Can we forecast induction failure in acute myeloid leukemia?

Felicitas Thol

Department of Hematology, Hemostasis, Oncology, and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany

E-mail: thol.felicitas@mh-hannover.de

doi:10.3324/haematol.2018.187575

Standard induction therapy for fit patients with acute myeloid leukemia (AML) consists of a combination therapy with anthracycline and cytarabine. This classical regimen, typically called "7+3", has not changed for several decades.[1] While many patients achieve a complete remission (CR) with standard induction therapy, approximately 10-40% of patients fail to respond to induction treatment.[2,3] These patients are classified as having primary refractory disease (RD) or treatment failure, defined as a failure to achieve CR or incomplete hematologic recovery (Cri) after two courses of induction treatment.[4] Unfortunately, treatment of patients with RD is extremely challenging, as even with salvage therapy followed by allogeneic stem cell transplantation, patient outcomes remain poor.[3]

It is still difficult for hematologists to reliably predict RD in newly diagnosed AML patients prior to initiation of therapy. At time of diagnosis, we typically risk stratify our patients based on their cytogenetic and molecular profile. A very helpful classification was introduced by the European Leukemia Net (ELN) in 2010,[5] (revised in 2017[4]) and this currently includes three prognostic groups integrating cytogenetics as well as the mutational status of *FLT3*-ITD (including mutational load), *NPM1*, *ASXL1*, *TP53*, *RUNX1*, *CEBPA* (biallelic mutants). However, this risk stratification is geared towards the estimation of overall survival (OS) and event-free survival (EFS), and not primarily towards forecasting RD.[4] Although there is a strong correlation between treatment failure and OS, they still present different outcome measures.[4,6]

Several groups have attempted to develop specific scores to predict induction failure in AML. A reliable score primarily focusing on the likelihood of treatment failure rather than OS could improve patient care and treatment in many ways. If we could reliably predict that a patient would not respond to "7+3" treatment prior to induction therapy, we would be compelled to search for alternatives at the time of diagnosis, potentially sparing the patient from the toxicity of treatments that prove to be ineffective. As several new agents are being studied front line (e.g. FLT3 and IDH1/2 inhibitors with intensive chemotherapy, BCL2-inhibitors in combination with low-dose cytarabine or azacitidine, etc.) alternatives for "7+3" might soon become a reality. In addition, a reliable RD score could allow us to identify those patients who require an urgent donor search at the time of diagnosis.[7-10] Furthermore, an RD score could become an important consideration when designing clinical trials that specifically target this high-risk patient group.

In this issue of *Haematologica*, Herold *et al.* introduce a 29-gene and cytogenetic score that can help to predict resistance to induction chemotherapy in adult AML patients.[11] Importantly, this score was developed on the basis of various categories of prognostic markers, considering clinical characteristics, laboratory variables, cytogenetics, mutational status of 68 genes that are frequently mutated in AML, and the expression profile of 29 genes known to be prognostic for AML. Their score estimates the likelihood of primary RD based on large independent clinical training sets. The first cohort (training set 1) included 407 patients of the AML Cooperative Group (AMLCG trials between 1999-2005), the second cohort (training set 2) consisted of 462 AML patients treated in the Haemato-Oncology Foundation for Adults in the Netherlands (HOVON) trials and the validation cohort was based on 210 AMLCG-2008 trial patients with the addition of 40 patients with RD from the AMLCG 1999 trial. The implementation of a large validation cohort is critical for assessing the reliability of any score, especially for clinical practice. The score was calculated as a weighted linear sum of the individual predictors. Interestingly, the final predictor by Herold *et al.* (predictive score 29 MRC or PS29MRC) included expression levels of 29 genes and the UK Medical Research Council (MRC) cytogenetic risk classification, while other parameters such as gene mutations were tested but were excluded from the final score.[12] Importantly, this predictive classifier proved to be significant for RD, both as a continuous variable as well as a dichotomous variable that divides patients into high and low risk. In the multivariate analysis, only PS29MRC, age and TP53 mutations remained independently significant for RD prediction. While the predictor was primarily designed to be associated with RD on day 16 after induction chemotherapy, the score also proved to be strongly associated with survival. When examining different groups of the current ELN 2017 classification, the predictive power of the score was shown in the intermediate and the unfavorable ELN groups, while it could not be shown in the favorable genetic group (likely related to low RD rate in patients with favorable cytogenetics). The validation cohort nicely reproduced the data of the training cohort. All these aspects are suggestive of a very reliable predictive score.

