ONLINE FIRST

This is a provisional PDF only. Copyedited and fully formatted version will be made available soon.

REPORTS OF PRACTICAL ONCOLOGY AND RADIOTHERAPY

**ISSN:** 1507-1367

**e-ISSN:** 2083-4640

# A cross-sectional study using the Surveillance, Epidemiology, and End Results Program database to estimate the prevalence of Myelodysplastic syndrome (MDS) in the United States

**Authors**: Shangyi Fu, Michel Adeniran, Zachrieh Alhaj, Diana Bonilla, Melissa Marchan-Martinez, Ibeth Caceres, Danny Huynh

**DOI:** 10.5603/rpor.103137

**Article type:** Letter to the Editor

**Published online:** 2024-11-05

# A cross-sectional study using the Surveillance, Epidemiology, and End Results Program database to estimate the prevalence of Myelodysplastic syndrome (MDS) in the United States

**DOI:** 10.5603/rpor.103137

Shangyi Fu[1, 2], Michel Adeniran[3], Zachrieh Alhaj[3], Diana Bonila[3], Melissa Marchan-Martinez[3], Ibeth Caceres[1], Danny Huynh[4]

[1]*School of Medicine, Baylor College of Medicine, Houston, TX, United States*
[2]*Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, United States*
[3]*John Sealy School of Medicine, The University of Texas Medical Branch, Galveston, TX, United States*
[4]*Cullen College of Engineering, University of Houston, Houston, TX, United States*

**Corresponding author:** Zachrieh Alhaj BA, BS, 301 University Blvd, Galveston, TX, 77555, United States; e-mail: ztalhaj@utmb.edu

## Abstract

Myelodysplastic syndrome (MDS) is a hematologic disorder characterized by ineffective blood cell production leading to cytopenias and a propensity for progression to acute myeloid leukemia (AML). In this study, we aimed to assess the prevalence of MDS in the United States population using data from the Surveillance, Epidemiology, and End Results (SEER) database. Our analysis revealed that the highest prevalence of MDS was observed in the over 85 age group, with an overall prevalence rate of 0.02%. Furthermore, our findings indicated that the White race had the highest prevalence rate compared to other racial groups. However, it should be noted that the SEER database, most likely due to sampling biases, has a potential underestimation of minority populations, but SEER and the US population are statistically similar enough for comparison. These results suggest a need for further research on the underlying factors contributing to the prevalence of MDS in the United States population; factors such as genetic, environmental,

comorbidities, and racial disparities should be explored. Early diagnosis and treatment of MDS are crucial for improving outcomes for patients with this condition.
**Key words:** cancer; cancer epidemiology

To the Editor

Myelodysplastic syndrome (MDS), also referred to as preleukemia, is a group of disorders of the bone marrow cells that results in abnormal hematopoiesis [1]. This morphologic dysplasia of hematopoietic cells results in peripheral cytopenias [2]. In this study, we aimed to estimate the prevalence of primary MDS by extracting data from the Surveillance, Epidemiology, and End Results (SEER) Program database corresponding with the following ICD-O-3 histology codes: 997, 998, and 999 and the following histology/behavior codes: 9975/3, 9986/3, 9989/3, 9993/3. SEER is a database supported by the Surveillance Research Program under the National Cancer Institute's (NCI) Division of Cancer Control and Population Sciences (DCCPS) that provides cancer statistics information among the U.S. population. Its goal is to reduce the cancer burden among U.S. populations. Estimating the prevalence of MDS is important for improving healthcare planning, understanding disease burden, and guiding early intervention strategies [3].

The population estimates in this study were done using the 2017 and 2018 SEER data, and we analyzed 26 years of data ending with January 1$^{st}$, 2018. At the date of data collection, the SEER database had 43,926,825 participants enrolled (Tab. 1). We estimated an overall count of 5,746.7 affected with MDS. The prevalence was estimated to be highest in the 85+ age group, showing a pattern of increasing with age. The prevalence was also estimated to be highest in White individuals at 0.03%. In Black, American Indian/Alaskan Native, and Asian or Pacific Islander populations, the prevalence was estimated to be 0.02% (Tab. 1). Moreover, it is important to note that the SEER database is composed of 71% White, 12% Black, 2% American-Indian, and 15% Asian-Pacific Islanders [4], which differs slightly from the demographic distribution of the United States, which is 76% White, 14% Black, 1% American-Indian, and 7% Asian [5]. Therefore, our calculation of the prevalence of AML may underestimate the White and Black

populations and overestimate the Hispanic and Asian populations. Moreover, there may be a significant number of unaccounted patients due to factors such as their residency status, limited access to healthcare, and census limitations. These limitations may affect the study's findings because they would not provide an accurate estimate of the perveance of MDS. The results of a chi-squared test of independence indicated that there was no significant difference between the SEER and the US Census populations ($X2$ (3, N = 304167848) = 3404209.8855, $p < 0.00001$) (Tab. 2). This implies that the two populations are statistically comparable, allowing for an estimation of the USA population using SEER.

Our data suggests that MDS is equally prevalent across various ethnic groups. It also suggests that its prevalence increases with age. Despite this, there were still some limitations, mentioned earlier, that led to several unaccounted patients. To improve these limitations, we suggest the use of advanced statistical methods, advocating for policy change, and forming a community based research program that allows diverse communities to participate in research studies and improves data collection. Altogether, with its prevalence in all races, we recommend increased awareness and education regarding MDS in minority patients. In addition, we suggest an increase in the development of screening programs and policies that help target the diverse population of MDS patients. We also encourage an increase in focus on barriers to health care access to allow for the diverse population of MDS patients to have equitable healthcare. It would be beneficial to conduct further epidemiological studies that are not limited by billing codes to validate our findings.

