Clinical relevance of an objective flow cytometry approach based on limit of detection and limit of quantification for measurable residual disease assessment in acute myeloid leukemia. A *post-hoc* analysis of the GIMEMA AML1310 trial

Francesco Buccisano,[1*] Raffaele Palmieri,[1*] Alfonso Piciocchi,[2] Valentina Arena,[2] Luca Maurillo,[1] Maria-Ilaria Del Principe,[1] Giovangiacinto Paterno,[1] Maria-Antonietta Irno-Consalvo,[1] Tiziana Ottone,[1] Mariadomenica Divona,[1] Consuelo Conti,[1] Daniela Fraboni,[1] Serena Lavorgna,[1] William Arcese,[1,3] Maria Teresa Voso[1] and Adriano Venditti[1]

[1]Ematologia, Dipartimento di Biomedicina e Prevenzione, "Tor Vergata" Università di Roma; [2]Centro Dati Fondazione GIMEMA Onlus and [3]Rome Transplant Network, Rome, Italy

**FB and RP contributed equally as co-first authors.*

Correspondence: F Buccisano
francesco.buccisano@uniroma2.it

Received: August 11, 2021.
Accepted: February 9, 2022.
Prepublished: March 17, 2022.

https://doi.org/10.3324/haematol.2021.279777

Abstract

Using a multiparametric flow cytometry assay, we assessed the predictive power of a threshold calculated applying the criteria of limit of detection (LOD) and limit of quantitation (LOQ) in adult patients with acute myeloid leukemia. This was a *post-hoc* analysis of 261 patients enrolled in the GIMEMA AML1310 prospective trial. According to the protocol design, using the predefined measurable residual disease (MRD) threshold of 0.035% bone marrow residual leukemic cells (RLC) calculated on mononuclear cells, 154 (59%) of the 261 patients were negative (MRD <0.035%) and 107 (41%) were positive (MRD ≥0.035%). Using LOD and LOQ, we selected the following categories of patients: (i) LOD^{neg} if RLC were below the LOD (74; 28.4%); (ii) LOD^{pos}-LOQ^{neg} if RLC were between the LOD and LOQ (43; 16.5%); and (iii) LOQ^{pos} if RLC were above the LOQ (144; 54.4%). Two-year overall survival of these three categories of patients was 75.4%, 79.8% and 66.4%, respectively (*P*=0.1197). Given their superimposable outcomes, the LOD^{neg} and LOD^{pos}-LOQ^{neg} categories were combined. Two-year overall survival of LOD^{neg}/LOD^{pos}-LOQ^{neg} patients was 77.0% *versus* 66.4% of LOQ^{pos} individuals (*P*=0.043). This figure was challenged in univariate analysis (*P*=0.046, hazard ratio=1.6, 95% confidence interval: 1.01-2.54) which confirmed the independent role of the LOD-LOQ approach in determining overall survival. In the AML1310 protocol, using the threshold of 0.035%, 2-year overall survival of patients with MRD <0.035% and MRD ≥0.035% was 74.5% *versus* 66.4%, respectively (*P*=0.3521). In conclusion, the use of the LOD-LOQ method results in more sensitive detection of MRD that, in turn, translates into a more accurate recognition of patients with different outcomes.

Introduction

Measurable residual disease (MRD) is being increasingly employed as a biomarker of quality of complete remission in patients with acute myeloid leukemia (AML) treated with intensive chemotherapy.[1] Multiparametric flow cytometry and reverse transcriptase quantitative polymerase chain reaction are the two leading techniques for MRD quantification. Recent studies indicate that, due to technical improvements and the availability of up to eight to ten color immunostains, the specificity and sensitivity of multiparametric flow cytometry may be reliably increased, provided that a sufficient number of relevant events is acquired.[2,3] In B-cell precursor acute lymphoid leukemia and multiple myeloma, the use of standardized panels and the acquisition of large numbers of events (>4x10^6) led to MRD assessment by multiparametric flow cytometry becoming at least as sensitive as that by polymerase chain reaction-based methods.[4,5] Likewise, the sensitivity of MRD determination in multiple myeloma and chronic lymphocytic

leukemia improved dramatically up to 10^{-5}-10^{-6} as soon as larger numbers of events (3-5x10^6) were acquired, in the context of so-called next-generation flow.[6–8]

In AML, the number of clustered events and the denominator of acquired events necessary for reliable MRD recognition are poorly standardized and may be affected by several technical and clinical variables. In myeloid bone marrow, particularly during regenerating phases after chemotherapy, the normal maturational patterns may interfere with the detection of leukemia-associated immunophenotypes generating a relevant background noise. Likewise, although to a lesser extent, this background noise may affect the identification of the putative "empty spaces" when MRD is detected by a "different-from-normal" approach.[9]

The consensus of the European LeukemiaNet (ELN) MRD working party suggests that a MRD threshold of 0.1% is informative for clinical decisions once 500,000-1,000,000 events are acquired.[10] Such a target of acquired events guarantees that the threshold of 0.1% has a reliable sensitivity and a sufficient specificity, because no leukemia-associated immunophenotypes have been detected above this threshold even in regenerating bone marrow.[11] Nonetheless, the same guidelines suggest that MRD tests with MRD quantified below 0.1% may still be consistent with residual leukemia; indeed several studies have shown prognostic significance of MRD levels below 0.1%.[12–16]

In the GIMEMA AML1310 protocol, post-remission therapy of young patients with AML was decided combining cytogenetic/genetic information and post-consolidation levels of MRD after consolidation as measured by multiparametric flow cytometry.[17] Intermediate-risk patients were to receive autologous or allogeneic stem cell transplant (SCT) depending on the post-consolidation levels of MRD. The threshold of negativity was set at 0.035% residual leukemic cells (RLC) as measured on mononuclear cells, with values below the threshold being considered negative. This threshold was selected after repeatedly validating it in retrospective, sequential cohorts of patients enrolled in former EORTC/GIMEMA protocols AML10, AML12, AML13, AML15 and AML17.[18–21] In the AML1310 protocol we confirmed, prospectively, that the threshold of 0.035% retained the same predictive value as in the retrospective analyses.[16]

However, since the previous EORTC/GIMEMA and AML1310 protocols had in common the same therapeutic schedule, either in induction or in consolidation, one could argue that the threshold of 0.035% may be protocol-specific so that it cannot be applied universally. In fact, thresholds in AML are often selected retrospectively based on their association with outcomes. Accordingly, confirmatory, prospective validations are required.[22,23] In an attempt to overcome such a "protocol-effect" and to reliably improve the statistical accuracy of MRD assessment, we revised the post-consolidation MRD determinations of the GIMEMA AML1310 protocol by calculating, for each case, the limit of detection (LOD) and limit of quantification (LOQ). As in multiple myeloma and chronic lymphocytic leukemia, the target of 20 and 50 relevant events in the final gate, respectively, were adopted. According to the ELN guidelines, the analysis was conducted on CD45-expressing elements.[3,24] The MRD status of patients was classified as negative (LOD^{neg}), positive not quantifiable (LOD^{pos}-LOQ^{neg}) and positive quantifiable (LOQ^{pos}). Due to the retrospective nature of the analysis, we were not able to establish a limit of blank to properly exclude the background noise of each aberrant phenotype selected for MRD assessment.

