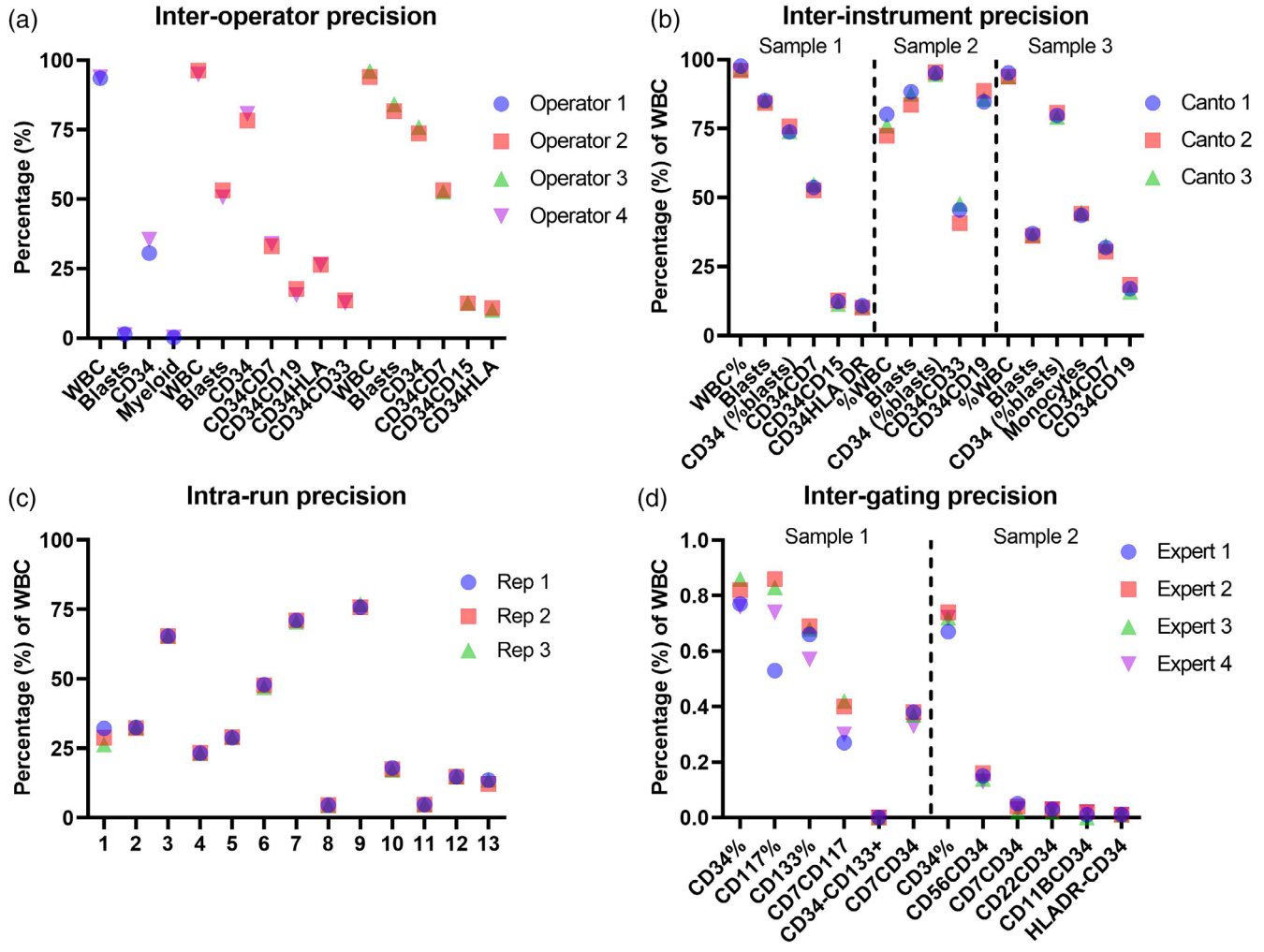


FIGURE 4 Results for the precision experiments. (a) Inter-operator precision was evaluated by comparing the measured results of all components for three AML samples, assessed by two different operators/technicians. A total of four different staff members were involved. Coefficient of variation (CV) ranged from 0.1% to 14%. (b) Inter-instrument precision was assessed by comparing the results obtained from three flow cytometry instruments (all FACS-Canto II) when measuring the same three AML samples (The %CV range was 0.4–8.2). (c) Thirteen AML samples were measured in triplicate, revealing a SD range of 0.0–2.9 and a %CV range of 0.1–10.0. (d) Inter-gating precision was determined by four different technicians who gated two AML samples. The comparison of all gates showed an SD range of 0.00–0.39 and a %CV range of 0.00–31.94. [Color figure can be viewed at wileyonlinelibrary.com]

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(CD34 + CD13 + CD7+ and CD34 + CD13 + CD11b+; range 0.0%–66.33%), which can be attributed to their low frequencies in the MRD-negative samples. In samples with frequencies below 0.1%, the CV is not considered an optimal parameter for assessing variability.

