Effect of age and treatment on predictive value of measurable residual disease: implications for clinical management of adult patients with acute myeloid leukemia

Francesco Mannelli,[1,2] Matteo Piccini,[1] Sara Bencini,[3] Giacomo Gianfaldoni,[1] Benedetta Peruzzi,[3] Roberto Caporale,[3] Barbara Scappini,[1] Laura Fasano,[1] Elisa Quinti,[1] Gaia Ciolli,[1] Andrea Pasquini,[1] Francesca Crupi,[1] Sofia Pilerci,[1] Fabiana Pancani,[1,2] Leonardo Signori,[1,2] Danilo Tarantino,[1,2] Chiara Maccari,[1,2] Vivian Paradiso,[1] Francesco Annunziato,[3] Paola Guglielmelli[1,2] and Alessandro M. Vannucchi[1,2]

[1]SOD Ematologia, Università di Firenze, AOU Careggi, [2]Centro Ricerca e Innovazione Malattie Mieloproliferative (CRIMM), AOU Careggi and [3]Centro Diagnostico di Citofluorimetria e Immunoterapia, AOU Careggi, Firenze, Italy

Correspondence: F. Mannelli
francesco.mannelli@unifi.it

Received: March 24, 2023.
Accepted: June 15, 2023.
Early view: June 22, 2023.

https://doi.org/10.3324/haematol.2023.283196

Abstract

Measurable residual disease (MRD) is a powerful predictor of outcome in acute myeloid leukemia. In the early phases of treatment, MRD refines initial disease risk stratification and is used for the allocation to allogeneic transplant. Despite its well-established role, a relatively high fraction of patients eventually relapses albeit achieving MRD^{neg} status. The aim of this work was to assess specifically the influence of baseline features and treatment intensity on the predictive value of an MRD^{neg} status, particularly focusing on MRD2, measured after two consecutive chemotherapy cycles. Among baseline features, younger MRD2^{neg} patients (<55 years) had a significantly longer disease-free survival (median not reached) compared to their older counterparts (median 25.0 months, *P*=0.013, hazard ratio=2.08). Treatment intensity, specifically the delivery of a high dose of cytarabine in induction or first consolidation, apparently had a pejorative effect on the outcome of MRD-2^{neg} patients compared to standard dose (*P*=0.048, hazard ratio=1.80), a finding also confirmed by the analysis of data extracted from the literature. The combination of age and treatment intensity allowed us to identify categories of patients, among those who reached a MRD2^{neg} status, characterized by significantly different disease-free survival rate. Our data showed that variables such as age and intensity of treatment administered can influence the predictive value of MRD in patients with acute myeloid leukemia. In addition to underscoring the need for further improvement of MRD analysis, these findings call for a reasoned application of MRD data, as currently available, to modulate consolidation therapy on adequately estimated relapse rates.

Introduction

The determination of measurable residual disease (MRD) relies on the detection of leukemic cells below the morphology-based threshold that defines complete remission.[1] The persistence of MRD is a powerful predictor of disease relapse and poor outcome in acute myeloid leukemia (AML),[2-6] a role formally recognized in the European LeukemiaNet (ELN) recommendations, which defines a specific response category of complete response (CR) without MRD.[1] The prognostic impact of MRD measurement in AML has been demonstrated regardless of the method applied to assess it,[6] although the ELN consensus suggests employing a molecular probe for core binding factor (CBF)-related and *NPM1*-mutated AML, and a multiparameter flow cytometry (MFC) approach in all other subgroups.[7]

In the early phases of treatment, MRD assessment refines initial disease risk stratification and is primarily used for the allocation of eligible patients to allogeneic hematopoietic stem cell transplant (HSCT), generally reserved for categories deemed to be at high risk of relapse.[8] From this perspective, the most informative time-point is set after two cycles of intensive chemotherapy (i.e., MRD2), as it has been shown to yield the greatest predictive value on outcome.[3,5,9] Such a strategy is stated as standard by ELN guidelines[7,10] and is increasingly becoming the basis for clinical decision-making in AML management.[9,11,12]

Despite its indisputable role, the prognostic value of MRD may be influenced by the clinical context, including age range[13] and genetic profile,[14] as well as by pre-MRD intensity

of treatment. In fact, the predictive value of MRD2 reflects the degree of chemosensitivity of leukemic cells: its positioning after two chemotherapy cycles provides a reliable estimation of the likelihood that further chemotherapy may lead to disease eradication. Overall, the reliability of MRD2 largely depends on previously delivered treatment, including the dosages of the backbone agents (cytarabine and anthracycline), the use of a third drug alongside (etoposide, fludarabine, gemtuzumab ozogamicin and, more recently for *FLT3*-mutant cases, an FLT3 inhibitor), as well as the regimen schedule (standard, double induction, sequential). Therefore, MRD2 status somehow captures information on disease sensitivity to the treatment actually delivered. Furthermore, MRD is measured after achievement of CR following induction therapy, and the intensity of chemotherapy in the first cycle is known to influence the rate of CR.[15-17] In other words, different treatment intensity in the first cycle contributes to shape the population of patients in which MRD is later assessed starting from CR achievement onward, and this might affect the ability of MRD itself to estimate outcome. Although the influence of such an effect is systematically stated in guidelines,[7] its impact has not been fully evaluated and is generally not taken into account in clinical decision-making. The aim of this work was to assess specifically the influence of baseline features (age, white blood cell count, ELN stratification), and treatment intensity on the predictive value of MRD, focusing specifically on MRD2-negative (MRD2^{neg}) status because of the key clinical information it can provide. For this purpose, we retrospectively analyzed our database and then interrogated the available literature to validate our findings.

Methods

Patients

Patients entering the study had a diagnosis of non-promyelocytic AML, based on morphological, immunophenotypic, and molecular criteria. Once eligible for intensive chemotherapy, they were consecutively assigned to treatment based on the availability of a clinical trial at the time of diagnosis and on local institutional treatment choices as previously reported[18] and detailed in the *Online Supplementary Data*. For the purpose of the study, the treatment intensity of the first two chemotherapy cycles was categorized according to the dosage of cytarabine (ARA-C): standard dose (SDAC) included "3+7" and "3+7"-like regimens sharing a single-dose infusion of 100-200 mg/m^2 and a cumulative dose of up to 1,400 mg/m^2 per cycle; high dose (HDAC) included several schedules sharing a single dose of at least 1,000 mg/m^2 and a cumulative dose of at least 6,000 mg/m^2 (*Online Supplementary Data*). CPX-351 was categorized as SDAC.

The study was approved by the local institutional review board (protocol number: 2013/0021560). Written informed consent was obtained from study patients in accordance with the Declaration of Helsinki. Enrollment criteria required were (i) intensive treatment, (ii) CR achievement after first induction therapy, and (iii) availability of MRD data after induction (MRD1) and first consolidation cycle (MRD2).

Measurable residual disease study

Multiparameter flow cytometry

MFC study files reporting individual leukemia-aberrant immunephenotype (LAIP) profiles were acquired locally according to pre-defined standard operating procedures. LAIP profiles were defined by antigen expression on blasts from fresh diagnostic bone marrow samples (or peripheral blood in the case of punctio sicca). Standard MFC methodologies for LAIP definition are detailed in the *Online Supplementary File* and reported elsewhere.[19] MRD was expressed as the percentage of LAIP$^+$ cells on CD45$^+$ cells. MRD was defined as positive for any detection ≥0.1%, in accordance with ELN recommendations.[7,20]

Polymerase chain reaction-based assessment

Sensitive real-time quantitative-polymerase chain reaction assays (RQ-PCR) were used for detection of MRD in patients with a suitable molecular probe (*RUNX1T1* and *CBFB-MYH11* transcripts and *NPM1* gene mutations). RQ-PCR was performed following the Europe Against Cancer (EAC) program recommendations[21] with a sensitivity of 10^{-5}. MRD was defined as positive for any detection ≥0.01%. Standard PCR methodologies are detailed in the *Online Supplementary File*.