The area under receiver-operating characteristic curve (AUC) can be used as a measure for the predictive ability of a score, with an AUC of 0.7-0.8 classified as fair and less than we would desire for primary treatment decisions.[13,14] The classifier by Herold *et al.* reached an AUC of 0.76 in the validation set. In contrast, Walter *et al.* developed a model for resistance prediction in AML based on the analysis of 4601 patients treated within European and US AML trials.[13] They found that age, performance status, white blood cell count, secondary disease, cytogenetic risk and *NPM1/FLT3-ITD* mutational status were strongly associated independently with primary resistance. Unlike Herold *et al.*, they did not include a complex mutational and gene expression profile in their analysis (Table 1). However, with their model, they achieved a similar AUC (0.78) to that of Herold *et al.*

Krug *et al.* also developed a model based on a cohort of 1406 patients aged over 60 years diagnosed with AML but otherwise medically fit, and who underwent treatment with two intense induction chemotherapy cycles within the

Table 1. Schematic overview of recent studies developing a model for response prediction to induction chemotherapy in intensively treated acute myeloid leukemia (AML) patients.

Publication (website for score)	Prediction for	Patient population	AUC	Variables considered	Variables in the final model
Herold *et al.*[11]	RD	1079 adult patients (including 210 patients in VC)	0.76 (VC)	Clinical characteristics, cytogenetics, laboratory variables, mutational status of 68 frequently mutated genes in AML, expression profile of 29 genes	Cytogenetics risk according to MRC, expression data of 29 genes
Krug *et al.*[15] (http://www.aml-score.org/)	CR+ ED	1406 patients (TC) + 801 patients (VC) (only ≥ 60 years)	0.68 (VC)	Body temperature, WBC, BM blasts, PB blasts, PB neutrophils, age, disease type, hemoglobin, platelet count, serum protein, ALT, bilirubin, BMI, extramedullary disease, fibrinogen, LDH, cytogenetics	Body temperature, age, disease type, hemoglobin, platelet count, fibrinogen, LDH and cytogenetics
Walter *et al.*[14]	RD	4601 adult patients	0.78	Age, PS,sex, WBC, platelet count, BM blast percentage, disease type, cytogenetic risk, FLT3-ITD and NPM1 mutation status	Age, PS, WBC, disease type, cytogenetic risk, FLT3-ITD/NPM1 mutation status
Gerstung *et al.*[16] (http://cancer.sanger.ac.uk/aml-multistage/)	Not primarily RD	1540 AML	N.A.	Clinical data, cytogenetics, mutational data of 111 frequently mutated genes	Age, sex, PS, WBC, platelet count, PB blasts, BM blasts, splenomegaly, disease type, hemoglobin, cytogenetics, mutational status of 58 genes

AUC: area under receiver-operating characteristic curve; CR: complete remission; RD: residual disease; ED: early death; VC: validation cohort; TC: training cohort; WBC: white blood cell count; BM: bone marrow; PB: peripheral blood; disease type: *de novo* leukemia *versus* leukemia secondary to cytotoxic treatment or an antecedent hematologic disease; ALT: alanine aminotransferase; BMI: Body Mass Index; LDH: serum concentration of lactate dehydrogenase; MRC: UK Medical Research Council; PS: Performance Status; N.A.: not applicable.

