A systematic review of higher-risk myelodysplastic syndromes clinical trials to determine the benchmark of azacitidine and explore alternative endpoints for overall survival

Jacqueline S Garcia  
*Dana-Farber Cancer Institute*

Ronan T Swords  
*AbbVie Inc.*

Gail J Roboz  
*Cornell University*

Meagan A Jacoby  
*Washington University School of Medicine in St. Louis*

Guillermo Garcia-Manero  
*University of Texas MD Anderson Cancer Center*

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4

Part of the Medicine and Health Sciences Commons

Please let us know how this document benefits you.
A systematic review of higher-risk myelodysplastic syndromes clinical trials to determine the benchmark of azacitidine and explore alternative endpoints for overall survival

Jacqueline S. Garcia a,*, Ronan T. Swords b, Gail J. Roboz c, Meagan A. Jacoby d, Guillermo Garcia-Manero e, Wan-Jen Hong f, Xiaoqing Yang b, Ying Zhou b, Uwe Platzbecker g, David P. Steensma a, Johannes E. Wolff b, Pierre Fenaux h

a Dana-Farber Cancer Institute, Boston, MA, USA
b AbbVie Inc., North Chicago, IL, USA
c Weill Medical College of Cornell University and New York-Presbyterian Hospital, NY, NY, USA
d Washington University School of Medicine, St. Louis, MO, USA
e University of Texas, M.D. Anderson Cancer Center, Houston, TX, USA
f Genentech Inc., South San Francisco, CA, USA
g Leipzig University Hospital, Leipzig, Germany
h Hôpital Saint-Louis, Assistance Publique - Hôpitaux de Paris, Université de Paris, Université Paris, France

**Keywords:**
- Myelodysplastic syndromes
- Azacitidine
- Higher-Risk
- Hypomethylating agent
- Overall survival

**ABSTRACT**

The hypomethylating agent azacitidine can prolong overall survival (OS) in patients with higher risk-myelodysplastic syndromes (HR-MDS) compared to conventional regimens. However, outcomes differ largely between studies, making it challenging to determine the contribution of novel therapies added to azacitidine. Further, a discrepancy is seen between complete (CR) or partial (PR) response rates and OS improvement with azacitidine, making it challenging to rely on earlier endpoints than OS.

We conducted a systematic literature search and study-level systematic review of 237 clinical studies to better understand outcomes for HR-MDS patients treated with azacitidine. Pooled marrow CR was 9% (N = 2654; 95% CI: 6–13 %), CR rate was 17 % (N = 6943; 95% CI: 15–20 %), and median OS (mOS) was 18.6 months (N = 2820; 95% CI: 15.3–21.9). A weak correlation to mOS was detected with CR rate (207 patient cohorts, Pearson $r = 0.315; P < 0.0005$), and a much stronger correlation with median progression-free survival (mPFS) ($r=0.88, P = 3 \times 10^{-14}$). Six-months progression-free survival rates correlated with 1-year OS rates but were only infrequently reported (N = 41 patient cohorts) therefore not allowing a robust recommendation for a surrogate to the established OS endpoint. Larger patient numbers and patient-level data appear necessary, especially for designing future clinical trials using azacitidine combinations.

**1. Introduction**

Myelodysplastic syndromes (MDS) are clonal hematopoietic stem cell disorders characterized by ineffective bone marrow hematopoiesis leading to peripheral blood cytopenias and frequent progression to acute myeloid leukemia (AML) [1]. Prognosis in MDS is largely based on the International Prognostic Scoring System (IPSS) that includes cytopenias, blast percentage, and cytogenetic risk to characterize MDS [2]. The revised system (IPSS-R) increases the relative contribution of cytogenetics and defines five categories. Patients with “high”, “very high”, and part of “intermediate” (with >3.5 points) risk categories are considered to have higher-risk (HR)-MDS, comprising 43 % of all MDS cases [3]. Patients with HR-MDS have poor median overall survival (mOS) ranging from 0.8 to 3 years [3].

Hypomethylating agents (HMAs) including azacitidine [4–6] or decitabine [5] are recommended in MDS as low-intensity therapy to reduce disease-associated symptoms and the risk of disease progression and death, thereby improving both quantity and quality of life [7]. Azacitidine [4–6] is the only drug that has demonstrated improved survival compared to conventional care regimens (24.5 months vs 15.0

* Corresponding author.
E-mail address: Jacqueline_garcia@dfci.harvard.edu (J.S. Garcia).

