Received: 1 June 2023 | Revised: 7 July 2023 | Accepted: 16 July 2023

DOI: 10.1002/ajh.27046

RESEARCH ARTICLE

AJH WILEY

# Cox proportional hazards deep neural network identifies peripheral blood complete remission to be at least equivalent to morphologic complete remission in predicting outcomes of patients treated with azacitidine—A prospective cohort study by the AGMT

Lisa Pleyer[1,2,3,4] | Marc Vaisband[2,3,4,5] | Manuel Drost[6] | Michael Pfeilstöcker[1,7,8] | Reinhard Stauder[1,9] | Sonja Heibl[1,10] | Heinz Sill[1,11] | Michael Girschikofsky[1,12] | Margarete Stampfl-Mattersberger[1,13] | Angelika Pichler[1,14] | Bernd Hartmann[1,15] | Andreas Petzer[1,16] | Martin Schreder[1,17] | Clemens A. Schmitt[1,18,19] | Sonia Vallet[1,20] | Thomas Melchardt[1,2,3,4] | Armin Zebisch[1,11,21] | Petra Pichler[1,22] | Nadja Zaborsky[2,3,4,23] | Sigrid Machherndl-Spandl[1,12] | Dominik Wolf[1,9] | Felix Keil[1,7,8] | Jan Hasenauer[5] | Julian Larcher-Senn[6] | Richard Greil[1,2,3,4]

[1]Austrian Group of Medical Tumor Therapy (AGMT) Study Group, Vienna, Austria

[2]Salzburg Cancer Research Institute (SCRI) Center for Clinical Cancer and Immunology Trials (CCCIT), Salzburg, Austria

[3]Cancer Cluster Salzburg (CCS), Salzburg, Austria

[4]3rd Medical Department with Hematology, Medical Oncology, Hemostaseology, Rheumatology and Infectiology, Oncologic Center, Paracelsus Medical University, Salzburg, Austria

[5]Life & Medical Sciences Institute, University of Bonn, Bonn, Germany

[6]Assign Data Management and Biostatistics GmbH, Innsbruck, Austria

[7]3rd Medical Department for Hematology and Oncology, Hanusch Hospital, Vienna, Austria

[8]Ludwig Boltzmann Institute for Hematology and Oncology, Medical University Vienna, Vienna, Austria

[9]Department of Hematology and Oncology, Comprehensive Cancer Center Innsbruck (CCCI), Medical University of Innsbruck (MUI), Innsbruck, Austria

[10]4th Medical Department of Internal Medicine, Hematology, Internistic Oncology and Palliative Medicine, Klinikum Wels-Grieskirchen GmbH, Wels, Austria

[11]Division of Hematology, Medical University of Graz, Graz, Austria

[12]1st Medical Department, Hematology with Stem Cell Transplantation, Hemostaseology and Medical Oncology, Ordensklinikum Linz GmbH Elisabethinen, Linz, Austria

[13]Department of Internal Medicine 2, Wiener Gesundheitsverbund, Klinik Donaustadt, Vienna, Austria

[14]Department of Internal Medicine, Hematology and Internal Oncology, LKH Hochsteiermark, Leoben, Austria

[15]Department of Internal Medicine 2, Landeskrankenhaus Feldkirch, Feldkirch, Austria

[16]Internal Medicine I: Medical Oncology and Hematology, Ordensklinikum Linz GmbH, Barmherzige Schwestern, Linz, Austria

[17]1st Department of Internal Medicine, Center for Oncology and Hematology, Wiener Gesundheitsverbund, Klinik Ottakring, Vienna, Austria

[18]Department of Hematology and Internal Oncology, Kepler University Hospital, Johannes Kepler University, Linz, Austria

[19]Charité—University Medical Center, Molecular Cancer Research Center, Berlin, Germany

[20]University Hospital Krems, Department of Internal Medicine 2, Karl Landsteiner Private University of Health Sciences, Krems, Austria

[21]Otto Loewi Research Center for Vascular Biology, Immunology and Inflammation, Division of Pharmacology, Medical University of Graz, Graz, Austria

[22]Clinical Department for Internal Medicine, University Hospital St Poelten, Karl Landsteiner University of Health Sciences, St Poelten, Austria

[23]Laboratory of Immunological and Molecular Cancer Research (LIMCR), Salzburg, Austria

**Correspondence**
Lisa Pleyer, 3rd Medical Department with Hematology, Medical Oncology, Hemostaseology, Rheumatology and Infectiology, Oncologic Center, Paracelsus Medical University, Müllnerhauptstrasse 48, 5020 Salzburg, Austria.
Email: dr.lisa.pleyer@gmail.com, l.pleyer@salk.at

## Abstract

The current gold standard of response assessment in patients with myelodysplastic syndromes (MDS), chronic myelomonocytic leukemia (CMML), and acute myeloid leukemia (AML) is morphologic complete remission (CR) and CR with incomplete count recovery (CRi), both of which require an invasive BM evaluation. Outside of clinical trials, BM evaluations are only performed in ~50% of patients during follow-up, pinpointing a clinical need for response endpoints that do not necessitate BM assessments. We define and validate a new response type termed "peripheral blood complete remission" (PB-CR) that can be determined from the differential blood count and clinical parameters without necessitating a BM assessment. We compared the predictive value of PB-CR with morphologic CR/CRi in 1441 non-selected, consecutive patients diagnosed with MDS ($n = 522$; 36.2%), CMML ($n = 132$; 9.2%), or AML ($n = 787$; 54.6%), included within the Austrian Myeloid Registry (aMYELOIDr; NCT04438889). Time-to-event analyses were adjusted for 17 covariates remaining in the final Cox proportional hazards (CPH) model. DeepSurv, a CPH neural network model, and permutation-based feature importance were used to validate results. 1441 patients were included. Adjusted median overall survival for patients achieving PB-CR was 22.8 months (95%CI 18.9–26.2) versus 10.4 months (95%CI 9.7–11.2) for those who did not; HR = 0.366 (95%CI 0.303–0.441; $p < .0001$). Among patients achieving CR, those additionally achieving PB-CR had a median adjusted OS of 32.6 months (95%CI 26.2–49.2) versus 21.7 months (95%CI 16.9–27.7; HR = 0.400 [95%CI 0.190–0.844; $p = .0161$]) for those who did not. Our deep neural network analysis-based findings from a large, prospective cohort study indicate that BM evaluations solely for the purpose of identifying CR/CRi can be omitted.

## 1 | INTRODUCTION

While myelodysplastic syndromes (MDS), chronic myelomonocytic leukemia (CMML), and acute myeloid leukemia (AML) are different disease entities, they share similar clinical features with considerable genetic, biologic, and phenotypic overlap. At least one-third of patients with MDS and CMML ultimately progress to AML, representing a disease continuum with differing prognoses along the trajectory and explaining why these diseases are often treated similarly.[1] Azacitidine (with or without venetoclax) is the recommended first-line treatment in almost all elderly patients who are ineligible for intensive chemotherapy. Complete remission (CR)/CR with incomplete blood count recovery (CRi) is deemed the major outcome associated with improved overall survival (OS) and is often used as primary endpoint in clinical AML trials, whereas red blood cell transfusion independence is often the endpoint used in MDS trials. Expert opinions questioning these endpoints are cumulating,[2–4] as is evidence from randomized clinical trials, real-world analyses, and treatment guidelines, that bone marrow (BM) blast clearance (and even BM evaluations) may not be mandatory to determine whether meaningful clinical response with OS prolongation is achieved.[5–12] Recently, an algorithm using a gradient-boosted model has been published, that even allows the diagnosis (positive predictive value 0.88) and exclusion (negative predictive value 0.94) of MDS in most patients without a BM evaluation.[13]

Morphologic BM evaluations are mandatory for (i) accurate diagnosis and disease classification, (ii) risk stratification, which typically includes the BM blast percentage as well as conventional cytogenetics and/or fluorescence in situ hybridization from BM specimen, and (iii) response assessment/disease monitoring during treatment.[14–24] However, outside of clinical trials, BM evaluations are not always feasible in clinical practice and, after having established the initial diagnosis, are only performed in approximately 50% of patients with MDS, CMML, or AML during follow-up.[25–28] This may be due to the fact that histologic and conventional cytogenetic results are time consuming and typically take at least 2 weeks. Other reasons are given in Figure S1. BM evaluations can also yield no results, even when performed, for example, in cases of fibrotic or hypocellular marrow resulting in dry taps, insufficient aspiration, or when the blood coagulates before an adequate smear can be performed. Hence, a real-world clinical need exists for an alternative sample and/or surrogate response

type to be able to obtain adequate information on a patient's response status without the requirement of repetitive BM evaluations, especially as frequent monitoring would be desirable, not only for adequate assessment of response, but also for determination of response duration, adequate risk stratification, prognostication, and clinical management. The pressing clinical relevance becomes clear in light of the fact that all current response criteria for MDS, CMML, and AML require a BM evaluation for determining adequate therapeutic response,[14–19] and that, particularly in AML, patients without morphologic CR or CRi or without a BM evaluation are deemed as non-responders.