Definitions

CR, disease-free survival (DFS), and overall survival (OS) were defined according to standard criteria.[1] As regards prognostic stratification, the criteria used for the therapeutic decision-making across different time periods, and particularly allocation to HSCT, are specified in the *Online Supplementary File*. In *post hoc* analysis, patients were stratified, according to the availability of genetic data, by Medical Research Council (MRC) criteria,[22] and ELN version 2010-2017.[1,23]

Analysis of literature

We carried out a search in PubMed for articles published between 2000 and 2021 by filtering for keywords (AML, acute myeloid leukemia, or acute myelogenous leukemia, and MRD, minimal residual disease, or measurable residual disease). The analytical method, the criteria for paper selection and the extracted data are detailed in the *Online Supplementary Materials and Methods S3, Online Supplementary Table S1.* Based on this assessment, a total of 33 articles were selected. We then extracted survival data from Kaplan-Meier curves by using commercial graph digitizer software (Digitizelt, version 2.1; Bormisoft) and applying a previously published algorithm to reconstruct survival data

for MRD^{pos} and MRD^{neg} cases.[24] The main characteristics of treatment in the first two cycles (drugs, ARA-C dosage, schedule) were obtained for each report and tabulated as in *Online Supplementary Table S2*.

Statistical methods

All statistical analyses were performed using R version 3.5.0 and SPPS version 26. Statistical methods are detailed in the *Online Supplementary File (Materials and Methods S5)*.

Results

Characteristics of patients and treatment flow

From April 2004 to January 2022, 194 patients affected by AML met the inclusion criteria. Considering the large enrollment period, we carried out an analysis of OS to check for an influence by the year of diagnosis (*Online Supplementary Figure S1*): no relevant impact on survival emerged. The clinical and biological characteristics of our patients are detailed in Table 1. In addition to baseline features, we classified patients according to treatment intensity (SDAC-based and HDAC-containing regimens) and MRD status after induction (MRD1) and first consolidation (MRD2); patients receiving HDAC in at least one of the first two cycles, in induction (cycle 1) or consolidation (cycle 2) (n=124, 63.9%), were compared with those receiving SDAC in both cycles (n=70, 36.1%) (Figure 1).

Measurable residual disease study

For CBF-related and *NPM1*-mutated AML, the categorization according to MRD was based on PCR data, available for all 33 CBF-related cases, and for 62/80 (77.5%) *NPM1*-mutated ones. In all other cases, MFC was used. After induction, 110 (56.7%) patients were defined as MRD1^{neg} and 84 (43.3%) as MRD1^{pos}. After first consolidation, a total of 121 (62.4%) cases reached MRD2^{neg} status, whereas 73 (37.6%) patients resulted MRD2^{pos} (Figure 1). MRD1 and MRD2 status showed an impact on DFS (*Online Supplementary Figures S2* and *3*), with MRD2 being more effective in discriminating outcome

Table 1. Characteristics of patients according to treatment group after induction and the first consolidation cycle.

	Overall N=194	SDAC N=70 (36.1%)	HDAC N=124 (63.9%)	*P*	
Age in years, median (range)	55 (16-73)	55 (22-70)	53 (16-73)	0.75	-
WBC, x10^9/L, median (range)	14.8 (0.6-435.0)	14.9 (0.6-191.0)	13.6 (0.9-435.0)	0.90	-
Hb, g/dL, median (range)	9.2 (3.9-14.9)	9.2 (5.4-14.5)	9.0 (3.9-14.9)	0.25	-
Platelets, x10^9/L, median (range)	53 (1-281)	47 (10-152)	57 (1-281)	0.053	-
Peripheral blasts, %, median (range)	53 (0-100)	58 (0-100)	52 (0-98)	0.30	-
Secondary AML, N (%)	12 (6.2%)	3 (4.3)	9 (7.3)	0.54	-
Karyotype, N (%) Favorable Normal Intermediate, non-normal Adverse Lack of growth	 33 (17.0) 110 (56.7) 26 (13.4) 15 (7.7) 10 (5.2)	 13 (18.6) 41 (58.6) 9 (12.9) 2 (2.9) 5 (7.1)	 20 (16.1) 69 (55.6) 17 (13.7) 13 (10.5) 5 (4.0)	 0.69 0.76 1.0 0.09 0.50	0.38
Molecular genetics, N (%) *NPM1*-mutated *FLT3*-ITD *CEBPA-bZIP*	 80 (41.2) 41 (21.1) 10 (5.2)	 39 (55.7) 16 (22.9) 4 (5.7)	 47 (37.9) 25 (20.2) 6 (4.8)	 0.23 0.71 0.74	-
ELN 2010 risk group, N (%) Favorable Intermediate-1 Intermediate-2 Adverse	 90 (46.4) 80 (41.2) 8 (4.1) 16 (8.2)	 39 (55.7) 27 (38.6) 2 (2.9) 2 (2.9)	 51 (41.1) 53 (42.7) 6 (4.8) 14 (11.3)	 0.06 0.64 0.71 0.06	**0.01**
ELN 2017 risk group, N (%) Favorable Intermediate Adverse Not assessable	 102 (52.5) 62 (32.0) 24 (12.4) 6 (3.1)	 41 (58.7) 19 (27.1) 5 (7.1) 5 (7.1)	 61 (49.2) 43 (34.7) 19 (15.3) 1 (0.8)	 0.23 0.33 0.11	0.05

Differences between treatment groups were evaluated using the Mann-Whitney test for continuous variables and Fisher exact test or χ^2 for categorical variables. *P* values <0.05 are statistically significant. SDAC: standard-dose cytarabine (both in induction and consolidation cycle); HDAC: high-dose cytarabine (at least in one cycle, induction or first consolidation); WBC: white blood cells; Hb: hemoglobin; AML: acute myeloid leukemia; ITD: internal tandem duplication; ELN: European LeukemiaNet.

categories (C-index 0.61 *vs*. 0.547, *P*=0.009). MRD2 status also significantly influenced OS (*P*=0.0014, hazard ratio [HR]=2.08) (*Online Supplementary Figure S3*).

Analysis according to baseline features

Due to its greater informativeness for prognosis, further analyses were focused on MRD2. We searched for any interaction between the prognostic impact of MRD2 and baseline features. First, patients were stratified for age: the observed median age was 55 years. By means of a receiving operating characteristic curve analysis using relapse as a binary endpoint, we confirmed an age threshold of 55.5 years as having the best performance (Youden score=0.224) on the basis of the combination of a sensitivity and specificity of 0.564 and 0.66, respectively. Accordingly, we separated patients for their age <55 years and ≥55 years. We also categorized patients for gender, white blood cell count (<30x10^9/L and ≥30x10^9/L), and ELN category (favorable and intermediate). ELN adverse category was too limited in size (n=24, of whom 13 MRD2^{neg}) for such an analysis. Significant differences between MRD2^{neg} and MRD2^{pos} patients were found as regarded DFS and OS in subgroups defined by age, white blood count, and ELN category (*Online Supplementary Figures S4-S6*) and within the female patient category (*Online Supplementary Figure S6*). In male patients, we observed a non-significant trend in favor of the MRD2^{neg} group (*Online Supplementary Figure S7*). Since we observed a still remarkable rate of relapse in the MRD2^{neg} group (DFS rate at 2 years, 60.2% *vs*. 32.1% in MRD2^{pos} patients, *P*=0.00013), we searched for variables having an impact on DFS by comparing the outcome of MRD2^{neg} patients in different disease categories. We found that younger (i.e., <55 years) MRD2^{neg} patients showed a significantly longer DFS than their older (i.e., ≥55 years) counterparts (*P*=0.013, HR=2.08) (*Online Supplementary Figure S6*), while no significant differences emerged for gender- (*P*=0.11), white blood count- (*P*=0.29) and ELN- (*P*=0.1) related strata (*Online Supplementary Figure S8*).