AML-CG.[15] The validation cohort consisted of an independent cohort of 801 patients aged over 60 years. Their score was based on body temperature, age, secondary disease, hemoglobin, platelet count, fibrinogen, serum concentration of lactate dehydrogenase and cytogenetics. Instead of RD, the achievement of CR and early death were the primary outcome parameters of this score (Table 1). Using CR prediction, the model of Krug *et al.* had an AUC of 0.68 in the validation set.[15]

Gerstung *et al.* have also developed a prognostic algorithm based on a knowledge bank of 1540 AML patients whose cytogenetic, molecular profile, and clinical data were analyzed in detail.[16,17] Here, a number of outcome parameters can be obtained (including death without remission, death without and after relapse, alive after relapse, alive in first CR and alive without CR), and RD can be indirectly calculated (Table 1).

Thus, prediction of RD remains complex, and these scoring systems have yet to find their way into routine clinical practice. The questions of when and how we employ them for everyday clinical evaluation and treatment decisions remain. Here, feasibility and predictability must be considered. It will not be feasible to use a score requiring far more laboratory evaluation (e.g. microarray data, etc.) than is routinely performed. For example, gene expression analysis is not routinely performed in clinical practice and the time required might become relevant for patients with a high leukemic burden in need of urgent therapy. Furthermore, unlike sequencing, gene expression analysis is not covered by the healthcare systems of many countries. However, with the advances being made in technologies, such evaluation could quickly become more feasible. Just as important as feasibility is the level of predictability. We can only justify primarily basing our treatment decisions on scoring systems with a sufficiently high predictability. That none of the proposed scoring systems reach an AUC close to 0.9, even when including all parameters currently known to be prognostic, underscores the challenges of reliably predicting patient outcome at the time of diagnosis. This is highlighted by Herold *et al.*, who used all prognostic parameters currently considered relevant, studied these parameters extensively in the context of RD prediction, and thus, rightfully described an "obstacle" to achieving a higher AUC that is difficult to overcome.

Herold *et al*. describe an innovative approach of how to tackle the pressing question of RD prediction. Independently of its clinical use, it can potentially help us to better understand the biology of primary refractory disease. It is still unknown why some patients with a molecularly more favorable risk profile still fail induction chemotherapy. The gene expression data that predict primary refractory disease might also lead the way to identifying novel targets for AML therapy. Even if the predictive classifier of Herold *et al.* may not find its way into clinical practice just yet, it carries the potential of becoming a tool for designing clinical trials and developing novel treatment strategies.

References

1. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med. 2015;373(12):1136-1152.
2. Burnett A. Treatment of acute myeloid leukemia: are we making progress? Hematology Am Soc Hematol Educ Program. 2012;2012:1-6.
3. Thol F, Schlenk RF, Heuser M, Ganser A. How I treat refractory and early relapsed acute myeloid leukemia. Blood. 2015;126(3):319-327.
4. 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.
5. Dohner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010;115(3):453-474.
6. 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.
7. Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. N Engl J Med. 2017;377(5):454-464.
8. Stein EM, DiNardo C, Altman JK, et al. Safety and Efficacy of AG-221, a Potent Inhibitor of Mutant IDH2 That Promotes Differentiation of Myeloid Cells in Patients with Advanced Hematologic Malignancies: Results of a Phase 1/2 Trial. Blood. 2015;126:323.
9. DiNardo CD, Pratz KW, Letai A, et al. Safety and preliminary efficacy of venetoclax with decitabine or azacitidine in elderly patients with previously untreated acute myeloid leukaemia: a non-randomised, open-label, phase 1b study. Lancet Oncol. 12 Jan 2018 [Epub ahead of print].
10. Wei AH, Tiong IS. Midostaurin, enasidenib, CPX-351, gemtuzumab ozogamicin, and venetoclax bring new hope to AML. Blood. 2017;130(23):2469-2474.
11. Herold T, Jurinovic V, Batcha AMN, et al. A 29-gene and cytogenetic score for the prediction of resistance to induction treatment in acute myeloid leukemia. Haematologica. 2017;103(3):000-000.
12. Grimwade D, Walker H, Oliver F, et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. Blood. 1998;92(7):2322-2333.
13. Walter RB, Othus M, Burnett AK, et al. Resistance prediction in AML: analysis of 4601 patients from MRC/NCRI, HOVON/SAKK, SWOG and MD Anderson Cancer Center. Leukemia. 2015;29(2):312-320.
14. Walter RB, Othus M, Paietta EM, et al. Effect of genetic profiling on prediction of therapeutic resistance and survival in adult acute myeloid leukemia. Leukemia. 2015;29(10):2104-2107.
15. Krug U, Rollig C, Koschmieder A, et al. Complete remission and early death after intensive chemotherapy in patients aged 60 years or older with acute myeloid leukaemia: a web-based application for prediction of outcomes. Lancet. 2010;376(9757):2000-2008.
16. Gerstung M, Papaemmanuil E, Martincorena I, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet. 2017;49(3):332-340.
17. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016;374(23):2209-2221.