https://doi.org/10.1016/j.leukres.2021.106555
Received 20 January 2021; Received in revised form 25 February 2021; Accepted 26 February 2021
Available online 2 March 2021
0145-2126/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
recorded both together and separately in the systematic review data. If subgroup analyses were reported with outcomes, they were number of events), 1-year PFS rate (rate, 95% CI), and mPFS (in months, clinical trial publications from January 1, 2009 through January 23, 2019. The start date was motivated by the intent to include azacitidine base. This resulted in overlapping sets of data (one line in the original systematic literature search was conducted in the multi-file data
base platform DIALOG. The search was limited to English language clinical trial publications from January 1, 2009 through January 23, 2019. The start date was motivated by the intent to include azacitidine data, five years after FDA approval, and the end date cutoff was when the search was completed. The databases used were Biosis Previews, Derwent Drug File, Embase, International Pharmaceutical Abstracts, Medline, and SciSearch. Keywords in the approach included: myelodysplastic syndrome, HR-MDS, LR-MDS, azacitidine, trials, and therapeutic outcome endpoints. Selection of the publications included in the analysis is described in the Supplementary Methods. 2. Methods 2.1. Systematic review Systematic review methodology followed the Cochrane Handbook with reporting according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The quantitative findings from separate but similar clinical studies were integrated to provide a numerical estimate of the overall treatment effect of interest, as has been done in other studies. A systematic literature search was conducted in the multi-file database platform DIALOG. The search was limited to English language clinical trial publications from January 1, 2009 through January 23, 2019. The start date was motivated by the intent to include azacitidine data, five years after FDA approval, and the end date cutoff was when the search was completed. The databases used were Biosis Previews, Derwent Drug File, Embase, International Pharmaceutical Abstracts, Medline, and SciSearch. Keywords in the approach included: myelodysplastic syndrome, HR-MDS, LR-MDS, azacitidine, trials, and therapeutic outcome endpoints. Selection of the publications included in the analysis is described in the Supplementary Methods. 2.2. Data extraction and data cleaning Data entry was performed using a predefined data dictionary. Data included 11 variables to identify the reference, 16 variables to describe the patient population (e.g., age), 7 variables to describe the treatment details (e.g., drugs and schedules), and 159 variables for outcomes. Among those variables were CR rate per IWG 2006 criteria, [23,24] (number of patients with CR, total number of patients, 95% CI), 1-year OS rate (with 95% CI), median overall survival (mOS, with 95% CI, number of events), 1-year PFS rate (rate, 95% CI), and mPFS (in months, 95% CI, number of events). Data entry aimed to extract the most granular level of data that was reported. In studies with multiple arms, every arm that was reported with an outcome was considered as one set of data. If subgroup analyses were reported with outcomes, they were recorded both together and separately in the systematic review database. This resulted in overlapping sets of data (one line in the original data spreadsheet), similar to bootstrapping of data subsets. Data cleaning was conducted as described in the Supplementary Methods. In total, 237 studies were identified in the final unselected dataset for subsequent statistical analysis. 2.3. Statistical analysis using the final unselected data set included 237 studies The first analysis used the entire data set for descriptive statistics, frequency of endpoints reported, average and standard deviations of quantitative variables. This included exploratory statistical tests to describe the relation of various endpoints to each other. Endpoints were then compared for their ability to detect validated MDS findings, which included: 1) treatment outcome has improved over the years; 2) younger patients have better outcomes than older patients; 3) HR-MDS has inferior outcomes compared to low-risk MDS; 4) relapsed and refractory disease has inferior outcomes compared to treatment-naive; and 5) azacitidine improves outcome when compared to supportive care. The exploratory analysis and comparison of endpoints was performed using Statistical Package of Social Studies (SPSS) v25.0. 2.4. Benchmark calculations for response rates and median OS by systematic review For calculating the benchmark response rates including CR and mCR, data were selected using the following criteria: a) the data set included a reported CR rate by IWG (or mCR rate, respectively); b) the patient cohort was treated with azacitidine monotherapy; and c) repetitive dataset describing the same patient populations were excluded. For calculating the benchmark of mOS using the “inverse standard error method” [25,26], the mOS had to be reported with 95% confidence interval. This resulted in different data sets for calculation of each of the calculated benchmarks. For studies with missing upper bound of 95% confidence interval, the variance was imputed using the method of Wan et al. [27]. The pooled CR rate and pooled mOS with 95% confidence intervals are reported as Forest plots. When heterogeneities across studies were not detected, pooled estimates were calculated by the fixed-effect method [28]. The fixed-effect model does not account for between-study variation. When substantial heterogeneity was identified, pooled estimates were calculated by the random-effect method. The random-effect model incorporates the between-study variations that exist with a strong signal. The between-study heterogeneity was tested and evaluated by Cochrane’s Q Statistics and $I^2$, with Cochrane’s $Q \leq 0.05$ indicating statistically significant between-study heterogeneity and $I^2$ values indicating: absence [0–25 %], low [25.1–50 %], moderate [50.1–75 %], or high [75.1–100 %] between-study heterogeneity [29]. Publication bias for the CR rates of this study in the HR-MDS cohorts was assessed, and a funnel plot was generated. 22 cohorts were analyzed for the pooled median OS. 2.5. Meta-Regression for CR and OS endpoints The correlation and potential surrogacy between CR and OS endpoints was explored by the weighted meta-regression. 28 cohorts were found for studies with both CR rate and median OS reported for HR-MDS. The weight was defined as the inverse of the standard error of the mOS [30,31]. A log-transformation of data was applied to fulfill the assumption of normal distribution. 2.6. Comparison of mOS by treatment A regression model was created using mOS as the dependent variable and 1-year OS as the independent variable: $\log(mOS) - \log(1/YOS)$. This model was used to compute the missing mOS. Weighted mean mOS was calculated using the formula: $\text{mean mOS} = \frac{\sum (\sqrt{n} \times mOS)}{\sum (\sqrt{n})}$, where $n$ is the number of patients in the cohort. One-way ANOVA was used to compare the mean mOS of three treatment groups. Weighted by square root of sample size was used in the ANOVA test, as previously described [31,32]. Statistical analyses were performed with R software version 3.4.1 (Vienna, Austria) and SAS version 9.4 (Cary, NC). A significance level of
0.05 was used for all analyses.