In a prospective setting, our group has recently shown a very strong correlation not only between the mutations detected ($r = .91$; $p < .0001$) but also between the variant allele frequencies found in paired peripheral blood and BM specimens ($r = .93$; $p < .0001$) analyzed with a next-generation sequencing 40-gene panel.[25] These data ascertained that molecular analyses from peripheral blood can safely be used as an alternative to BM samples to reliably molecularly classify and monitor gene mutations, as well as to guide treatment decisions in patients with myeloid neoplasms, without loss of sensitivity or specificity. In a similar fashion, we aimed to (i) use our large database to prospectively define a new response type termed "peripheral blood complete remission" (PB-CR) for patients with MDS, CMML, and AML that would not necessitate an invasive BM evaluation, and (ii) identify whether PB-CR provides added value to and is at least as good at predicting outcome measures as morphologic CR and CR/CRi, which are the internationally accepted gold standards of response assessment in these diseases.

## 2 | METHODS

### 2.1 | Study design and participants

In this prospective cohort study, data from non-selected, consecutive patients participating in the Austrian Myeloid Registry (aMYELOIDr; Table S1) of the Austrian Group for Medical Tumor Therapy (AGMT) Study Group (NCT04438889; ethics committee approval 415-E/2581/9-2020; first patient in 13.07.2020), which partially encompasses the former Austrian Registry of Hypomethylating Agents (NCT01595295; ethics committee approval 415-EP/39/11-2009; first patient in 09.02.2009, last patient in 21.01.2021, last visit 23.08.2022; details published previously[7,12,29–32]).

The only inclusion criteria for this study were the diagnosis of MDS, AML, or CMML according to WHO 2016 criteria and documentation of treatment with azacitidine. No patients were excluded. Diagnosis was independently verified based on submitted data. Written informed consent was obtained for all patients alive at the time of data entry. Due to the non-interventive nature of the Austrian Myeloid Registry (aMYELOIDr), the choice of treatment (medications, schedule, dose, duration, and sequence), and the choice of if and when to perform BM evaluations were entirely at the respective treating physician's discretion.

### 2.2 | Missing data

Missing baseline data (Tables S2-S5), missing azacitidine response and treatment characteristics (Tables S6–S9), and missing data per azacitidine treatment cycle (Tables S11–S15) were mostly rare, that is, between 0.0 and <5.0%, except for certain laboratory values that are not routinely assessed, such as for example, albumin. No imputations were performed for missing data. Only observed values were analyzed.

### 2.3 | Outcomes

OS was defined as the time from day 1 cycle 1 of azacitidine to death from any cause. Time to next treatment was defined as the time from day 1 cycle 1 of azacitidine to day 1 cycle 1 of the next (ensuing) treatment line or death. Transfusions, growth factors, and iron chelators did not count as treatment lines. Time on treatment was used as an additional endpoint and was defined as the time from day 1 cycle 1 of azacitidine to the end of treatment date of azacitidine.

Response to azacitidine was assessed according to European LeukemiaNet (ELN) criteria for AML[16] and according to International working group (IWG) criteria for MDS and CMML.[20] In this publication, the term complete remission with incomplete blood count recovery (CRi) is used to describe both complete remission with incomplete neutrophil count as well as complete remission with incomplete platelet count (CRi/CRp). Delayed blood count recovery (i.e., after BM evaluation) was assessed separately and is indicated as such. To ensure uniformity and reduce human errors, all response types were calculated from data entered in the electronic case report form.

### 2.4 | Definition of PB-CR

PB-CR was defined by adopting all of the criteria required for CR in all of the current response criteria published at the time of analyses for MDS,[20] CMML,[23] and AML,[16,21,22] excluding BM parameters. This includes achievement of a normal differential blood count (defined by peripheral blood blasts = 0%, absolute neutrophil count ≥1.0 × $10^3$/μL, white blood cell count ≥1.5 × $10^3$/μL and <10.0 × $10^3$/μL, monocytes ≤1.0 × $10^3$/μL, hemoglobin ≥11.0 g/dL, platelet count ≥100 × $10^3$/μL and <450 × $10^3$/μL), lack of transfusion dependence (for both red blood cells and platelets), and no treatment with growth factors (i.e., no treatment with erythrocyte stimulating agents, thrombocyte stimulating agents, or granulocyte colony stimulating factors) (Table 1). Thus, PB-CR is determined by a differential blood count and clinical parameters and does not necessitate a BM evaluation.

### 2.5 | Timepoints of response assessment

All response (sub)types necessitating a BM evaluation (particularly CR and CRi) were assessed at each BM evaluation performed. As stated above, the timing of BM evaluations was solely at the treating

**TABLE 1** Definition of PB-CR in context with other response criteria for MDS, CMML, and MPN.

| Recommendations | IWG 2006[20] | IWG 2015[23] | IWG 2003[21] | ELN 2022[16] | AMR 2023 |
|---|---|---|---|---|---|
| **Entity** | **MDS** | **MDS/CMML** | **AML** | **AML** | **MDS/CMML/AML** |
| **Response type** | **CR** | **CR** | **CR** | **CR** | **PB-CR** |
| PB blasts = 0% | ✓ | ✓ | - | ✓ | ✓ |
| Absolute neutrophil count ≥1.0 × $10^3$/μL | ✓ | ✓ | ✓ | ✓ | ✓ |
| White blood cell count ≥1.5 × $10^3$/μL | - | ✓ | - | - | ✓ |
| White blood cell count <10.0 × $10^3$/μL | - | ✓ | - | - | ✓ |
| Monocytes ≤1.0 × $10^3$/μL | - | ✓ | - | - | ✓ |
| Hemoglobin ≥11.0 g/dL | ✓ | ✓ | - | - | ✓ |
| Platelet count ≥100 × $10^3$/μL | ✓ | ✓ | ✓ | ✓ | ✓ |
| Platelet count <450 × $10^3$/μL | - | ✓ | - | - | ✓ |
| Red blood cell transfusion independence | ✓ | ✓ | ✓ | - | ✓ |
| Platelet transfusion independence | - | ✓ | - | - | ✓ |
| No myeloid growth factors | - | - | - | - | ✓ |
| No erythropoietin stimulating agents | ✓ | - | - | - | ✓ |
| No thrombopoietin stimulating agents | - | - | - | - | ✓ |
| Bone marrow blasts <5% | ✓ | ✓ | ✓ | ✓ | - |
| Bone marrow fibrosis ≤ grade 1 | - | ✓ | - | - | - |
| No extramedullary disease | - | ✓ | ✓ | ✓ | - |
| No dysplasia | - | - | - | - | - |
| Duration of response | ≥8 weeks | ≥8 weeks | - | - | ≥8 weeks |

Abbreviations: AML, acute myeloid leukemia; AMR, Austrian Myeloid Registry (aMYELOIDr); CMML, chronic myelomonocytic leukemia; CR, complete remission; ELN, European Leukemia Net; IWG, International Working Group; MDS, myelodysplastic syndromes; PB-CR, peripheral blood CR.

physician's discretion. Transfusion independence, hematologic improvement, and PB-CR were assessed at day 1 of each azacitidine treatment cycle for all patients.

## 2.6 | Descriptive statistics

OS (patients still alive or lost to follow-up were censored at the last follow-up), time to next treatment as a surrogate for cessation of treatment benefit (patients still alive or lost to follow-up and who had received no next treatment were censored at the last follow-up), and time on treatment (patients still on azacitidine treatment were censored at the last day of azacitidine application) were analyzed using the Kaplan–Meier method.

Baseline and treatment-related factors were compared using the $\chi^2$ test for categorical variables and the Wilcoxon test for continuous variables. Patient subgroups were compared using the log-rank test. All *p*-values and 95%CIs are two-sided. The threshold for statistical significance was .05.

## 2.7 | Cox proportional hazards model

To minimize selection bias, Cox proportional hazards (CPH) models for time-to-event endpoints were applied to identify and account for possible interrelations between predictors. Details on variable selection and the CPH model itself are provided in Tables S16 and S17.

## 2.8 | Likelihood ratio test

To identify which response type achieved by which azacitidine cycle had the highest impact on time-to-event endpoints, multivariate-adjusted likelihood ratios (LHRs) of the CPH model for OS or time to next treatment were calculated using the respective response types as covariates. Results were visualized with heatmaps. Further details are provided in the legend of Figure 1.

## 2.9 | Added value of PB-CR

Proceeding analogously to Efficace et al.,[33] who demonstrated that self-reported fatigue provided added value to the IPSS and R-IPSS in patients with MDS, the multivariate-adjusted LHR test was used to determine whether PB-CR provided added value to CR or CR/CRi.

