

The Danish Myelodysplastic Syndromes Database: Patient Characteristics and Validity of Data Records

Tine Bichel Lauritsen¹
 Jan Maxwell Nørgaard¹
 Kirsten Grønbæk^{2,4}
 Anders Pommer Vallentin⁵
 Syed Azhar Ahmad⁶
 Louise Hur Hannig⁷
 Marianne Tang Severinsen^{8,9}
 Kasper Adelborg^{10,11}
 Lene Sofie Granfeldt Østgård^{11,12}

¹Department of Hematology, Aarhus University Hospital, Aarhus, Denmark; ²Department of Hematology, Rigshospitalet, Copenhagen, Denmark; ³Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark; ⁴Novo Nordisk Foundation Center for Stem Cell Biology (DanStem), Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark; ⁵Zealand University Hospital, Roskilde, Denmark; ⁶Department of Hematology, Herlev Hospital, Herlev, Denmark; ⁷Department of Hematology, Vejle Hospital, Vejle, Denmark; ⁸Department of Hematology, Clinical Cancer Research Center, Aalborg University Hospital, Aalborg, Denmark; ⁹Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; ¹⁰Department of Clinical Biochemistry, Aarhus University Hospital, Aarhus, Denmark; ¹¹Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; ¹²Department of Hematology, Odense University Hospital, Odense, Denmark

Correspondence: Tine Bichel Lauritsen
 Department of Hematology, Aarhus University Hospital, Palle Juul-Jensens Boulevard 99, Aarhus N, 8200
 Email tinelaur@rm.dk

Background: The Danish Myelodysplastic Syndromes Database (DMDS) comprises nearly all patients diagnosed with myelodysplastic syndromes (MDS) in Denmark since 2010. The DMDS has not yet been used for epidemiological research and the quality of registered variables remains to be investigated.

Objective: To describe characteristics of the patients registered in the DMDS and to calculate predictive values and the proportion of missing values of registered data records.

Methods: We performed a nationwide cross-sectional validation study of recorded disease and treatment data on MDS patients during 2010–2019. Patient characteristics and the proportion of missing values were tabulated. A random sample of 12% was drawn to calculate predictive values with 95% confidence intervals (CIs) of 48 variables using information from medical records as a reference standard.

Results: Overall, 2284 patients were identified (median age: 76 years, men 62%). Of these, 10% had therapy-related MDS, and 6% had an antecedent hematological disease. Hemoglobin level was less than 6.2 mmol/L for 59% of patients. Within the first two years of treatment, 59% received transfusions, 35% received erythropoiesis-stimulating agents, and 15% were treated with a hypomethylating agent. For the majority of variables (around 80%), there were no missing data. A total of 260 medical records were available for validation. The positive predictive value of the MDS diagnosis was 92% (95% CI: 88–95). Predictive values ranged from 64% to 100% and exceeded 90% for 36 out of 48 variables. Stratification by year of diagnosis suggested that the positive predictive value of the MDS diagnosis improved from 88% before 2015 to 95% after.

Conclusion: In this study, there was a high accuracy of recorded data and a low proportion of missing data. Thus, the DMDS serves as a valuable data source for future epidemiological studies on MDS.

Keywords: myelodysplastic syndromes, cohort, validation, accuracy, database

Introduction

Myelodysplastic syndromes (MDS) encompasses a heterogeneous group of myeloid neoplasms characterized by inefficient hematopoiesis, morphologic dysplasia, and variable degrees of cytopenias.^{1,2} MDS can be indolent or rapidly progressive, and may transform into secondary acute myeloid leukemia (sAML).^{1,2} The annual age-adjusted incidence is approximately 4 per 100,000 persons, increasing markedly with advancing age.^{3–5} Despite improvements in treatment and increasingly use of allogeneic bone marrow transplantation (allo-HSCT), the five-year overall survival remains around 30%.⁵

Denmark is known worldwide for its many high quality population-based health-care registries.⁶ The Danish Myelodysplastic Syndromes Database (DMDS) was launched in 2010 as a nationwide population-based database, currently covering at least 98% of all Danish MDS patients.⁷ It contains detailed clinical data on MDS subtypes, comorbidity, laboratory records, cytogenetic alterations, treatment and progression to acute myeloid leukemia (sAML). Through linkage to the extensive network of Danish health-care registries, the DMDS holds several epidemiological research possibilities. The use of data may add to the understanding of the epidemiology of MDS, including time trends in incidence and survival, clinical course, and effectiveness and safety of different treatment modalities.

Given that validation of a data source is a prerequisite prior to its use in epidemiological research, the objective of this study was to describe patient characteristics, the proportion of missing values, and the validity of data records in the DMDS.

Methods

Setting

The Danish health-care system is universal and tax-supported, providing free of charge and equal access to primary and hospital care for all Danish citizens.⁶ Patients with MDS are treated with chemotherapy at six specialized hematological departments in Denmark and four hematological departments treat MDS patients with best supportive care (transfusions, growth-factors, and antibiotics). Two hospital departments are accredited to perform allo-HSCT. Cytogenetic and mutational analyses are performed at three laboratories in Denmark and karyotypes are documented according to the International System for Human Cytogenetic Nomenclature (ISCN).⁸

The Danish Myelodysplastic Syndrome Database

The DMDS was founded by the Danish Society of Hematology as a nationwide clinical quality database with prospectively collected data on patients with incident MDS starting from January 2010. The main aim of the DMDS is to register all cases of MDS, including baseline characteristics, and treatment information to ensure comparable diagnostic work-up and treatment of MDS in Denmark. The database can also be used for research purposes.⁹

Data are submitted to the DMDS through an online registration system. Up to five standardized registration forms are used to collect data for the DMDS (Table 1). The first form is completed at diagnosis. The second and third forms are completed two and five years after diagnosis or earlier in case of progression to sAML, death, or termination of outpatient follow-up. A follow-up form is completed at death or termination of outpatient follow-up. The result of the cytogenetic evaluation is currently reported according to the Revised International Prognostic Scoring System (IPSS-R) by a clinical cytogeneticist at a separate registration form at diagnosis. Prior cytogenetic results reported according to the International Prognostic Scoring System (IPSS) were revised and updated according to IPSS-R by clinical cytogeneticists in 2020. IPSS-R data are, however, currently not available to researchers, and therefore they are not validated in this study, but IPSS characteristics are shown in Table 2.