3.4.3 | Inter-instrument precision

Three AML de novo patient samples were acquired using three different FACS Canto II instruments operated by the same individual. The obtained results were analyzed and checked by two different operators. The frequencies of WBC, blasts, primitive compartments (CD34+ cells) and LAIPs were compared across the samples. The calculated %CV for all cell populations were below 20% (range 0.9%–8.2%). For further details, see Figure 4b and Table S16.

3.4.4 | Intra-assay precision

Thirteen patient specimens were divided into three parts and processed separately. The sample processing was performed by a single operator and the data were acquired on a single instrument. The analysis of the data was performed by the same expert. Across the triplicate measurements of the 13 patient specimens, all LAIP frequencies were found to be below 20% (range 0.1%–10.0%). For detailed information, see Figure 4c and Table S17.

3.4.5 | Inter-laboratory precision

A total of 82 fresh diagnostic AML samples were acquired for MRD analysis at the laboratory of Amsterdam UMC, location Vrije Universiteit. Of these samples, 64 were also acquired simultaneously at Erasmus MC, 14 at Radboudumc and 4 at MST laboratories. In 77 of 82 samples (93.9%), both laboratories identified a LAIP above or below the 0.1% cut-off, generating a similar result in both centers. Table S18 provides a detailed overview of these findings. To ensure comprehensive inter-laboratory precision, five AML diagnosis samples were selected, containing 18 different LAIPs and representing a range of tumor cell frequencies. These samples were initially analyzed at the Amsterdam UMC laboratory and duplicate specimens were frozen and sent to Navigate BioPharma laboratory in the United States for concurrent assessment, to mimic a global clinical trial. The results of these five samples are shown separately due to the possible influence of the freeze–thaw process. Antibody reagents and flow cytometer settings were standardized across all laboratories. The percentages of LAIP are shown in Table S19. Across the two laboratories, the frequencies of LAIPs acquired demonstrated a CV below 20% for 17 LAIPs. The only exception was the CD45 + CD34 + CD13 + CD11b+ LAIP, with results ranging from 0.39% to 4.01% across laboratories.

3.5 | Linearity

Linearity demonstrates an assay's ability to obtain results that are directly or indirectly proportional to the concentration of an aberrant subpopulation in a sample within a specified range. We employed two methods to assess linearity. In the first method, LAIP-positive cells were spiked into the total WBC compartment of normal bone marrow specimens, and the measured values were compared with the expected (calculated) values. The obtained results were plotted, and the regression analysis (PCC) was performed based on the mean of all measurements (see Figure 5a). A total of 11 BM samples and one PB

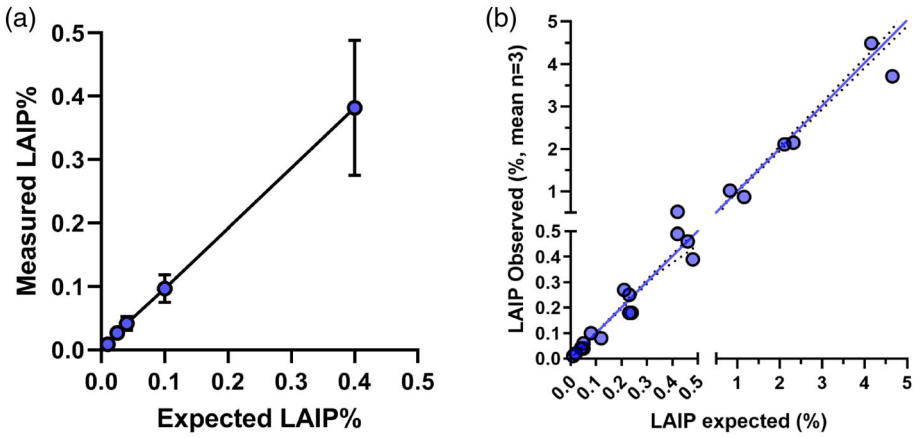


FIGURE 5 Linearity assessment. (a) Correlation plot with percentage of expected LAIPs (*x*-axis) versus measured LAIPs (*y*-axis) at five different concentration levels (0.01%, 0.025%, 0.04%, 0.1% and 0.4%, respectively). The Pearson correlation coefficient (PCC) is $r = 0.99$ ($p < 0.001$), indicating a strong correlation between expected and measured values. (b) Percentage of expected LAIP-positive cells (*x*-axis) compared to the mean percentage of three measurements (*y*-axis). One of 22 measurements (Expected% 16.64, observed% 16.98) is excluded from the plot for better visualization. The overall PCC is $r = 0.995$ ($p < 0.001$), again demonstrating a high correlation between the expected and observed values. [Color figure can be viewed at wileyonlinelibrary.com]

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(relapse) sample were used to achieve 12 serial dilution of LAIP at five concentrations. The overall PCC was determined to be $r = 0.99$ ($p < 0.001$), with the largest deviation observed at 0.4% dilution (SD = 0.107). For the second method, linearity was assessed by spiking BMMCs from four AML patients into three BM and one PB obtained from four healthy donors. This approach aimed to cover a tumor target range of 20%–0.1% of total WBC. A total of 22 LAIPs were detected across the four specimens, with each specimen tested in triplicate. Pearson's correlation analysis was performed by comparing the observed averages (from triplicates) with the expected normalized values obtained from the sensitivity (LLoQ) study. The overall PCC of the 66 measurements was determined to be $r = 0.995$ ($p < 0.001$), as illustrated in Figure 5b.