3. Results

3.1. Study selection and data extraction

A total of 725 unique publications reporting azacitidine treatment for HR-MDS were identified by the ProQuest Dialog literature search and citation analysis. After screening and ranking for relevance, the final unselected data set included 237 studies with analyzable data from 1493 patient cohorts with at least one reported outcome variable (Fig.S1 for data extraction). The final selected data set included a total of 10,119 HR-MDS patients treated with azacitidine monotherapy with at least one endpoint addressed in the analysis.

3.2. Endpoints reported in the statistical analysis using unselected data

We first evaluated the surrogate endpoints for correlation with the gold standard OS among unselected data. From a total of 1493 patient cohorts, azacitidine was the most commonly used drug (1146), followed by decitabine (210), and best supportive care (45). The most commonly reported response endpoints were defined using IWG 2006 criteria (632). Although ORR was the most commonly reported endpoint (838), this endpoint was the least standardized as hematologic improvement (HI) [24] and mCR were inconsistently included. Marrow CR was not captured in data sets that used IWG 2000 criteria. Other reported endpoints included mOS (660), 1-year OS rate (1Y-OS: 446), CR rate (438), 3-year OS rate (3Y-OS: 283), overall HI rate (249), HI rate platelets (93), rate of transformation to AML (60), median PFS time (mPFS: 56), 1-year PFS (1Y-PFS: 43), 6-months PFS (6M-PFS: 41), median time to AML transformation (32), 3-year PFS rate (3Y-PFS: 17), and 5-year PFS rate (5Y-PFS: 3). We asked if OS data was frequently paired with objective responses among these patient cohorts: mOS and ORR (335), and mOS and CR rate (207). Other pairs of data were less frequent, including those with PFS, highlighting the challenges of establishing potential surrogates.

3.3. Correlation among endpoints using unselected data

Among unselected data, we sought to identify a surrogate endpoint for mOS in HR-MDS and performed a Pearson’s correlation analysis (Pearson’s correlation coefficient “r”). We found CR rate and ORR to have a weak positive correlation [33] with mOS (CR rate r = 0.315, P < 0.0005; ORR r=0.239, P = 4 × 10^-5). ORR moderately correlated to 1Y-OS (Pearson correlation of r = 0.49) (Fig.1A). OS and PFS showed a strong positive correlation (Table 1), with a Pearson correlation coefficient of r = 0.88 for mOS and mPFS (P < 0.0005) (Fig.1B), and r = 0.80 for 1Y-OS and 6M-PFS (P < 0.0005). ORR rate correlated with 6M-PFS (r=0.88, P = 0.001) and less so did CR rate with 6M-PFS (r=0.62, P = 0.031). The correlations between HI categories and OS rate were weak to moderate (Pearson’s correlations coefficients between r=-0.48 and r=-0.17; Table 1). As expected, the median time to AML were strongly correlated with 1Y-OS and the rate of hematologic improvement (Pearson’s correlation coefficient of r = 0.79 and r = 0.83, respectively).