Two Danish health-care registries are of central importance to the DMDS. The Civil Registration System provides data about vital status, date of death, and residential area.¹⁰ The Danish National Patient Registry (DNPR) provides data on primary and secondary discharge diagnoses (eg MDS ICD-10 codes) from all Danish hospitals since 1977.¹¹ Complete and valid linkage between these registries and the DMDS is done via the unique central personal registry number assigned to all Danish citizens.¹⁰

To ensure that the DMDS captures all newly diagnosed MDS patients, lists of patients with a newly registered MDS ICD-10 code according to the DNPR are provided for all hematological departments to confirm the diagnosis. If the diagnosis is correct, the patient is recorded in the DMDS and the registration form at diagnosis is completed. To ensure inclusion of incident MDS patients only, a look back period to 2001 is applied. Reminders to fill out the registration forms including 2-year data, 5-year data, follow-up-data and cytogenetic details, are also sent to the clinical departments. It is mandatory for all clinical departments to register MDS patients in the DMDS and according to Danish law patient consent is not required.

Patient Population and Study Design

We performed a nationwide cross-sectional validation study. We identified all patients with a diagnosis of MDS registered in the DMDS from January 1st 2010 until October 1st 2019. Then, a random validation sample of 12% of all patients diagnosed with MDS over a 10-year

Table I Data Recorded on Five Registration Forms Used by the Danish Myelodysplastic Syndromes Database

Registration Form and Time of Registration	Variables
At diagnosis To be filled out at diagnosis	Name, civil registration number, and demographic data Diagnosis according to WHO(2008/2016)/ICD-10 Date of diagnosis Prior hematological disease, prior chemotherapy or prior radiotherapy (yes/no/unknown) Results of cytogenetic evaluation (grouped according to IPSS-R) Cytopenia Anemia (< 6.2 mmol/l (yes/no)** Thrombocytopenia (< 100 x10 ⁹ /L) (yes/no)** Neutropenia (< 1.8 x10 ⁹ /L) (yes/no)** Blast percentage in blood and bone marrow Treatment with blood transfusion prior to diagnosis (yes/no/unknown) Presence of material in biobank (yes/no/unknown) Laboratory records*
Cytogenetic registration form To be filled out at diagnosis by a clinical Cytogeneticist	Date of diagnostic bone marrow sample Diagnoses according to WHO (2008/2016)/ICD-10 Cytogenetic result (normal, clonal changes, not done, no mitoses) Cytogenetic material (peripheral blood, bone marrow, others) Number of analyzed metaphases, abnormal metaphases and clonal changes ISCN karyotype IPSS-R
2-year data and 5-year data Registration form to be filled out two and five Years after diagnosis or in case of progression to AML, death or termination of outpatient follow-Up	Diagnosis according to WHO (2008/2016)/ICD-10 Date of follow-up Treatment Transfusion (yes/no)** Erythropoiesis-stimulating agents (yes/no)** Granulocyte-colony-stimulating-factors (yes/no)** Iron chelating agents (yes/no)** Hydroxyurea (yes/no)** Hypomethylating agents (Azacitidine) (yes/no)** AML-like chemotherapy (yes/no)** Immunosuppressive treatment (yes/no)** Lenalidomide (yes/no)** Other treatment (yes/no)** No treatment (yes/no)** Progression to acute myeloid leukemia (yes/no). If yes, date of progression Participation in protocol (yes/no) Bone marrow transplantation (yes/no). If yes, date and type
Follow up form At death or termination of follow-up as an Out-patient	Status (dead/alive) Date of death or end of follow-up Did the patient die at the hospital (yes/no) Was follow-up terminated from the department (yes/no) Referred to follow-up at another department (yes/no)

Notes: *Middle cell volume, hemoglobin, platelet count, white blood cell count, monocyte count, lactate dehydrogenase, ferritin, urate, erythropoietin, CD 34 positive cells, myoblasts, packed cell volume. **In order to register a "yes" you have to tick a "box". If nothing is ticked off a "no" is automatically registered.

Abbreviations: AML, acute myeloid leukemia; ICD, International Classification of Diseases; ISCN, International System for Human Cytogenetic Nomenclature; IPSS-R, International Prognostic Scoring System; WHO, World Health Organization.

Table 2 Accepted Values and Gold Standards for the Validation of 48 Variables from the Danish Myelodysplastic Syndromes Database*

Variables	Correct Specification of Registered Variables	Gold Standards
Diagnosis	MDS unclassified or specification of MDS subtype	Patobank/Medical records
Date of diagnosis	Date of diagnostic bone marrow sample (± 14 days)	Patobank
Prior hematological disease (>3months)	Yes/no/unknown	Medical records
Prior chemotherapy	Yes/no/unknown	Medical records
Prior radiotherapy	Yes/no/unknown	Medical records
Anemia (< 6.2 mmol/L)	Yes/no **	Laboratory information system
Thrombocytopenia ($< 100 \times 10^9/L$)	Yes/no **	Laboratory information system
Neutropenia ($< 1.8 \times 10^9/L$)	Yes/no **	Laboratory information system
Blast percentage in bone marrow	Exact value	Patobank/Medical records
Blood transfusion prior to diagnosis	Yes/no/unknown	Medical records
MCV (fL)	***	Laboratory information system
Hemoglobin (mmol/L)	***	Laboratory information system
Platelet count ($10^9/L$)	***	Laboratory information system
White blood cell count ($10^9/L$)	***	Laboratory information system
Leucocyte count ($10^9/L$)	***	Laboratory information system
Monocyte count ($10^9/L$)	***	Laboratory information system
Lactate dehydrogenase (U/L)	***	Laboratory information system
Ferritin ($\mu g/L$)	***	Laboratory information system
Transfusion	Yes/no**	Medical records
Erythropoiesis-stimulating agents	Yes/no**	Medical records
G-CSF	Yes/no**	Medical records
Iron-chelating agents	Yes/no**	Medical records
Hypomethylating agents	Yes/no**	Medical records
Hydroxurea	Yes/no**	Medical records
AML-like chemotherapy	Yes/no**	Medical records
Immunosuppressive agents	Yes/no**	Medical records
Lenalidomide	Yes/no**	Medical records
Protocol participation	Yes/no/unknown	Medical records
Progression to sAML	Yes/no/unknown	Patobank/medical records
Date of progression to sAML	Date specified (± 7 days)	Patobank/medical records
Bone marrow transplantation	Yes/no	Medical records
Date of bone marrow transplantation	Exact date of transplantation	Medical records
Type of bone marrow transplantation	Myeloablative/reduced intensity conditioning or autologous	Medical records