3.6 | Stability

Stability of specimens and reagents is essential for obtaining reliable MRD results, as logistical constrains may prevent samples from being acquired within 24 h of collection, likewise, preparing fresh antibody cocktails for each specimen is impractical. To assess specimen stability, the LAIP percentage was measured in BM (at diagnosis and at follow-up) and PB samples (at diagnosis only) over consecutive days up to 7 days, comparing each measurement to the initial measurement on day 1. The specific LAIP can be found in the Supplementary Tables. In addition, a fixable viability dye (BD, Fixable Viability Stain 780) was utilized to assess the percentage of viable cells at each time point. The results were used to identify the time range in which samples remain stable, ensuring reliable identification of LAIPs. The specimens were stored at room temperature to mimic an ordinary diagnostic laboratory setting.

To assess stability of antibody reagents used in the MRD assay; the MFIs of each CD marker on relevant aberrant cell populations were quantified in three cryopreserved AML BM samples. These samples were stained and acquired with the antibody cocktail on Days 1, 7, 14, 28 and 56 days after the initial measurement on the same cryopreserved specimens.

3.6.1 | Stability of bone marrow diagnosis specimens with higher tumor burden

For the specimen stability assessments, five specimens were selected with LAIPs above the 10% threshold on day 1. In these samples, 10 of the 11 LAIPs were still measurable above the 10% cut-off after 5–7 days, as indicated in Table S20. Eight out of eleven (73%) LAIPs remained stable until day 5, with a CV below 20% compared to day 1. However, only 4 out of 11 LAIPs remained stable for 7 days, as shown in Table S20. Interestingly, in one sample, the CD34/CD15 double positive LAIP increased over time, surpassing the 10% threshold 6 days after the initial measurement. To verify the adequacy of LAIPs in specimens with low-cell count, five additional BM samples were diluted 1:10 with AB plasma on day 1 and processed similarly to the undiluted samples. In comparison to the undiluted specimens described earlier, the diluted samples showed even less variation, and all LAIPs (except for CD15) could be detected after 7 days. The LAIP frequencies remained stable until day 4, with 7 out of 10 LAIPs having a %CV below 20%, as detailed in Table S21.

3.6.2 | Stability of post-treatment bone marrow samples with low tumor burden

Over the course of five successive days, six AML BM follow-up specimens and one healthy donor BM were tested. These specimens contained a total of 15 LAIPs, with six LAIPs above the 0.1% MRD cut-off on day 1 of testing. All six LAIPs remained MRD-positive on subsequent days (see Table S22). On day 7, only one LAIP became MRD-positive, while it had been MRD-negative in the previous measurements (specifically, CD34+/CD15+; 0.06% on day 1 vs. 0.14% on day 7). The MRD-negative LAIPs decreased with a maximum of 0.02% within the first 3 days after the initial measurement and 0.06% after 5 days. In general, the LAIP cells remained stable within the first 4 days and no change in MRD status was observed during the first 6 days of testing.

3.6.3 | Stability of peripheral blood specimens

The same experiment was repeated for five PB specimens collected at the time of diagnosis. These samples are particularly important as they allow for LAIP determination upon diagnosis, while the availability of a BM sample might be intermittent. Out of the 12 LAIPs identified above the 10% cut-off, nine LAIPs remained above the cut-off 6 days after initial measurement. However, it is worth noting that PB specimens exhibited less stability compared to BM specimens, with only 6 out of the 12 LAIPs stable on day 6 (with a change of less than 20% from the day 1 measurement). Detailed values for each LAIP can be found in Table S23.

3.6.4 | Stability of MRD at low WBCs

As variation in cell yields is a common occurrence in specimens, we assessed the impact of total acquired WBC events on accurate MRD interpretation. Three independent AML patient specimens were spiked separately into healthy donor peripheral blood at a target concentration of 0.1% (MRD cut-off) of total WBC and stained with the respective LAIP antibody tube. Varying number of total WBC events ranging from 0.025 to 1.0 million were acquired for each LAIP tube. Across the three AML patients, six LAIPs were identified. The final frequencies of all LAIPs remained consistent at 0.1% (with an acceptable CV of 20%) across the range of WBC enumerated (see Figure 6). However, for two out of a six LAIPs, the absolute count fell below the threshold of 50 events (lower limit for MRD interpretation), even when 100,000 WBC were acquired. Hence, to ensure reliable MRD-