3.4. Detection of known MDS findings with potential surrogate endpoints

We next asked how well the endpoints of interest (mOS, 1Y-OS, mPFS, 6M-PFS, ORR, CR rate, and HI rate) could detect known MDS findings using the unselected data set. We first asked if any of these endpoints correlated with improvement in MDS outcomes over time. Boxplots used to visualize the improvement in CR over the years did not find any statistical correlation (Fig.S2). The only endpoints that detected improvement in MDS outcomes over the years were mPFS (P = 0.002) and 6M-PFS (P = 0.015). Next, we asked if any of these potential surrogate endpoints correlated with patient age and could show younger patients having better outcomes than older patients. Superior outcome of younger patients was detectable by both OS and PFS variables, as well as by CR rate (Table S1). Among unselected data, first-line treatment, relapsed or refractory (R/R) populations, and mixtures of various lines of therapy were described in 213, 137, and 501 individual records (lines of data entry), respectively. As R/R disease results in shorter outcomes among MDS patients, we tested the ability of our candidate endpoints to detect this; inferior outcome could only be confirmed when considering mOS (P = 0.004) or 6M-PFS (P = 0.02). Other endpoints had either marginal associations (1Y-OS) or even opposite trends (ORR) (Table S1). Last, we confirmed that mOS, 1Y-OS, mPFS, 6M-PFS, CR, and ORR endpoints were significantly better among cohorts treated with azacitidine therapy compared to cohorts that received best supportive care. Though the 6M-PFS rate was the only endpoint that reliably detected all of the above known MDS findings (Table S1), this endpoint was rarely reported (N = 41 cohorts), rendering it unsuitable for creating a benchmark in the context of this analysis (Table 1).

3.5. Study characteristics and population for benchmark calculations using selected data from systematic review

Characteristics of the final selected set of studies included in the systematic review are summarized in Table S2. From these 237 publications, 68 cohorts from 60 studies were identified with 6,943 HR-MDS patients treated with azacitidine monotherapy for inclusion in the benchmark calculation of CR rate. The final analysis cohort included HR-MDS patients that were either treatment naive or R/R. Reasons for not including a set of extracted data in the analysis were: 1) CR rate was not reported or 2) AML, CMMI, or LR-MDS patients were part of the reported cohort. The characteristics of studies used for benchmark analysis of response rate and median OS are presented in Table 2.

The pooled CR rate among all HR-MDS patient cohorts treated with
Table 1
Relation of endpoint variables to each other in unselected data.