Analysis according to treatment groups

We then analyzed the baseline characteristics and outcomes of patients stratified according to treatment intensity (SDAC

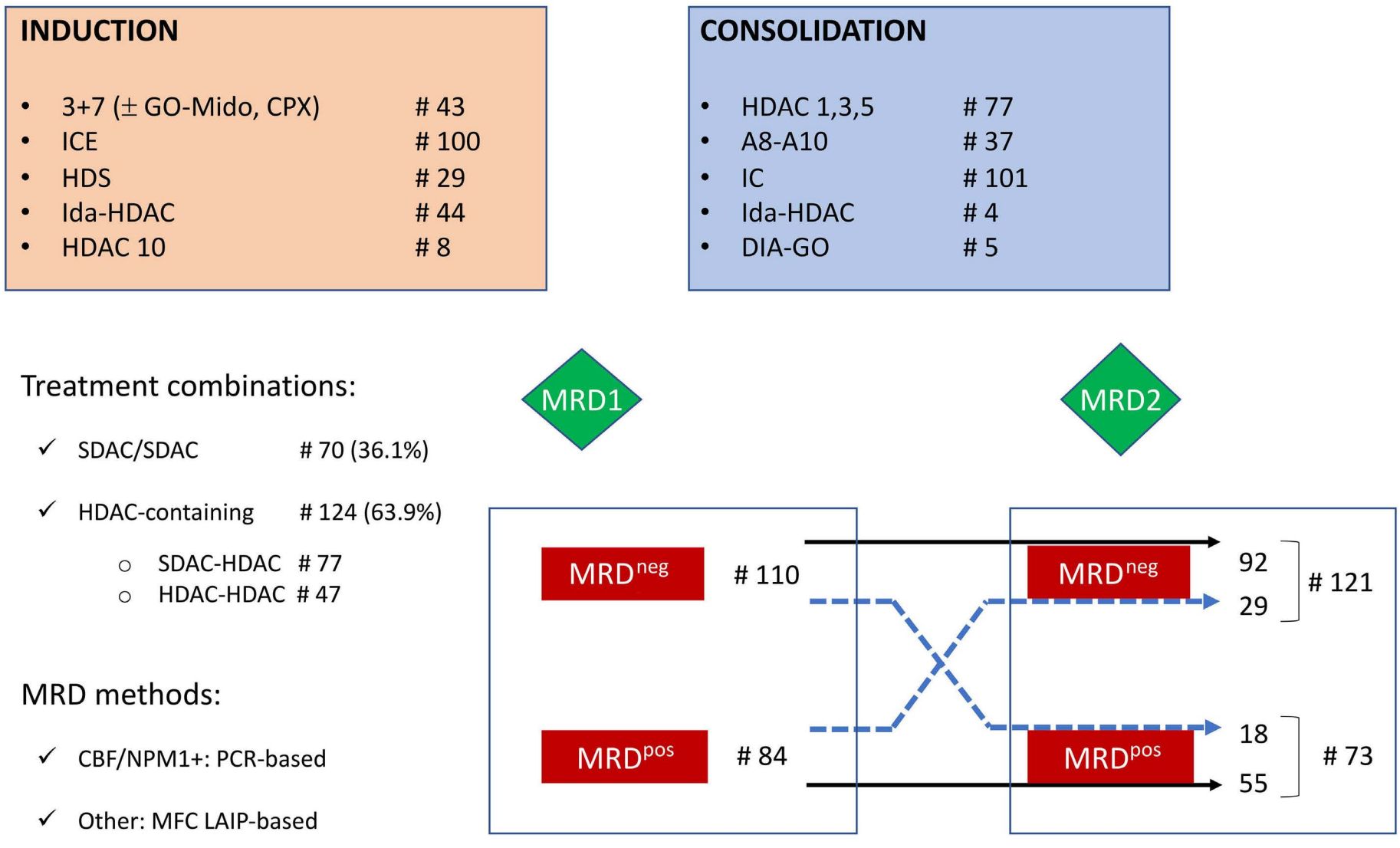


Figure 1. Kinetics of measurable residual disease according to timepoint of assessment. Flow of data for measurable residual disease (MRD) status (negative, MRD^{neg}; positive, MRD^{pos}) at the post-induction (MRD1) and post-consolidation (MRD2) timepoints. The chemotherapy regimens and the number of patients treated with each scheme are listed for induction (pink panel) and consolidation (blue panel) cycle. The cycles were classified according to the dose of cytarabine as standard dose (SDAC) or high dose (HDAC). Treatment details for each chemotherapy scheme are provided in the *Online Supplementary File*. GO: gemtuzumab ozogamicin; Mido: midostaurin; CPX: CPX-351; ICE: idarubicin, cytarabine, etoposide; HDS: high-dose sequential; Ida: idarubicin; HDAC 10: high-dose cytarabine 10 g/m^2; HDAC 1,3,5: high-dose cytarabine days 1, 3, and 5; DIA: intermediate-dose cytarabine; PCR: polymerase chain reaction; LAIP: leukemia-associated immunophenotype.

vs. HDAC). The only statistically significant difference was younger age in patients treated with HDAC-containing induction (47 *vs*. 57 years; *P*<0.0001) (*Online Supplementary Table S3*). No significant difference in outcome was observed according to cytarabine dose in induction (*Online Supplementary Figure S9*).

We further stratified patients according to the cumulative effect of induction and first consolidation, i.e., those who received at least one HDAC-containing cycle *versus* those treated only with SDAC. We only observed a trend towards enrichment in 2010 ELN adverse-risk categories within the HDAC-treated group (Table 1). However, no significant outcome benefit was highlighted for patients receiving HDAC during induction or consolidation (*Online Supplementary Figure S9*).

Effects of treatment intensity on MRD2 predictive capability

We searched for interactions between MRD2 status and

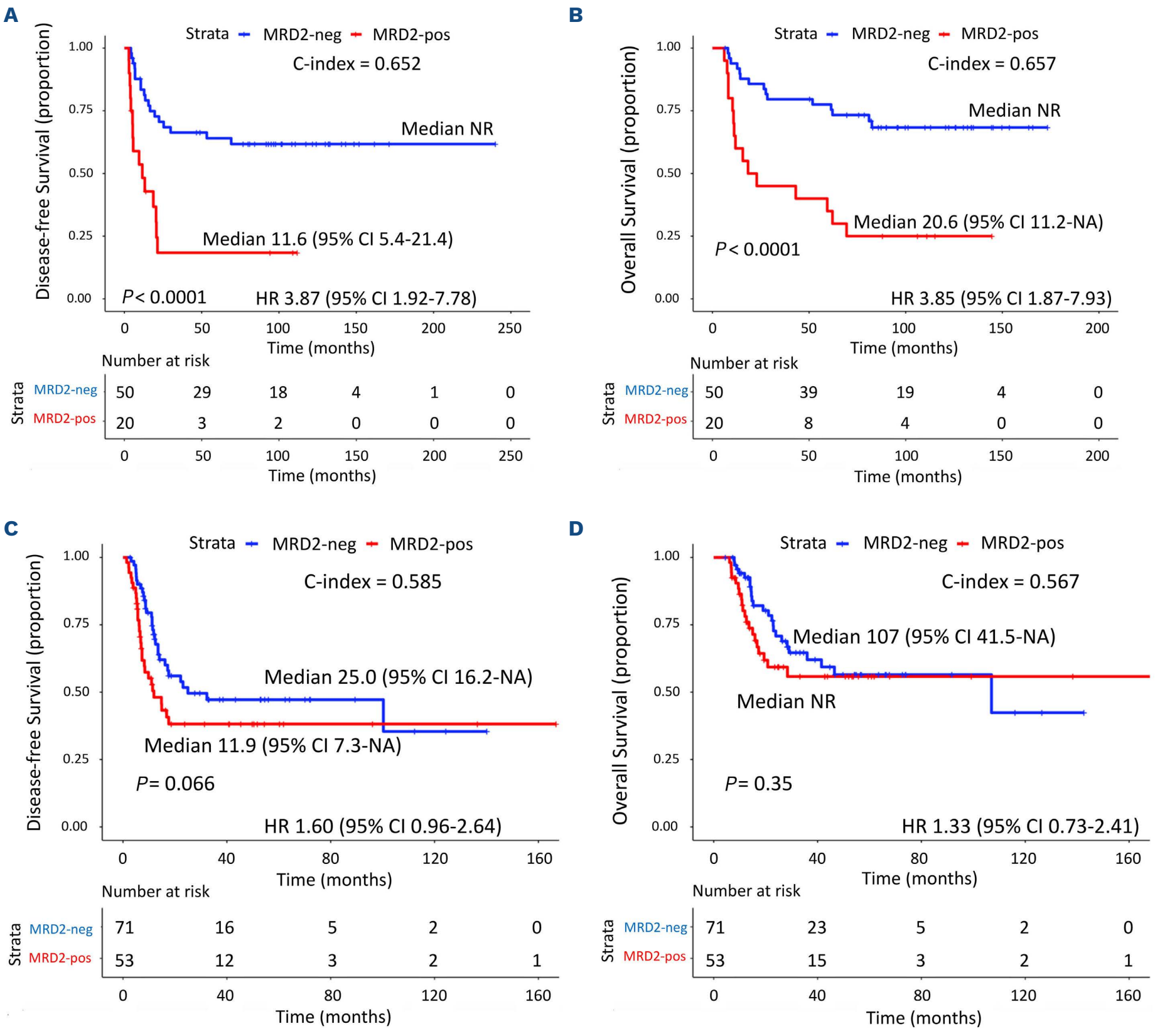


Figure 2. Analysis of survival according to measurable residual disease status after first consolidation in different treatment groups. (A, C) Disease-free survival and (B, D) overall survival according to measurable residual disease status after first consolidation (MRD2) in patients treated with standard-dose cytarabine in both induction and consolidation (A, B) or with at least one cycle including high-dose cytarabine (C, D). The curves of patients with MRD2^{neg} status are depicted in blue; the curves of patients with MRD2^{pos} status are depicted in red. Median survival estimates are reported in months. NR: not reached; HR: hazard ratio; 95% CI: 95% confidence interval; NA: not available.