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

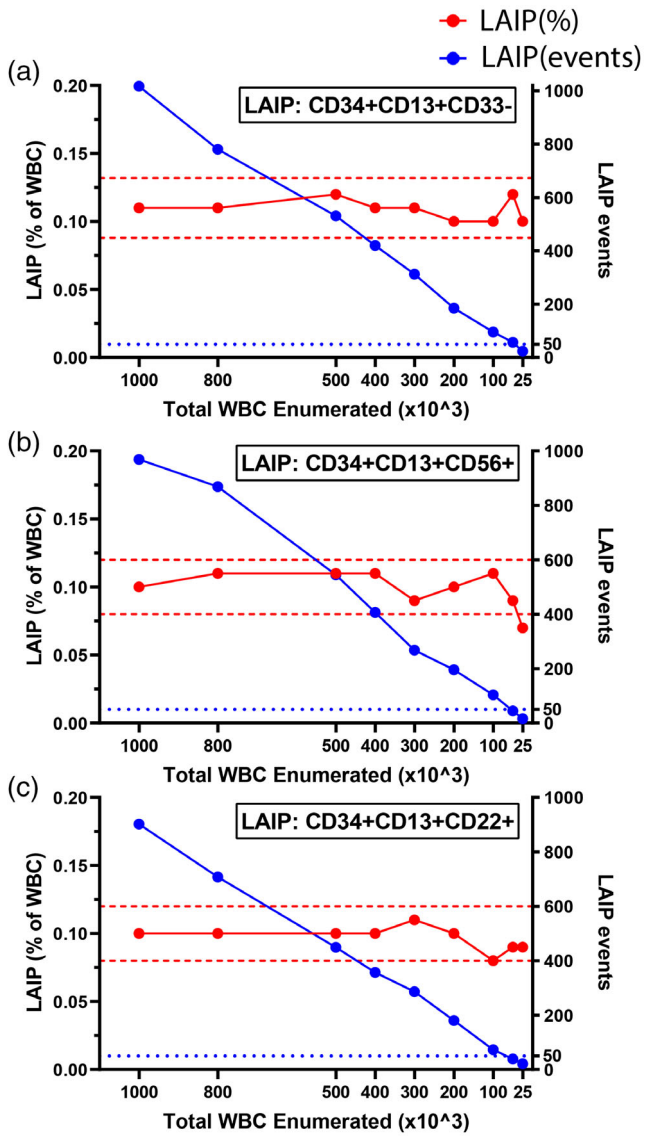


FIGURE 6 Stability of MRD at low white blood cells levels. This figure depicts the frequencies (solid red line) and total events (solid blue line) of representative LAIPs observed in each of three AML subjects across a range of total WBC enumerated. The green dotted lines represent the ±20% change from the expected value, indicating the stability range. The blue dotted line represents the cut point for a minimum of 50 LAIP events, ensuring sufficient sensitivity for MRD detection. [Color figure can be viewed at wileyonlinelibrary.com]

positive interpretation, we recommend evaluating a minimum of 200,000 total WBCs and at least 50 clustered LAIP events.

3.6.5 | Reagent stability

The stability assessment of the MRD antibody panel involved the measurement of MFI for specific MRD-defining cell populations in cryopreserved AML BM samples collected at time of diagnosis. Multiple vials from the sample were cryopreserved, and on day 1, all samples were stained with the same combination of antibodies. Samples were thawed on day 1, 7, 14, 28 and 56. Figure S2 provides the median MFI values for all markers used to identify abnormal populations within the blast population of each patient. MFI values for all CD markers in all samples, at all time-points, showed a high level of reproducibility (<20% CV) until Day 56, except for CD15, which remained stable until Day 28, as indicated in Table S24.

3.7 | Reference range

To establish reference ranges observed with this assay in both healthy donors (negative) and AML patients (positive), we analyzed LAIP data from 10 healthy donor BM and 300 post-treatment AML BM samples from patients in CR(i) who were participating in the HOVON 102 trial (Zeijlemaker, Grob, et al., 2019). The healthy donor samples have been measured only once and are different from the samples used for the LoB experiments. The LAIP percentages in the healthy donor BM samples were below the MRD interpretation cut-off of 0.1%, ranging from 0% to 0.04%). In contrast, the LAIP frequencies in the positive AML patient samples varied from 0.1% to 30%. These results confirm the reliability of our assay for wider application (see Figure 7 for visualization of the results).

4 | DISCUSSION

For accurate clinical decision making using a MFC-based MRD assay for AML patients, it is essential to use a rigorously validated and reliable assay that meets current regulatory expectations (Armbruster & Pry, 2008; der Strate et al., 2017; Du et al., 2015; Selliah et al., 2019; Wood et al., 2013). Our results demonstrate that our standardized MRD assay has met all predefined acceptance criteria (Table 1). However, it is important to note that additional parameters such as qualitative and quantitative specificity, which were not reported in this study, may be required for compliance with the IVDR. Our experiments have some limitations. One notable limitation is the challenge in conducting accuracy experiments due to the absence of an appropriate reference assay. Although techniques like next-generation sequencing (NGS) can be utilized to measure MRD, they are complementary to flow cytometry assays and not entirely interchangeable. This is because not all LAIP+ cells contain mutations, rendering NGS less suitable as reference assay (Jongen-Lavrencic et al., 2018; Patkar et al., 2021). However, once further understanding of these discrepancies is achieved, NGS may serve as a valuable reference for future accuracy experiments. Due to the quasi-quantitative character of the assay and the lack of reference standard, the direct applicability of the assay accuracy is limited. It has to be emphasized that not all experiments could be performed with native specimens, most notably samples containing very low levels of LAIPs required for sensitivity experiments, as obtaining such samples proved challenging. Hence, we had to resort to diluting samples with higher MRD burden in order to establish the LLoD and limits of quantitation (MRD cut-off). Furthermore, it is worth noting that the inter-instrument experiments