<table>
<thead>
<tr>
<th>Pearson’s Correlation</th>
<th>1Y OS rate</th>
<th>2Y OS rate</th>
<th>3Y OS rate</th>
<th>5Y OS rate</th>
<th>mOS 6-mo PFS rate</th>
<th>1Y FFS rate</th>
<th>2Y FFS rate</th>
<th>3Y FFS rate</th>
<th>mPFS Time</th>
<th>% ORR</th>
<th>% CR</th>
<th>% HI-Platelets</th>
<th>% HI-Neutrophils</th>
<th>% HI-Hgb</th>
<th>% HI-overall</th>
<th>% Transformation to AML</th>
<th>median time to AML Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Y OS rate</td>
<td>1.00</td>
<td>.825**</td>
<td>.714**</td>
<td>.475**</td>
<td>.708**</td>
<td>.800**</td>
<td>.865**</td>
<td>.792**</td>
<td>.751**</td>
<td>.770**</td>
<td>.493**</td>
<td>.207*</td>
<td>.01</td>
<td>.22</td>
<td>.07</td>
<td>.04</td>
<td>.12</td>
</tr>
<tr>
<td>2Y OS rate</td>
<td>.825**</td>
<td>1.00</td>
<td>.910**</td>
<td>.805**</td>
<td>.770**</td>
<td>.802**</td>
<td>.890**</td>
<td>.805**</td>
<td>.598**</td>
<td>.749**</td>
<td>.394**</td>
<td>.253*</td>
<td>.18</td>
<td>.07</td>
<td>.03</td>
<td>.01</td>
<td>.49</td>
</tr>
<tr>
<td>3Y OS rate</td>
<td>.714**</td>
<td>.910**</td>
<td>1.00</td>
<td>.917**</td>
<td>.818**</td>
<td>.520</td>
<td>.805**</td>
<td>.806**</td>
<td>.906**</td>
<td>.815**</td>
<td>0.10</td>
<td>0.23</td>
<td>0.16</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>.08</td>
</tr>
<tr>
<td>5Y OS rate</td>
<td>.475**</td>
<td>.805**</td>
<td>.917**</td>
<td>1.00</td>
<td>.852**</td>
<td>.692**</td>
<td>.858**</td>
<td>.850**</td>
<td>.946**</td>
<td>.876**</td>
<td>.239**</td>
<td>.315**</td>
<td>0.07</td>
<td>0.21</td>
<td>0.07</td>
<td>0.17</td>
<td>.34</td>
</tr>
<tr>
<td>mOS 6-mo PFS rate</td>
<td>.708**</td>
<td>.770**</td>
<td>.818**</td>
<td>.852**</td>
<td>1.00</td>
<td>.692**</td>
<td>.858**</td>
<td>.850**</td>
<td>.946**</td>
<td>.876**</td>
<td>.239**</td>
<td>.315**</td>
<td>0.07</td>
<td>0.21</td>
<td>0.07</td>
<td>0.17</td>
<td>.34</td>
</tr>
<tr>
<td>1Y PFS rate</td>
<td>.800**</td>
<td>.802**</td>
<td>.520</td>
<td>.990</td>
<td>.692**</td>
<td>1.00</td>
<td>.836**</td>
<td>.754**</td>
<td>0.46</td>
<td>.729**</td>
<td>.875**</td>
<td>.620*</td>
<td>0.16</td>
<td>0.16</td>
<td>0.11</td>
<td>0.65*</td>
<td>0.02</td>
</tr>
<tr>
<td>2Y PFS rate</td>
<td>.865**</td>
<td>.890**</td>
<td>.805**</td>
<td>.806**</td>
<td>.852**</td>
<td>1.00</td>
<td>.823**</td>
<td>.786**</td>
<td>.870**</td>
<td>.621*</td>
<td>0.41</td>
<td>0.61</td>
<td>0.66</td>
<td>0.69</td>
<td>0.02</td>
<td>0.02</td>
<td>0.32</td>
</tr>
<tr>
<td>3Y PFS rate</td>
<td>.792**</td>
<td>.805**</td>
<td>.806**</td>
<td>.850**</td>
<td>.754**</td>
<td>.823**</td>
<td>1.00</td>
<td>.882**</td>
<td>.889**</td>
<td>0.16</td>
<td>0.65</td>
<td>0.69</td>
<td>0.65</td>
<td>0.69</td>
<td>0.02</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>4Y PFS rate</td>
<td>.751**</td>
<td>.598**</td>
<td>.906**</td>
<td>.946**</td>
<td>0.46</td>
<td>.786**</td>
<td>.882**</td>
<td>1.00</td>
<td>.943**</td>
<td>0.28</td>
<td>0.51</td>
<td>0.44</td>
<td>0.33</td>
<td>0.04</td>
<td>0.13</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>5Y PFS rate</td>
<td>.770**</td>
<td>.749**</td>
<td>.815**</td>
<td>.876**</td>
<td>.729**</td>
<td>.870**</td>
<td>.889**</td>
<td>.943**</td>
<td>1.00</td>
<td>.617**</td>
<td>0.44</td>
<td>0.33</td>
<td>0.03</td>
<td>0.04</td>
<td>0.13</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>% ORR</td>
<td>.493**</td>
<td>.394**</td>
<td>0.10</td>
<td>.12</td>
<td>.239**</td>
<td>.875**</td>
<td>.621**</td>
<td>0.16</td>
<td>0.28</td>
<td>.617**</td>
<td>1.00</td>
<td>.527**</td>
<td>.506**</td>
<td>.457**</td>
<td>0.13</td>
<td>.452**</td>
<td>0.01</td>
</tr>
<tr>
<td>% CR</td>
<td>.207**</td>
<td>.253**</td>
<td>0.23</td>
<td>0.04</td>
<td>.315**</td>
<td>.620**</td>
<td>0.41</td>
<td>0.65</td>
<td>0.51</td>
<td>.44</td>
<td>0.44</td>
<td>.527**</td>
<td>1.00</td>
<td>0.10</td>
<td>.437**</td>
<td>0.20</td>
<td>0.11</td>
</tr>
<tr>
<td>% HI-Platelets</td>
<td>0.01</td>
<td>0.18</td>
<td>0.16</td>
<td>0.07</td>
<td>0.07</td>
<td>0.16</td>
<td>0.61</td>
<td>0.33</td>
<td>0.10</td>
<td>1.00</td>
<td>0.706</td>
<td>.617**</td>
<td>.785**</td>
<td>0.54</td>
<td>0.830*</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>% HI-Neutrophils</td>
<td>-0.22</td>
<td>-0.07</td>
<td>0.01</td>
<td>-0.21</td>
<td>0.16</td>
<td>-0.66</td>
<td>-0.53</td>
<td>0.457**</td>
<td>-0.796**</td>
<td>1.00</td>
<td>0.372</td>
<td>.498**</td>
<td>1.00</td>
<td>0.04</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>% HI-Hgb</td>
<td>-0.14</td>
<td>-0.35</td>
<td>-0.48</td>
<td>0.07</td>
<td>-0.11</td>
<td>-0.69</td>
<td>-0.04</td>
<td>0.13</td>
<td>0.20</td>
<td>0.617**</td>
<td>0.372</td>
<td>.633**</td>
<td>1.00</td>
<td>0.02</td>
<td>0.22</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>% HI-overall</td>
<td>0.12</td>
<td>0.03</td>
<td>0.13</td>
<td>0.10</td>
<td>0.17</td>
<td>0.655*</td>
<td>0.02</td>
<td>0.32</td>
<td>-0.74</td>
<td>0.13</td>
<td>0.452</td>
<td>0.11</td>
<td>0.785**</td>
<td>0.498**</td>
<td>1.00</td>
<td>0.22</td>
<td>0.38</td>
</tr>
<tr>
<td>Transformation to AML</td>
<td>-0.41</td>
<td>-0.49</td>
<td>-0.48</td>
<td>-0.27</td>
<td>-0.34</td>
<td>0.29</td>
<td>0.01</td>
<td>-0.07</td>
<td>0.05</td>
<td>0.32</td>
<td>-0.44</td>
<td>0.22</td>
<td>0.07</td>
<td>0.38</td>
<td>0.22</td>
<td>1.00</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Abbreviations: AML, acute myeloid leukemia; CR, complete remission; Hgb, hemoglobin; HI, hematologic improvement; mPFS, median progression-free survival; mOS, median overall survival; ORR, overall remission rate; OS, overall survival; PFS, progression-free survival; Y, year.

