Overview of Risk Stratification Methodologies for MDS and AML (IPSS-M, IPSS-R, ELN 2022)

April 12, 2025

Contents

1 Introduction

This document provides an overview of the methodologies employed for risk stratification in patients diagnosed with Myelodysplastic Syndromes (MDS) and Acute Myeloid Leukemia (AML). Specifically, it details the implementation logic for the Molecular International Prognostic Scoring System (IPSS-M), the Revised International Prognostic Scoring System (IPSS-R) for MDS, and the European LeukemiaNet (ELN) 2022 risk stratification for AML. These models integrate clinical, cytogenetic, and molecular data to predict patient outcomes and guide treatment decisions. The logic described reflects the implementation within our diagnostic system.

Methodologies based on (Bernard et al. 2022; Greenberg et al. 2012; Döhner et al. 2022).

2 IPSS-M Methodology for MDS

The IPSS-M represents a significant advancement in MDS prognostication by incorporating molecular data alongside traditional clinical and cytogenetic variables.

2.1 Input Variables

The IPSS-M calculation requires the following input parameters:

- **Clinical Variables:**
  - Hemoglobin (Hb, g/dL)
  - Platelet count (PLT, Giga/L or 10^9/L)
  - Bone marrow blast percentage (BM_BLAST, %)
- **Cytogenetics:** IPSS-R cytogenetic risk category (Very Good, Good, Intermediate, Poor, Very Poor), represented numerically (`CYTOVEC`: 0–4).
- **Key Gene Mutations (Status: 0=wild-type, 1=mutated):** *SF3B1, ASXL1, SRSF2, DNMT3A, RUNX1, U2AF1, EZH2, CBL, NRAS, IDH2, KRAS, MLL/KMT2A*-PTD (`MLL_PTD`), *ETV6, NPM1, FLT3.*
- **TP53 Status:**
  - Number of *TP53* mutations (`TP53mut`: "0", "1", or "2 or more").
  - Maximum *TP53* Variant Allele Frequency (`TP53maxvaf`, %).
  - Evidence of *TP53* Loss of Heterozygosity (`TP53loh`: 0 or 1, derived from VAF $> 55\%$ or presence of del(17p)/-17).
- **Residual Gene Mutations (Status: 0=wild-type, 1=mutated):** *BCOR, BCORL1, CEBPA, ETNK1, GATA2, GNB1, IDH1, NF1, PHF6, PPM1D, PRPF8, PTPN11, SETBP1, STAG2, WT1.*

2.2 Preprocessing Steps

Several input variables require preprocessing before calculation:

- `SF3B1_5q`: A binary variable set to 1 if *SF3B1* is mutated AND del(5q) is present AND neither del(7q)/-7 nor complex karyotype are present. Otherwise, it is 0 if *SF3B1* is wild-type, del(5q) is absent, or del(7q)/-7/complex are present.
- `SF3B1_alpha`: A binary variable set to 1 if *SF3B1* is mutated AND `SF3B1_5q` is 0 AND none of the adverse mutations (*SRSF2, STAG2, BCOR, BCORL1, RUNX1, NRAS*) are present. Otherwise, it is 0.
- `TP53multi`: A binary variable representing biallelic *TP53* inactivation. It is set to 1 if `TP53mut` is "2 or more", OR if `TP53mut` is "1" AND `TP53loh` is 1. Otherwise, it is 0.
- `HB1`: Directly uses the Hemoglobin value.
- `BLAST5`: Bone marrow blast percentage divided by 5 (capped at 20% original value, resulting in a max `BLAST5` of 4).
- `TRANSF_PLT100`: Platelet count divided by 100 (capped at 250 Giga/L original value, resulting in a max `TRANSF_PLT100` of 2.5).
- `CYTOVEC`: Numerical mapping of IPSS-R cytogenetic category (0=Very Good, 1=Good, 2=Intermediate, 3=Poor, 4=Very Poor).
- `Nres2`: Calculated count of mutated residual genes (from the list above), capped at a maximum value of 2. Missing gene data is handled by imputing based on a reference frequency (`n_ref` = 0.388), generating mean, best-case (missing=0), and worst-case (missing=1) scenarios.