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

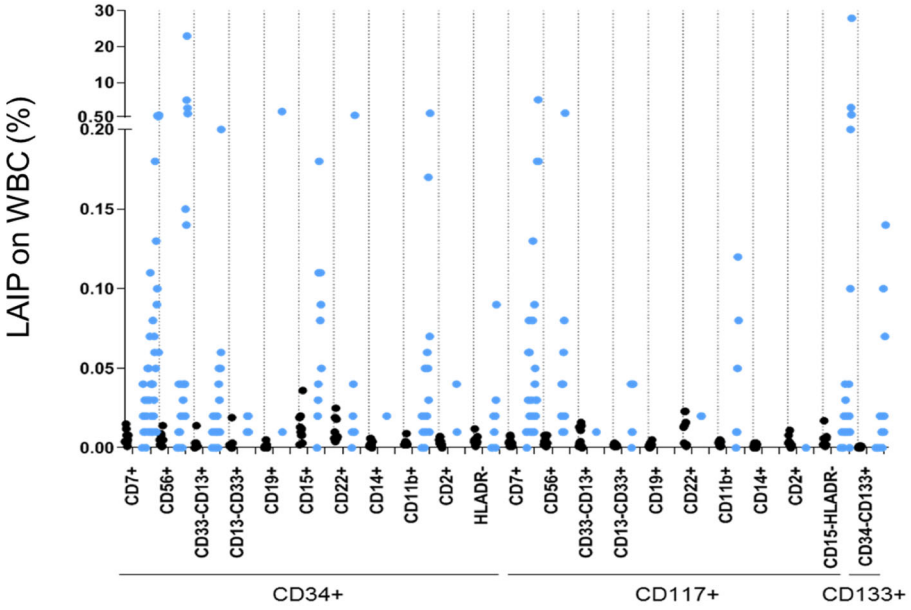


FIGURE 7 Reference LAIP percentages for different LAIPs. The LAIP percentage represents LAIP events gated on different primitive markers (CD34, CD117 and CD133). The black dots represent LAIP percentages from individual NBM samples, while the blue dots represent the LAIPs from AML samples. It is important to note that a single dot can represent multiple measurements if the results overlap. [Color figure can be viewed at wileyonlinelibrary.com]

were carried out using de novo samples, which may potentially underestimate the observed effect. While the majority of validation acceptance criteria are often based on %CV, this parameter may not be optimal for assessing very low levels of LAIPs. Although certain individual comparisons showed higher variation than the desired CV of ≤20%, these were mainly in the very low range. Importantly, none of the experiments resulted in misclassification of MRD results based on a cut point of 0.1%. Differences in gating strategy among individual operators could introduce some variability that cannot be eliminated in each measurement. In addition, pipetting errors may contribute to small differences between samples. The limit set by health authorities to demonstrate detectability of LAIPs at a 10-fold lower level than the MRD cut-off could not be achieved, potentially due to presence of non-leukemic hematopoietic precursors in regenerating bone marrow at very low levels, as observed in healthy donors included in our reference range study (Figure 7). Finding appropriate samples for LLoD/LLoQ experiments posed a challenge. Firstly, a sample needed to be assessed to determine if it was MRD-positive with a MRD percentage in the lower range. Subsequently, the sample needed to have sufficient remaining cells for the LLoD/LLoQ experiments. Inter-laboratory experiments showed a discrepancy in 5 out of 82 (6.1%) MRD results, which urged further investigation. After discussion and data re-analyses, we concluded that the discrepancy was most likely due to hemodilution, as each laboratory received different pulls of the BM sample and %LAIP differences were in the same ratio between the pulls. We are currently conducting further investigations to identify potential markers that may indicate hemodilution and to assess the extent of its influence on MRD results.

These validation experiments have led to the establishment of best quality practices that are now prerequisite in our laboratory for releasing MRD results for clinical decision making. These practices include the following; (1) Post-treatment samples should be tested within 7 days after collection; (2) caution is warranted for samples collected in tubes with anticoagulant other than heparin, as specimen stability in other anticoagulants may vary (e.g., EDTA); and (3) when a sample has limited cell availability, the tube with the LAIP observed at diagnosis should be tested first. Notably, the results indicated that acquiring 200,000 CD45 expressing cells at follow-up still resulted in an accurate MRD%, which may be used when limited material is available, in particular for MRD-positive samples. The consensus remains that a reliable MRD-negative result requires the acquisition of 500,000 events in all tubes. Additionally, it is becoming increasingly relevant to state the achieved sensitivity in MRD reporting to the clinic. When the data are deemed unreliable based on any of the criteria but is critical for clinical decision making, a second sample taken after 2 weeks can be used to confirm the results. It has to be emphasized that in addition to having a validated assay, care should be taken to ensure all required steps are in place prior to clinical specimen testing. This included instrument set up, critical reagent qualification, verification of specimen quality via viability and hemodilution assessments and utilization of harmonized criteria for interpreting (gating) raw data (Tettero et al., 2022). In addition, it should be noted that new therapies/agents targeting specific CD markers present on AML cells, which are used in the current assay, can potentially impact the reliability of the assay. Therefore, the potential influence of new agents on the LAIP and, consequently, the accuracy of the assay should always be investigated before incorporating them into clinical decision-making. Additionally, the specific treatment received and time point at which the sample is collected can influence the regeneration patterns, potentially mimicking a LAIP post-treatment and affecting the background levels of the LAIPs. Therefore, it is advisable to measure both normal bone marrow and regenerating bone marrow to gain knowledge about transient clonal shifts during therapy. This is particularly important for achieving proper LoB/LLoD. Moreover, ensuring a correct gating strategy, including the use of back gates and careful avoidance of gates passing through populations, is essential.