Pearson’s Correlation Coefficient (r) is shown for each pair of endpoints. Significance values (two-tailed) are marked: ** P < 0.01 and * P < 0.05.
Table 2

<table>
<thead>
<tr>
<th>Total Cohorts, n (%)</th>
<th>Studies included in OS analysis</th>
<th>Studies included in CR analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Cohorts, n (%)</td>
<td>22 (100)</td>
<td>68 (100)</td>
</tr>
<tr>
<td>Publication Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002-2011</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>2012-2014</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>2015-2017</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>2018-2019</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Median Age, years (range)</td>
<td>71 (66.5–77)</td>
<td>72 (67–77)</td>
</tr>
<tr>
<td>% Male (range)</td>
<td>65.7 (44–74)</td>
<td>65 (44–84)</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azacitidine</td>
<td>22 (100)</td>
<td>68 (100)</td>
</tr>
<tr>
<td>CR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IWG Year 2000</td>
<td>2 (9)</td>
<td>5 (7)</td>
</tr>
<tr>
<td>IWG Year 2006</td>
<td>14 (64)</td>
<td>48 (71)</td>
</tr>
<tr>
<td>Unknown</td>
<td>6 (27)</td>
<td>15 (22)</td>
</tr>
<tr>
<td>% Patients with MDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR-MDS</td>
<td>22 (100)</td>
<td>68 (100)</td>
</tr>
</tbody>
</table>

Abbreviations: CR, complete remission; OS, overall survival; HR-MDS, higher-risk myelodysplastic syndromes; IWG, International Working Group.

azacitidine that met inclusion criteria in our systematic review was 17% (95% CI: 15–20 %) (Fig. 2A). The pooled mCR rate was 9% (95% CI: 6–13 %) (Fig.2B). To assess for inter-study heterogeneity among included cohorts, the Cochran’s Q P-values and I² for CR rate, marrow CR rate and mOS were <0.0001/<0.0001/<0.0001 and 75.6%/85.5%/95.1%, respectively. There was no indication of publication bias in this analysis (Fig.S3).

To establish the mOS benchmark we only included patient cohorts that reported mOS with a 95% confidence interval compared to those included to determine the CR rate (N = 22 vs 68 patient cohorts, respectively). The mOS was 18.6 months (95% CI: 15.3–21.9) for 2820 patients among 22 cohorts of HR-MDS patients treated with azacitidine only (Fig.3A). Fewer cohorts were included for mOS analysis compared to CR rate (N = 22 vs 68 patient cohorts, respectively) because most of the studies did not report mOS with a 95% confidence interval.

3.6. Meta-regression for CR and OS endpoints

The preliminary results from the meta-regression (Fig.4) of 28 cohorts indicate a moderate to strong positive correlation between CR rate (p value <0.0001) and median OS for HR-MDS, with adjusted R-Squared of 0.61. CR rate potentially is predictive of the median OS. Further analyses are needed to demonstrate the strong correlation in CR rate and OS in “treatment effect” to consolidate the conclusion of potential surrogacy.

3.7. Clinical outcomes and correlations among selected cohorts

We next repeated the analyses for surrogate endpoints using the same patient datasets for benchmark calculations. Clinical outcome was superior in cohorts that included azacitidine in the treatment regimen compared with cohorts that received best supportive care only. This was true regardless of the endpoint used, but the various endpoints differed in how clearly the findings appeared. The endpoints associated with greatest significance were mOS (18 months vs 12 months), 1-year OS (90 % vs 80 %), and mPFS (12 months vs 10 months) in azacitidine-treated versus best supported care treated patients. The CR rate among treatment-naive patients was similarly ~16 %. As treatment naive patients were only found in 9 cohorts, patients with R/R MDS were also included in the analysis.