In our exploratory analysis, the new MRD categories were compared to the protocol reference threshold of 0.035%, the genetic/cytogenetic subgroups and the post-remission treatments. To the best of our knowledge, this is the first time that an absolute threshold based on LOD and LOQ has been applied to assess MRD in AML by multiparametric flow cytometry. We believe that the analysis of a prospective series of homogeneously treated patients, represents a unique chance to corroborate the robustness of the LOD and LOQ approach in MRD determination in AML.

Methods

Patients

Previously untreated patients with a diagnosis of *de novo* AML according to the World Health Organization diagnostic criteria[25] were eligible for the GIMEMA AML1310 study (EudraCT number 2010-023809-36; ClinicalTrials.Gov Identifier NCT01452646) (*Online Supplementary Methods*).[16,26] The present analysis was performed with different purposes on a subgroup of 261 patients whose MRD status was determined after the consolidation cycle of treatment. The study was approved by the ethical committees of the participating hospitals or academic institutions and was conducted in accordance with the Declaration of Helsinki. All participants gave their informed consent.

Limits of detection and quantification calculations

There are numerous studies demonstrating that 20 events are a conservative value for the smallest (homogeneous) population that can be detected in a given flow cytometric list mode data file by experienced operators. This implies that the LOD can be estimated as (20/total number of cells analyzed) × 100%.[26] Similarly, it is also widely accepted that more than 50 events is a reasonable threshold for reproducible enumeration of a cell population by experienced operators; consequently, the LOQ

can be estimated as (50/total number of cells analyzed) × 100%.[27] Thus, the LOD and the LOQ will both be typically dependent on the total number of cells analyzed. The LOD and LOQ were established at 20 and 50 clustering events expressing a leukemia-associated immunophenotype, respectively, and counted on CD45-expressing events according to the ELN recommendations.[10] Based on such an approach, patients were classified as MRD-negative if RLC were below the LOD (LOD^{neg}), MRD-positive non-quantifiable if RLC were between the LOD and LOQ (LOD^{pos}-LOQ^{neg}) and MRD-positive quantifiable if RLC were above the LOQ (LOQ^{pos}). Samples were acquired by a FacSCanto II (Becton Dickinson, Mountain View, CA, USA). Data were analyzed using Infinicyt-software version 1.7 (Cytognos SL, Salamanca, Spain).

Statistical analysis

Overall survival (time elapsed from the start of treatment to death) and disease-free survival (time from complete remission to relapse or death in remission) were calculated using the Kaplan-Meier product limit estimator. Differences in terms of overall and disease-free survival were evaluated by means of a log-rank test in univariate analysis and by means of a Cox regression model in multivariate analysis, after assessment of proportionality of hazards. All variables with a *P*-value less than 0.15 in univariate analysis were considered in the multivariate models. The influence of the transplant on the survival outcome was evaluated in the Cox model by means of a time-dependent covariate. The cumulative incidence of relapse was estimated by cumulative incidence curves using the proper non-parametric method. Patients' and disease characteristics were summarized by means of cross-tabulations for categorical variables or by quintiles for continuous variables. Differences between categorical variables or response rates in subgroups were tested by the χ^2 or Fisher exact tests, as appropriate. Confidence intervals were calculated at the 95% level and all tests were two-sided, accepting $P \leq 0.05$ as indicating a statistically significant difference. All covariates were evaluated in univariate models and all factors with univariate association with a *P*-value <0.1 were considered in the multivariate models as potential parameters. Backward and stepwise methods were applied to identify the multivariate models with a step-by-step iterative construction that involved the selection of independent variables to be considered in the final model. All analyses were performed using SAS (version 9.4) and R (R Foundation for Statistical Computing, Vienna, Austria) software. Study data were collected and managed using the REDCap20 electronic data capture tools hosted at the GIMEMA Foundation.

Results

The present analysis includes 261 patients from whom a post-consolidation bone marrow sample was collected and sent to the central laboratory for MRD determination. Clinical characteristics of the patients are summarized in Table 1. Subjects with a percentage of RLC ≥0.035% the total number of acquired mononuclear cells qualified as $MRD^{\geq 0.035\%}$. In the same 261 patients, the LOD and LOQ were calculated on CD45-expressing elements. The median number of mononuclear cells acquired was 559,197 (range, 100,450-1,561,221) and the median number of CD45-expressing cells was 538,527 (range, 88,040-1,548,172). Overall, of the

Table 1. General characteristics of the study population.

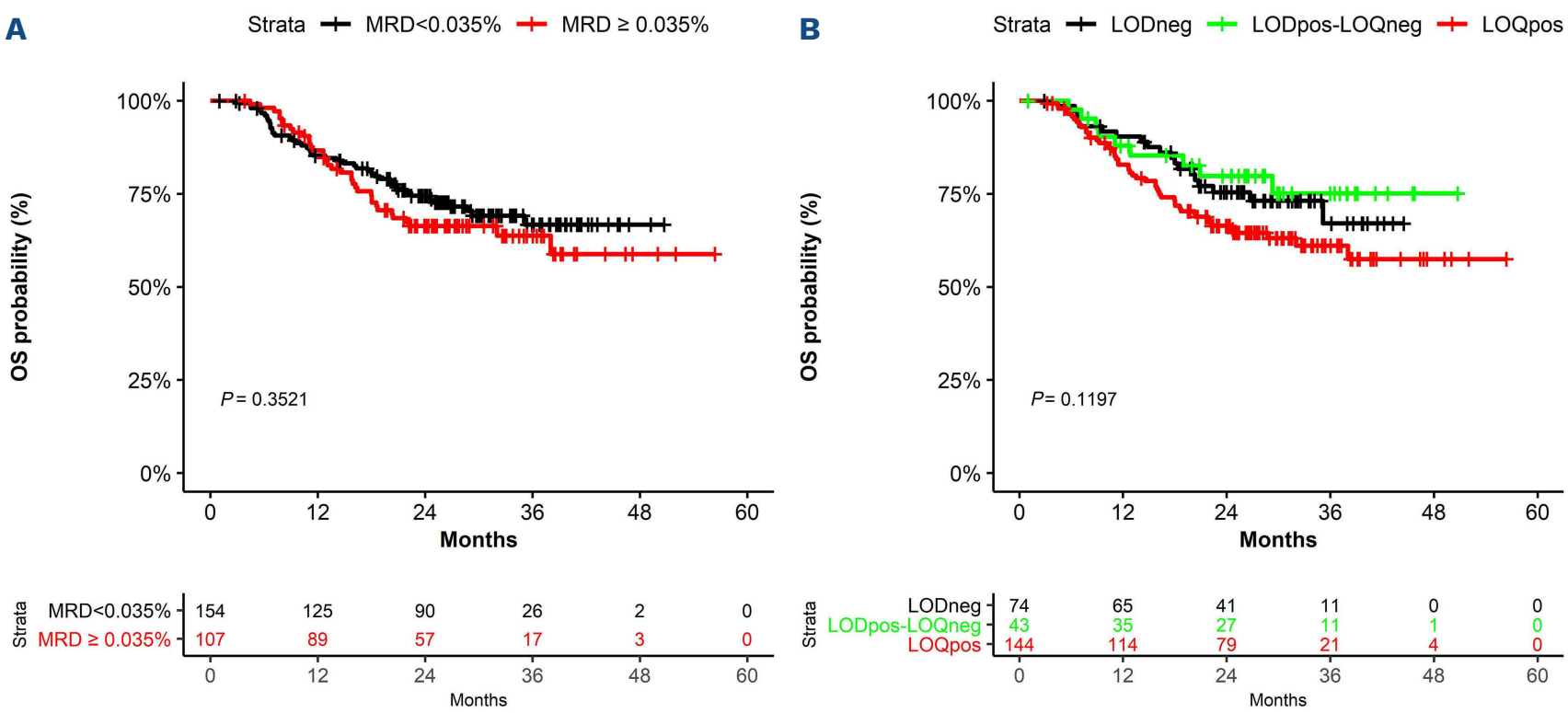
	Level	Overall
Number		261
Sex, N (%)	Male	139 (53.3)
	Female	122 (46.7)
Age in years, median (range)		49.39 (18.32-60.95)
White blood cells x10^9/L, median (range)		12.66 (0.16-186.00)
Platelets x10^9/L, median (range)		55.00 (7.00-1020.00)
Risk category, N (%)*	NCCN-FR	87 (33.3)
	NCCN-IR	77 (29.5)
	NCCN-PR	97 (37.2)
Cytogenetic risk, N (%)**	Favorable risk	28 (12.3)
	Poor risk	29 (12.8)
	Intermediate risk	170 (74.9)
FLT3 ITD, N (%)	Negative	190 (73.1)
	Positive	70 (26.9)
NPM1, N (%)	Negative	145 (55.6)
	Positive	115 (44.1)
Graft, N (%)	No graft	85 (32.6)
	Allo-SCT	93 (35.6)
	Auto-SCT	83 (31.8)