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

These guidelines are especially important for acute monocytic leukemia, which can be challenging to gate due to the high overlap with regenerating bone marrow. We recommend utilizing back gating in the CD45 versus SSC-plot to reduce false-positivity. Yet, the LLoD/LLoQ may be higher for this more mature AML phenotype.

To conclude, our extensive dataset demonstrates that the MFC assay described in this manuscript fulfills all validation requirements and is capable of reporting reliable and accurate MRD results. These results include assay accuracy, specificity, sensitivity, precision, linearity and stability, which are fundamental for the correct identification of rare subpopulations. We anticipate that our approach will assist other laboratories expediting their validation activities to meet recent health authority guidelines and facilitate the implementation of novel effective therapeutic modalities once MRD is approved as a surrogate endpoint.

ACKNOWLEDGMENTS

We thank all laboratory staff from the participating centers for performing the experiments and analyzing the data.

FUNDING INFORMATION

This study was partly funded by Novartis and Dutch Cancer Society Coherent project: 13650/2021-Infra.

CONFLICT OF INTEREST STATEMENT

All authors declared that they have no conflicts of interest to this work.

ORCID

Jesse M. Tettero https://orcid.org/0000-0002-0811-0824
Vincent H. J. van der Velden https://orcid.org/0000-0001-9457-3763
Costa Bachas https://orcid.org/0000-0001-5983-6193
Jacqueline Cloos https://orcid.org/0000-0001-9150-8026

REFERENCES

Armbruster, D. A., & Pry, T. (2008). Limit of blank, limit of detection and limit of quantitation. *The Clinical Biochemist Reviews*, *29*(Suppl 1), S49–S52.

Buccisano, F., Hourigan, C. S., & Walter, R. B. (2017). The prognostic significance of measurable ("minimal") residual disease in acute myeloid leukemia. *Current Hematologic Malignancy Reports*, *12*(6), 547–556.

Cloos, J., Harris, J. R., Janssen, J. J. W. M., Kelder, A., Huang, F., Sijm, G., Vonk, M., Snel, A. N., Scheick, J. R., Scholten, W. J., Carbaat-Ham, J., Veldhuizen, D., Hanekamp, D., Oussoren-Brockhoff, Y. J. M., Kaspers, G. J. L., Schuurhuis, G. J., Sasser, A. K., & Ossenkoppele, G. (2018). Comprehensive protocol to sample and process bone marrow for measuring measurable residual disease and leukemic stem cells in acute myeloid leukemia. *Journal of Visualized Experiments: JoVE*, *133*, 56386.

CLSI. (2021). *Validation of assays performed by flow cytometry*. CLSI guideline H62. Clinical and Laboratory Standards Institute.

Cremers, E. M., Westers, T. M., Alhan, C., Cali, C., Visser-Wisselaar, H. A., Chitu, D. A., van der Velden, V., te Marvelde, J. G., Klein, S. K., Muus, P., Vellenga, E., de Greef, G. E., Legdeur, M. C., Wijermans, P. W., Stevens-Kroef, M. J., Silva-Coelho, P. D., Jansen, J. H., Ossenkoppele, G. J., van de Loosdrecht, A., & A Study on Behalf of the HOVON89 Study Group. (2017). Implementation of erythroid lineage analysis by flow cytometry in diagnostic models for myelodysplastic syndromes. *Haematologica*, *102*(2), 320–326.

der Strate, B. V., Longdin, R., Geerlings, M., Bachmayer, N., Cavallin, M., Litwin, V., et al. (2017). Best practices in performing flow cytometry in a regulated environment: Feedback from experience within the European bioanalysis forum. *Bioanalysis*, *9*(16), 1253–1264.

Dohner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Buchner, T., et al. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*, *129*(4), 424–447.

Dohner, H., Weisdorf, D. J., & Bloomfield, C. D. (2015). Acute myeloid leukemia. *The New England Journal of Medicine*, *373*(12), 1136–1152.

Donnenberg, A. D., & Donnenberg, V. S. (2007). Rare-event analysis in flow cytometry. *Clinics in Laboratory Medicine*, *27*(3), 627–652. viii.

Du, L., Grover, A., Ramanan, S., & Litwin, V. (2015). The evolution of guidelines for the validation of flow cytometric methods. *International Journal of Laboratory Hematology*, *37*(Suppl 1), 3–10.