## 2.10 | CPH neural network and permutation-based feature importance

To strengthen and validate the results obtained with conventional statistical methods, we utilized DeepSurv,[34] a validated CPH model with a parametrization of the log-hazards through a deep neural network. DeepSurv functions by optimizing the neural network incorporated

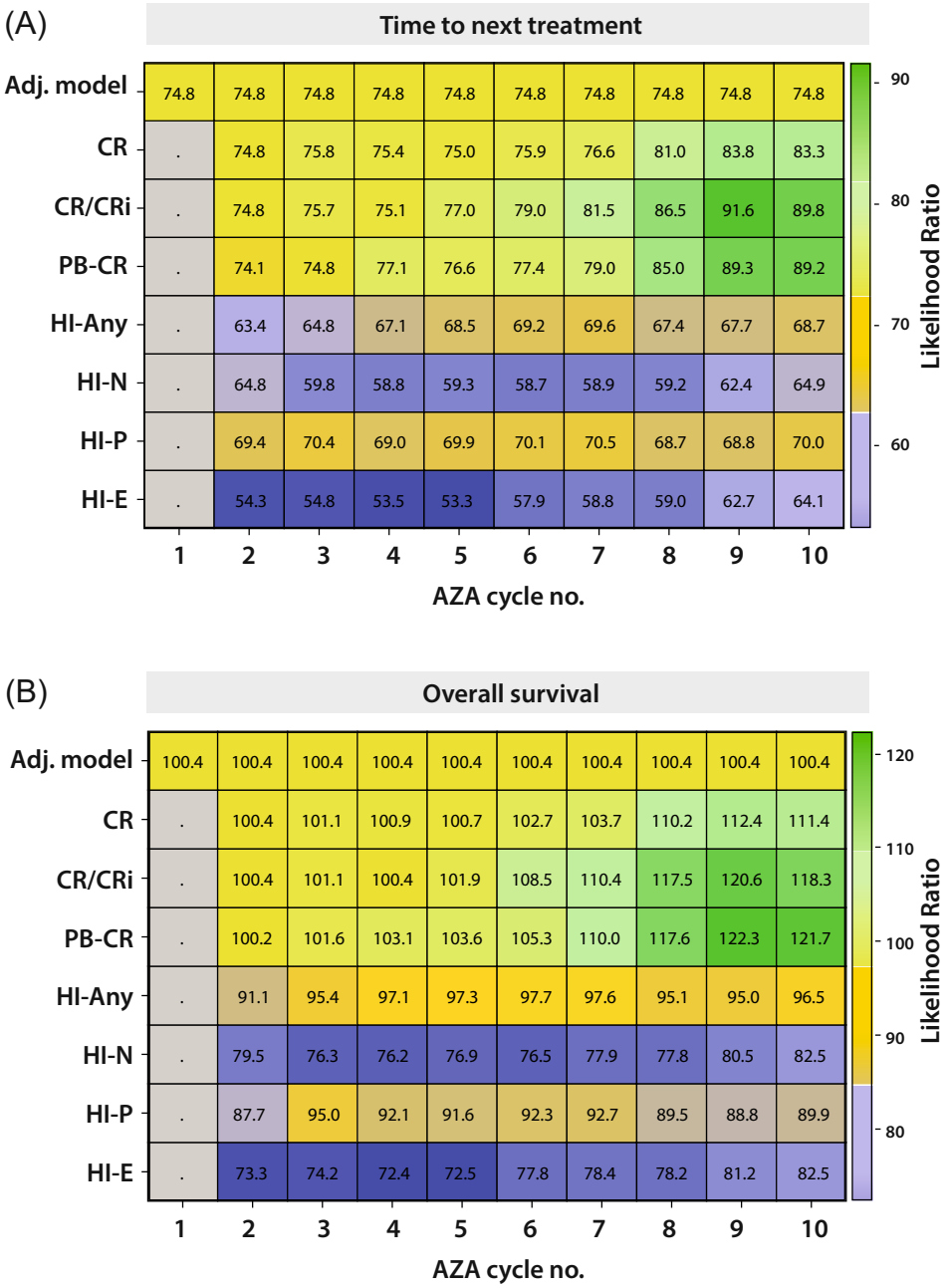


**FIGURE 1** Heatmaps of likelihood ratios. Multivariate-adjusted likelihood ratios of response types (by endpoint). (A) Time to next treatment. (B) Overall survival. Adjusted for all 17 baseline covariates at azacitidine treatment start remaining in the final CPH model (Table S17). Only patients who received ≥10 azacitidine cycles ($n = 444$) and for whom all 17 covariates were available were included ($n = 436$). 3CR and CR/CRi were assessed according to European LeukemiaNet (ELN) criteria for AML[16] and International Working Group (IWG) criteria for MDS/CMML.[20] Hematologic improvement (HI) was assessed according to IWG-2006 criteria for myelodysplastic syndromes and chronic myelomonocytic leukemia.[20] The *Y*-axis contains the analyzed response type. For each response type "Patients with abnormal baseline values" (i.e., hemoglobin <11.0 g/dL, platelet <100 or >450 $\times$ $10^3$/μL, absolute neutrophil count <1.0 $\times$ $10^3$/μL, PB blasts >0%, monocytes >1.0 $\times$ $10^3$/μL or white blood cell count <1.5 or >10.0 $\times$ $10^3$/μL) were included, and responders and non-responders were compared. The adjusted model itself was included to be able to identify whether the response types offered additional value to the multivariate model. The *X*-axis represents the timepoint of response analyses, that is, day 1 of the first 10 azacitidine treatment cycles. Response was assessed for each azacitidine cycle separately. Hematologic improvement (HI) and PB-CR were assessed at day 1 of each azacitidine cycle. CR and CRi were assessed at each BM evaluation performed, the timing of which was solely at the treating physician's discretion. The individual boxes contain the likelihood ratio (LHR) of the Cox-regression model for the time-to-event endpoint (OS or time to next treatment) using the respective response variable as covariate. The higher the LHR, the better the model fit. CR, complete morphologic remission; CRi, CR with incomplete blood count recovery; PB-CR, peripheral blood complete remission; HI, hematologic improvement; HI-any indicates HI in at least one of the three lineages: HI-erythrocytes, HI-platelets, and/or HI-neutrophils. [Color figure can be viewed at wileyonlinelibrary.com]

into the model with respect to the negative partial log-likelihood, the canonical energy functional for survival models. The parametrization of the log-hazards through a deep neural network instead of a linear combination of the input features allows for much more complex relationships between input variables and outcomes to be captured.

Due to its nature as a dense neural network, the DeepSurv model no longer offers explicit coefficients for the input features. To assess which of the response types (CR, CR/CRi, or PB-CR) has greater predictive value, we used a permutation-based feature importance approach (which was repeated for each time-to-event endpoint). Only patients who underwent a BM evaluation during treatment with azacitidine and for whom all 17 baseline covariates remaining in the final CPH model (Table S17) were available were included. To assess generalization performance, 5-fold cross-validation was employed. Sex, age-group (</≥75 years), and diagnosis at azacitidine treatment start were used as stratification factors for fold splitting (Figure S2). Harrell's Concordance Index (C-index), which quantifies the quality of rankings and is the standard goodness of fit measure for risk model assessment in survival analysis. The average validation C-index of 1000 random permutations was reported (with the cross-validation employed), and the scores between the PB-CR-permutated and either the CR-corrupted or the CR/CRi-permutated models were compared by using the Wilcoxon signed-rank test. Bonferroni correction was applied to correct for multiple testing for each endpoint separately. Results were visualized with violin plots. Further details are provided in the legends of Figures S3 and 3.

### 2.11 | Sensitivity analyses

Sensitivity analyses were performed to check the general conclusions by using differing definitions of PB-CR, assessing different endpoints (response subtypes, OS, time to next treatment), applying different regression models, assessing several subgroups, and varying the inner network architecture of the neural network model. Assign Data Management and Biostatistics GmbH performed statistical analyses with SAS® 9.3 and 9.4. Life & Medical Sciences Institute, University of Bonn, performed statistical analyses, including neural network modeling, with Python 3.8.12.

## 3 | RESULTS

### 3.1 | Patient characteristics

Data from non-selected, consecutive 1441 patients diagnosed with MDS, CMML, or AML and who were treated with azacitidine between December 6th 2006 and July 5th 2021, were prospectively collected by 20 Austrian centers and included in the Austrian Myeloid Registry (aMYELOIDr; Table S1). Patient characteristics at azacitidine treatment start are shown in Table S2 for the whole cohort and by diagnosis. At the start of azacitidine treatment, 522 of 1441 (36.2%) patients were diagnosed with MDS, 132 (9.2%) had CMML, and 787 (54.6%) had AML, respectively. Median age was 73.0 (IQR 67.0–78.0) years, 61% were male, 24% had performance status ≥2, and 85% had higher risk disease according to the Revised International Prognostic Scoring System (R-IPSS) (Table S2). 592 of 1441 (41%) patients were red blood cell transfusion dependent, and 287 (20%) were platelet transfusion dependent (Table S3). Serum ferritin levels and serum erythropoietin levels were only assessed in 726 (50%) and 325 (23%) of patients (Table S4). The most common comorbidities were cardiac comorbidity (37%), diabetes mellitus (19%), impaired renal function (17%), prior solid tumor (17%), pulmonary comorbidity (14%), and an additional hematologic malignancy (11%) (Table S5).