outcome within different treatment groups. We observed statistically significant longer DFS in MRD2^{neg} *versus* MRD-2^{pos} patients who received SDAC during induction and first consolidation (HR=3.87, *P*<0.0001, C-index=0.652) (Figure 2A). Contrariwise, MRD2 status failed to significantly resolve DFS in patients treated with at least one HDAC-containing cycle, in induction or consolidation (HR=1.60, *P*=0.066, C-index=0.585) (Figure 2C). Similar findings concerned the impact of MRD2 on OS, which was associated with a significant value in patients treated with SDAC (*P*<0.0001, HR=3.85, 95% confidence interval [95% CI]: 1.87-7.93), unlike what was observed for HDAC (*P*=0.35, HR=1.33, 95% CI: 0.73-2.41) (Figure 2B-D). The above differences were also maintained in the category of intermediate-risk karyotype: the favorable impact of MRD2^{neg} status on DFS was more evident in patients receiving SDAC (HR=3.98, *P*=0.00078, C-index=0.65) than HDAC (HR=2.13, *P*=0.01, C-index=0.613) (*Online Supplementary Figure S10*), although statistical significance was reached also in the latter case. Conversely, among MRD2^{pos} patients, DFS was largely unaffected by treatment arm (median 11.6 and 11.4 months in SDAC- and HDAC-treated patients, respectively) and the worse performance of MRD2 in the HDAC group was mainly the consequence of a relatively high rate of relapse in MRD2^{neg} patients. OS estimates followed the same pattern in the intermediate-risk karyotype tier (*Online Supplementary Figure S10*).

Characteristics of MRD2^{neg} patients according to treatment arm

We then analyzed MRD2^{neg} patients according to the intensity of the first two chemotherapy cycles. At baseline, there was a statistically significant enrichment of ELN adverse-risk patients in HDAC-treated (16.9% according to 2017 version) compared to SDAC-treated MRD2^{neg} cases (2.0%, *P*=0.014) (Table 2). When comparing patients

Table 2. Characteristics of MRD2^{neg} patients according to treatment group after induction and first consolidation cycle.

	MRD2^{neg} overall N=121	MRD2^{neg} SDAC N=50 (41.3%)	MRD2 HDAC N=71 (58.7%)	*P*	
Age in years, median (range)	55 (22-73)	55 (22-69)	55 (22-73)	0.23	-
WBC, x10^9/L, median (range)	8.9 (0.6-380.0)	11.8 (0.6-191.0)	7.5 (0.9-380.0)	0.30	-
Hb, g/dL, median (range)	9.3 (3.9-14.1)	9.3 (5.4-13.3)	9.3 (3.9-14.1)	0.55	-
Platelets, x10^9/L, median (range)	56 (9-271)	47 (10-152)	71 (9-271)	**0.03**	-
Peripheral blasts, %, median (range)	47 (0-100)	57 (0-100)	40 (0-98)	0.06	-
Secondary AML, N (%)	10 (8.3)	7 (14.0)	3 (4.2)	0.09	-
Karyotype, N (%) Favorable Normal Intermediate, non-normal Adverse Lack of growth	 13 (10.7) 72 (59.5) 19 (15.7) 9 (7.5) 8 (6.6)	 8 (16.0) 31 (62.0) 6 (12.0) 1 (2.0) 4 (8.0)	 5 (7.0) 41 (57.7) 13 (18.3) 8 (11.3) 4 (5.6)	 0.14 0.71 0.45 0.07 0.72	0.148
Molecular genetics, N (%) *NPM1*-mutated *FLT3*-ITD *CEBPA-bZIP*	 50 (41.3) 20 (16.5) 10 (8.3)	 25 (50.0) 9 (18.0) 4 (8.0)	 25 (35.2) 11 (15.5) 6 (8.5)	 0.13 0.19 1.0	-
ELN 2010 risk group, N (%) Favorable Intermediate-1 Intermediate-2 Adverse	 55 (45.5) 50 (41.3) 6 (5.0) 10 (8.2)	 30 (60.0) 17 (34.0) 2 (4.0) 1 (2.0)	 25 (35.2) 33 (46.5) 4 (5.6) 9 (12.7)	 0.009 0.19 1.0 0.045	**0.003**
ELN 2017 risk group, N (%) Favorable Intermediate Adverse Not assessable	 62 (51.3) 42 (34.7) 13 (10.7) 4 (3.3)	 32 (64.0) 13 (24.0) 1 (2.0) 4 (10.0)	 30 (42.2) 29 (40.8) 12 (16.9) 0 (-)	 **0.026** 0.12 **0.014** **0.027**	**0.001**

Differences between treatment groups were evaluated using the Mann-Whitney test for continuous variables and Fisher exact test or χ^2 for categorical variables. *P* values <0.05 are statistically significant. MRD2: measurable residual disease assessed after induction therapy and first consolidation cycle; SDAC: standard-dose cytarabine (both in induction and consolidation cycle); HDAC: high-dose cytarabine (at least in one cycle, induction or first consolidation); WBC: white blood cells; Hb: hemoglobin; AML: acute myeloid leukemia; ITD: internal tandem duplication; ELN: European LeukemiaNet.

reaching MRD2^{neg} status after SDAC or HDAC, significantly different DFS (*P*=0.048, HR=1.80, 95% CI: 1.0-3.23) and OS (*P*=0.049, HR=1.94, 95% CI: 0.99-3.8) were highlighted (*Online Supplementary Figure S11*). After being superimposable at 6 months (93.9% and 90.0% for SDAC- and HDAC-treated patients, respectively), DFS rates started to diverge at 12 months (83.4% *vs*. 71.5%) and were 70.6% *vs*. 51.7% at 24 months. Similar trends in DFS were observed in HDAC- *vs*. SDAC-treated patients with intermediate-risk karyotype (*P*=0.092) and ELN 2017 (*P*=0.23) categories (*Online Supplementary Figure S12*).