*Genetic/cytogenetic risk group was attributed according to National Comprehensive Cancer Network clinical practice guidelines (version 2009) as follows: "favorable" risk [cases with Inv(16), t(8;21), t(16;16), *RUNX1/RUNXT1* without c-Kit mutations, *CBFB/MYH11* without c-Kit mutations, *NPM1* mutation without *FLT3* mutations]; "intermediate" risk [cases with normal karyotype, isolated +8, isolated t(9;11), other karyotypic abnormalities not listed as favorable or adverse, *RUNX1/RUNXT1* with c-Kit mutations, *CBFB/MYH11* with c-Kit mutations, no *NPM1* mutations, no *FLT3*-ITD mutations]; "adverse" risk [cases with complete karyotype e.g. >3 abnormalities, -5/5q-, -7/7q-, abnormalities of 11q23 excluding t(9;11), inv(3), t(3;3), t(6;9), *FLT3*-ITD mutations]. **Patients were stratified according to the refined Medical Research Council (MRC) classification of cytogenetic risk, as follows: "favorable" risk [cases with t(8;21), t(15;17) or inv(16)/t(16;16)]; "adverse" risk [cases with complex cytogenetic changes (>3 unrelated abnormalities), -5, add(5q)/del(5q), -7/add(7q), t(6;11), t(10;11), t(9;22), -17, abn(17p) with other changes, 3q abnormalities excluding t(3;5), inv(3)/t(3;3)]; and "intermediate" risk [cases with normal karyotype and other non-complex]. NCCN: National Comprehensive Cancer Network; FR: favorable risk; IR: intermediate risk; PR: poor risk; Allo-SCT: allogeneic stem cell transplant; Auto-SCT: autologous stem cell transplant.

261 cases, 74 (28.4%) were classified as LOD^{neg}, whereas 43 (16.5%) and 144 (55.2%) were classified as LOD^{pos}-LOQ^{neg} and LOQ^{pos}, respectively (*Online Supplementary Table S1*). The target of 500,000 processed CD45^{+} events was reached in 158 (60.5%) of the 261 patients. The calculated median LOD and LOQ values were 0.0037 (0.0013-0.0227) and 0.0093 (0.0032-0.0568), respectively (*Online Supplementary Table S1*).

According to the protocol MRD threshold of 0.035%, 107 (41.0%) of the 261 patients were MRD$^{\geq 0.035\%}$ and 154 (59.0%) MRD$^{<0.035\%}$. The interactions between the different MRD estimates are summarized in *Online Supplementary Table S2*. Overall, 105/107 (98.1%) MRD$^{\geq 0.035\%}$ patients were LOQ^{pos} whereas only 74/154 (48.1%) MRD$^{<0.035\%}$ ones were LOD^{neg}-LOQ^{neg} (P<0.001). In fact, 41 (26.6%) and 39 (25.3%) of 154 MRD$^{<0.035\%}$ patients were reclassified as LOD^{pos}-LOQ^{neg} and LOQ^{pos}, respectively.

In the whole population, 2-year overall and disease-free survival rates were 71.2% and 57.5%, respectively. No difference was observed in duration of overall survival between MRD$^{<0.035\%}$ and MRD$^{\geq 0.035\%}$ patients (74.5% *vs*. 66.4%, P=0.3521) (Figure 1A). When the survival analysis was conducted according to the new categories that we identified, patients who were LOD^{neg} or LOD^{pos}-LOQ^{neg} had a superior overall survival as compared to LOQ^{pos} patients (75.4% and 79.8% *vs*. 66.4%), although the difference was not statistically significant (P=0.119). The equivalent outcome of LOD^{neg} and LOD^{pos}-LOQ^{neg} patients (Figure 1B) persuaded us to aggregate these subgroups. Accordingly, we sorted two categories of patients, (LOD^{neg}/LOD^{pos}-LOQ^{neg}) and