1262 (88%) of 1441 patients had died at the last follow-up, median follow-up was 18.8 months (IQR 9.8–37.4) from diagnosis and 10.7 months (IQR 4.1–21.1) from azacitidine treatment start, median duration of azacitidine treatment was 5.0 months (IQR 1.9–12.0), median time from azacitidine stop to death was 2.4 months (IQR 0.9–6.3), early mortality at 30 days was low (79 (<6%) of 1441 patients, 722 (50%), 382 (27%), 269 (19%), and 229 (16%) of patients were alive 1, 2, 3, and 4 years after azacitidine start, respectively (Table S6). A total of 13 971 cycles of azacitidine were documented. The median number of azacitidine cycles administered to all patients was 5.0 (IQR 2.0–12.0) and 6.0 (IQR 3.0–14.0), 8.0 (IQR 4.0–16.5), and 4.0 (IQR 2.0–10.0) for patients with MDS, CMML, and AML, respectively (Table S7). The median number of azacitidine treatment days per cycle was 7 (IQR 5–7), and the median azacitidine dose per cycle was 875 (IQR 700–1000) mg (Table S7). Of the 699 of 1441 (49%) patients in whom a BM evaluation was performed during treatment with azacitidine, a CR was observed in 17% and a CRi in an additional 9%). These numbers increased to 20% and 9% with the allowance of delayed blood count recovery after a median of 4 (IQR 3–7) cycles of azacitidine (Table S8). Hematologic improvement was observed in 43% of patients, and 13% of patients achieved PB-CR after a median of 4.6 (IQR 2.8–6.7) months. Median duration of PB-CR was 7.0 (IQR 1.9–15.6) months (Table S9). Of 119 patients who achieved PB-CR and who had a concomitant BM evaluation, 88 (74.0%) had a CR, 12 (1.0%) had a PR, 19 (1.6%) had a stable disease in the BM, and 0 (0.0%) had a CRp, CRi, CRh, MLFS, or progressive disease, respectively (Table S10). Information regarding differential blood count, number of transfusions received, treatment with growth factors, and hematologic improvement was available in almost all cycles (Tables S11–S15).

### 3.2 | Multivariable model

Our first aim was to build a multivariable model. Univariate Cox regression was performed for all baseline variables at azacitidine start that were available in ≥96% of patients (Table S16). After bidirectional stepwise selection, 17 baseline variables available at azacitidine treatment start remained in the final model: diagnosis, treatment-related disease, age, body mass index, Revised International Prognostic Scoring System (R-IPSS) risk category, R-IPSS cytogenetic risk, Eastern Cooperative Oncology Group Performance Status (ECOG-PS), number

of comorbidities, hemoglobin level, platelet count, mean corpuscular hemoglobin, white blood cell count, absolute neutrophil count, peripheral blood blasts, serum lactate dehydrogenase (LDH), azacitidine treatment line, and number of azacitidine treatment days in cycle 1 (Table S17). It should be noted that transfusion dependence and BM blast count were removed by the model.

### 3.3 | Likelihood ratios

Next, we aimed to identify which response type had the highest impact on time-to-event endpoints and how this was influenced by the azacitidine cycle at which it was achieved. Only patients who received ≥10 azacitidine cycles ($n = 444$) and for whom all baseline covariates remaining in the final CPH model were available were included ($n = 436$).

Among all response types, and after MVA, the highest prognostic impact on time to next treatment was observed when achieving CR/CRi (LHR 91.6) or PB-CR (LHR 89.3) by cycle 9, respectively (Figure 1A). Similar results were observed for the endpoint OS, however, with much higher LHR values: the highest prognostic impact on OS was observed when achieving PB-CR (LHR 122.3) or CR/CRi (LHR 120.6) by cycle 9, respectively (Figure 1B). The LHR of CR/CRi and PB-CR were higher than for CR, higher than for the CPH model alone, and much higher than for any hematologic improvement subtype for both endpoints, respectively (Figure 1A,B).

In addition, we compared CR, CR/CRi, and PB-CR with the achievement of absolute blood counts (rather than hematologic improvement). In this context, achievement of PB-CR by cycle 10 had the highest prognostic impact of all response types for the endpoints time to next treatment (LHR 96.2) and OS (LHR 127.8) (Figure S4).

### 3.4 | Multivariate-adjusted Kaplan–Meier curves

After having identified PB-CR as the best hematologic improvement response in terms of LHR, we wanted to know whether this translates into observed outcomes. Only patients for whom all 17 covariates were available were included ($n = 1161$). Patients achieving PB-CR had +10.5 months longer adjusted time to next treatment (18.9 [95% CI 16.5–21.6] vs. 8.4 [95%CI 7.8–9.0] months; $p < .0001$) and a 64% reduced risk of requiring a next treatment or death (HR 0.361 [95%CI 0.301–0.433]; Figure 2A), and + 12.4 months longer adjusted OS (22.8 [95%CI 18.9–26.2] vs. 10.4 [95%CI 9.7–11.2] months; $p < .0001$) and a 63% reduced risk of death (HR 0.366 [95%CI 0.303–0.441]; Figure 2B) than patients who did not. Similar results were observed when selectively analyzing patients with (Figure S5A,B) or without (Figure S5C,D) BM evaluations.

Among patients achieving morphologic CR, those additionally achieving PB-CR had +10.4 months longer adjusted time to next treatment (25.4 [95%CI 21.1–32.6] vs. 15.0 [95%CI 11.8–20.5] months; $p = .0032$) and a 67% reduced risk of requiring a next treatment or death (HR 0.334 [95%CI 0.161–0.693]; Figure 2C), and + 10.9 months longer adjusted OS (32.6 [95%CI 26.2–49.2] vs. 21.7 [95%CI 16.9–27.7] months; $p = .0161$) and a 60% reduced risk of death (HR 0.400 [95%CI 0.190–0.844]; Figure 2D). Similarly, among patients achieving morphologic CR/CRi, those additionally achieving PB-CR had +7.3 months longer adjusted time to next treatment (24.1 [95%CI 20.5–27.7] vs. 16.8 [95%CI 13.8–22.0] months; $p = .0037$) and a 48% reduced risk of requiring a next treatment or death (HR 0.517 [95%CI 0.331–0.808]; Figure 2E), and + 7.5 months longer adjusted OS (29.5 [95%CI 24.7–32.8] vs. 22.0 [95%CI 17.7–25.6] months, $p = .0043$) and a 49% reduced risk of death (HR 0.509 [95%CI 0.320–0.810]; Figure 2F).

Among patients not achieving morphologic CR, those achieving PB-CR had +5.1 months longer adjusted time to next treatment (17.7 [95%CI 15.0–19.8] vs. 12.6 [95%CI 10.4–12.6] months; $p < .0001$) and a 49% reduced risk of requiring a next treatment or death (HR 0.506 [95%CI 0.399–0.643]; Figure S6A), and + 5.7 months longer adjusted OS (19.9 [95%CI 17.6–22.4] vs. 14.2 [95%CI 13.1–16.1] months; $p < .0001$) and a 44% reduced risk of death (HR 0.556 [95%CI 0.435–0.709]; Figure S6B). Similarly, among patients achieving morphologic CR/CRi, those additionally achieving PB-CR had +6.7 months longer adjusted time to next treatment (17.5 [95%CI 14.0–19.7] vs. 10.8 [95%CI 9.8–12.1] months; $p < .0001$) and a 52% reduced risk of requiring a next treatment or death (HR 0.482 [95%CI 0.364–0.639]; Figure S6C), and + 5.5 months longer adjusted OS (19.3 [95%CI 16.5–28.2] vs. 13.8 [95%CI 12.5–15.0] months, $p < .0001$) and a 45% reduced risk of death (HR 0.550 [95%CI 0.412–0.736]; Figure S6D). Similar results were achieved when limiting the analyses to patients with or without BM evaluations during treatment with azacitidine, respectively (data not shown).

### 3.5 | Added value of PB-CR

Among individual response types, the multivariate-adjusted LHR was highest for PB-CR for the endpoints time to next treatment (LHR 224.9) and OS (LHR 209.3), as compared to morphologic CR (LHR 194.0 for time to next treatment and 189.7 for OS) or CR/CRi (LHR 210.7 for time to next treatment and 202.5 for OS; Table S18).

For the endpoint time to next treatment, significant increases of the LHR were observed after addition of the PB-CR to either morphologic CR (LHR increase from 194.0 to 234.4; $p < .0001$) or CR/CRi (LHR increase from 210.7 to 246.7; $p < .0001$; Table S18, gray shaded columns). Similarly, for the endpoint OS, significant increases of the LHR were observed after addition of (i) the PB-CR to morphologic CR (LHR increase from 189.7 to 221.1; $p < .0001$) or CR/CRi (LHR increase from 202.5 to 229.5; $p < .0001$; Table S18, gray shaded columns), indicating that PB-CR provides added value to these individual response types for both endpoints.