Combined model including age and treatment intensity

Based on the observed influence of age and treatment on

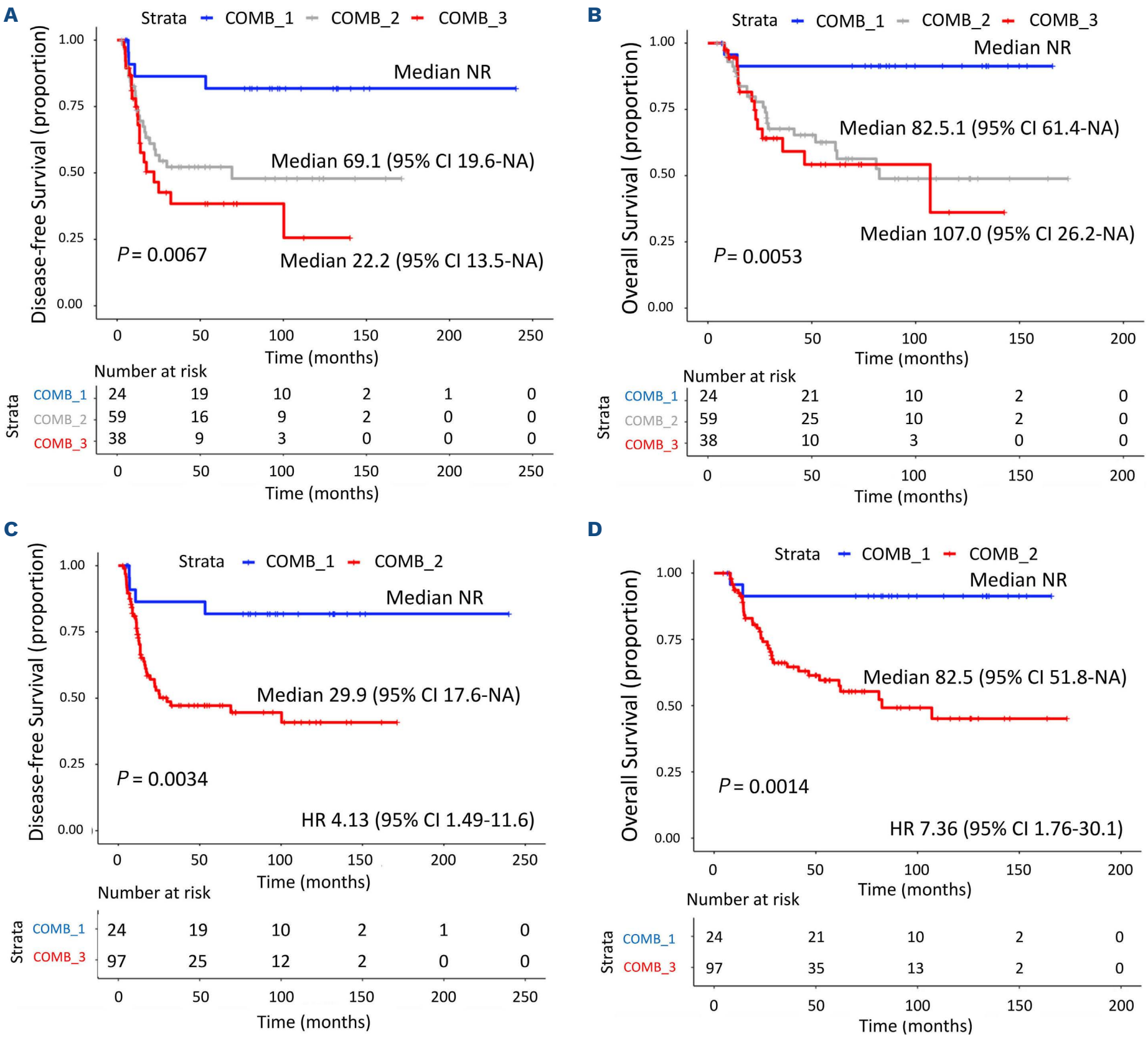


Figure 3. Analysis of major outcomes for patients with negative measurable residual disease after first consolidation according to the combined model. (A, C) Disease-free survival and (B, D) overall survival according to the combined model including age range and treatment intensity in patients with measurable residual disease negativity after first consolidation (MRD2^{neg}): the curves of patients aged less than 55 years and treated with standard-dose cytarabine in both induction and consolidation (COMB_1) are depicted in blue. In the upper panels, the curves of patients aged more than or equal to 55 years and treated with high-dose cytarabine in at least one of the first two cycles (COMB_3) are depicted in red; the remaining patients (COMB_2) are depicted in gray. In the lower panels, COMB_2 and COMB_3 categories are grouped, and their survival curves depicted in red. Median survival estimates are reported in months. NR: not reached; HR: hazard ratio; 95% CI: 95% confidence interval; NA: not available.

the prognostic value of MRD2^{neg} status, we combined these two variables and stratified MRD2^{neg} patients accordingly. Twenty-four (out of 121, 19.8%) patients were younger than 55 years and reached MRD2 negativity after two SDAC-based cycles (COMB_1): they showed a DFS rate of 86.4% at 3 years. At the opposite, the group of 38 (31.4%) elderly patients who had received at least one HDAC-based cycle (COMB_3) had a median DFS of 22.2 months and a 3-year DFS rate of 46.6% (Figure 3A, B). The remaining patients (COMB_2) displayed an intermediate behavior, but with DFS largely superimposable to the latter group in the first 2 years after CR achievement (Figure 3A). After combining COMB_2 and _3 patients, survival estimates for DFS (*P*=0.0034, HR=4.13) and OS (*P*=0.0014, HR=7.36) were significantly different with respect to that of COMB-1 patients (Figure 3C, D). The significant effect of the model on DFS was maintained in intermediate-risk stratification tiers according to karyotype (*P*=0.013) (*Online Supplementary Figure S13*) and ELN 2017 category (*P*=0.016) (*Online Supplementary Figure S13*).

Allogeneic hematopoietic stem cell transplant

Further analysis included censoring for patients who received HSCT. The differential effect of age and treatment intensity on DFS of MRD2^{neg} patients was confirmed in this analysis, either as single variables (*Online Supplementary Figure S14*) or in the combined model (*Online Supplementary Figure S15*). A non-significant trend for a lower actual rate of HSCT in first CR was observed for MRD2^{neg} (32/121, 26.4%) *versus* MRD2^{pos} groups (28/73, 38.3%, *P*=0.108). We highlighted a similar pattern within MRD2^{neg} patients for SDAC- (10/50, 20.0%) and HDAC-treated cases (22/71, 31.0%, *P*=0.212). Then, to estimate the benefit from HSCT in MRD2^{neg} cases according to the combined model (age + treatment) category, we used the Mantel-Byar method and Simon-Makuch plots to correct for pre-HSCT events considering HSCT as a time-dependent variable. In this analysis of DFS, the protection from relapse was not significant in the COMB_1 group (*P*=0.64, HR=0.60) (*Online Supplementary Figure S15*), whereas it benefitted the COMB_2-3 patients (*P*=0.00433, HR=0.26) (*Online Supplementary Figure S16*).

Analysis of literature

Our data indicated an interaction between the predictive value of MRD2^{neg} status and the intensity of treatment. We thus carried out an analysis of available literature to validate our findings. We selected a set of papers as detailed in the Methods section; studies selected for analysis of DFS in MRD2^{neg} cases based on treatment intensity were processed as in conventional meta-analyses and extracted MRD data are summarized in a forest plot (*Online Supplementary Data S5*). Then we performed DFS analysis after merging extracted data annotated for MRD status after two chemotherapy cycles, AML subset and treatment schedule. We selected cases that resulted negative after two chemotherapy cycles (i.e., MRD2^{neg}), stratifying patients according to the intensity of treatment (SDAC- *vs*. HDAC-based regimens).

First, we carried out an analysis unselected for AML subset, and we observed a difference in DFS in favor of the SDAC-treated group (*P*=0.014) (Figure 4A). To adjust for the fact that CBF-related cases were restricted to the group receiving HDAC, we performed further analyses after excluding CBF-related cases (Figure 4B) and focusing on intermediate-risk karyotype (Figure 4C). In both analyses, the SDAC-treated category displayed longer DFS (*P*<0.0001 and *P*=0.0018, respectively), an observation consistent with the findings in our dataset.