We last explored mOS outcomes for patients treated with azacitidine monotherapy versus azacitidine-based combination regimens. In general, the outcomes of studies using azacitidine as monotherapy were superior to those from combination regimens containing azacitidine plus a second drug (monotherapy mOS 16.0 months vs combination regimens mOS 11.1 months, in a non-weighted analysis). Of note, we had excluded patients who received azacitidine-based treatment combinations when calculating the benchmark. We also performed a weighted analysis of mOS that included both HR-MDS and non-HR-MDS studies. The weighted average mOS was 15.7 months for patients treated with azacitidine monotherapy (90 cohorts), 11.7 months for patients treated with azacitidine combined with other drugs (6 cohorts), and 9.0 months for patients treated with best supportive care only (Fig.3B).

4. Discussion

Survival rates reported among HR-MDS patients who underwent azacitidine treatment are largely inconsistent across clinical studies. Using one of the largest contemporary sample sets of HR-MDS patients treated with azacitidine monotherapy reported to date, our analyses determined the pooled mOS was 18.6 months (N = 2820 patients; 95% CI: 15.3–21.9). Improved survival in large pivotal trials, such as AZA-001, compared to real-world evidence data is often attributed to patient selection bias and may account for the differences in mOS observed. These pooled mCR, CR, and OS data will serve as the benchmark and future comparator for HMA-based combinations under development for this disease. Of note, data from the ongoing prospective studies such as the randomized, double-blind, phase 3 VERONA trial comparing azacitidine plus placebo to azacitidine plus venetoclax (ClinicalTrials.gov number, NCT04401748) will likely add to this dataset.

Validating surrogate endpoints remains a challenge and has the potential to be misleading due to the actual effects of treatment interventions on a patient’s overall health [34,35]. Particularly in the case of higher-risk MDS, it is difficult to fully discriminate disease-related versus treatment-related effects, which may be better assessed through long-term trials. By definition, a robust surrogate endpoint requires data to demonstrate that the surrogate is prognostic for the true endpoint (overall survival) independent to treatment and that the effect of treatment (azacitidine) on the surrogate consistently predicts its effect on the true endpoint [36]. We found it difficult to formally conclude that either CR or PFS endpoints would be a strong surrogate for OS. Though clinically meaningful, a high rate of blast reduction does not always translate into survival improvement, as we have learned in other disease settings including for second-line MDS therapy with the rigosertib ONTIME study [37] and front-line elderly AML therapy with clofarabine vs low-dose cytarabine [38] as part of the AML16 trial. On the other hand, although CR remains the main treatment goal in elderly MDS patients treated with HMA, it may not be a prerequisite for survival benefit in MDS. Significantly improved OS was observed for patients with hematologic improvement who had never achieved complete or partial remission (hazard ratio 0.19 [95%CI: 0.08–0.46], P = 0.001) in sensitivity analyses from the AZA-001 trial. [39]. Achievement of stable disease in both azacitidine-treated and conventional care-treated patients in AZA-001 was also associated with a significantly reduced risk of death (hazard ratio 0.09, [95%CI: 0.06–0.15]; P < 0.001). Additional caveats for considering CR as an early surrogate endpoint includes variability in time to achieve CR among studies and achievement of CR, by definition, requires a 4-week confirmation period. In clinical practice the treatment might be effective, yet the following treatment cycle might start before the criteria for neutrophil recovery have been met, thus never reaching a point to document CR. Nonetheless, azacitidine was originally approved in the US based upon response data. CR may have value as an alternative endpoint for OS. The CR rate calculated in this systematic review was remarkably similar to the CR rate in the pivotal azacitidine study [4]. Earlier reports described lower CR rates likely due to the stricter hematologic parameters defining CR [40]. More recent literature is consistent with IWG 2006 criteria, which makes CR rate a useful benchmark. The CR definition has the further advantage over other response definitions as it is determined independent of
baseline hematologic values. CR rate was more sensitive in detecting the clinical efficacy of azacitidine as compared to mOS (\(P = 3 \times 10^{-12}\) for CR rate, and \(P = 0.005\) for mOS), although the magnitude of the Pearson’s correlation coefficient with mOS was weak. This suggests that further data are necessary before defining CR rate as an acceptable early surrogate endpoint for OS.

Six month-PFS and mPFS appeared to be potentially interesting relatively early endpoints (Table S1) and correlated well with OS endpoints (\(r = 0.88, P = 3 \times 10^{-14}\). N = 42 patient cohorts for mPFS); but these were rarely reported, preventing strong conclusions. In studies with control arms, PFS can be used as continuous parameter without the need of a single time point benchmark. As most treatment failures are due to lack of remission within 1 year after treatment initiation, PFS can be assessed much faster than OS. As decreasing the rate of transformation to AML is the treatment goal of HMA therapy for HR-MDS patients, delaying disease progression is a valuable and clinically meaningful endpoint. Unlike OS, PFS is not influenced by therapy given after initial treatment failure and so may provide a better assessment of benefit from a monotherapy when evaluating for improvement from a combination regime. Nonetheless, our analysis suggests that the 6M-PFS endpoint should be incorporated into prospective trials, as it may prove to be a surrogate early endpoint for survival. If this is true, then this would translate into a rapid read out of outcome compared to OS and potentially more opportunities for patients to access newer therapies. A future challenge also includes determining if surrogate endpoints for OS are also valid for those who undergo transplantation.