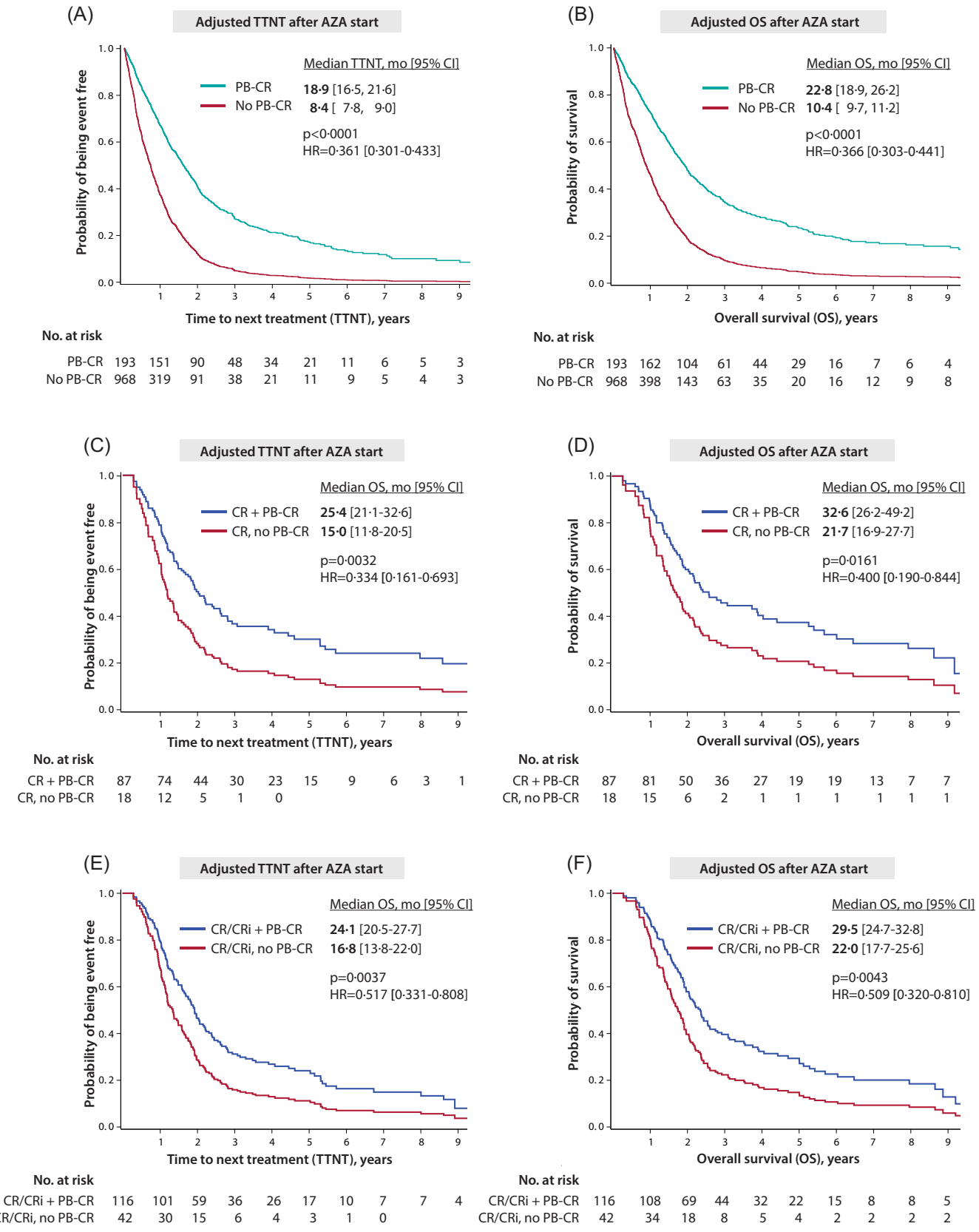


**FIGURE 2** Legend on next page.

### 3.6 | CPH neural network model and permutation-based feature importances

After having identified PB-CR to be at least equivalent to CR or CR/CRi with conventional statistical models, and to have added value to these conventional response types, we aimed to strengthen these results by applying DeepSurv,[34] a CPH neural network (Figure S3; Table S19, left columns). Only patients for whom (a) all 17 baseline covariates remaining in the final CPH model (Table S17), (b) who had a BM evaluation during treatment with azacitidine, and (c) for whom both response variables were available, were included ($n = 624$).

For the whole cohort, after 5-fold cross-validation (Figure S2), and after Bonferroni correction for multiple testing, the concordance indices with intact PB-CR and randomly shuffled (and thus deprived of information content) CR (blue violin plots) were significantly higher than for intact CR and shuffled PB-CR (orange violin plots) for the endpoints time on treatment (C-index 0.62 [IQR 0.61–0.63] vs. 0.59 [IQR 0.58–0.60], $p < .0001$), time to next treatment (C-index 0.65 [IQR 0.64–0.66] vs. 0.62 [IQR 062–0.63], $p < .0001$), and OS (C-index 0.62 [IQR 0.62–0.6 vs. 0.61] [IQR 0.60–0.62], $p < .0001$), indicating a significantly higher information content for the PB-CR variable as compared to morphologic CR (Figure 3A–C; Table S19, left columns). This also held true when stratifying the cohort by diagnosis at azacitidine treatment start. Permutation of PB-CR (orange plots) resulted in a significantly higher drop in model performance, than shuffling of CR (blue plots), for all 3 time-to-event endpoints: time on treatment, time to next treatment, and OS for patients diagnosed with AML (Figure 3D–F), MDS (Figure 3G–I), or CMML (Figure 3J–L; Table S19, left columns).

Similarly, concordance indices with intact PB-CR and shuffled CR/CRi (blue violin plots) were significantly higher than for intact CR and shuffled PB-CR (orange violin plots) for the endpoints time on treatment (C-index 0.62 [IQR 0.61–0.63] vs. 0.60 [IQR 0.60–0.61], $p < .0001$), time to next treatment (C-index 0.65 [IQR 0.65–0.66] vs. 0.63 [IQR 0.62–0.63], $p < .0001$), and OS (C-index 0.62 [IQR 0.62–0.63] vs. 0.61 [IQR 0.60–0.62], $p < .0001$), indicating a significantly higher information content for the PB-CR variable as compared to CR/CRi (Figure S7A–C; Table S19, right columns). This also held true when stratifying the cohort by diagnosis at azacitidine treatment start. Permutation of PB-CR (orange plots) resulted in a significantly higher drop in model performance, than shuffling of CR/CRi (blue plots), for all 3 time-to-event endpoints: time on treatment, time to next treatment, and OS for patients diagnosed with AML (Figure S7D–F), MDS (Figure S7G–I), or CMML (Figure S7J–L; Table S19, right columns). These results remained nearly identical for other network architectures (data not shown).

## 4 | DISCUSSION

To our knowledge, our data are the only information on an evidence-based comparison between CR and CR/CRi, and the new response type PB-CR defined by our group. The latter incorporates and combines all requirements for morphologic CR for patients with MDS, CMML, and AML,[16,18,20–24] without BM findings and therefore does not necessitate an invasive BM evaluation. Using a conventional CPH model adjusting for 17 baseline variables at azacitidine treatment start, we found that achievement of PB-CR confers significantly longer adjusted time to next treatment (+10.5 months, $p < .0001$, HR 0.361) and longer OS (+12.4 months, $p < .0001$, HR 0.366), irrespective of BM blast count. In patients achieving morphologic CR, those additionally achieving PB-CR had a longer adjusted time to next treatment (+10.4 months, $p = .0032$, HR 0.334) and longer adjusted OS (+10.9 months, $p = .0161$, HR 0.400). Among patients in whom a BM evaluation had been performed during azacitidine treatment and who had received at least 10 cycles of treatment, heatmaps of LHRs identified PB-CR to be at least as good as, or better than, CR and CR/CRi and much better than hematologic improvement responses (including hematologic improvement in the erythrocyte, platelet, and/or neutrophil lineage(s)) in predicting time to next treatment and OS.

Standard CPH models have been deemed too simplistic to accurately predict individual survival and other outcomes.[34–36] We thus aimed to strengthen our CPH-based results by applying advanced machine learning techniques. Machine learning is a branch of artificial intelligence in which complex, nonlinear interacting variables can be acquired, thus minimizing the error gap between predictions and observations.[35] DeepSurv is a multiply validated machine learning method of survival analysis that predicts a patient's risk of death by parametrizing the survival hazard through a deep neural network, whose flexibility allows it to outperform, and to provide an improved predictive ability relative to, other common methods of parametric survival analyses, including Weibull, exponential, Gaussian, logistic, loglogistic, and log Gaussian.[34–36] DeepSurv has been shown to be superior to several machine learning and canonical regression survival models and to have the best discriminative performance and calibration at providing accurate predictions of individual survival and at predicting prognosis and risk stratification.[36] Since publication of the method in 2018, 70 publications have applied this technique to data

**FIGURE 2** Multivariate-adjusted KM-curves by response type and endpoint. (A) Stratification by PB-CR in all pts, irrespective of BM evaluation. Endpoint time to next treatment (TTNT). (B) Endpoint overall survival (OS). (C) Stratification by CR and PB-CR at any time in patients with a BM evaluation during azacitidine treatment. Endpoint TTNT. (D) Endpoint OS. (E) Stratification by CR/CRi and PB-CR at any time in patients with a BM evaluation during azacitidine treatment. Endpoint TTNT. (F) Endpoint OS. Adjusted for all 17 baseline covariates at azacitidine treatment start remaining in the final CPH model (Table S17). Only patients for whom all 17 covariates were available were included ($n = 1161$). Only patients for whom all 17 covariates were available, who had a BM evaluation during treatment with azacitidine, and who experienced the respective response type were included. CR, complete morphologic remission; CRi, CR with incomplete blood count recovery; PB-CR, peripheral blood complete remission; mo, months. [Color figure can be viewed at wileyonlinelibrary.com]