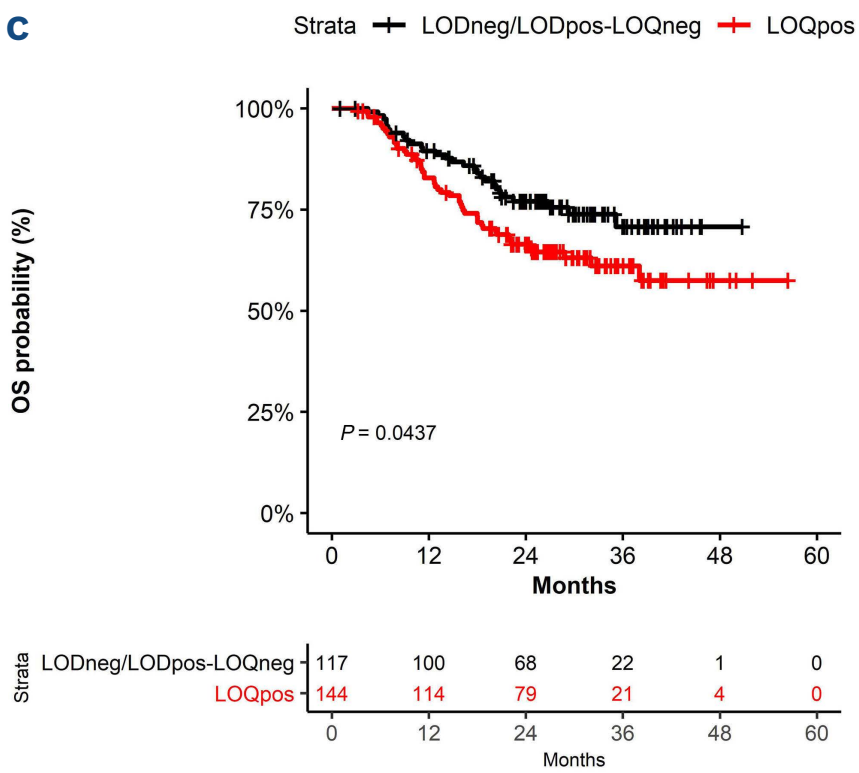


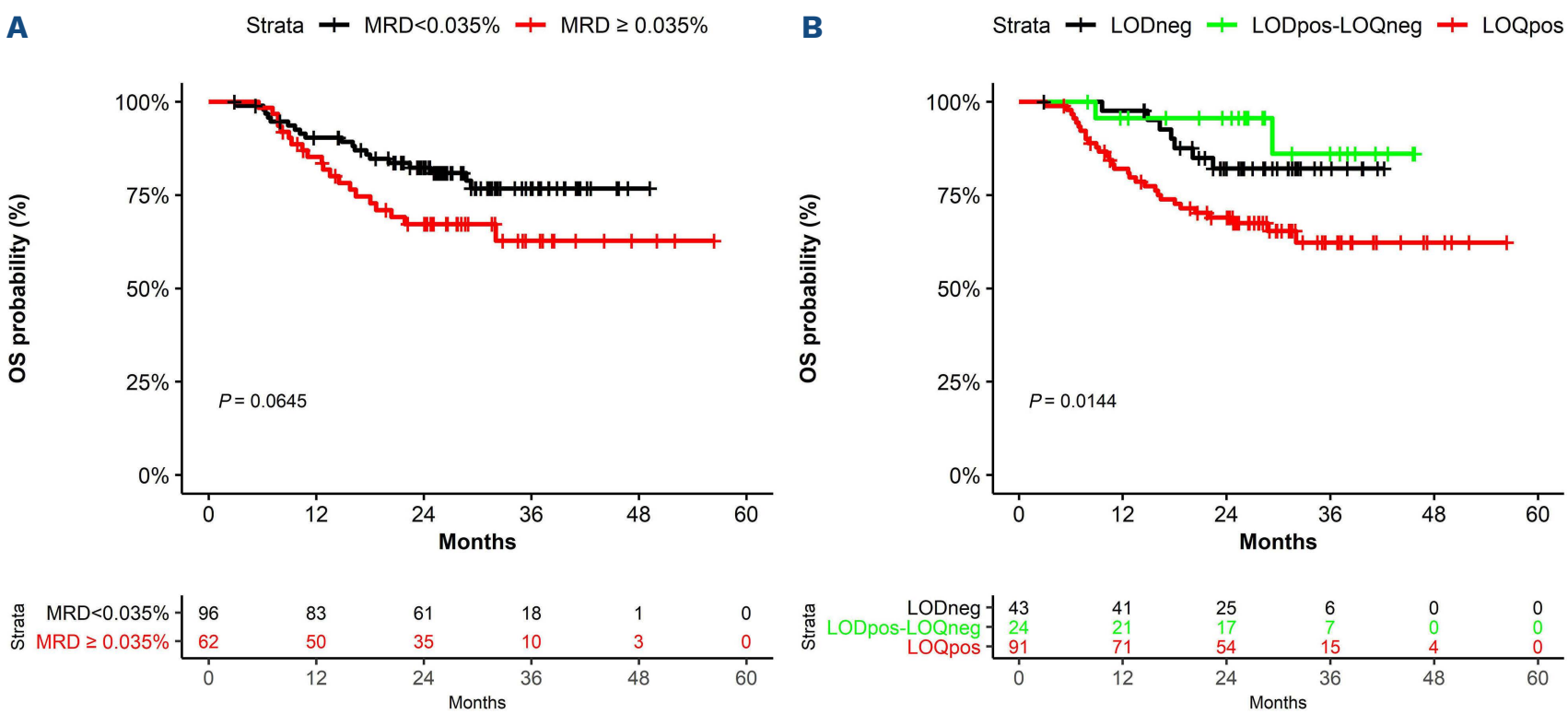
Figure 1. Overall survival analysis of the whole series of 261 patients according to different measurable residual disease estimates. Measurable residual disease stratification according to the AML1310 threshold (0.035%) was not statistically different (A). LOD^{neg}, LOD^{pos}-LOQ^{neg} and LOQ^{pos} are analyzed separately (B) and merging LOD^{neg} and LOD^{pos}-LOQ^{neg} (C), with only the latter reaching a statistically significant difference (P=0.043). OS: overall survival; MRD: measurable residual disease; LOD: limit of detection; LOQ: limit of quantification.

LOQ^{pos} whose duration of overall survival was statistically different (77.0% *vs*. 66.4%, *P*=0.0437), as depicted in Figure 1C.

As a further step of investigation, we repeated our analysis on the 158 (60.5%) of 261 patients in whom ≥500,000 CD45^{+} events were acquired. This was to test whether a more numerically robust denominator enhanced specificity and then prognostic power of the LOD-LOQ estimate. The threshold-based MRD allocation (MRD$^{<0.035\%}$ 82.4% *vs*. 67.2% of MRD$^{\geq 0.035\%}$, *P*=0.064) (Figure 2A) was less effective in discriminating patients with different 2-year overall survival rates, whereas 2-year overall survival rates of LOD^{neg} and LOD^{pos}-LOQ^{neg} patients were superior to that of LOQ^{pos} patients (82.1% and 95.7% *vs*. 69.0%, *P*=0.014) with a significant difference between LOQ^{pos} and both LOD^{neg} and LOD^{pos}-LOQ^{neg} patients (*P*=0.038 and *P*=0.024, respectively) (Figure 2B). The LOD^{neg}/LOD^{pos}-LOQ^{neg} category identified a subset of patients with a strongly favorable outcome as compared to the LOQ^{pos} subgroup (2-year overall survival of 86.7% *vs*. 69.0%, *P*=0.004) (Figure 2C).

We then tried to integrate the MRD and LOD-LOQ models. By doing so, we generated three categories of patients (MRD$^{<0.035\%}$LOD^{neg}/LOD^{pos}-LOQ^{neg}, MRD$^{<0.035\%}$LOQ^{pos}, and MRD$^{\geq 0.035\%}$LOQ^{pos}), whose features are shown in Table 2. A fourth category (MRD$^{\geq 0.035\%}$LOD^{neg}/LOD^{pos}-LOQ^{neg}) was dropped from the analysis because it was represented by only two patients.

Notably, MRD$^{<0.035\%}$LOD^{neg}/LOD^{pos}-LOQ^{neg} patients had a better 2-year overall survival not only when compared to MRD$^{\geq 0.035\%}$LOQ^{pos} patients but also when compared to