Our systematic review has strengths and weaknesses. The large volume of data analyzed was a strength of this study, and inter-study heterogeneity was taken into account. However, as these analyses include aggregate data rather than individual patient-level data, missing a correlation does not necessarily mean that the data sets are unrelated. When determining the usefulness of an endpoint in comparing outcomes of patient cohorts, the cohorts need to differ in that variable. Another limitation is that the patient population studied may include those with...
MDS and low-blast count AML (20–30 %) and those who were treatment naïve and HMA refractory. The question of bridging therapy and its necessity remains under debate [41–43]. Within this analysis, it was not possible to separate patients who received transplant from those who did not. HMA therapy is commonly used to prevent AML progression while waiting for transplantation, which often takes weeks to months to organize (5.7 months was the median time from diagnosis to transplant) [44]. A prospective phase 3 study (VidazaAllo) from the German MDS and Cooperative Transplant Study Group suggested that many elderly HR-MDS patients (55–70 years) who were candidates for allogeneic stem cell transplantation (allo–HCT) never received the graft after HMA. Induction therapy with azacitidine prior to allo–HCT was associated with a considerable drop out rate (33 %) due to disease progression, adverse events, and mortality prior to allo–HCT [45]. Although the role of cytoreduction prior to transplantation remains in question, univariate analysis from registry-based data including patients with MDS who underwent allogeneic transplantation demonstrates the presence of 5–9 % excess blasts pre-transplant does not clearly impact survival though >10 % blasts portends shorter overall [46]. In addition, the recent availability of an oral decitabine/cedazuridine (ASTX727) may influence physician/patient preference of decitabine versus azacitidine as first-line therapy for MDS. Further follow up will be needed to determine if this oral alternative becomes the de facto standard of care due to ease of administration.

In summary, in a systematic review of patients with HR-MDS treated with azacitidine, PFS may be a potential surrogate for OS but additional studies are required to confirm this finding. As aggregate data allows evaluation of interstudy variation, this study however generated useful clinical benchmarks for planning future studies. To hasten drug approvals for successful azacitidine-based combination therapies, future...
studies must clearly capture well-defined outcomes including response, type of response, 6M-PFS, and median PFS to validate them as potential surrogate markers for OS.

CRediT authorship contribution statement

JE Wolff: conceived the study. RT Swords and JE Wolff: collected the data. X Yang and Y Zhou performed statistical analysis. All authors: interpreted the data; all authors wrote the manuscript.

Role of the funding source

AbbVie Inc. provided financial support for this study and participated in the design, study conduct, analysis and interpretation of the data, as well as the writing, review, and approval of this manuscript. No honoraria or payments were made for authorship.

Declaration of Competing Interest

JS Garcia: Research support from AbbVie. Clinical trial support (institutional) from AbbVie, Genentech, Lilly, and Pfizer. Serves on SAB for AbbVie.

GJ Roboz: Consultancy or Advisory Board or Data and Safety Monitoring Committee: AbbVie, Actinium, Agios, Amphilvena, Argenx, Array Biopharma, Astex, Astellas, AstraZeneca, Bayer, Celgene, Celltrion, Daiichi Sankyo, Eisai, Epizyme, E succeeding, Janssen, Janssen, Jaspier Therapeutics, Jazz, MEI Pharma – IDMC Chair, Novartis, Orixen, Otsuka, Pfizer, Roche/Genentech, Sandzol, Takeda – IRC Chair, Trovagen. Research Support: Cellectis and AbbVie.

M Jacoby: Clinical trial funding from Jazz Pharmaceuticals. Research support from AbbVie. Co-investigator (non-funded) on an AbbVie clinical trial.

G Garcia-Manero: Research support from AbbVie.

U Platzecker: Data monitoring committee: Onconova, Takeda, Novartis, Celgene/BMS. Consulting: Celgene/BMS, Novartis, BMS, Daiichi Sankyo, Agios. Research support from AbbVie.


P Fenaux: Research support, as GFM chairperson, from AbbVie, Celgene, Janssen, Novartis.

WJ Hong: Genentech/Roche employee and may own Roche stock or stock options.

Y Zhou, JE Wolff: AbbVie employees and may own stock.

X Yang, RT Swords: Former AbbVie employee and may own stock.

Acknowledgements

We thank Marceline S. Corbin and William A. Bryce (AbbVie Library Information Service) for their contributions to the manuscript. Medical writing support was provided by Ana Mrejeru, Ph.D. (AbbVie employee).

Appendix A. Supplementary data

Supplementary material related to this article can be found in the online version, at doi:https://doi.org/10.1016/j.leukres.2021.106555.

References