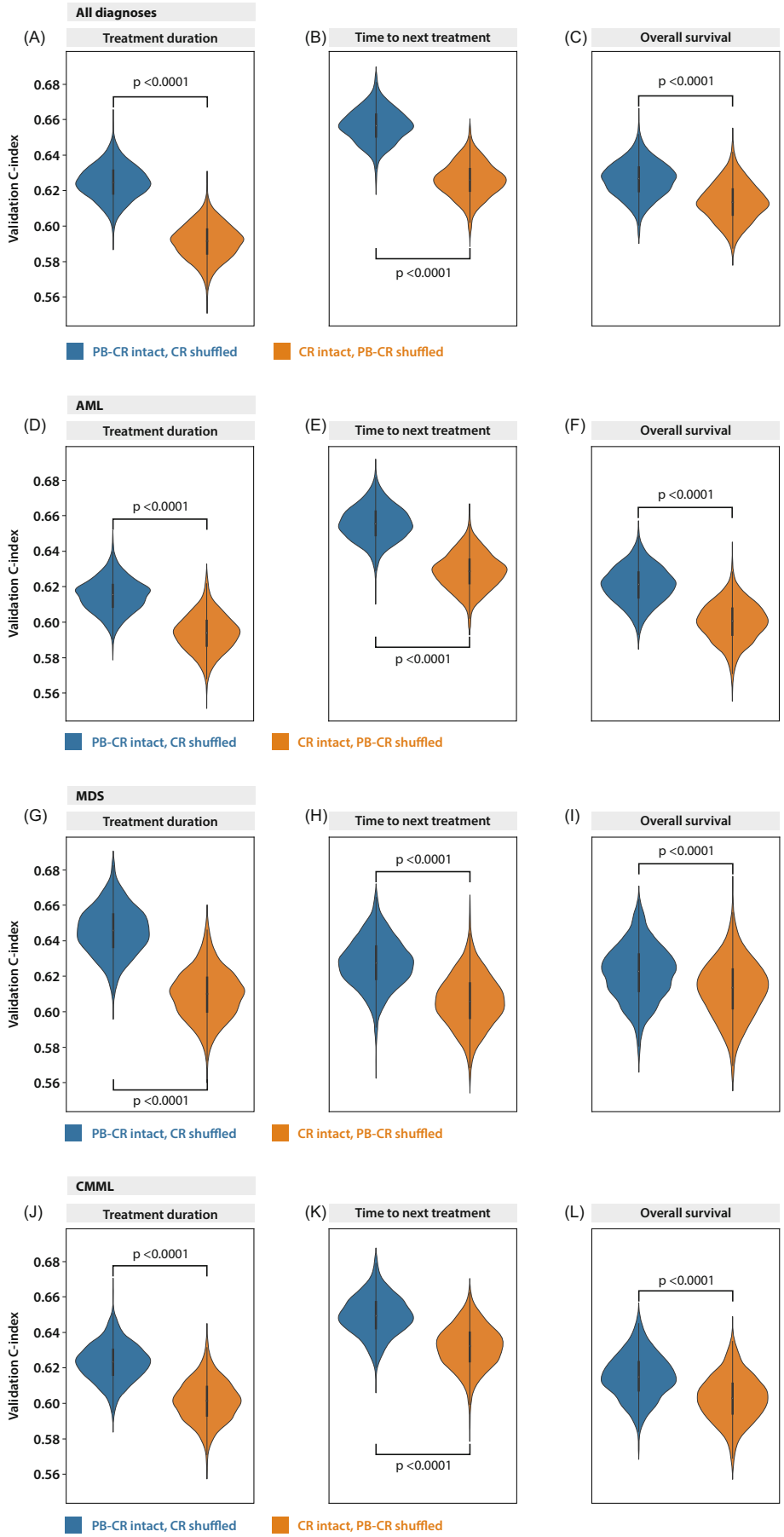


**FIGURE 3** Violin plots. Permutation-based feature importances of PB-CR and CR are significantly different by diagnosis at azacitidine start and by endpoint. Only patients for whom all 17 baseline covariates remaining in the final CPH model (Table S17), who had a BM evaluation during treatment with azacitidine, and for whom both response variables were available, and who were with the respective diagnosis at azacitidine treatment start were included: All diagnoses ($n = 624$), AML ($n = 309$), MDS ($n = 243$), CMML ($n = 72$). Violin plots represent the average validation Concordance Index (C-index) of 1000 random permutations (=destruction of information content). Random permutation (or reshuffling) removes (or destroys) the information content of the respective response type. Blue violin plots represent a corrupted version of the dataset in which the CR response column was randomly reshuffled according to the permutation, with PB-CR remaining intact. Orange violin plots represent a corrupted version of the dataset where the PB-CR response column was randomly reshuffled according to the permutation with the information content of CR remaining intact. For each dataset corrupted by random permutation, 5-fold cross-validation was performed, as described in Figure S2. This yielded a pair of validation C-indices for each permutation, that is, the scores between either the CR-permutated (blue violin plots) or the PB-CR-permutated models (orange violin plots) were compared by using the Wilcoxon signed-rank test. Bonferroni correction was applied to correct for multiple testing for each endpoint separately. Permutation of PB-CR (orange plots) resulted in a significantly higher drop in model performance (lower median C-index), than permutation of CR (blue plots), indicating that PB-CR has a significantly higher information content than morphologic CR. This could be observed for all three time-to-event endpoints analyzed. [Color figure can be viewed at wileyonlinelibrary.com]

from patients, mostly with solid tumours (e.g., References[35–46]). DeepSurv has the potential to supplement traditional survival analysis and become a standard method for medical practitioners to study and recommend personalized treatment options,[34–36] which is why we chose this method to strengthen and validate our results.

To our knowledge, we are the first to apply DeepSurv to response data of patients with myeloid neoplasias in general, and to patients with MDS, CMML, and AML treated with azacitidine in particular. DeepSurv and permutation-based feature importances confirmed PB-CR to have a higher information content than either CR or CR/CRi, and thus to be at least equivalent to these responses with regards to all time-to-event endpoints analyzed, including time on treatment, time to next treatment, and OS.

Thus, both conventional statistical and machine learning techniques show that PB-CR can identify patients with adequate responses to non-intensive therapies without requiring BM assessments. These data thus ascertain that achievement of PB-CR can safely be used as an alternative to BM samples to identify an adequate clinical response and guide treatment decisions. Hence, BM evaluations solely for the purpose of identifying patients with or without CR/CRi are no longer necessary in patients treated with non-intensive strategies. This allows for less frequent follow-up of BM evaluations, which may perhaps be entirely omitted in the future. This is extremely relevant information for both the treating physicians and patients, as samples of peripheral blood can be drawn easily, nearly painlessly, and can be quickly assessed at multiple timepoints.

PB-CR should be used (i) in patients in whom a BM evaluation is not performed due to decisions made by the treating physician and/or due to the patients' wish, (ii) in patients in whom a performed BM evaluation remains uninformative (e.g., due to dry tap, hypocellular marrow, coagulation of the specimen before an adequate smear could be performed, or due to sampling errors), and (iii) to provide added information in patients in whom BM evaluations are performed. PB-CR provides critical information for patients who would formerly have been deemed non-responders and taken off therapy. Achievement of PB-CR, with or without a BM evaluation, with or without achievement of CR/CRi, identifies patients with a clinically meaningful response, and who should remain on treatment. Importantly, lack of achievement of PB-CR should not be used to identify patients who no longer require or profit from treatment. Further analysis of patients not achieving PB-CR will be required to determine whether patients can be identified who no longer profit from treatment.

We acknowledge limitations inherent to real-world evidence. One important limitation of this study is that in order to be able to compare both response types (i.e., PB-CR with either CR or CR/CRi), patients could only be included in the analyses if both bone marrow and peripheral blood samples were obtained during treatment, as the permutation-based feature importance approach cannot be extended to accommodate missing bone marrow information. This may introduce a bias, as the decision of whether or not to perform a bone marrow evaluation might not have been taken at random. However, our data show that achievement of PB-CR results in significant improvement of adjusted OS and TTNT, irrespective of whether a BM evaluation was performed or not. Selection bias and residual confounding cannot be excluded completely but were minimized by multivariate adjustment for 17 baseline variables. Sensitivity analyses were performed to check the general conclusions by assessing different endpoints (time on treatment, time to next treatment, OS), applying different regression models, applying artificial intelligence-based methods, patient stratification, using 5-fold cross-validation, and using different neural network inner architectures. Findings remained significant after application of all these measures and after Bonferroni correction for multiple testing.

Our database has previously been used to confirm phase-3 randomized clinical trial data of the AZA-AML-001 study (NCT01074047) by direct comparison of patient-level data, showing the high quality and utility of our database.[7] CR and CR/CRi are the internationally accepted gold standards of response assessment in patients with AML, MDS, and CMML and are often used as primary endpoints in randomized clinical trials. All current response criteria for require an invasive BM evaluation.[16,18,20–24] Outside of clinical trials, however, BM evaluations are only performed in ~50% of patients during follow-up,[25–27] thus pinpointing a clinical need to be able to identify patients that have adequate response to therapy without repeated BM assessments.