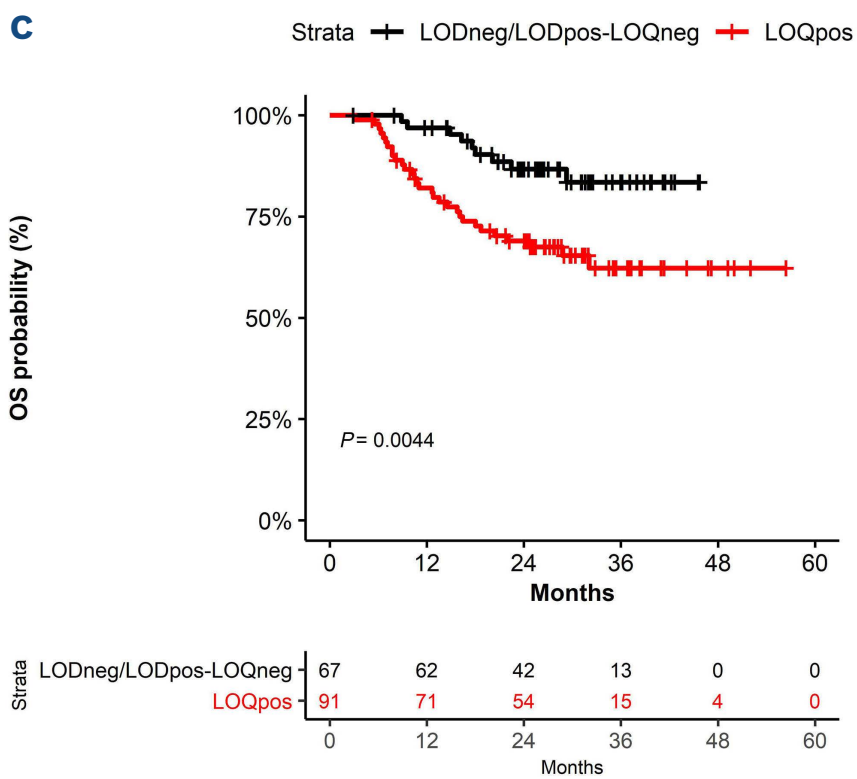


Figure 2. Overall survival analysis of 158 patients from whom >500,000 CD45^{+} cells were acquired. Stratification according to the AM1310 MRD threshold showed a lower power of discrimination in terms of 2-year overall survival (*P*=0.064) (A). LOD^{neg}, LOD^{pos}-LOQ^{neg} and LOQ^{pos} are analyzed separately (B) and merging LOD^{neg} and LOD^{pos}-LOQ^{neg} (C); both tests stratified patients with a statistical significance (*P*=0.023 and *P*=0.009, respectively). OS: overall survival; MRD: measurable residual disease; LOD: limit of detection; LOQ: limit of quantification.

Table 2. Integration of the "relative" 0.035% and "absolute" limit of detection/limit of quantification approaches for measurable residual disease determination.

	Level	MRD$^{<0.035\%}$ LOD^{neg}/LOD^{pos}-LOQ^{neg}	MRD$^{<0.035\%}$ LOQ^{pos}	MRD$^{\geq 0.035\%}$ LOQ^{pos}	*P*
Number		115	39	105	
Sex, N (%)	Male	61 (53.0)	19 (48.7)	57 (54.3)	0.837
	Female	54 (47.0)	20 (51.3)	48 (45.7)	
Age in years, median (range)		48.7 (18.3-60.3)	44.5 (21.9-60.7)	52.3 (19.4-60.9)	0.066
WBC x10^9/L, median (range)		9.60 (0.16-181.38)	11.70 (0.74-186.00)	16.73 (0.48-158.30)	0.078
Risk category, N (%)*	NCCN-FR	42 (36.5)	7 (17.9)	38 (36.2)	<0.001
	NCCN-IR	32 (27.8)	24 (61.5)	20 (19.0)	
	NCCN-PR	41 (35.7)	8 (20.5)	47 (44.8)	
Citogenetic risk, N (%)**	Favorable-risk	19 (18.8)	2 (6.7)	7 (7.4)	0.141
	Poor-risk	11 (10.9)	4 (13.3)	13 (13.8)	
	Intermediate-risk	71 (70.3)	24 (80.0)	74 (78.7)	
FLT3-ITD, N (%)	Negative	83 (72.8)	35 (89.7)	71 (67.6)	0.028
	Positive	31 (27.2)	4 (10.3)	34 (32.4)	
NPM1, N (%)	Negative	64 (55.7)	30 (76.9)	50 (48.1)	0.008
	Positive	51 (44.3)	9 (23.1)	54 (51.9)	
Graft, number, N (%)	No graft	38 (33.0)	9 (23.1)	37 (35.2)	0.629
	Allo-SCT	38 (33.0)	17 (43.6)	37 (35.2)	
	Auto-SCT	39 (33.9)	13 (33.3)	31 (29.5)	

*Genetic/cytogenetic risk group was attributed according to National Comprehensive Cancer Network clinical practice guidelines (version 2009) as follows: "favorable" risk [cases with Inv(16), t(8;21), t(16;16), *RUNX1/RUNXT1* without c-Kit mutations, *CBFB/MYH11* without c-Kit mutations, *NPM1* mutation without *FLT3* mutations]; "intermediate" risk [cases with normal karyotype, isolated +8, isolated t(9;11), other karyotypic abnormalities not listed as favorable or adverse, *RUNX1/RUNXT1* with c-Kit mutations, *CBFB/MYH11* with c-Kit mutations, no *NPM1* mutations, no *FLT3*-ITD mutations]; "adverse" risk [cases with complete karyotype e.g. >3 abnormalities, -5/5q-, -7/7q-, abnormalities of 11q23 excluding t(9;11), inv(3), t(3;3), t(6;9), *FLT3*-ITD mutations]. **Patients were stratified according to the refined Medical Research Council (MRC) classification of cytogenetic risk, as follows: "favorable" risk [cases with t(8;21), t(15;17) or inv(16)/t(16;16)]; "adverse" risk [cases with complex cytogenetic changes (>3 unrelated abnormalities), -5, add(5q)/del(5q), -7/add(7q), t(6;11), t(10;11), t(9;22), -17, abn(17p) with other changes, 3q abnormalities excluding t(3;5), inv(3)/t(3;3)]; and "intermediate" risk [cases with normal karyotype and other non-complex]. LOD: limit of detection; LOQ: limit of quantification; MRD: measurable residual disease; WBC: white blood cells; NCCN: National Comprehensive Cancer Network; FR: favorable risk; IR: intermediate risk; PR: poor risk; Allo-SCT: allogeneic stem cell transplant; Auto-SCT: autologous stem cell transplant.

MRD$^{<0.035\%}$LOQ^{pos} patients, whose median MRD percentage was 0.016% (range, 0.006-0.032). This comparison did not reach statistical significance when the overall series was analyzed (76.7% *vs*. 67.5% and 65.9%, *P*=0.116) but was clearly significant when patients with at least 500,000 events were taken into account. More in detail, among patients in whom ≥500,000 CD45-expressing events were acquired, those who were MRD$^{<0.035\%}$LOD^{neg}/LOD^{pos}-LOQ^{neg} had a longer duration of overall survival as compared to those who were MRD$^{<0.035\%}$LOQ^{pos} and MRD$^{\geq 0.035\%}$LOQ^{pos} (86.7%, 72.5% and 67.0%, respectively, *P*=0.018). Furthermore, MRD$^{<0.035\%}$ patients had a statistically different overall survival if they tested LOD^{neg}/LOD^{pos}-LOQ^{neg} or LOQ^{pos} (86.7% *vs*. 72.5%, *P*=0.007) (Figure 3). To avoid a possible bias deriving from the original design of the protocol, in which MRD was used to address treatment only in the intermediate-risk category, we conducted the same analysis in the 77 patients belonging to this category. The results (*Online Supplementary Figure S1*) were completely superimposable (*P*=0.0286).