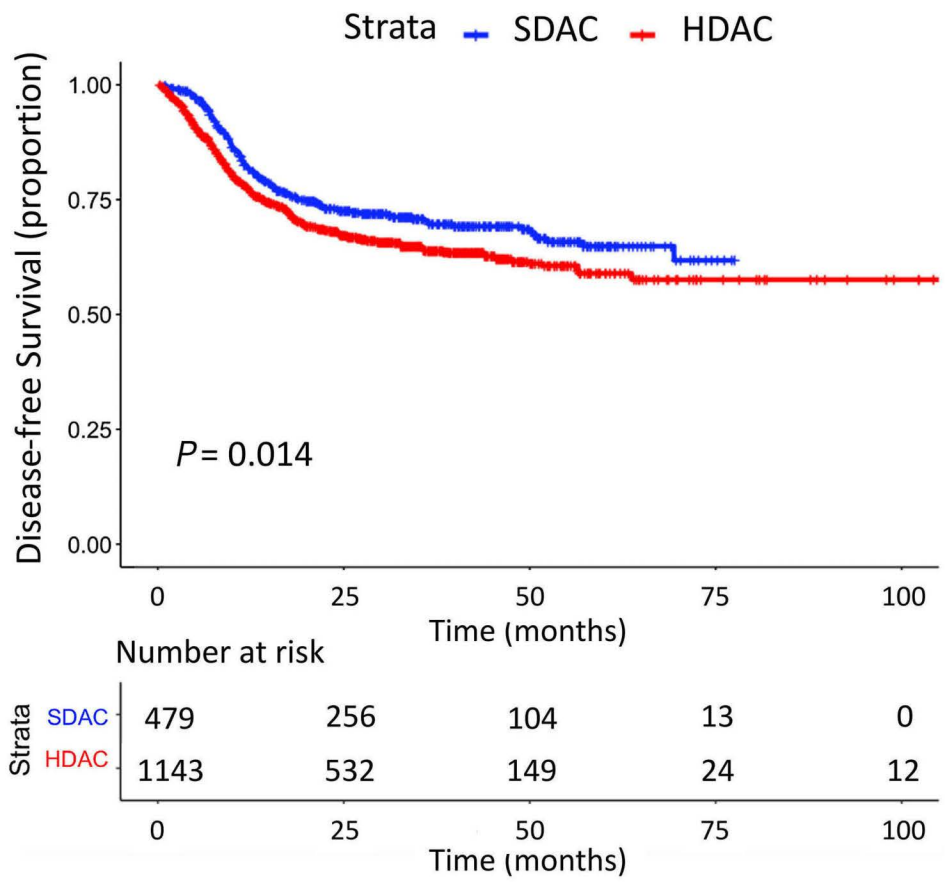
Discussion

MRD assessment is an essential tool for the management of patients with AML. In fact, it integrates the baseline and pre-treatment prognostic variables by conveying information on the sensitivity of the disease to treatment in the early phases of therapy.[2-5,9,12,25] In terms of decision-making, the persistence of MRD after intensive treatment correlates with a very high risk of disease recurrence, and patients are immediately allocated to allogeneic HSCT.[26] An undetectable MRD is far trickier to interpret and translate into a clinical decision. Although with some variability based on the methodology used, the rate of patients who relapse despite the achievement of an MRD^{neg} status is not negligible, ranging from 20 to 40% in the first 2 years from CR.[6,27] The reasons for the failure of MRD^{neg} to predict for maintenance of CR in such a relevant fraction of patients are not fully understood. Relapse is supposed to derive from a minor population of leukemic cells that escape MRD detection, likely due to a combination of factors. These include technical issues in the MRD measurement, as supported by the improved prognostic performance obtained with the use of more sensitive molecular methods[28-30] or by increasing the number of parameters and cells analyzed by MFC.[8,10] Indeed, it has been shown that introduction of more stringent criteria for the limit of detection and quantitation resulted in a more accurate ability to predict patients' outcome.[31] Furthermore, residual AML cells might escape detection as the result of changes in their phenotypic profile or confinement in more immature cell compartments. To address these specific issue, some investigators have proposed advancements to standard approaches, aimed at extending MRD evaluation to deviations from physiological hematopoiesis (i.e., so-called "different from normal")[7] or by highlighting abnormalities in leukemic stem cells,[32,33] but both strategies remain to be validated.

Whatever the reasons, failure of MRD^{neg} status to appropriately predict for maintenance of CR has relevant clinical consequences. Particularly when a relapse occurs in the first year after completion of treatment, it can be challenging to obtain a second response, outside the CBF-related subset.[34]

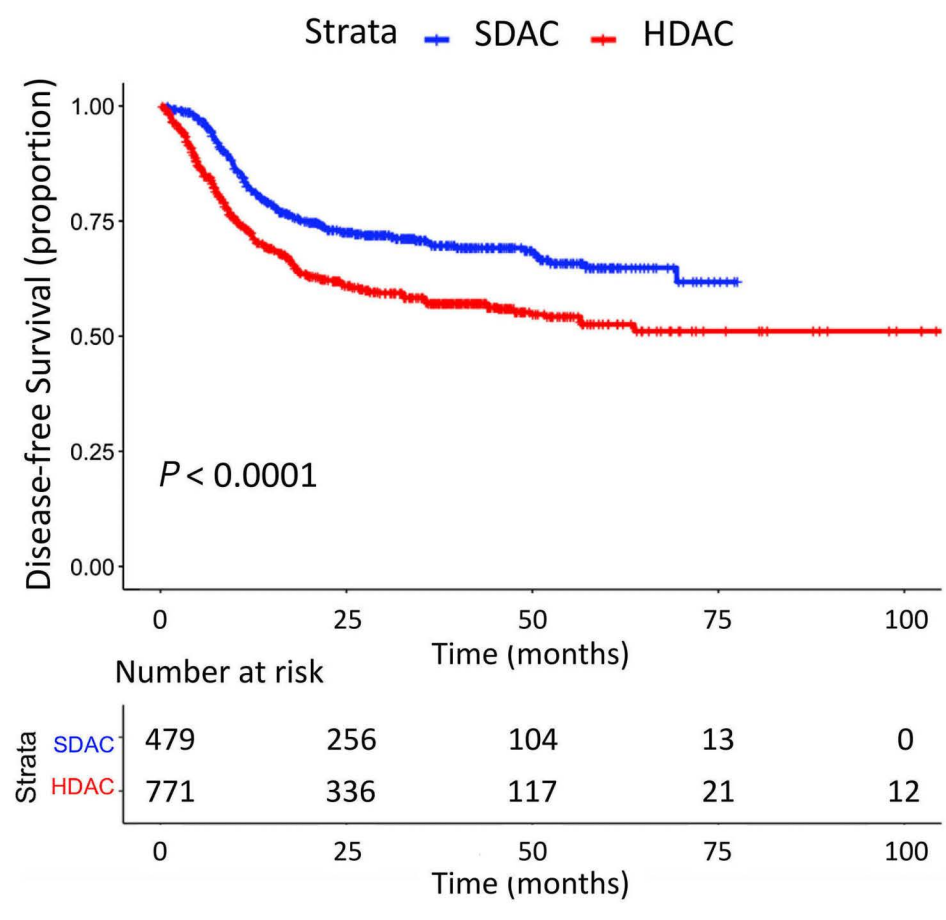
In these patients, a negative MRD status can thus misguide the treatment planning, including the timely decision to exploit allogeneic HSCT when the disease burden is low. Pending further methodological improvements (e.g., next-generation sequencing), the clinical application of current MRD^{neg} data should take into an account those limitations, and efforts should be devoted to shape the settings in which MRD^{neg} predictive value is robust enough or, contrariwise, relatively weak to influence important therapeutic decisions. With such an aim in mind, we searched for interactions between MRD, baseline characteristics (age, gender, white cell count, ELN stratification) and the intensity of treatment in a series of AML patients from our Institution. We focused on the predictive value of MRD2^{neg} status after two chemotherapy cycles (MRD2) in view of its greater capability to discriminate DFS in comparison with earlier assessment (i.e., after induction, MRD1), consistent with previous experiences.[4,35]

A



Papers:
- ✓ Buccisano 2010
- ✓ Chou 2010
- ✓ Jourdan 2013
- ✓ Narimatsu 2008
- ✓ Onecha 2019
- ✓ Ravandi 2017
- ✓ Shayegi 2013
- ✓ Terwjin 2012
- ✓ Terwjin 2013
- ✓ Zhang 2013
- ✓ Bataller 2021
- ✓ Wei 2021
- ✓ Freeman 2018
- ✓ Hubmann 2014
- ✓ Ivey 2016
- ✓ Jongen-Lavrencic 2020
- ✓ Kapp-Schwoerer 2016
- ✓ Willekens 2016
- ✓ Zeijlemaker 2015

B



Papers:
- ✓ Buccisano 2010
- ✓ Chou 2010
- ✓ Onecha 2019
- ✓ Ravandi 2017
- ✓ Shayegi 2013
- ✓ Terwjin 2012
- ✓ Terwjin 2013
- ✓ Freeman 2018
- ✓ Hubmann 2014
- ✓ Ivey 2016
- ✓ Jongen-Lavrencic 2018
- ✓ Kapp-Schwoerer 2020
- ✓ Zeijlemaker 2015

Continued on following page.