All current response criteria were based on expert opinions, personal experience, and reports analyzing certain aspects, components, or thresholds of response criteria, rather than on results generated from a large dataset. Due to this lack of experimental data, knowledge on which precise thresholds of individual and combined response parameters (such as BM blasts, peripheral blood blasts, hemoglobin levels, platelet count, white blood cell, and/or neutrophil count) are required for optimal prediction of time-to-event endpoints is non-existent. In addition, data on the use of growth factors and their potential interference with arbitrary thresholds chosen for the response criteria are completely lacking. Indeed, many of the published response criteria do not mention whether, or how, the application of growth factors is to be considered when determining patients' response.

The availability of comprehensive clinical data represents the greatest strength of our cohort. To our knowledge, the Austrian Myeloid Registry (aMYELOIDr) is the only real-world evidence database worldwide to have systematically and prospectively documented (a) all information necessary to be able to calculate responses from entered data, (b) information whether growth factors were applied (i.e. treatment with erythrocyte stimulating agents, thrombocyte stimulating agents, and/or granulocyte colony stimulating factors), (c) detailed information from BM evaluations, in addition to (d) detailed treatment information (including absolute dose, dose per body surface area, and days of application) for each documented azacitidine cycle. In this prospective cohort study, 1441 patients included within the aMYELOIDr, 13 971 treatment cycles, 13 971 differential blood counts assessed at the start of each treatment cycle, 16 730 applied red blood cell transfusions, 7652 applied platelet transfusions, and 1228 BM evaluations were analyzed. Analyses were adjusted for 17 baseline variables that remained significant in the final CPH model.

While the uniform treatment of this cohort with azacitidine is a strength regarding the interpretability of our results with regards to patients treated with this drug, it can also be seen as a drawback regarding the generalizability of our results to other treatments. We want to stress that all our conclusions refer to patients treated with hypomethylating agents and perhaps other non-intensive treatment strategies. The generalizability of our results to other treatment strategies needs to be tested and validated separately. Further evidence in patients treated with intensive chemotherapy would be desirable. These analyses are planned when sufficient data on other treatments has been entered into the Austrian Myeloid Registry (aMYELOIDr). Validation of our results with randomized clinical trial data, which represent the highest level of clinical evidence, would be highly desirable, and in this regard, we are open for collaborations with other research groups and/or pharmaceutical companies to achieve the best possible scientifically valid results for both patients and treating physicians.

In conclusion, our data confirm CR/CRi to be a valid and useful clinical endpoint, but more importantly, they show that PB-CR not only adds value to CR/CRi but is at least as good, if not better, at predicting patient outcomes without necessitating an invasive BM evaluation. These data could change clinical practice and raise the current evidence level in future expert opinions and guidelines. We propose and anticipate the rapid and widespread adoption of PB-CR as a new response type in clinical practice, the incorporation of PB-CR as secondary endpoint in randomized clinical trials, the assessment of whether these findings hold true for other treatments, as well as the consideration of PB-CR in future response criteria.

## AUTHOR CONTRIBUTIONS

**Lisa Pleyer:** research design, data collection and assembly, data consistency check and data cleaning, supervision of data analysis, interpretation of the data, writing of the manuscript, preparation of the tables, supervision of figure assembly and formatting, provision of patient data, and approval of the final manuscript for submission; **Manuel Drost, Julian Larcher-Senn, Marc Vaisband, and Jan Hasenauer**: performed all statistical analyses, reviewed the manuscript, made a significant contribution to the manuscript, and approved the final manuscript for submission. **Lisa Pleyer, Manuel Drost, Julian Larcher-Senn, Marc Vaisband, and Jan Hasenauer**: had access to and verified the data. **Lisa Pleyer, Manuel Drost, Julian Larcher-Senn, Marc Vaisband, and Jan Hasenauer**: verified all results. **All other coauthors**: collection and provision of patient data, critical review of the manuscript, significant contribution to the manuscript, and approval of the final manuscript for submission. All authors had full access to the results reported in this manuscript and accepted responsibility to submit for publication.

## FUNDING INFORMATION

The Austrian Group for Medical Tumor Therapy (AGMT) is the sponsor for the Austrian Myeloid Registry (aMYELOIDr) and the Austrian Registry of Hypomethylating Agents and has received funding from Celgene/BMS, AbbVie, and Janssen Cilag. The AGMT is a non-for-profit organization and an academic study group. The group performed administrative and legal management as well as funding acquisition. No pharmaceutical company and no other funding source were involved in any way or had a role in the study design, data collection, data analysis, data interpretation, or writing of the report. No medical writer or editor was involved.

## CONFLICT OF INTEREST STATEMENT

**Lisa Pleyer:** Honoraria from AbbVie, BMS/Celgene, Novartis, and Otsuka Pharmaceutical Co.; Travel support from AbbVie, AstraZeneca, BeiGene, and Takeda. **Marc Vaisband**: No potential conflicts of interest; **Manuel Drost**: No potential conflicts of interest; **Michael Pfeilstöcker**: Honoraria from Abbvie, Astellas, BMS, Jazz, Novartis, and Takeda; **Reinhard Stauder**: Honoraria from BMS/Celgene. **Sonja Heibl**: Honoraria from AbbVie, AOP, BMS, Janssen Cilag, Novartis, and Roche; **Heinz Sill**: AdBoard—Celgene, AbbVie, Pfizer, Astellas; Consultancy: Celgene/BMS, AbbVie, Astellas; **Petra Pichler:** No potential conflicts of interest; **Michael Girschikofsky**: No potential conflicts of interest; **Margarete Stampfl-Mattersberger**: No potential conflicts of interest; **Angelika Pichler**: No potential conflicts of interest; **Bernd Hartmann**: Honoraria from Celgene, Amgen, Janssen, and AbbVie; **Angelika Pichler**: Honoraria from Celgene; **Martin Schreder**: Honoraria and consultancy from BMS/Celgene; **Clemens A. Schmitt**: Travel support, honoraria, and consulting fees from Abbvie, AstraZeneca, Bayer, Bristol-Myers Squibb/Celgene, Gilead/Kite, Janssen-Cilag, MSD, Novartis, Octapharma, Pierre Fabre, Roche, Sanofi, Takeda, and TissUse; **Sonia Vallet**: Honoraria from Bristol Myers Squibb, Merck, MSD, and Pfizer; consultancy fees from Roche, MSD, EUSA Pharma, and Merck; Travel support from Pfizer, Roche, Pierre Fabre, and Angelini; **Thomas Melchardt**: Honoraria from AbbVie, and Celgene/BMS; **Armin Zebisch**: Honoraria from Abbvie, Roche, BMS-Celgene, Novartis, and Jazz; **Sigrid Machherndl-Spandl**: Honoraria from Abbvie, Celgene/BMS, and Pfizer; **Dominik Wolf**: Research Funding: BMS/Celgene, MSD, Novartis, Pfizer, and Roche; Honoraria: BMS/Celgene, GEMOAB, Gilead, Incyte, MSD, Novartis, Pfizer, and Roche; **Felix Keil**: No potential conflicts of interest; **Jan Hasenauer**: Research Funding: Böhringer-Ingelheim; **Julian Larcher-Senn**: No potential conflicts of interest; **Richard Greil**: Honoraria from AbbVie, Amgen, AstraZeneca, BMS/Celgene, Daiichi Sankyo, Gilead, Merck, Novartis, Roche, Takeda, BMS, MSD, Sandoz, and Gilead; Research funding from Celgene, Roche, Merck, Novartis, MSD, Sandoz, and Takeda; Consulting: AbbVie, Astra Zeneca, BMS/Celgene, Novartis, Roche, Takeda, Janssen, MSD, Merck, Gilead, and Daiichi Sankyo; Travel support from AbbVie, Amgen, Astra Zeneca, BMS/Celgene, Daiichi Sankyo, Gilead, Janssen Cilag, MSD, Novartis, and Roche; The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## DATA AVAILABILITY STATEMENT

The datasets supporting the conclusions of this article are included within the article and the on-line supplementary material. Data

sharing of patient level data collected for the study is not planned. However, we are open to research questions asked by other researchers, and we are also open to data contributions by others. Participation requests or potential joint research proposals can be made at any timepoint to the corresponding author via email (dr.lisa.pleyer@gmail.com) and are subject to approval by the AGMT and its collaborators.