Notes: *Variables regarding treatment, progression to sAML, protocol participation and bone marrow transplantation were validated twice (at a registration form including 2-year data and 5-year data respectively, accounting for 30 variables in total). **In order to register a “yes” you have to tick a “box”. If nothing is ticked off a “no” is automatically registered. ***Exact value measured on day of diagnostic bone marrow sample (± 30 days).

Abbreviations: G-CSF, granulocyte-colony-stimulating-factors; MCV, middle cell volume; sAML, secondary acute myeloid leukemia.

period was drawn. This number was chosen to ensure that we included a sample size of at least 10% of the whole MDS population, as we expected that some of the medical records could not be retrieved.

Record Review

For validation, we selected 48 variables from the DMDS (18 from the registration form at diagnosis and 15 from each of the registration forms including 2-year data and 5-year-data), which we regarded most important for studying MDS (Table 2). These were validated using information from medical records, the Danish Pathology Databank

(Patobank), and the clinical laboratory information system as the gold standard.¹² Patobank is a nationwide database that contains descriptions of pathology examinations since 1997.¹³ It is accessible online at all Danish hospitals. Accepted values and gold standards for the 48 variables are shown in Table 2. For example, laboratory values were accepted from blood samples conducted ± 30 days from date of diagnostic bone marrow examination and correct date of progression to AML was accepted with a range of ± 7 days. Data on vital status and date of death were not validated in this study, as this information is routinely drawn from the Civil Registration System and linked to

the DMDS. During the validation process, one of four outcomes were recorded: 1) consistent with reference standard, 2) not consistent with reference standard, 3) missing value (when information was present in reference, but not in the DMDS, and 4) not relevant (eg type of allo-HSCT, when the patient was not transplanted).

Validation was performed during March through August 2020. All patients were confirmed to have MDS by assessment of the morphological bone-marrow descriptions in Patobank, assessment of cytopenias, evaluation of chromosomal aberrations, and mutational markers according to “The World Health Organizations classification of myeloid neoplasms ICD-10 2008/2016”.² All medical records were reviewed by one specialist registrar in hematology (TBL), and any questions about interpretation of data were discussed with two consultants in hematology specialized in MDS (JMN and LSGØ).

Statistical Analysis

Descriptive characteristics for the total MDS population and the validation sample (with and without a MDS diagnosis) including the proportion of missing data were tabulated. Continuous variables were reported as medians with 25th and 75th percentiles. Patients for whom we were unable to retrieve information from medical records or who did not fulfill the diagnostic criteria for a MDS-diagnosis were excluded from the study. Predictive values were calculated as the number of patients with correct registration divided by the number of patients registered. Predictive values for categorical variables were computed within strata of “yes”, “no”, and overall. If one or more variables from a patient could not be evaluated because of missing information in the reference (eg no laboratory measurements available), the patient was not included when calculating predictive values for that specific variable. Further, we examined if the predictive values varied according to subgroup of patients defined by departments (university hospitals vs regional hospitals) and date of diagnosis (before and after 2014). Estimates were given with binomial 95% confidence intervals (CIs). To clarify whether incorrect registration was represented by few patients or equally distributed among all patients, we calculated a median sum score of correctly coded variables within the registration form filled out at diagnosis (18 variables).

In addition, the annual incidence rate for MDS was computed as the number of incident MDS patients registered in the DMDS from 2014–2018 divided by the

Danish midyear population in 2014–2018 (per 100.000).¹⁴ A CI around the estimate was calculated with a Poisson approximation. Data were entered into a Redcap database^{15,16} and Stata version 16 was used for the sampling process and the statistical analyses.

The study was approved by the Danish Data Protection Agency (record number: 1–16-02-321-18) and the Danish Patient Safety Authority (record number 3–3013- 2960/1). According to Danish law, no approval from an ethical committee was required.

Results

The overall MDS cohort comprised 2284 individuals. Of these, we randomly sampled 272 patients for the validation study. We were unable to obtain medical records on 12 patients and they were excluded. Patient characteristics and the proportion of missing values are presented in Table 3 for the registration form completed at diagnosis and the registration form including 2-year data. The median age at diagnosis was 76 years with a male predominance (62%). Of the 2284 patients, 140 (6%) patients had an antecedent hematological disease and 222 (10%) patients had a history of chemotherapy or radiotherapy prior to the diagnosis of MDS (therapy-related MDS). The median hemoglobin level was 6.1 mmol/L, the median platelet count was $125 \times 10^9/L$ and the median white blood cell count was $1.8 \times 10^9/L$. The median blast count in the bone marrow was 2%. Karyotype was available for 2020 patients (88%). Overall characteristics for the validation sample were broadly similar to the total MDS population, whereas the misclassified MDS patients were older, more frequently diagnosed before 2015, more often had the WHO subtype “MDS unclassifiable”, and had less pronounced changes in laboratory values (except from white blood cell count).

For 15 out of 18 variables on the registration form filled out at diagnosis, we observed no missing data. “Blood transfusion prior to diagnosis” and “monocyte count” had a high proportion of missing values (53% for both). For 13 out of 18 variables on the registration form filled out at diagnosis a possible registration was “unknown”. Ferritin had the highest proportion of “unknown” registrations (15%) followed by “prior radiotherapy” (8%) and “prior chemotherapy” (7%). None of the patients with an “unknown” registration for prior radiotherapy or chemotherapy had, however, received this treatment according to information in medical records.