Dutch Cytometry Association. (2022). Nederlandse Vereniging voor Cytometrie. https://www.cytometrie.nl/

Grimwade, D., & Freeman, S. D. (2014). Defining minimal residual disease in acute myeloid leukemia: Which platforms are ready for "prime time"? *Hematology. American Society of Hematology. Education Program*, *2014*(1), 222–233.

Heuser, M., Freeman, S. D., Ossenkoppele, G. J., Buccisano, F., Hourigan, C. S., Ngai, L. L., Tettero, J. M., Bachas, C., Baer, C., Béné, M. C., Bücklein, V., Czyz, A., Denys, B., Dillon, R., Feuring-Buske, M., Guzman, M. L., Haferlach, T., Han, L., Herzig, J. K., … Cloos, J. (2021). 2021 update measurable residual disease in acute myeloid leukemia: European LeukemiaNet working party consensus document. *Blood*, *138*, 2753–2767.

Hou, H.-A., & Tien, H.-F. (2020). Genomic landscape in acute myeloid leukemia and its implications in risk classification and targeted therapies. *Journal of Biomedical Science*, *27*(1), 81.

Hu, G.-h., Cheng, Y.-f., Lu, A.-d., Wang, Y., Zuo, Y.-x., Yan, C.-h., et al. (2020). Allogeneic hematopoietic stem cell transplantation can improve the prognosis of high-risk pediatric t(8;21) acute myeloid leukemia in first remission based on MRD-guided treatment. *BMC Cancer*, *20*(1), 553.

Jongen-Lavrencic, M., Grob, T., Hanekamp, D., Kavelaars, F. G., Al Hinai, A., Zeilemaker, A., et al. (2018). Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *New England Journal of Medicine*, *378*(13), 1189–1199.

Kalina, T., Flores-Montero, J., van der Velden, V. H. J., Martin-Ayuso, M., Böttcher, S., Ritgen, M., et al. (2012). EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia*, *26*(9), 1986–2010.

Keeney, M., Halley, J. G., Rhoads, D. D., Ansari, M. Q., Kussick, S. J., Karlon, W. J., Mehta, K. U., Dorfman, D. M., & Linden, M. A. (2015). Marked variability in reported minimal residual disease lower level of detection of 4 Hematolymphoid neoplasms: A survey of participants in the College of American Pathologists Flow Cytometry Proficiency Testing Program. *Archives of Pathology & Laboratory Medicine*, *139*(10), 1276–1280.

Lacombe, F., Bernal, E., Bloxham, D., Couzens, S., Porta, M. G. D., Johansson, U., Kern, W., Macey, M., Matthes, T., Morilla, R., Paiva, A., Palacio, C., Preijers, F., Ratei, R., Siitonen, S., Allou, K., Porwit, A., & Béné, M. C. (2016). Harmonemia: A universal strategy for flow cytometry immunophenotyping—A European LeukemiaNet WP10 study. *Leukemia*, *30*(8), 1769–1772.

Löwenberg, B., Pabst, T., Maertens, J., Gradowska, P., Biemond, B. J., Spertini, O., Vellenga, E., Griskevicius, L., Tick, L. W., Jongen-Lavrencic, M., van Marwijk Kooy, M., Vekemans, M. C., van der Velden, W. J. F. M., Beverloo, B., Michaux, L., Graux, C., Deeren, D., de

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Weerdt, O., van der Esser, J. W. J., … Ossenkoppele, G. J. (2021). Addition of lenalidomide to intensive treatment in younger and middle-aged adults with newly diagnosed AML: The HOVON-SAKK-132 trial. *Blood Advances*, *5*(4), 1110–1121.

Morita, K., Kantarjian, H. M., Wang, F., Yan, Y., Bueso-Ramos, C., Sasaki, K., Issa, G. C., Wang, S., Jorgensen, J., Song, X., Zhang, J., Tippen, S., Thornton, R., Coyle, M., Little, L., Gumbs, C., Pemmaraju, N., Daver, N., DiNardo, C. D., … Takahashi, K. (2018). Clearance of somatic mutations at remission and the risk of relapse in acute myeloid leukemia. *Journal of Clinical Oncology*, *36*(18), 1788–1797.

Narayanan, D., & Weinberg, O. K. (2020). How I investigate acute myeloid leukemia. *International Journal of Laboratory Hematology*, *42*(1), 3–15.

Ossenkoppele, G. J., & Schuurhuis, G. J. (2014). MRD in AML: It is time to change the definition of remission. *Best Practice & Research. Clinical Haematology*, *27*(3–4), 265–271.

Ossenkoppele, G. J., van de Loosdrecht, A. A., & Schuurhuis, G. J. (2011). Review of the relevance of aberrant antigen expression by flow cytometry in myeloid neoplasms. *British Journal of Haematology*, *153*(4), 421–436.

Patkar, N., Kakirde, C., Shaikh, A. F., Salve, R., Bhanshe, P., Chatterjee, G., Rajpal, S., Joshi, S., Chaudhary, S., Kodgule, R., Ghoghale, S., Deshpande, N., Shetty, D., Khizer, S. H., Jain, H., Bagal, B., Menon, H., Khattry, N., Sengar, M., … Gujral, S. (2021). Clinical impact of panel-based error-corrected next generation sequencing versus flow cytometry to detect measurable residual disease (MRD) in acute myeloid leukemia (AML). *Leukemia*, *35*(5), 1392–1404.