Finally, we explored the interaction of LOD^{neg}, LOD^{pos}-LOQ^{neg} and LOQ^{pos} categories with the post-remission treatment received (autologous SCT, allogeneic SCT and no graft). As shown in *Online Supplementary Figure S2*, LOD^{neg}/LOD^{pos}-LOQ^{neg} patients submitted to autologous SCT had the best 2-year overall survival (88.9%) as compared to the other categories (*P*=0.026). Notably, these

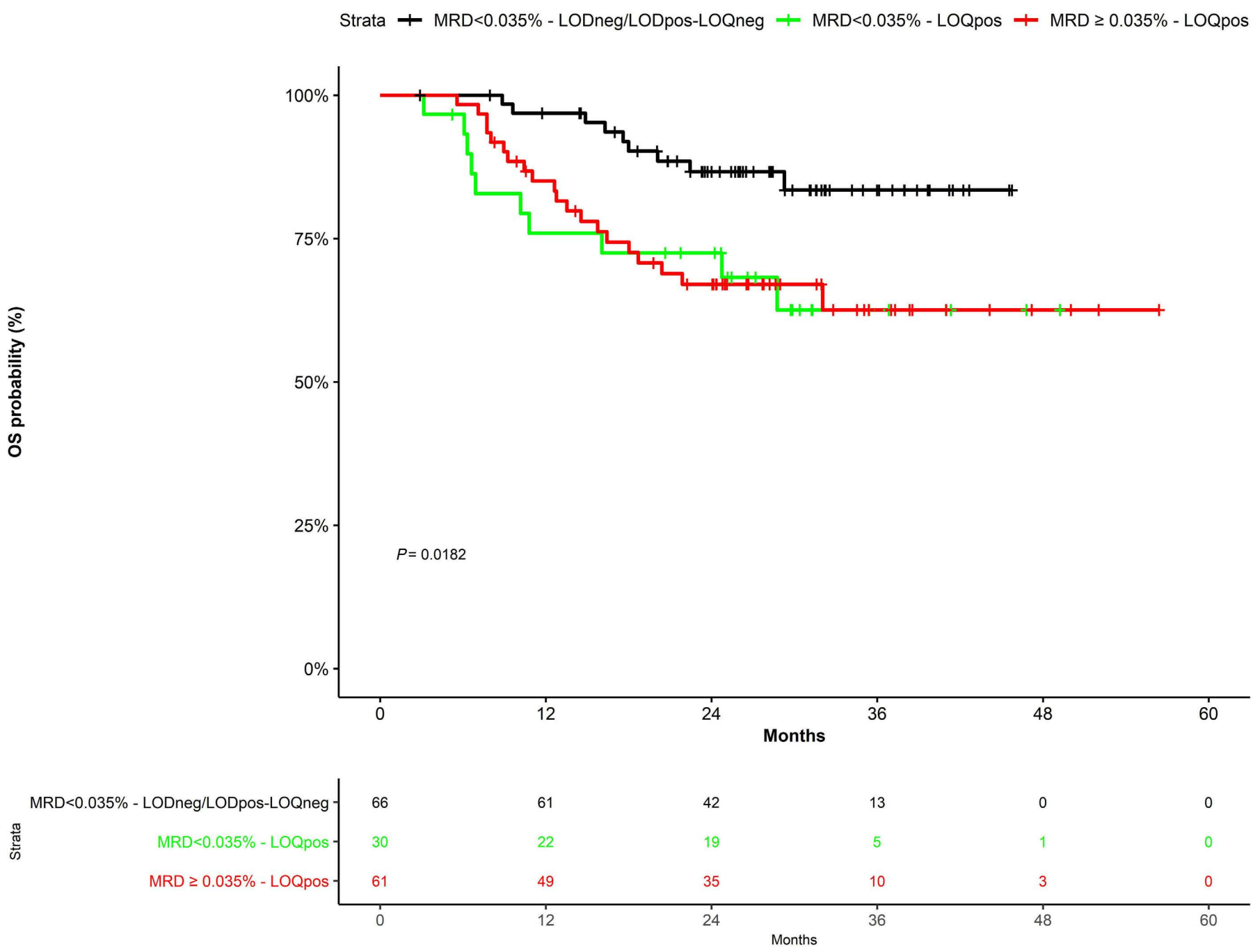


Figure 3. Overall survival analysis of MRD$^{<0.035\%}$ and MRD$^{\geq 0.035\%}$ patients according to limit of detection and limit of quantification status. MRD$^{<0.035\%}$LOD^{neg}/LOD^{pos}-LOQ^{neg} patients had a longer duration of overall survival as compared to MRD$^{<0.035\%}$LOQ^{pos} and MRD$^{\geq 0.035\%}$LOQ^{pos} (P=0.018). Even more, MRD$^{<0.035\%}$ patients had a statistically different overall survival if they tested LOD^{neg}/LOD^{pos}-LOQ$^{neg\ neg}$ or LOQ^{pos} (P=0.007). OS: overall survival; MRD: measurable residual disease; LOD: limit of detection; LOQ: limit of quantification.

patients, benefitted more from autologous SCT (88.9%) than from no-graft (55.9%, P=0.017) or allogeneic SCT (76.5%, P=0.089).

All clinical variables testing significant in univariate analysis were entered in the multivariate model (Table 3). The multivariate analysis confirmed the independent impact on overall survival of poor-risk upfront classification (P<0.001, hazard ratio [HR]=5.02, 95% confidence interval [CI]: 2.31-10.9), allogeneic SCT (P=0.005, HR=0.47, 95% CI: 0.28-0.80) and MRD$^{<0.035\%}$LOQ^{pos} status (P=0.021, HR=2.19, 95% CI: 1.13-4.27). The multivariate model including LOD-LOQ stratification and transplant as a time-dependent covariate resulted in achievement of significant P values in both univariate (P<0.001, HR=5.02, 95% CI: 2.31-10.9) and multivariable analyses (P=0.048, HR=0.628, 95% CI: 0.396-0.997) for LOD-LOQ stratification but not for transplantation.

Discussion

In this preliminary study, we demonstrated that an MRD estimate based on LOD and LOQ of CD45-expressing cells predicts survival of AML patients more accurately than the pre-established threshold of 0.035% RLC of mononuclear cells, which was used in the AML1310 protocol. Moreover, we observed that the predictive power of the LOD-LOQ approach increases proportionally with the number of events acquired (higher or lower than 500,000).

The search for the most informative value of MRD for clinical use remains a matter of debate in AML. The general experience indicates that many technical, biological and clinical confounding factors interfere with the identification of the “absolute threshold” below or above which the prognosis is more accurately predicted.[28] In fact, the background noise due to the normal maturational curves

Table 3. Univariate and multivariate Cox regression models for overall survival.

	Univariate analysis			Multivariate analysis		
Characteristic	**HR**	**95% CI**	***P*-value**	**HR**	**95% CI**	***P*-value**
White blood cells	1.00	0.99-1.00	0.30			
FLT3 ITD						
Negative	—	—				
Positive	2.40	1.53-3.77	<0.001			
Risk category						
NCCN-FR	—	—		—	—	
NCCN-IR	1.95	1.00-3.81	0.051	1.99	0.95-4.15	0.068
NCCN-PR	3.73	2.04-6.82	<0.001	5.02	2.31-10.9	<0.001
LOD LOQ stratification						
LOD^{neg}-$LOD^{pos}LOQ^{neg}$	—	—				
LOQ^{pos}	1.60	1.01-2.54	0.046			
BM MRD status after consolidation						
Negative	—	—	0.35			
Positive	1.23	0.79-1.92				
MRD_LODLOQ						
$MRD^{<0.035\%}/LOD^{neg}$-$LOD^{pos}LOQ^{neg}$	—	—		—	—	
$MRD^{>0.035\%}/LOD^{neg}$-$LOD^{pos}LOQ^{neg}$	0.00	0.00-Inf	>0.99	0.00	0.00-Inf	>0.99
$MRD^{<0.035\%}/LOQ^{pos}$	1.82	0.97-3.41	0.061	2.19	1.13-4.27	0.021
$MRD^{>0.035\%}/LOQ^{pos}$	1.49	0.91-2.44	0.11	1.29	0.78-2.13	0.33
Graft						
No graft	—	—		—	—	
Allo-SCT	0.77	0.47-1.27	0.30	0.47	0.28-0.80	0.005
Auto-SCT	0.41	0.23-0.75	0.003	0.72	0.36-1.46	0.37