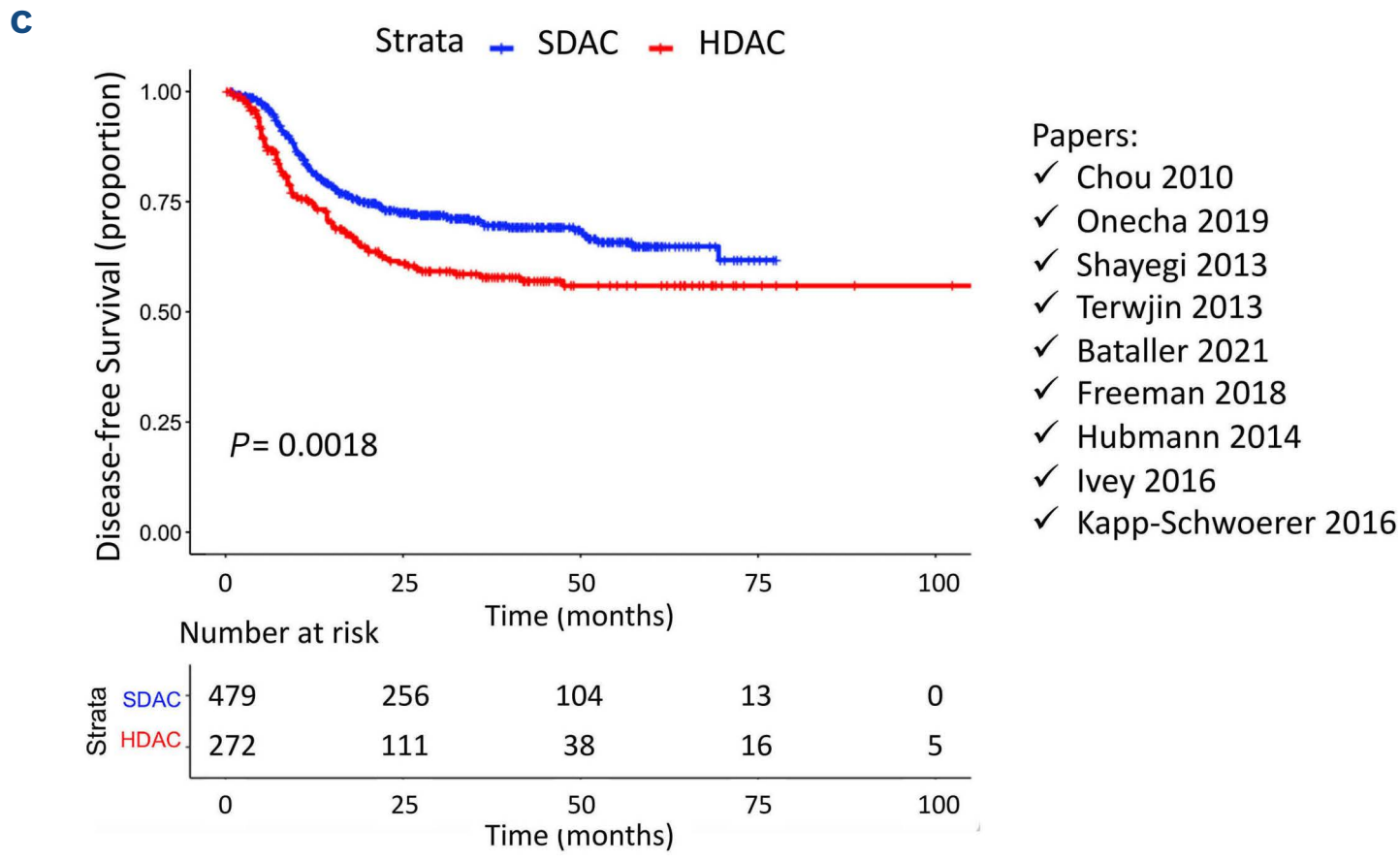


Figure 4. Analysis of disease-free survival according to treatment intensity based on data extracted from literature. Disease-free survival in patients with negative measurable residual disease status after first consolidation (MRD2^{neg}) according to treatment intensity, standard-dose cytarabine (SDAC, blue curve) *versus* high-dose cytarabine (HDAC, red curve) in different acute myeloid leukemia subsets: (A) unselected; (B) after exclusion of CBF-related acute myeloid leukemia; (C) intermediate-risk karyotype. Next to each Kaplan-Meier curve, the papers from which data were extracted for each analysis are listed by first author's name and year of publication.

Among baseline features, we found that younger MRD2^{neg} patients (i.e., aged less than 55 years) had a significantly longer DFS (median not reached) compared to their elderly counterparts (median 25.0 months, *P*=0.013, HR=2.08, 95% CI: 1.15-3.77). Treatment intensity, specifically the delivery of HDAC in either induction or first consolidation, apparently had a pejorative effect on the outcome of MRD2^{neg} patients. This finding can likely be explained by the progressive selection of patients with unfavorable characteristics exerted by HDAC, as supported by the enrichment in ELN 2017 adverse-risk category for HDAC-treated in comparison to SDAC-treated patients from after induction (19.0% *vs*. 11.2, respectively, *P*=0.46) to after consolidation (16.9% *vs*. 2.0%, *P*=0.014). In other words, one might surmise that the early delivery of HDAC could promote CR attainment, eventually making eligible for MRD assessment a category of patients featured by inferior chemosensitivity, who would otherwise have resulted refractory (or MRD^{pos}) when treated with SDAC. HDAC could then transiently mask intrinsically chemorefractory disease, thereby correlating with a higher relapse rate than that with SDAC, despite both treatments resulting in MRD2^{neg} status.

Our findings on the effect of treatment on the significance of MRD are consistent with a previous report by Maurillo *et al*: aiming to use MRD as a biomarker for optimal ARA-C dosing, they also described lower reliability of MRD in discriminating prognosis in patients treated with HDAC compared to SDAC.[36]

To validate our observations, we selected and interrogated the available literature on the interaction between treatment effect and MRD2^{neg} status. The results of this analysis confirmed the pattern in our patients, with longer DFS for those receiving SDAC in induction and first consolidation (Figure 4). The combination of age and treatment effects allowed us to identify categories of patients with remarkably different rates of relapse, in spite of the achievement of MRD2 negativity. The group of elderly, HDAC-receiving (COMB_2-3) patients displayed a DFS rate of around 50% at 3 years. This figure challenges the usual implications of MRD^{neg} status, as current ELN guidelines recommend consolidation with allogeneic HSCT as the preferred post-CR option for patients with an estimated relapse risk exceeding 35-40%.[1,37] Of note, this category of patients showed the greatest benefits from the delivery of HSCT in first CR, as emerging from a specific analysis using the Mantel-Byar method (*Online Supplementary Figure S16*). This finding is consistent with the established influence of disease burden, usually estimated by MRD, on the efficacy of the antileukemic effect of HSCT.[26,38,39] Indeed, it is plausible that, although both characterized by MRD2^{neg} status, COMB_1 patients achieved a more profound level of response than COMB_2-3 patients, likely explaining the difference in outcome in the post-HSCT setting.

The longer DFS for SDAC- *vs*. HDAC-treated MRD2^{neg} patients remained in the intermediate-risk tier, as assessed by karyotype and ELN (*Online Supplementary Figure S13*), the

category in which MRD2 assessment is generally employed for the clinical management of AML patients.

These data emphasize some of the current limitations of MRD assessment. Although MRD represents the strongest predictive parameter for long-term survival overall, in some settings it actually lacks the robustness needed to support key clinical decisions. In particular, the translation of an MRD2^{neg} status into a consolidation program including or omitting HSCT should take into account other contributing variables. To this regard, we herein provide evidence that age and treatment intensity may help to effectively delineate settings (younger age, SDAC-treated) in which the application of MRD2^{neg} data to support clinical decisions is justified based on its correlation with an excellent outcome, unlike in others instances (older patients who were HDAC-treated) in which it did not add to the baseline stratification and should not be used as a key decision-making parameter. As with the dose of ARA-C, other treatment variables (e.g., gemtuzumab ozogamicin, FLT3 inhibitors) may also have an effect on MRD, an issue that deserves to be evaluated in large series of patients.

Our work has some limitations that should be acknowledged. We recognize the limits of a retrospective study with a long enrollment period, with changes in risk assessment and treatment allocation (in particular with regards to HSCT), although we checked for an impact of year of diagnosis on OS (*Online Supplementary Figure S1*). Another point concerns the limited number of patients in the adverse-risk category according to karyotype and ELN, preventing specific analyses. This finding is clearly due to the selection of patients who had a response to chemotherapy, at least initially, and in any case represents the disease subset in which MRD is generally not crucial in clinical decisions. The lack of validation of our results in an independent cohort is a further weakness that we tried to address with the analysis of data extracted from the available literature, which confirmed our findings on the impact of treatment.