2.3 Calculation Formula

The IPSS-M risk score is calculated as the sum of contributions from each variable. The contribution of each variable (i) is determined by the formula:

$$\text{Contribution}_i = \frac{(\text{Value}_i - \text{Mean}_i) \times \beta_i}{\ln(2)}$$

Where:

- Value_i is the preprocessed value of the variable for the patient.
- Mean_i is the reference mean value for that variable in the model's derivation cohort.
- β_i is the coefficient (beta weight) for that variable.
- $\ln(2) \approx 0.6931$ is the natural logarithm of 2, used for scaling.

The coefficients (β) and mean values used are listed in Table 1. Missing data for any variable results in using the corresponding best, worst, or mean value for that variable (as defined in the model parameters) depending on the scenario being calculated.

Table 1: IPSS-M Coefficients (β) and Mean Values

Variable Name (`name`)	Coefficient (`coeff`, β)	Mean (`means`)
CYTOVEC	0.287	1.39
BLAST5	0.352	0.922
TRANSF_PLT100	-0.222	1.41
HB1	-0.171	9.87
SF3B1_alpha	-0.0794	0.186
SF3B1_5q	0.504	0.0166
ASXL1	0.213	0.252
SRSF2	0.239	0.158
DNMT3A	0.221	0.161
RUNX1	0.423	0.126
U2AF1	0.247	0.0866
EZH2	0.27	0.0588
CBL	0.295	0.0473
NRAS	0.417	0.0362
IDH2	0.379	0.0429
KRAS	0.202	0.0271
MLL_PTD	0.798	0.0247
ETV6	0.391	0.0216
NPM1	0.43	0.0112
TP53multi	1.18	0.071
FLT3	0.798	0.0108
Nres2	0.231	0.388

2.4 Risk Categories and Cutpoints

The final IPSS-M score determines the risk category based on the following cutpoints:

- **Very Low:** Score ≤ -1.5
- **Low:** Score > -1.5 and ≤ -0.5
- **Moderate Low:** Score > -0.5 and ≤ 0
- **Moderate High:** Score > 0 and ≤ 0.5
- **High:** Score > 0.5 and ≤ 1.5
- **Very High:** Score > 1.5

2.5 Survival Data

Approximate median survival outcomes associated with each IPSS-M risk category are provided in Table 2.

IPSS-M methodology based on (Bernard et al. 2022).

Table 2: IPSS-M Risk Categories and Associated Median Survival Outcomes

Risk Category	Leukemia-Free Survival (LFS)	Overall Survival (OS)
Very Low	9.7 years	10.6 years
Low	5.9 years	6.0 years
Moderate Low	4.5 years	4.6 years
Moderate High	2.3 years	2.8 years
High	1.5 years	1.7 years
Very High	0.76 years	1.0 years

Note: Survival times are approximate medians; ranges are provided in the source publication (Bernard *et al.* 2022).

3 IPSS-R Methodology for MDS

The IPSS-R refines the original IPSS by adjusting the weighting of prognostic variables and redefining cytogenetic risk groups.

3.1 Input Variables

The IPSS-R requires the following parameters:

- **Cytogenetics:** IPSS-R cytogenetic risk category (Very Good, Good, Intermediate, Poor, Very Poor).
- **Bone marrow blast percentage** (BM_BLAST, %).
- **Hemoglobin** (Hb, g/dL).
- **Platelet count** (PLT, Giga/L or 10^9/L).
- **Absolute Neutrophil Count** (ANC, Giga/L or 10^9/L).
- **Age** (Years, optional for age-adjusted IPSS-RA).