HR: hazard ratio; 95% CI: 95% confidence interval; NCCN: National Comprehensive Cancer Network; FR: favorable risk; IR: intermediate risk; PR: poor risk; LOD: limit of detection; LOQ: limit of quantification; BM: bone marrow; MRD: measurable residual disease; Allo-SCT, allogeneic stem cell transplant; Auto-SCT, autologous stem cell transplant.

of bone marrow precursors has forced researchers to define the MRD status as above or below a given level, which is able to anticipate a different clinical outcome, rather than as negative or positive.[23,29] Finally, the multifaceted interpretation of MRD is made even more complicated as a consequence of the therapy delivered. Different treatment schedules may have different thresholds of prognostic significance. These thresholds are currently selected by different approaches, in some cases applying empirical logarithmic scales or quartile segregation, in others applying specific statistical methods (e.g., receiver operating characteristic curve analysis or maximally-selected log-rank statistics).[23] A comprehensive review of the literature[30] prompted the ELN panel to recommend a threshold of 0.1% not because it was the most predictive but because it was used and found relevant in the majority of the published studies.[10] Nevertheless, the panel of experts was well aware that levels of MRD below 0.1% are consistent with residual leukemia and that further efforts should be made to identify and validate lower thresholds. In theory, the validation of MRD as a clinical biomarker should rely on the well-designed analysis of retrospective case series, leading to the identification of informative thresholds. Subsequently, these thresholds should be validated in prospective, MRD-oriented trials.[22] Despite these attempts, doubts will still persist because of the many different therapeutic contexts that can hamper the universal applicability of the selected thresholds. Indeed, the last Food and Drug Administration MRD guidance for the development of novel agents raised concerns about the role of MRD as a surrogate endpoint. Such concerns were due to the biological heterogeneity of AML and the lack of prospective studies having MRD negativity as a primary endpoint.[31,32] Furthermore, the putative threshold of sensitivity of the MRD assay should be at least 10-fold (1-log) below the clinical decision-making threshold.[31]

At variance, in other pathologies (e.g., acute lymphoblastic leukemia, multiple myeloma, and chronic lymphocytic leukemia), MRD assessment by multiparametric flow cytometry is highly standardized and reproducible in different treatment scenarios, so that it is proposed as a surrogate endpoint in clinical trials.[31] In these diseases, an innovative approach called next-generation flow has substantially improved the performance of standard multiparametric flow cytometry which now reaches levels of sensitivity comparable to those of reverse transcriptase quantitative polymerase chain reaction (10^{-4}–10^{-6}).[4–6,27] Such an approach requires the application of a minimum

of an eight-color panel and the acquisition of several million relevant events.[4,24] Using this approach in chronic lymphocytic leukemia, it was demonstrated that an MRD threshold of 0.01% (10^{-4}) was an independent predictor of progression-free survival in patients treated with either chemo-immunotherapy or novel agents.[33]

In the GIMEMA AML1310 trial, patients with intermediate risk, defined according to the NCCN 2009,[34] were addressed to allogeneic or autologous SCT if MRD-positive or -negative, respectively, after the consolidation cycle.[17] The threshold defining MRD negativity (0.035%) was validated in several retrospective analyses of previous EORTC/GIMEMA trials. In those analyses the threshold of 0.035% allowed discrimination of patients with clearly distinct long-term prognoses across different genetic/cytogenetic subgroups.[14,18–20] This threshold was prospectively validated in the AML1310 trial, in which delivery of allogeneic SCT prolonged the overall survival of MRD-positive intermediate-risk patients to equalize that of MRD-negative intermediate-risk patients, who underwent autologous SCT.[16]

The working hypothesis leading to the current analysis was that an MRD estimate based on the LOD-LOQ approach might further refine the outcome prediction of the 0.035% threshold. In the AML1310 trial, we assumed that the MRD-oriented post-remission strategy (allogeneic *vs.* autologous SCT) used for patients belonging to the intermediate-risk category, nullified the poor prognostic weight of MRD positivity. This resulted in an equivalent duration of overall survival and disease-free survival of MRD-negative and MRD-positive patients, with MRD positivity losing its independent prognostic value in multivariate analysis, as compared to genetic-cytogenetic risk and post-remission treatment.[16] In contrast, the LOD-LOQ calculation of MRD discriminated two populations of patients (LOD^{neg}/LOD^{pos}-LOQ^{neg} and LOQ^{pos}) with statistically significant different durations of overall survival. Multivariable analysis confirmed the independent prognostic role of the LOD-LOQ approach.

The power of the LOD-LOQ outcome prediction increased when the analysis included only samples in which the count of CD45-expressing events was at least 500,000. This observation confirms that, when dealing with the identification of rare events, the larger the denominator of relevant events the more accurate the target population estimation, provided that an adequate number of relevant events is collected (i.e., 20 for LOD and 50 for LOQ calculations). Furthermore, the availability of a marker allowing an easier extrapolation of the cell population under study (e.g. CD45) increases the accuracy of the measurement. This has also been proven true by others when MRD was determined only on the population defined by immature markers.[35]

Based on this, we assume that the LOD-LOQ MRD estimate is more accurate than the $MRD^{0.035\%}$ threshold because it enabled a superior discrimination within the $MRD^{<0.035\%}$ category. In fact, among $MRD^{<0.035\%}$ patients, the LOD^{neg} and LOD^{pos}-LOQ^{neg} status identified "true negative" or "non-quantifiable" cases with a better outcome. These patients might have been cured of their disease without allogeneic SCT, as demonstrated in a further subgroup analysis in which patients submitted to autologous SCT showed a very favorable outcome (*Online Supplementary Figure S2*). Interestingly, in our hands, LOD^{neg} and LOD^{pos}-LOQ^{neg} patients showed the same overall survival. We hypothesized at least two technical explanations. First, the median number of $CD45^{+}$ events acquired may not be sufficient. In fact, the category of LOD^{pos}-LOQ^{neg} patients might be progressively narrowed if a very high number of relevant events is acquired. Second, LOD sensitivity may have been affected by the lack of limit-of-blank subtraction, whereas the LOQ value may not have been, maintaining its predictive value.

We are aware of the preliminary nature of our report and of its possible weaknesses. The observation that an MRD estimation system independent of a pre-established threshold performs as well as in retrospective and prospective contexts is *per se* relevant, even though far from representing the identification of an absolute threshold. This proof of principle will become standard of care when its predictive value is demonstrated in different series of patients, treated with different schedules. Meanwhile, all MRD-driven clinical studies should rigorously comply with the procedures recommended for the acquisition of rare events. In our analysis, increasing the numbers of events acquired (>500,000) and refining the population under investigation (gating $CD45^{+}$ cells) resulted in a significantly enhanced predictive power of the test.

Thresholds for MRD estimation are likely to change in the near future but making them clinically informative requires that for every individual determination, the detection and quantification limits are described. Along this direction, multiparametric flow cytometry analyses in AML would possibly reach values of sensitivity comparable to those of polymerase chain reaction, as demonstrated in acute lymphoblastic leukemia and multiple myeloma.[4]

Disclosures

No conflicts of interest to disclose.