In conclusion, we showed a variable reliability of MRD in different AML settings, as defined by age and intensity of treatment. In addition to underscoring the need for further improvement of MRD approaches, these findings call for a reasoned application of MRD information, as currently available, to modulate consolidation therapy on adequately estimated relapse rates.

Disclosures

No conflicts of interest to disclose.

Contributions

FM designed and performed the research, analyzed, and interpreted the data, and wrote the manuscript. MP performed the research, analyzed, and interpreted the data. SB performed flow cytometric analysis and collected the data. BP, RC performed flow cytometric analysis. GG, BS, LS, EQ, GC, FC, SP, and VP collected clinical data. FP, LS, CM, and PG performed molecular analysis. FA and AMV interpreted the data and critically reviewed the manuscript.

Acknowledgments

This work is dedicated to the memory of Dr. Franco Leoni, for his invaluable mentorship in the care of patients with acute leukemias.

Funding

The work was supported by AIRC 5x1000 call "Metastatic disease: the key unmet need in oncology" to MYNERVA project, #21267 (MYeloid NEoplasms Research Venture AIRC) and by Legato Zottola. A detailed description of the MYNERVA project is available at http://www.progettomynerva.it

Data-sharing statement

The data that support the findings of this study are available on request from the corresponding author.

References

1. Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 ELN recommendations from an international expert panel. Blood. 2022;140(12):1345-1377.
2. Miguel JFS, Martínez A, Macedo A, et al. Immunophenotyping investigation of minimal residual disease is a useful approach for predicting relapse in acute myeloid leukemia patients. Blood. 1997;90(6):2465-2470.
3. Venditti A, Buccisano F, Poeta DG, et al. Level of minimal residual disease after consolidation therapy predicts outcome in acute myeloid leukemia. Blood. 2000;96(12):3948-3952.
4. Terwijn M, Putten WL van, Kelder A, et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. J Clin Oncol. 2013;31(31):3889-3897.
5. Ivey A, Hills RK, Simpson MA, et al. Assessment of minimal residual disease in standard-risk AML. N Engl J Med. 2016;374(5):422-433.
6. Short NJ, Zhou S, Fu C, et al. Association of measurable residual disease with survival outcomes in patients with acute myeloid leukemia. JAMA Oncol. 2020;6(12):1890-1899.
7. Heuser M, Freeman SD, Ossenkoppele GJ, et al. 2021 update measurable residual disease in acute myeloid leukemia: European LeukemiaNet Working Party Consensus Document. Blood. 2021;138(26):2753-2767.
8. Buccisano F, Maurillo L, Principe MID, 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.
9. Freeman SD, Hills RK, Virgo P, et al. Measurable residual disease at induction redefines partial response in acute myeloid

leukemia and stratifies outcomes in patients at standard risk without NPM1 mutations. J Clin Oncol. 2018;36(15):1486-1497.
10. Tettero JM, Freeman S, Buecklein V, et al. Technical aspects of flow cytometry-based measurable residual disease quantification in acute myeloid leukemia: experience of the European LeukemiaNet MRD Working Party. Hemasphere. 2021;6(1):e676.
11. Zhu HH, Zhang HX, Qin ZY, et al. MRD-directed risk stratification treatment may improve outcomes of t(8;21) AML in the first complete remission: results from the AML05 multicenter trial. Blood. 2013;121(20):4056-4062.
12. 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.
13. Buccisano F, Maurillo L, Piciocchi A, et al. Minimal residual disease negativity in elderly patients with acute myeloid leukemia may indicate different postremission strategies than in younger patients. Ann Hematol. 2015;94(8):1319-1326.
14. Jentzsch M, Grimm J, Bill M, et al. Clinical value of the measurable residual disease status within the ELN2017 risk groups in AML patients undergoing allogeneic stem cell transplantation. Am J Hematol. 2021;96(7):E237-E239.
15. Burnett AK, Goldstone A, Hills RK, et al. Curability of patients with acute myeloid leukemia who did not undergo transplantation in first remission. J Clin Oncol. 2013;31(10):1293-1301.
16. Castaigne S, Chevret S, Archimbaud E, et al. Randomized comparison of double induction and timed-sequential induction to a "3 + 7" induction in adults with AML: long-term analysis of the Acute Leukemia French Association (ALFA) 9000 study. Blood. 2004;104(8):2467-2474.
17. Bassan R, Intermesoli T, Masciulli A, et al. Randomized trial comparing standard vs sequential high-dose chemotherapy for inducing early CR in adult AML. Blood Adv. 2019;3(7):1103-1117.
18. Mannelli F, Bencini S, Piccini M, et al. Multilineage dysplasia as assessed by immunophenotype in acute myeloid leukemia: a prognostic tool in a genetically undefined category. Cancers. 2020;12(11):3196.
19. Mannelli F, Gianfaldoni G, Bencini S, et al. Early peripheral blast cell clearance predicts minimal residual disease status and refines disease prognosis in acute myeloid leukemia. Am J Hematol. 2020;95(11):1304-1313.
20. Schuurhuis GJ, Heuser M, Freeman S, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2018;131(12):1275-1291.
21. Gabert J, Beillard E, van der Velden VHJ, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia – a Europe Against Cancer program. Leukemia. 2003;17(12):2318-2357.
22. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood. 2010;116(3):354-365.
23. 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.
24. Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol. 2012;12(1):9.
25. Balsat M, Renneville A, Thomas X, et al. Postinduction minimal residual disease predicts outcome and benefit from allogeneic stem cell transplantation in acute myeloid leukemia with NPM1 mutation: a study by the Acute Leukemia French Association group. J Clin Oncol. 2016;35(2):185-193.
26. Buccisano F, Maurillo L, Piciocchi A, et al. Pre-transplant persistence of minimal residual disease does not contraindicate allogeneic stem cell transplantation for adult patients with acute myeloid leukemia. Bone Marrow Transpl. 2017;52(3):473-475.
27. Paiva B, Vidriales M-B, Sempere A, et al. Impact of measurable residual disease by decentralized flow cytometry: a PETHEMA real-world study in 1076 patients with acute myeloid leukemia. Leukemia. 2021;35(8):2358-2370.
28. Pastore F, Levine RL. Next-generation sequencing and detection of minimal residual disease in acute myeloid leukemia: ready for clinical practice? JAMA. 2015;314(8):778-780.
29. Klco JM, Miller CA, Griffith M, et al. Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. JAMA. 2015;314(8):811-822.
30. Jongen-Lavrencic M, Grob T, Hanekamp D, et al. Molecular minimal residual disease in acute myeloid leukemia. N Engl J Med 2018;378(13):1189-1199.
31. Buccisano F, Palmieri R, Piciocchi A, et al. 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. Haematologica. 2022;107(12):2823-2833.
32. Li S-Q, Xu L-P, Wang Y, et al. An LSC-based MRD assay to complement the traditional MFC method for prediction of AML relapse: a prospective study. Blood. 2022;140(5):516-520.
33. Zeijlemaker W, Kelder A, Oussoren-Brockhoff YJM, et al. A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia. Leukemia. 2016;30(2):439-446.
34. Breems DA, Putten WLJV, Huijgens PC, et al. Prognostic index for adult patients with acute myeloid leukemia in first relapse. J Clin Oncol. 2005;23(9):1969-1978.
35. Maurillo L, Buccisano F, Principe M, et al. Toward optimization of postremission therapy for residual disease–positive patients with acute myeloid leukemia. J Clin Oncol. 2008;26(30):4944-4951.
36. 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.
37. Cornelissen JJ, Blaise D. Hematopoietic stem cell transplantation for patients with AML in first complete remission. Blood. 2016;127(1):62-70.
38. Walter RB, Gooley TA, Wood BL, et al. Impact of pretransplantation minimal residual disease, as detected by multiparametric flow cytometry, on outcome of myeloablative hematopoietic cell transplantation for acute myeloid leukemia. J Clin Oncol. 2011;29(9):1190-1197.
39. Hourigan CS, Dillon LW, Gui G, et al. Impact of conditioning intensity of allogeneic transplantation for acute myeloid leukemia with genomic evidence of residual disease. J Clin Oncol. 2019;38(12):1273-1283.