Table 3 Characteristics of Patients Registered in the Danish Myelodysplastic Syndromes Database. Data are Shown for the Whole Population, the Validation Sample and the Misclassified MDS Patients

	Total MDS Population	Validation Sample	Misclassified as MDS Patients
At diagnosis (number of patients)	2284	238	22
Male sex	1417 (62%)	131 (55%)	15 (68%)
Median age, years (IQR)	76 (68–82)	76 (68–82)	78 (63–82)
Enrolment year			
2010–2014	1185 (52%)	117 (49%)	16 (73%)
2015–2019	1099 (48%)	121 (51%)	6 (27%)
Diagnostic WHO subtype			
MDS with single lineage dysplasia	241 (11%)	29 (12%)	–
MDS with ringed sideroblasts	238 (10%)	26 (11%)	0
MDS with excess blasts	563 (25%)	57 (24%)	0
MDS with multi-lineage dysplasia	665 (29%)	70 (29%)	–
MDS with isolated del(5q)	77 (3%)	11 (4%)	0
MDS unclassifiable	500 (22%)	46 (19%)	16 (72%)
Prior hematological disease (>3 months)	140 (6%)	19 (8%)	0
Missing	0	0	0
Unknown	59 (3%)	6 (2%)	0
Prior chemotherapy	126 (6%)	13 (5%)	0
Missing	0	0	0
Unknown	164 (7%)	14 (6%)	4 (18%)
Prior radiotherapy	96 (4%)	9 (4%)	–
Missing	0	0	0
Unknown	175 (8%)	15 (6%)	–
Blood transfusion prior to diagnosis	106 (5%)	14 (6%)	0
Missing	1218 (53%)	122 (51%)	15 (68%)
Unknown	128 (6%)	16 (7%)	–
IPSS			
Good-risk	1013 (44%)	107 (45%)	12 (55%)
Intermediate-risk	746 (33%)	80 (34%)	–
Poor-risk	261 (11.5%)	22 (9%)	–
Unknown	254 (11%)	29 (12%)	–
Missing	10 (0.5%)	0	0
Anemia (hemoglobin < 6.2 mmol/L)	1353 (59%)	135 (56%)	10 (45%)
Thrombocytopenia (platelets < 100 × 10 ⁹ /L)	905 (40%)	83 (35%)	6 (27%)
Neutropenia (WBC < 1.8 × 10 ⁹ /L)	1026 (45%)	97 (41%)	11 (50%)
Blast count marrow, median (IQR)	2 (0–7)	2 (0–5)	1 (0–2)
Missing	0	0	0
Unknown	133 (6%)	6 (2.5%)	4 (18%)
MCV, fL, median (IQR)	99 (91–106)	101 (94–107)	97 (88–103)
Missing	0	0	0
Unknown	130 (6%)	8 (3%)	1 (5%)

(Continued)

Table 3 (Continued).

	Total MDS Population	Validation Sample	Misclassified as MDS Patients
Hemoglobin, mmol/L, median (IQR)	6.1 (5.5–6.9)	6.2 (5.5–7.0)	7 (5.6–8.2)
Missing	0	0	0
Unknown	11 (0.5%)	0	0
Platelet count, $\times 10^9/L$, median (IQR)	125 (68–235)	140 (76.0–252)	155 (99–250)
Missing	0	0	0
Unknown	15 (0.6%)	0	0
WBC, $\times 10^9/L$, median (IQR)	1.8 (0.9–3.5)	2.1 (1.1–3.7)	1.6 (1.1–4.2)
Missing	0	0	0
Unknown	45 (2%)	-	0
Leucocyte count, $\times 10^9/L$, median (IQR)	3.8 (2.5–6.1)	4 (2.4–6.4)	3.9 (6.3)
Missing	0	0	0
Unknown	13 (0.6%)	-	0
Monocyte count, $\times 10^9/L$, median (IQR)	0.3 (0.2–0.6)	0.3 (0.2–0.7)	0.7 (0.2–0.8)
Missing	1218 (53%)	122 (51%)	14 (64%)
Unknown	44 (2%)	-	0
LDH, U/L, median (IQR)*	208 (175–252)	205 (171–241)	200 (170–236)
Missing	0	0	0
Unknown	88 (4%)	9 (4%)	0
Ferritin ($\mu g/L$)	362 (174–691)	366 (186–698)	168 (79–398)
Missing	0	0	0
Unknown	331 (15%)	26 (11%)	2 (9%)
2- year data (number of patients)	1910 (84%)	197 (83%)	21 (95%)
Treatment			
Transfusion	936 (49%)	100 (51%)	5 (24%)
Erythropoiesis-stimulating agents	667 (35%)	73 (37%)	4 (19%)
Granulocyte-colony-stimulating-factors	220 (12%)	22 (11%)	–
Hypomethylating agents (Azacitidine)	291 (15%)	25 (13%)	0
Iron-chelating agents	61 (3%)	7 (4%)	–
Hydroxyurea	88 (5%)	7 (4%)	–
AML-like chemotherapy	106 (6%)	12 (6%)	0
Immune suppressive agents	63 (3%)	9 (5%)	0
Lenalidomide	28 (2%)	–	0
Protocol participation	111 (5%)	6 (3%)	0
Missing	374 (16%)	0	0
Unknown	54 (2%)	5 (2%)	0
Progression to sAML	238 (13%)	31 (16%)	0
Missing	55 (3%)	5 (2.5%)	0
Unknown	40 (2%)	-	–
Date of progression to sAML	139 (58%)	21 (68%)	–
Missing	99 (42%)	10 (32%)	–
Unknown	0	0	–
Allo-HSCT	109 (5%)	17 (9%)	0
Missing	374 (16%)	0	0
Unknown	0	0	0

(Continued)

Table 3 (Continued).

	Total MDS Population	Validation Sample	Misclassified as MDS Patients
Date of allo-HSCT	109 (100%)	17 (100%)	–
Missing	0	0	–
Unknown	0	0	–
Type of allo-HSCT	109 (100%)	17 (100%)	–
Missing	0	0	–
Unknown	0	0	–

Notes: *For patients < 70 years old normal range for LDH is 105 to 205 U/L and for patients > 70 years normal range for LDH is 115 to 255 U/L. – Data not shown due to small numbers.