## ORCID

*Lisa Pleyer* https://orcid.org/0000-0002-2790-5067

*Heinz Sill* https://orcid.org/0000-0003-0993-4371

## REFERENCES

1. Pleyer L, Sekeres MA. An early glimpse at azacitidine plus venetoclax for myelodysplastic syndromes. *Lancet Haematol*. 2022;9(10):e714-e716.
2. Sekeres MA, Steensma DP. Rethinking clinical trial endpoints in myelodysplastic syndromes. *Leukemia*. 2019;33(3):570-575.
3. Kim N, Pavletic S, Norsworthy KJ. Meaningful response criteria for myelodysplastic syndromes. *Br J Haematol*. 2022;196(5):1137-1148.
4. Chandhok NS, Sekeres MA. What constitutes meaningful improvement in myelodysplastic syndromes? *Leuk Lymphoma*. 2022;63:1-8.
5. Bloomfield CD, Estey E, Pleyer L, et al. Time to repeal and replace response criteria for acute myeloid leukemia? *Blood Rev*. 2018;32(5):416-425.
6. Schuh AC, Dohner H, Pleyer L, Seymour JF, Fenaux P, Dombret H. Azacitidine in adult patients with acute myeloid leukemia. *Crit Rev Oncol Hematol*. 2017;116:159-177.
7. Pleyer L, Dohner H, Dombret H, et al. Azacitidine for front-line therapy of patients with AML: reproducible efficacy established by direct comparison of international phase 3 trial data with registry data from the Austrian Azacitidine registry of the AGMT study group. *Int J Mol Sci*. 2017;18(2):415-432.
8. Schuh ADH, Sandhu I, JSeymour JF, et al. Overall survival (OS) without complete remission (CR) in older patients with acute myeloid leukemia (AML): azacitidine (AZA) vs conventional care regimens (CCR) in the AZA AML-001 study. *Haematologica*. 2015;100:219-220.
9. Pleyer L, Burgstaller S, Stauder R, et al. Is it time to redefine response in elderly patients with World Health Organization-defined acute myeloid leukemia unfit for intensive chemotherapy? *Blood*. 2015;126(23):126.
10. Pleyer L, Pfeilstoecker M, Stauder R, et al. Expanding on current definitions of hematologic improvement in MDS, CMML and AML: landmark analyses of 1301 patients treated with Azacitidine in the Austrian registry of hypomethylating agents by the AGMT-study group. *Blood*. 2019;134(Supplement_1):3182.
11. Pleyer L, Pfeilstocker M, Stauder R, et al. Peripheral blood complete remission provides added value to the classical definition of morphologic complete remission—a prospective cohort study of 1441 patients with MDS, CMML and AML treated within the Austrian Azacitidine registry. *Blood*. 2021;138:3387.
12. Pleyer L, Burgstaller S, Girschikofsky M, et al. Azacitidine in 302 patients with WHO-defined acute myeloid leukemia: results from the Austrian Azacitidine registry of the AGMT-study group. *Ann Hematol*. 2014;93(11):1825-1838.
13. Oster HS, Crouch S, Smith A, et al. A predictive algorithm using clinical and laboratory parameters may assist in ruling out and in diagnosing MDS. *Blood Adv*. 2021;5(16):3066-3075.
14. Arber DA, Orazi A, Hasserjian RP, et al. International consensus classification of myeloid neoplasms and acute leukemias: integrating morphologic, clinical, and genomic data. *Blood*. 2022;140(11):1200-1228.
15. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization classification of Haematolymphoid Tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*. 2022;36(7):1703-1719.
16. Dohner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood*. 2022;140(12):1345-1377.
17. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*. 2012;120(12):2454-2465.
18. Zeidan AM, Platzbecker U, Bewersdorf JP, et al. Consensus proposal for revised international working group response criteria for higher risk myelodysplastic syndromes. *Blood*. 2023;141:2047-2061.
19. Tazi Y, Arango-Ossa JE, Zhou Y, et al. Unified classification and risk-stratification in acute myeloid leukemia. *Nat Commun*. 2022;13(1):4622.
20. Cheson BD, Greenberg PL, Bennett JM, et al. Clinical application and proposal for modification of the international working group (IWG) response criteria in myelodysplasia. *Blood*. 2006;108(2):419-425.
21. Cheson BD, Bennett JM, Kopecky KJ, et al. Revised recommendations of the international working Group for Diagnosis, standardization of response criteria, treatment outcomes, and reporting standards for therapeutic trials in acute myeloid leukemia. *J Clin Oncol*. 2003;21(24):4642-4649.
22. Dohner 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.
23. Savona MR, Malcovati L, Komrokji R, et al. An international consortium proposal of uniform response criteria for myelodysplastic/myeloproliferative neoplasms (MDS/MPN) in adults. *Blood*. 2015;125(12):1857-1865.
24. Platzbecker U, Fenaux P, Ades L, et al. Proposals for revised IWG 2018 hematological response criteria in patients with MDS included in clinical trials. *Blood*. 2019;133(10):1020-1030.
25. Jansko-Gadermeir B, Leisch M, Gassner FJ, et al. Myeloid NGS analyses of paired samples from bone marrow and peripheral blood yield concordant results: a prospective cohort analysis of the AGMT study group. *Cancers (Basel)*. 2013;15(8):2305-2325.
26. Leisch M, Pfeilstocker M, Stauder R, et al. Adverse events in 1406 patients receiving 13,780 cycles of Azacitidine within the Austrian registry of hypomethylating agents-a prospective cohort study of the AGMT study-group. *Cancers (Basel)*. 2022;14(10):2459-2478.
27. Dinmohamed AG, van Norden Y, Visser O, et al. Effectiveness of azacitidine for the treatment of higher-risk myelodysplastic syndromes in daily practice: results from the Dutch population-based PHAROS MDS registry. *Leukemia*. 2015;29(12):2449-2451.
28. Dinmohamed AG, van Norden Y, Visser O, et al. The use of medical claims to assess incidence, diagnostic procedures and initial treatment of myelodysplastic syndromes and chronic myelomonocytic leukemia in The Netherlands. *Leuk Res*. 2015;39(2):177-182.
29. Pleyer L, Leisch M, Kourakli A, et al. Outcomes of patients with chronic myelomonocytic leukaemia treated with non-curative therapies: a retrospective cohort study. *Lancet Haematol*. 2021;8(2):e135-e148.
30. Pleyer L, Burgstaller S, Stauder R, et al. Azacitidine front-line in 339 patients with myelodysplastic syndromes and acute myeloid leukaemia: comparison of French-American-British and World Health Organization classifications. *J Hematol Oncol*. 2016;9:39.
31. Pleyer L, Heibl S, Tinchon C, et al. Health-related quality of life as assessed by the EQ-5D-5L predicts outcomes of patients treated with Azacitidine-a prospective cohort study by the AGMT. *Cancers (Basel)*. 2023;15(5):1388-1408.

32. Pleyer L, Stauder R, Burgstaller S, et al. Azacitidine in patients with WHO-defined AML—results of 155 patients from the Austrian Azacitidine registry of the AGMT-study group. *J Hematol Oncol*. 2013;6:32.
33. Efficace F, Gaidano G, Breccia M, et al. Prognostic value of self-reported fatigue on overall survival in patients with myelodysplastic syndromes: a multicentre, prospective, observational, cohort study. *Lancet Oncol*. 2015;16(15):1506-1514.
34. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18(1):24.
35. Hatano Y, Ishihara T, Hirokawa S, Onodera O. Machine learning approach for the prediction of age-specific probability of SCA3 and DRPLA by survival curve analysis. *Neurol Genet*. 2023;9(3):e200075.
36. Zeng J, Li K, Cao F, Zheng Y. Development and validation of survival prediction model for gastric adenocarcinoma patients using deep learning: a SEER-based study. *Front Oncol*. 2023;13:1131859.
37. Wang D, Pan B, Huang JC, et al. Development and validation of machine learning models for predicting prognosis and guiding individualized postoperative chemotherapy: a real-world study of distal cholangiocarcinoma. *Front Oncol*. 2023;13:1106029.
38. Li B, Yang L, Jiang C, et al. Integrated multi-dimensional deep neural network model improves prognosis prediction of advanced NSCLC patients receiving bevacizumab. *Front Oncol*. 2023;13:1052147.
39. Kim DY, Choi KH, Kim JH, et al. Deep learning-based personalised outcome prediction after acute ischaemic stroke. *J Neurol Neurosurg Psychiatry*. 2023;94(5):369-378.
40. Deng Y, Liu L, Jiang H, et al. Comparison of state-of-the-art neural network survival models with the pooled cohort equations for cardiovascular disease risk prediction. *BMC Med Res Methodol*. 2023;23(1):22.
41. Yang B, Liu C, Wu R, et al. Development and validation of a DeepSurv nomogram to predict survival outcomes and guide personalized adjuvant chemotherapy in non-small cell lung cancer. *Front Oncol*. 2022;12:895014.
42. Huang C, Dai Y, Chen Q, et al. Development and validation of a deep learning model to predict survival of patients with esophageal cancer. *Front Oncol*. 2022;12:971190.
43. Bao L, Wang YT, Zhuang JL, et al. Machine learning-based overall survival prediction of elderly patients with multiple myeloma from multicentre real-life data. *Front Oncol*. 2022;12:922039.
44. Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine learning-guided adjuvant treatment of Head and neck cancer. *JAMA Netw Open*. 2020;3(11):e2025881.
45. She Y, Jin Z, Wu J, et al. Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA Netw Open*. 2020;3(6):e205842.
46. Jobczyk M, Stawiski K, Kaszkowiak M, et al. Deep learning-based recalibration of the CUETO and EORTC prediction tools for recurrence and progression of non-muscle-invasive bladder cancer. *Eur Urol Oncol*. 2022;5(1):109-112.


## SUPPORTING INFORMATION

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

**How to cite this article:** Pleyer L, Vaisband M, Drost M, et al. Cox proportional hazards deep neural network identifies peripheral blood complete remission to be at least equivalent to morphologic complete remission in predicting outcomes of patients treated with azacitidine—A prospective cohort study by the AGMT. *Am J Hematol*. 2023;98(11):1685-1698. doi:10.1002/ajh.27046