Schuurhuis, G. J., Heuser, M., Freeman, S., Béné, M.-C., Buccisano, F., Cloos, J., Grimwade, D., Haferlach, T., Hills, R. K., Hourigan, C. S., Jorgensen, J. L., Kern, W., Lacombe, F., Maurillo, L., Preudhomme, C., van der Reijden, B. A., Thiede, C., Venditti, A., Vyas, P., … Ossenkoppele, G. J. (2018). Minimal/measurable residual disease in AML: A consensus document from the European LeukemiaNet MRD working party. *Blood*, *131*(12), 1275–1291.

Selliah, N., Eck, S., Green, C., Oldaker, T., Stewart, J., Vitaliti, A., & Litwin, V. (2019). Flow cytometry method validation protocols. *Current Protocols in Cytometry*, *87*(1), e53.

Short, N. J., Rytting, M. E., & Cortes, J. E. (2018). Acute myeloid leukaemia. *Lancet*, *392*(10147), 593–606.

Short, N. J., Zhou, S., Fu, C., Berry, D. A., Walter, R. B., Freeman, S. D., Hourigan, C. S., Huang, X., Nogueras Gonzalez, G., Hwang, H., Qi, X., Kantarjian, H., & Ravandi, F. (2020). Association of measurable residual disease with survival outcomes in patients with acute myeloid leukemia: A systematic review and meta-analysis. *JAMA Oncology*, *6*(12), 1890–1899.

Terwijn, M., van Putten, W. L., Kelder, A., van der Velden, V. H., Brooimans, R. A., Pabst, T., et al. (2013). High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: Data from the HOVON/SAKK AML 42A study. *Journal of Clinical Oncology*, *31*(31), 3889–3897.

Tettero, J. M., Freeman, S., Buecklein, V., Venditti, A., Maurillo, L., Kern, W., Walter, R. B., Wood, B. L., Roumier, C., Philippé, J., Denys, B., Jorgensen, J. L., Bene, M. C., Lacombe, F., Plesa, A., Guzman, M. L., Wierzbowska, A., Czyz, A., Ngai, L. L., … Buccisano, F. (2022). Technical aspects of flow cytometry-based measurable residual disease quantification in acute myeloid leukemia: Experience of the European LeukemiaNet MRD working party. *HemaSphere*, *6*(1), e676.

U.S. Department of Health and Human Services OCoE, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. (2020). *Hematologic malignancies: Regulatory considerations for use of minimal residual disease in development of drug and biological products for treatment*. Food and Drug Administration.

Venditti, A., Piciocchi, A., Candoni, A., Melillo, L., Calafiore, V., Cairoli, R., de Fabritiis, P., Storti, G., Salutari, P., Lanza, F., Martinelli, G., Luppi, M., Mazza, P., Martelli, M. P., Cuneo, A., Albano, F., Fabbiano, F., Tafuri, A., Chierichini, A., … Amadori, S. (2019). GIMEMA AML1310 trial of risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia. *Blood*, *134*(12), 935–945.

Wood, B., Jevremovic, D., Béné, M. C., Yan, M., Jacobs, P., Litwin, V., & ICSH/ICCS Working Group. (2013). Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS – Part V—Assay performance criteria. *Cytometry Part B: Clinical Cytometry*, *84*(5), 315–323.

Zeijlemaker, W., Grob, T., Meijer, R., Hanekamp, D., Kelder, A., Carbaat-Ham, J. C., Oussoren-Brockhoff, Y. J. M., Snel, A. N., Veldhuizen, D., Scholten, W. J., Maertens, J., Breems, D. A., Pabst, T., Manz, M. G., van der Velden, V. H. J., Slomp, J., Preijers, F., Cloos, J., van de Loosdrecht, A. A., … Schuurhuis, G. J. (2019). CD34(+)CD38(−) leukemic stem cell frequency to predict outcome in acute myeloid leukemia. *Leukemia*, *33*(5), 1102–1112.

Zeijlemaker, W., Kelder, A., Cloos, J., & Schuurhuis, G. J. (2019). Immunophenotypic detection of measurable residual (stem cell) disease using LAIP approach in acute myeloid leukemia. *Current Protocols in Cytometry*, *91*(1), e66.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tettero, J. M., Dakappagari, N., Heidinga, M. E., Oussoren-Brockhoff, Y., Hanekamp, D., Pahuja, A., Burns, K., Kaur, P., Alfonso, Z., van der Velden, V. H. J., te Marvelde, J. G., Hobo, W., Slomp, J., Bachas, C., Kelder, A., Nguyen, K., & Cloos, J. (2023). Analytical assay validation for acute myeloid leukemia measurable residual disease assessment by multiparametric flow cytometry. *Cytometry Part B: Clinical Cytometry*, 1–14. https://doi.org/10.1002/cyto.b.22144

15524957, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cyto.b.22144 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [12/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License