Abbreviations: allo-HSCT, allogeneic hematopoietic stem cell transplantation; IPSS, International Prognostic Scoring System; IQR, interquartile range (25th to 75th percentiles); LDH, lactate dehydrogenase; MCV, middle cell volume; sAML, secondary acute myeloid leukemia; WBC, white blood cell count; WHO, World Health Organization.

Predictive values for each of the 18 variables reported at diagnosis are shown in [Figure 1](#). The positive predictive value of the MDS diagnosis was 91.5% (95% CI: 87.5–94.7). Of the 22 patients who did not fulfil the criteria for MDS, the diagnoses were unsettled cytopenia, chronic myelomonocytic leukemia, chronic idiopathic neutropenia, reactive anemia, chronic myeloproliferative neoplasia, idiopathic thrombocytopenic purpura, and cytopenia caused by treatment with cytotoxic drugs. The overall predictive values ranged from 86.9% for “blood transfusion prior to diagnosis” (95% CI: 78.6–92.8) to 97.8% (95% CI: 94.8–99.3) for “prior radiotherapy” and were higher than 90% for 16 of 18 variables.

Of the 1910 patients with a completed registration form including 2-year data, 936 (49%) received treatment with transfusion, 667 (35%) received treatment with erythropoiesis-stimulating agents, and 291 (15%) received treatment with a hypomethylating agent. Progression to sAML occurred in 238 (13%) patients and 109 patients (5%) were treated with allo-HSCT. Four of 15 variables contained missing data. “Date of progression to AML” had the highest proportion of missing data (42%) followed by “participation in protocol” (16%) and “allo-HSCT” (16%).

The overall predictive values for variables included in the 2-year data registration form ranged from 71.4% (95% CI: 47.7–87.3) for “date of progression to AML” to 99.5% (95% CI: 97.2–99.9) for “treatment with hydroxyurea” and were greater than 94% for 11 out of 15 variables ([Figure 2](#)). Due to relatively few observations, no solid conclusions could be derived from predictive values on patients receiving iron-chelating treatment, hydroxyurea, AML-like-chemotherapy, immune modulating agents, and lenalidomide. The predictive values were, however,

generally high for patients not receiving the aforementioned treatment modalities.

The registration form including 5-year data was completed for 449 patients (20%) and characteristics and predictive values are shown in [Supplemental Table 1 and Figure 1](#).

Stratification by departments (university hospitals vs regional hospitals) revealed broadly comparable predictive values for the 18 variables from the registration form filled out at diagnosis ([Supplemental Table 2](#)). The stratified analysis suggested, however, that the positive predictive value of the MDS diagnosis improved from 87.7% before 2015 to 95.3% thereafter. The overall median sum score for a correctly filled registration form at diagnosis was 16 (25th–75th: 15–17). The overall incidence of MDS was 6.0 per 100,000 persons (95% CI: 5.7–6.3).

Discussion

Our study provided patient characteristics of 2284 unselected MDS patients registered in the DMDS during 2010–2019. Further, we showed that data in the DMDS were highly valid with a positive predictive value of the MDS diagnosis of 91.5% and predictive values exceeding 90% for the majority of study variables.

Descriptive characteristics, including the age- and sex composition, laboratory values, and proportion of therapy-related MDS-patients, reported in our study, are broadly comparable to findings from the Swedish MDS Registry,¹⁷ the Düsseldorf MDS Registry¹⁸ and six cross-sectional physician surveys from the United States.¹⁹

In the Swedish MDS Registry, approximately 50% of the patients from 2009–2013 received transfusions at diagnosis.¹⁷ In the DMDS, information on blood