Contributions

FB, RP, and AV designed the study, FB, AP and VA collected and analyzed the data, FB, RP, AP, VA, LM, MIDP, GP, MAIC, TO, MD, CC, DF, SL, WA, MTV and AV wrote the paper and approved its final version.

Acknowledgments

The authors are indebted to Prof. Bruno Brando and Dr. Ari-

anna Gatti for their critical review of the manuscript and helpful suggestions.

Funding

This work was supported by a grant from the Ministero della Salute, Rome, Italy (Finalizzata 2018, NET-2018-12365935, Personalized medicine program on myeloid neoplasms: characterization of the patient's genome for clinical decision making and systematic collection of real world data to improve quality of health care)

Data-sharing statement

For original, anonymized data, please contact the corresponding author (francesco.buccisano@uniroma2.it).

References

1. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129(4):424-447.
2. Wood B, Jevremovic D, Béné MC, Yan M, Jacobs P, Litwin V. Validation of cell-based fluorescence assays: practice guidelines from the ICSH and ICCS - part V - performance criteria. Cytometry B Clin Cytom. 2013;84(5):315-323.
3. Wood BL. Principles of minimal residual disease detection for hematopoietic neoplasms by flow cytometry. Cytometry B Clin Cytom. 2016;90(1):47-53.
4. Theunissen P, Mejstrikova E, Sedek L, et al. Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia. Blood. 2017;129(3):347-357.
5. Paiva B, Puig N, Cedena MT, et al. Measurable residual disease by next-generation flow cytometry in multiple myeloma. J Clin Oncol. 2020;38(8):784-792.
6. Flores-Montero J, Sanoja-Flores L, Paiva B, et al. Next generation flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia. 2017;31(10):2094-2103.
7. Rawstron AC, Gregory WM, de Tute RM, et al. Minimal residual disease in myeloma by flow cytometry: independent prediction of survival benefit per log reduction. Blood. 2015;125(12):1932-1936.
8. Rawstron AC, Fazi C, Agathangelidis A, et al. A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. Leukemia. 2016;30(4):929-936.
9. Wood BL. Acute myeloid leukemia minimal residual disease detection: the difference from normal approach. Curr Protoc Cytom. 2020;93(1):e73.
10. Schuurhuis GJ, Heuser M, Freeman S, et al. Minimal/measurable residual disease in AML: consensus document from ELN MRD Working Party. Blood. 2018;131(12):1275-1291.
11. Hanekamp D, Bachas C, van de Loosdrecht A, Ossenkoppele G, Cloos J. Re: myeloblasts in normal bone marrows expressing leukaemia-associated immunophenotypes. Pathology. 2020;52(2):289-291.
12. Venditti A, Buccisano F, Del Poeta G, et al. Level of minimal residual disease after consolidation therapy predicts outcome in acute myeloid leukemia. Blood. 2000;96(12):3948-3952.
13. Buccisano F, Maurillo L, Gattei V, et al. The kinetics of reduction of minimal residual disease impacts on duration of response and survival of patients with acute myeloid leukemia. Leukemia. 2006;20(10):1783-1789.
14. Maurillo L, Buccisano F, Del Principe MI, et al. Toward optimization of postremission therapy for residual disease-positive patients with acute myeloid leukemia. J Clin Oncol. 2008;26(30):4944-4951.
15. Buccisano F, Maurillo L, Spagnoli A, et al. Cytogenetic and molecular diagnostic characterization combined to postconsolidation minimal residual disease assessment by flow cytometry improves risk stratification in adult acute myeloid leukemia. Blood. 2010;116(13):2295-2303.
16. Venditti A, Piciocchi A, Candoni A, et al. GIMEMA AML1310 trial of risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia. Blood. 2019;134(12):935-945.
17. Venditti A, Piciocchi A, Candoni A, et al. GIMEMA AML1310 trial of risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia. Blood. 2019;134(12):935-945.
18. Buccisano F, Maurillo L, Spagnoli A, et al. Cytogenetic and molecular diagnostic characterization combined to postconsolidation minimal residual disease assessment by flow cytometry improves risk stratification in adult acute myeloid leukemia. Blood. 2010;116(13):2295-2303.
19. Maurillo L, Buccisano F, Piciocchi A, et al. Minimal residual disease as biomarker for optimal biologic dosing of ARA-C in patients with acute myeloid leukemia. Am J Hematol. 2015;90(2):125-131.
20. Buccisano F, Maurillo L, Gattei V, et al. The kinetics of reduction of minimal residual disease impacts on duration of response and survival of patients with acute myeloid leukemia. Leukemia. 2006;20(10):1783.
21. Maurillo L, Buccisano F, Del Principe MI, et al. Toward optimization of postremission therapy for residual disease-positive patients with acute myeloid leukemia. J Clin Oncol. 2008;26(30):4944-4951.
22. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: Theoretical considerations and practical challenges. J Clin Oncology. 2009;27(24):4027-4034.
23. Buccisano F, Maurillo L, Del Principe MI, et al. Prognostic and therapeutic implications of minimal residual disease detection in acute myeloid leukemia. Blood. 2012;119(2):332-341.
24. Arroz M, Came N, Lin P, et al. Consensus guidelines on plasma cell myeloma minimal residual disease analysis and reporting. Cytom B Clin Cytom. 2016;90(1):31-39.
25. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937-951.
26. Hedley BD, Keeney M. Technical issues: flow cytometry and rare event analysis. Int J Lab Hematol. 2013;35(3):344-350.
27. Rawstron AC, Böttcher S, Letestu R, et al. Improving efficiency and sensitivity: European Research Initiative in CLL (ERIC) update on the international harmonised approach for flow cytometric residual disease monitoring in CLL. Leukemia. 2013;27(1):142-149.

28. Othus M, Gale RP, Hourigan CS, Walter RB. Statistics and measurable residual disease (MRD) testing: uses and abuses in hematopoietic cell transplantation. Bone Marrow Transplant. 2020;55(5):843-850.
29. Buccisano F, Maurillo L, Del Principe MI, et al. Minimal residual disease as a biomarker for outcome prediction and therapy optimization in acute myeloid leukemia. Expert Rev Hematol. 2018;11(4):307-313.
30. Ossenkoppele G, Schuurhuis GJ. MRD in AML: does it already guide therapy decision-making? Hematol Am Soc Hematol Educ Progr. 2016;2016(1):356-365.
31. Hematologic Malignancies: Regulatory Considerations for Use of Minimal Residual Disease in Development of Drug and Biological Products for Treatment | FDA. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/hematologic-malignancies-regulatory-considerations-use-minimal-residual-disease-development-drug-and (accessed May 2, 2020).
32. Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukaemia. Leukemia. 2017;31(7):1482-1490.
33. Zalcberg I, D'Andrea MG, Monteiro L, Pimenta G, Xisto B. Multidisciplinary diagnostics of chronic lymphocytic leukemia: European Research Initiative on CLL - ERIC recommendations. Hematol Transfus Cell Ther. 2020;42(3):269-274.
34. Acute Myeloid Leukemia, Version 1.2009, NCCN Clinical Practice Guidelines in Oncology. https://www.nccn.org/professionals/physician_gls/pdf/aml.pdf [Accessed 6 May, 2018].
35. Hanekamp D, Tettero JM, Ossenkoppele GJ, et al. AML/normal progenitor balance instead of total tumor load (MRD) accounts for prognostic impact of flowcytometric residual disease in AML. Cancers (Basel). 2021;13(11):2597.