3.2 Scoring System

Scores are assigned based on predefined breakpoints for each variable:

- **Cytogenetics (`CYTOVEC`):** Very Good=0, Good=1, Intermediate=2, Poor=3, Very Poor=4.
- **BM Blasts (%):** $\leq 2\% = 0$; > 2 to $< 5\% = 1$; 5 to $10\% = 2$; $> 10\% = 3$.
- **Hemoglobin (g/dL):** $\geq 10 = 0$; 8 to $< 10 = 1$; $< 8 = 1.5$.
- **Platelets (Giga/L):** $\geq 100 = 0$; 50 to $< 100 = 0.5$; $< 50 = 1$.
- **ANC (Giga/L):** $\geq 0.8 = 0$; $< 0.8 = 0.5$.

3.3 Calculation and Categories

The IPSS-R score is the sum of the individual scores from the five variables above. The risk category is determined by the total score based on these cutpoints:

- **Very Low:** Score ≤ 1.5
- **Low:** Score > 1.5 and ≤ 3
- **Intermediate:** Score > 3 and ≤ 4.5
- **High:** Score > 4.5 and ≤ 6
- **Very High:** Score > 6

3.4 Age Adjustment (IPSS-RA)

If the patient's age is provided, an age-adjusted score (IPSS-RA) can be calculated using the formula:

$$\text{IPSS-RA} = \text{IPSS-R Score} + (\text{Age} - 70) \times (0.05 - \text{IPSS-R Score} \times 0.005)$$

The resulting IPSS-RA score is then categorized using the same cutpoints as the IPSS-R.

IPSS-R methodology based on (Greenberg et al. 2012).

4 ELN 2022 Risk Stratification for AML

The ELN 2022 guidelines provide a framework for risk stratification in adult AML based primarily on genetic features identified at diagnosis. This stratification guides treatment intensity and prognosis estimation.

4.1 Favorable Risk Markers

Cases are classified as Favorable Risk if *any* of the following are present, in the absence of any Adverse Risk markers and typically without FLT3-ITD:

- t(8;21)(q22;q22.1); *RUNX1::RUNX1T1*
- inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB::MYH11*
- Mutated *NPM1* in the absence of *FLT3*-ITD
- Biallelic mutated *CEBPA* (specifically, bZIP in-frame mutations)

(Note: Acute Promyelocytic Leukemia with t(15;17) is typically managed separately and not included in this general risk stratification).

4.2 Adverse Risk Markers

Cases are classified as Adverse Risk if *any* of the following markers are present, regardless of other findings:

- t(6;9)(p23;q34.1); *DEK::NUP214*
- t(9;22)(q34.1;q11.2); *BCR::ABL1*
- inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2, MECOM(EVI1)*
- -5 or del(5q)

- -7
- -17 or abnormal 17p
- Complex karyotype (3 unrelated chromosomal abnormalities)
- Monosomal karyotype
- Mutated *TP53* (typically biallelic or VAF 10%)
- Mutated *RUNX1*
- Mutated *ASXL1*
- Mutated *BCOR*
- Mutated *EZH2*
- Mutated *SF3B1*
- Mutated *SRSF2*
- Mutated *STAG2*
- Mutated *U2AF1*
- Mutated *ZRSR2*
- Other specific KMT2A rearrangements (excluding t(9;11))
- t(8;16)(p11;p13); KAT6A::CREBBP
- Hyperdiploid karyotype with 3 trisomies (without other structural abnormalities - can be considered adverse in some contexts, though ELN 2022 is less explicit than prior versions). *Note: Check primary source for precise placement.*

4.3 Intermediate Risk

Cases are classified as Intermediate Risk if they do not meet criteria for Favorable or Adverse risk. This includes:

- Mutated *NPM1* and *FLT3*-ITD positive.
- Wild-type *NPM1* and *FLT3*-ITD positive.
- Wild-type *NPM1* and *FLT3*-ITD negative (without adverse or favorable risk markers).
- t(9;11)(p21.3;q23.3); *MLLT3::KMT2A*.
- Other cytogenetic or molecular abnormalities not classified as Favorable or Adverse.

4.4 Classification Hierarchy

The ELN 2022 risk stratification follows a hierarchy:

1. Check for any Adverse Risk marker. If present, the risk is Adverse.

2. If no Adverse markers, check for any Favorable Risk marker. If present AND *FLT3*-ITD is negative, the risk is Favorable.