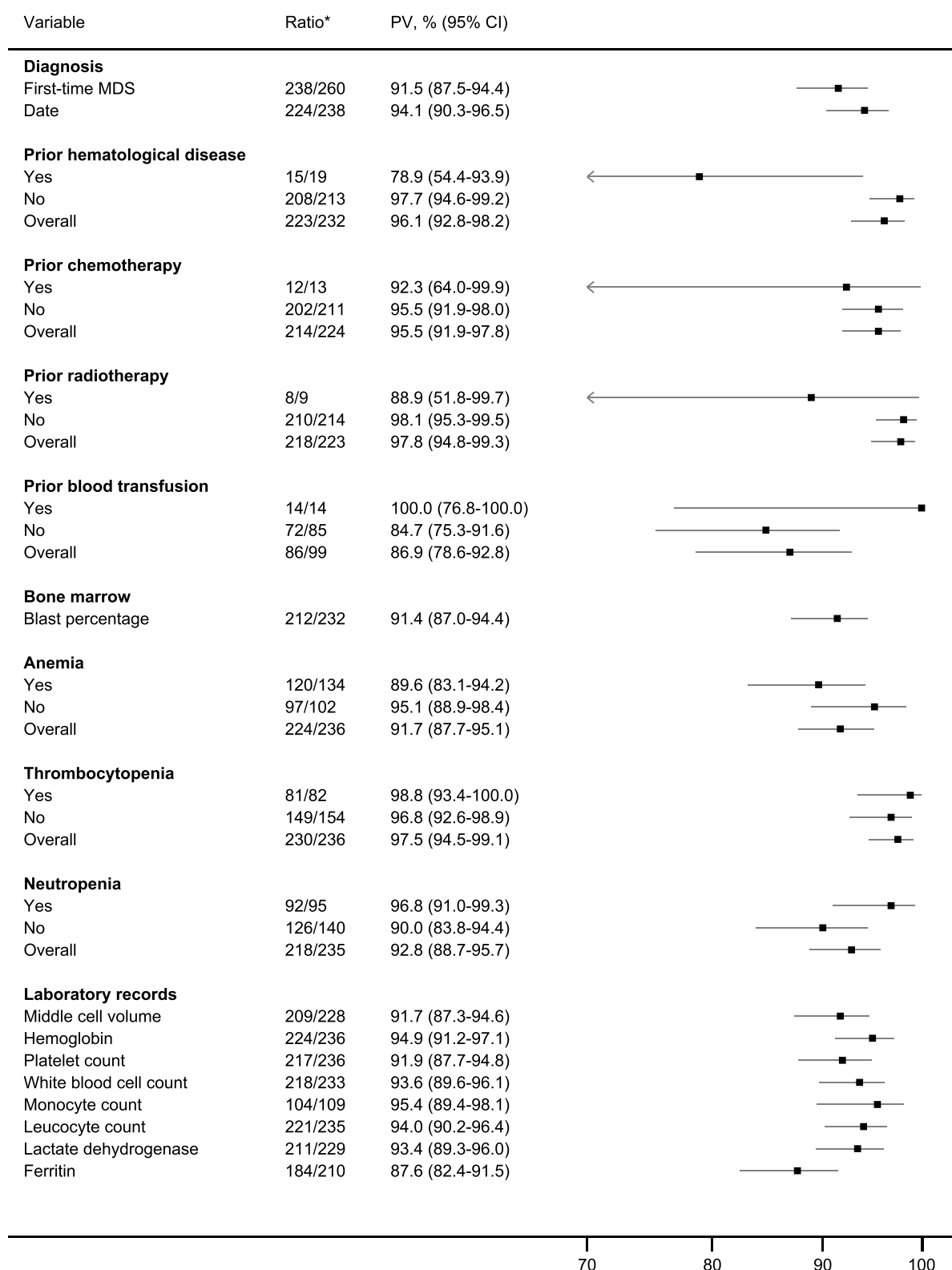


Figure 1 Predictive values for 18 variables from the registration form completed at diagnosis from the Danish Myelodysplastic Syndromes Database.

Notes: *Number of correctly coded records/number of relevant records reviewed (missing/unknown values excluded).

Abbreviations: BM, bone marrow; CI, confidence interval.

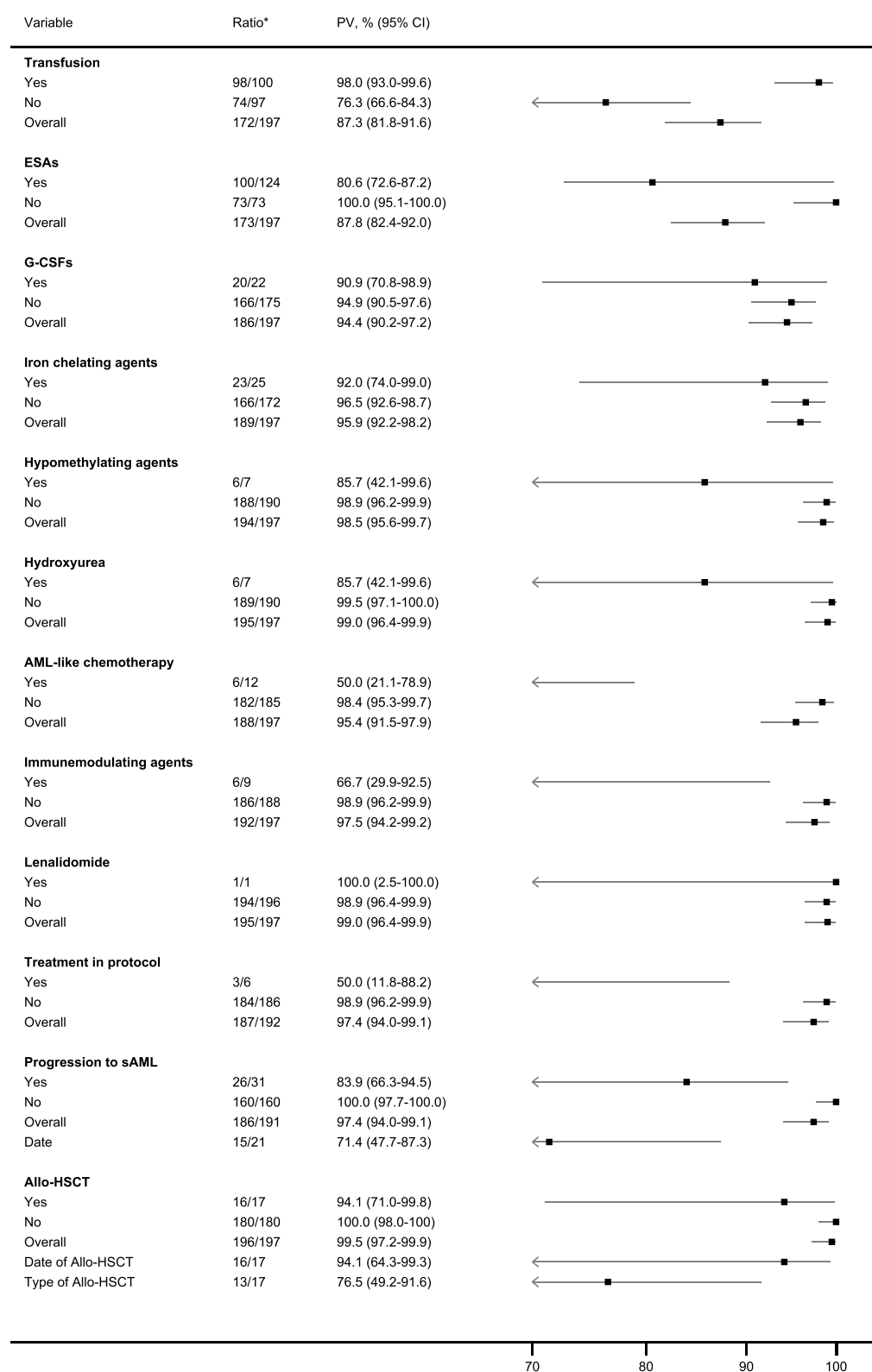


Figure 2 Predictive values for 15 variables from the registration form including 2-year data from the Danish Myelodysplastic Syndromes Database.

Notes: *Number of correctly coded records/number of relevant records reviewed (missing/unknown values excluded).