**Conflict of interests**

Authors declare no conflict of interests.

**Funding**

None declared.

**References**

1. Hasserjian RP. Myelodysplastic Syndrome Updated. Pathobiology. 2019; 86(1): 7–13, doi: 10.1159/000489702, indexed in Pubmed: 30041243.

2. Garcia-Manero G. Myelodysplastic syndromes: 2023 update on diagnosis, risk-stratification, and management. Am J Hematol. 2023; 98(8): 1307–1325, doi: 10.1002/ajh.26984, indexed in Pubmed: 37288607.
3. Gondek LP, DeZern AE. Assessing clonal haematopoiesis: clinical burdens and benefits of diagnosing myelodysplastic syndrome precursor states. Lancet Haematol. 2020; 7(1): e73–e81, doi: 10.1016/S2352-3026(19)30211-X, indexed in Pubmed: 31810765.
4. SEER Cancer Statistics Review, 1975–2018 [Internet]. https://seer.cancer.gov/csr/1975_2018/index.html (2023 Mar 2).
5. Bureau UC. 2018 Population Estimates by Age, Sex, Race and Hispanic Origin [Internet]. https://www.census.gov/newsroom/press-kits/2019/detailed-estimates.html (2023 Mar 30).

**Table 1.** The prevalence of myelodysplastic syndrome in the United States, stratified by age and race

| Group | Estimated prevalence percent | Estimated prevalence count | Population at prevalence date | Known alive | Lost | Lost estimated alive | Dead prior to prevalence date |
|---|---|---|---|---|---|---|---|
| White | 0.03% | 4455.4 | 31240899 | 4335 | 182 | 120.4 | 13873 |
| Black | 0.02% | 499.1 | 5224726 | 484 | 21 | 15.1 | 1125 |
| American Indian/Alaska Native | 0.02% | 47 | 897021.5 | 46 | 1 | 1 | 106 |
| Asian or Pacific Islander | 0.02% | 661.9 | 6564178.5 | 628 | 60 | 33.9 | 1643 |
| Unknown | | 83.3 | 0 | 60 | 30 | 23.3 | 40 |
| 00 years at prev date | 0.00% | 1 | 522150 | 1 | 0 | 0 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 01–04 years at prev date | 0.00% | 6 | 2159900.5 | 6 | 0 | 0 | 0 |
| 05–09 years at prev date | 0.00% | 10.5 | 2732193.5 | 7 | 4 | 3.5 | 3 |
| 10–14 years at prev date | 0.00% | 20.6 | 2786556.5 | 17 | 4 | 3.6 | 4 |
| 15–19 years at prev date | 0.00% | 30.5 | 2781405 | 28 | 3 | 2.5 | 15 |
| 20–24 years at prev date | 0.00% | 25.9 | 2954646.5 | 25 | 1 | 0.9 | 15 |
| 25–29 years at prev date | 0.00% | 30.5 | 3391241 | 29 | 2 | 1.5 | 18 |
| 30–34 years at prev date | 0.00% | 39.4 | 3191280.5 | 36 | 4 | 3.4 | 15 |
| 35–39 years at prev date | 0.00% | 51.8 | 3049214.5 | 46 | 7 | 5.8 | 23 |
| 40–44 years at prev date | 0.00% | 57.3 | 2779268.5 | 52 | 6 | 5.3 | 26 |
| 45–49 years at prev date | 0.00% | 117.6 | 2889978.5 | 111 | 7 | 6.6 | 65 |
| 50–54 years at prev date | 0.01% | 172.1 | 2837532.5 | 161 | 13 | 11.1 | 119 |
| 55–59 years at prev date | 0.01% | 272.7 | 2873440 | 258 | 18 | 14.7 | 226 |
| 60–64 years at prev date | 0.01% | 422.1 | 2591284.5 | 405 | 19 | 17.1 | 417 |
| 65–69 years at prev date | 0.03% | 667.3 | 2124581 | 645 | 30 | 22.3 | 695 |
| 70–74 years at prev date | 0.04% | 866 | 1598623.5 | 838 | 38 | 28 | 1114 |
| 75–79 years at prev date | 0.06% | 824.1 | 1077416 | 803 | 34 | 21.1 | 1633 |
| 80–84 years at prev date | 0.09% | 836 | 736037 | 821 | 31 | 15 | 2177 |
| 85+ years at prev date | 0.13% | 1295.3 | 850075.5 | 1264 | 73 | 31.3 | 10222 |

**Table 2.** A Chi-square test of independence was performed using 2018 estimated population data comparing both the Surveillance, Epidemiology, and End Results (SEER) and the United States. The observed population was recorded in each cell, while the expected population was indicated in parentheses. The individual chi statistics are presented in brackets. The overall chi-square

statistic was computed as 3404209.8855, and the p-value was found to be less than 0.00001 at a significance level of less than 0.05

| | **Estimated SEER Population in 2018** | **Estimated USA Population in 2018** | | | ***Row Totals*** |
|---|---|---|---|---|---|
| White | 31240899 (33049304.38) [98953.07] | 197606407 (195798001.62) [16702.57] | | | 228847306 |
| Black | 5224726 (6661487.98) [309883.47] | 40902223 (39465461.02) [52306.12] | | | 46126949 |
| American-Indian | 897022 (478652.71) [365678.20] | 2417371 (2835740.29) [61723.87] | | | 3314393 |
| Asian-Pacific Islander | 6564179 (3737380.93) [2138071.41] | 19315021 (22141819.07) [360891.18] | | | 25879200 |
| **Column Totals** | 43926826 | 2.6E+8 | | | **304167848 (Grand Total)** |