Bortezomib for prevention of acute graft-*versus*-host disease: a conclusion reached

Paul J. Martin

Fred Hutchinson Cancer Research Center, Seattle and University of Washington, Seattle, WA, USA

E-mail: pmartin@fredhutch.org

doi:10.3324/haematol.2018.188052

Prevention of acute graft-*versus*-host disease (GvHD) after allogeneic hematopoietic cell transplantation (HCT) has long posed a major challenge for the field. In this issue, Koreth *et al.*[1] report results of a randomized three-arm phase II trial testing two different immunosuppressive regimens using bortezomib to prevent acute GvHD after allogenic hematopoietic cell transplantation. Contrary to expectations, the results did not show major improvement in the experimental groups as compared to the control group.

The impetus to explore the use of bortezomib to prevent GvHD came from its mechanism of action to prevent signaling through nuclear factor (NF)κB in activated T cells. In resting T cells, the inhibitor (I)-κB binds to NFκB as a complex that is sequestered in the cytoplasm.[2] In activated T cells, ubiquitin moieties attach to I-κB, which is delivered to proteasomes. The NFκB molecules released from I-κB translocate to the nucleus where they activate the transcription of genes involved in immune responses. Among other possible mechanisms of action, bortezomib inhibits proteasome activity, allowing I-κB to prevent NFκB-mediated activation of T cells.

Experimental results showed that administration of bortezomib early after allogeneic HCT could prevent acute GvHD in mice.[1] These results led to a phase I/II study demonstrating that bortezomib can be combined with tacrolimus and methotrexate in a tolerable post-transplant immunosuppressive regimen after HCT with human leukocyte antigen (HLA)-mismatched donors. Results of the completed study were published in 2012[3] and are summarized in Figure 1. The interpretation of the 2012 study was initially informed by historical experience showing a 46% incidence of grade II – IV GvHD in patients with HLA-mismatched unrelated donors.[4] The 22% incidence of grade II – IV GvHD in the 2012 study was indeed encouraging when compared against this benchmark, although the comparability of demographic and treatment characteristics of patients in the phase I/II study and the historical group[5] was not well documented.

In the current study of HLA-matched unrelated HCT, the benchmark for grade II – IV GvHD was set at 40%.[1] The observed 33% incidence of grade II – IV GvHD in patients treated with tacrolimus and methotrexate (Arm A) was somewhat lower than this benchmark, while the 29% incidence in patients treated with bortezomib added to tacrolimus and methotrexate (Arm B) was somewhat higher than the 22% incidence observed in HLA-mismatched recipients in the phase I/II study (Figure 1). As a result, the current study did not demonstrate a statistically significant improvement following the addition of bortezomib to tacrolimus and methotrexate in patients with HLA-matched unrelated donors.

Results of the current study with HLA-matched unrelated recipients are similar to those observed in the BMT CTN 1203 PROGRESS study, which enrolled a mixed cohort of HLA-matched related and unrelated recipients and a small proportion of HLA-mismatched unrelated recipients. In this study, the day 180 cumulative inci-