Abbreviations: Allo-HSCT, allogeneic hematopoietic stem cell transplantation; CI, confidence interval, ESAs, erythropoiesis-stimulating agents; G-CSFs, granulocyte-colony-stimulating-factors; sAML, secondary acute myeloid leukemia.

transfusion prior to diagnosis had a high proportion of missing data (53%), and only 5% of the patients were registered as transfusion-dependent at diagnosis. Conversely, 59% of the patients in the DMDS had pronounced anemia (less than 6.2 mmol/L) and 49% of the patients were treated with transfusion within the first two years. This suggests that information about blood transfusion is underreported at the time of diagnosis in the DMDS. However, complete transfusion data, including type of product and time of transfusion can be supplemented with data from the Danish Transfusion Database.²⁰ Information about date of progression to sAML was also missing for a considerably proportion of patients. However, this information can easily be retrieved from the Danish National Acute Leukemia Registry.²¹

The overall use of erythropoiesis-stimulating agents, granulocyte-colony-stimulating factors, hypomethylating agents, and lenalidomide in an unselected MDS-population is sparsely reported in prior studies.^{19,22} Within the first two years of diagnosis, 35% of patients in the DMDS were treated with an erythropoiesis-stimulating agent and only 2% of patients in the DMDS were treated with lenalidomide. This is somewhat lower than in two studies from the United States where erythropoiesis-stimulating agents were used in 50–60% of recently diagnosed MDS patients, while lenalidomide was used for around 7% of recently diagnosed MDS patients.^{19,22} The use of granulocyte-colony-stimulating-factors (12%) and hypomethylating agents (15%) reported in the DMDS were however in line with results from the two aforementioned studies from the United States. As such, our study confirms findings from the United States; indicating that disease-modifying therapies (hypomethylating agents and lenalidomide) approved for MDS are used for a minority of MDS patients only. Allo-HSCT and clinical trial enrolment were only registered for a minority of MDS patients (5%) which also is in line with data from the United States.¹⁹

To the best of our knowledge, no prior validation studies of MDS registries have been performed. Although direct comparison to our findings therefore cannot be made, two validation studies of The Danish National Acute Leukemia Registry and The Danish National Lymphoma Registry showed high completeness and accuracy of registered variables.^{21,23} The positive predictive value of an MDS diagnosis in our study (91.5%) was lower than the positive predictive value of

an acute myeloid leukemia diagnosis in the Danish National Acute Leukemia Registry [99.6% (95% CI: 98.1–100)]²¹ and of a lymphoma subtype diagnosis in the Danish National Lymphoma Registry [98.5% (95% CI: 97.8–99.1)].²⁴ The misclassified cases in the DMDS reflect the challenges of establishing a correct diagnosis of MDS.^{25,26} Cases in which secondary causes of cytopenia are excluded but the karyotype is normal, blast-count, and dysplastic changes are few and other features convincing for an MDS diagnosis are lacking often present diagnostic difficulties. Most of the misclassified patients in the present study, had “idiopathic cytopenia of undetermined significance”¹ or clonal cytopenia of unknown significance.¹ Both these conditions are potential pre-phases of MDS or other hematological malignancies, but can also persist without clinical manifestation or progression.^{27–29} Mutational profiling and flow cytometry are used increasingly among patients with MDS and may contribute to the diagnosis of MDS. No generally accepted minimal criteria or classification using mutational markers or flow cytometry are however yet available.³⁰ Reassuringly, our data suggested that the positive predictive value of the MDS diagnosis has improved over time. Furthermore, the positive predictive value is presumably higher in the DMDS if analyses are restricted to intermediate- to high-risk MDS patients as they often have more evident changes in the bone marrow, more pronounced cytopenias and more severe cytogenetic alterations which makes the diagnosis more obvious.

The DMDS has several strengths. It is population-based, it contains detailed and valid clinical information on more than 2200 MDS patients, and it can be linked with other Danish health-care registries.⁶ The Swedish MDS-registry is comparable to the DMDS regarding registered variables but it has a larger sample size.³¹ The European MDS Registry (EUMDS) also contains detailed information on patients with MDS of IPSS low and intermediate-1 subtypes and has recently been extended to enroll all subtypes of MDS. Reporting of MDS-patients to the EUMDS is however optional, which may cause selection bias. The Surveillance, Epidemiology, and End Results (SEER) Program contains a large sample size, but has numerous limitations including lack of validation of the diagnosis, selection bias, and lack of clinical information.⁴

The DMDS has important limitations. It lacks detailed treatment information, eg start/end date of treatment, evaluation of treatment response according to the International Working Group Criteria, performance

status, lifestyle factors, family history of myeloid disease and comorbidity.^{32,33} However, data from the DMDSD are easily and cost-effectively linked at the individual level to local and national registries including the Danish National Patient Registry to obtain a medical history of the patient. This nationwide registry holds complete information and valid discharge diagnoses on comorbidities since 1977 and in-hospital treatments, eg chemotherapy starting from 1999.¹¹ Antineoplastic procedurecodes in the Danish National Patient Registry were recently validated showing an overall positive predictive value of 95% (95% CI: 94.0–95.0).³⁴ Detailed information about allo-HSCT can furthermore be supplemented using information from the Danish transplant centers.³⁵ Finally, an important limitation is the lack of information regarding molecular alterations. These data are currently stored in local databases, but will likely be available to researchers in the near future.

The present study has several strengths. We conducted a comprehensive review of individual medical records at all the departments of hematology in Denmark during the entire study period, and we selected a high number of study variables. For validation, we used the same data sources as the ones used by the clinicians when reporting patients to the DMDSD. Our study also had some limitations. We lacked a validation of the individual World Health Organization MDS subtypes and of cytogenetic alterations. The latter were however revised and reported according to the IPSS-R in 2020 by clinical cytogeneticists and will be available to researchers during 2021. The validation was performed by one physician only. However, we pre-specified the gold standards of the selected study variables to minimize the impact of this issue. Last, our validation sample comprised few patients with submitted information on the registration form including 5-year data, resulting in imprecise predictive values, which warrants careful interpretation.

Conclusion

In this nationwide study, characteristics of the MDS patients broadly resembled that of other MDS registries. For the majority of variables there were no missing data and data were generally valid with a predictive value exceeding 90% for most study variables. Thus, the DMDSD represents a valuable source for future research.

Acknowledgment

This study was supported by grants from The Danish Cancer Society (Kræftens Bekæmpelse, grant number R223-A13094-18-S68), The Danish Research Center for Equity in Cancer (COMPAS), the Dagmar Marshalls Foundation, and the Danish Acute Leukemia Group. Kirsten Grønbaek is supported by grants from The Danish Cancer Society (Kræftens Bekæmpelse, grant no. R223-A13071 and the Danish Research Center for Precision Medicine in Blood Cancers).