3. If no Adverse markers are present, but a Favorable marker is present alongside *FLT3*-ITD, the risk is Intermediate.

4. If no Adverse or Favorable markers are present (or only Favorable markers with *FLT3*-ITD), the risk is Intermediate. This includes cases with specific intermediate markers like t(9;11) or isolated *FLT3*-ITD.

Secondary/therapy-related AML is noted but does not override the genetic risk group.

4.5 Associated Survival Data

The ELN 2022 risk groups are associated with distinct prognoses:

- **Favorable Risk:** Median overall survival (OS) often not reached or exceeds 5 years (>60 months).
- **Intermediate Risk:** Median OS typically in the range of 16–24 months.
- **Adverse Risk:** Median OS generally poor, often around 8–10 months.

ELN 2022 risk stratification based on (Döhner et al. 2022).

5 Implementation Details

The methodologies described above are implemented in Python within our system. Key aspects include:

- **Modularity:** Separate functions handle preprocessing, core calculation (IPSS-M, IPSS-R, ELN), and data retrieval (survival data, explanations).
- **Data Handling:** Functions are designed to accept patient data as Python dictionaries. Preprocessing functions normalize variable names and formats (e.g., converting cytogenetic categories to numerical vectors, calculating derived variables like `TP53multi`).
- **Missing Data (IPSS-M):** The IPSS-M calculation explicitly handles missing genetic data by computing 'best', 'worst', and 'mean' scenarios, using defined default values or imputation methods (especially for residual genes `Nres2`).
- **Risk Categorization:** Utility functions map raw scores to final risk categories based on established cutpoints.
- **Output Options:** Calculations can optionally return detailed contributions of each variable to the final score (IPSS-M) or component scores (IPSS-R) for transparency and interpretation.
- **CLI and Integration:** The code includes a command-line interface for direct use and functions (‘prepare$_i$pssm$_i$nput‘, ‘*parse*$_f$*or*$_i$*pssm*‘)*designedtointegratewithupstreamdataparsingmodules*(*likeNLPoutput*).

References

References

[1] Arber, D.A., Orazi, A., Hasserjian, R.P., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M. and Vardiman, J.W. (2022). The International Consensus Classification of Myeloid Neoplasms and Acute Leukemia: integrating morphological, clinical, and genomic data. *Blood*, 140(11), pp.1200–1228. doi: 10.1182/blood.2022015850.

[2] Bernard, E., Tuechler, H., Greenberg, P.L., Hasserjian, R.P., Arber, D.A., Della Porta, M.G., Hosono, S., Mian, H.S., Naqvi, K., Fuchida, S. *et al.* (2022). Molecular International Prognostic Scoring System for Myelodysplastic Syndromes. *NEJM Evidence*, 1(7). doi: 10.1056/EVIDoa2200008.

[3] Döhner, H., Wei, A.H., Appelbaum, F.R., Craddock, C., DiNardo, C.D., Dombret, H., Ebert, B.L., Fenaux, P., Godley, L.A., Hasserjian, R.P. *et al.* (2022). Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood*, 140(12), pp.1345–1377. doi: 10.1182/blood.2022016867.

[4] Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Solé, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F. *et al.* (2012). Revised International Prognostic Scoring System for Myelodysplastic Syndromes. *Blood*, 120(12), pp.2454–2465. doi: 10.1182/blood-2012-03-420489.

[5] Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Merson, J.A., Burtt, N.P., Chavez, A. *et al.* (2014). Age-Related Clonal Haematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*, 371(26), pp.2488–2498. doi: 10.1056/NEJMoa1408617.

[6] World Health Organization Classification of Tumours Editorial Board (2022). *Haematolymphoid Tumours*. 5th edn. Lyon: International Agency for Research on Cancer (WHO classification of tumours, volume 11). ISBN: 978-928324518-7. Available at: https://tumourclassification.iarc.who.int/chapters/63 [Accessed 10 April 2025].