Disclosure

LH declares research grants from Celgene. The authors report no other conflicts of interest in this work.

References

1. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391–2405. doi:10.1182/blood-2016-03-643544
2. Stephen H, Swerdlow EC, Harris NL, et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. World Health Organization; 2017.
3. Rollison DE, Howlader N, Smith MT, et al. Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001–2004, using data from the NAACCR and SEER programs. *Blood*. 2008;112(1):45–52. doi:10.1182/blood-2008-01-134858
4. Ma X, Does M, Raza A, Mayne ST. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer*. 2007;109(8):1536–1542. doi:10.1002/cncr.22570
5. Zeidan AM, Shallis RM, Wang R, Davidoff A, Ma X. Epidemiology of myelodysplastic syndromes: why characterizing the beast is a prerequisite to taming it. *Blood Rev*. 2019;34:1–15. doi:10.1016/j.blre.2018.09.001
6. Schmidt M, Schmidt SAJ, Adelborg K, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol*. 2019;11:563–591. doi:10.2147/CLEP.S179083
7. Akut Leukæmi Gruppen, annual report 2018. Available from: www.hematology.dk/index.php/om-alg/arsrapporter. Accessed April 30, 2021.
8. Simons A, Shaffer LG, Hastings RJ. Cytogenetic nomenclature: changes in the ISCN 2013 Compared to the 2009 Edition. *Cytogenet Genome Res*. 2013;141(1):1–6. doi:10.1159/000353118
9. Nørgaard M, Johnsen SP. How can the research potential of the clinical quality databases be maximized? The Danish experience. *J Intern Med*. 2016;279(2):132–140. doi:10.1111/joim.12437
10. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541–549. doi:10.1007/s10654-014-9930-3
11. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490. doi:10.2147/CLEP.S91125
12. Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, Adelborg K. Existing data sources in clinical epidemiology: laboratory information system databases in Denmark. *Clin Epidemiol*. 2020;12:469–475. doi:10.2147/CLEP.S245060

13. Bjerregaard B, Larsen OB. The Danish pathology register. *Scand J Public Health*. 2011;39(7 Suppl):72–74. doi:10.1177/1403494810393563
14. Statistics Denmark. Available from: <https://www.statbank.dk/statbank5a/default.asp?w=1366>. Accessed April 30, 2021.
15. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208. doi:10.1016/j.jbi.2019.103208
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381. doi:10.1016/j.jbi.2008.08.010
17. Moreno Berggren D, Folkvaljon Y, Engvall M, et al. Prognostic scoring systems for myelodysplastic syndromes (MDS) in a population-based setting: a report from the Swedish MDS register. *Br J Haematol*. 2018;181(5):614–627. doi:10.1111/bjh.15243
18. Neukirchen J, Schoonen WM, Strupp C, et al. Incidence and prevalence of myelodysplastic syndromes: data from the Düsseldorf MDS-registry. *Leuk Res*. 2011;35(12):1591–1596. doi:10.1016/j.leukres.2011.06.001
19. Sekeres MA, Schoonen WM, Kantarjian H, et al. Characteristics of US patients with myelodysplastic syndromes: results of six cross-sectional physician surveys. *J Natl Cancer Inst*. 2008;100(21):1542–1551. doi:10.1093/jnci/djn349
20. The Danish Transfusion Database. Available from: <https://www.rkkp.dk/om-rkkp/de-kliniske-kvalitetsdatabaser/transfusionsdatabase/>. Accessed April 30, 2021.
21. Ostgård LS, Nørgaard JM, Severinsen MT, et al. Data quality in the Danish National Acute Leukemia Registry: a hematological data resource. *Clin Epidemiol*. 2013;5:335–344. doi:10.2147/CLEP.S48411
22. Ma X, Steensma DP, Scott BL, Kiselev P, Sugrue MM, Swern AS. Selection of patients with myelodysplastic syndromes from a large electronic medical records database and a study of the use of disease-modifying therapy in the United States. *BMJ Open*. 2018;8(7):e019955. doi:10.1136/bmjopen-2017-019955
23. Arboe B, Josefsson P, Jørgensen J, et al. Danish National Lymphoma Registry. *Clin Epidemiol*. 2016;8:577–581. doi:10.2147/CLEP.S99470
24. Arboe B, El-Galaly TC, Clausen MR, et al. The Danish National Lymphoma Registry: coverage and Data Quality. *PLoS One*. 2016;11(6):e0157999. doi:10.1371/journal.pone.0157999
25. Tanaka TN, Bejar R. MDS overlap disorders and diagnostic boundaries. *Blood*. 2019;133(10):1086–1095. doi:10.1182/blood-2018-10-844670
26. Steensma DP. Dysplasia has A differential diagnosis: distinguishing genuine myelodysplastic syndromes (MDS) from mimics, imitators, copycats and impostors. *Curr Hematol Malig Rep*. 2012;7(4):310–320. doi:10.1007/s11899-012-0140-3
27. Malcovati L, Cazzola M. The shadowlands of MDS: idiopathic cytopenias of undetermined significance (ICUS) and clonal hematopoiesis of indeterminate potential (CHIP). *Hematology Am Soc Hematol Educ Program*. 2015;2015:299–307. doi:10.1182/asheducation-2015.1.299
28. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9–16. doi:10.1182/blood-2015-03-631747
29. Valent P. ICUS, IDUS, CHIP and CCUS: diagnostic criteria, separation from MDS and clinical implications. *Pathobiology*. 2019;86(1):30–38. doi:10.1159/000489042
30. Valent P, Orazi A, Steensma DP, et al. Proposed minimal diagnostic criteria for myelodysplastic syndromes (MDS) and potential pre-MDS conditions. *Oncotarget*. 2017;8(43):73483–73500. doi:10.18632/oncotarget.19008
31. Nationell rapport från kvalitetsregistret Myelodysplastiskt syndrom. Rapport för diagnosår 2009–2017. Available from: <https://www.cancercentrum.se/samverkan/cancerdiagnoser/blod-lymfom-myelom/myelodysplastiskt-syndrom-mds/kvalitetsregister/rapporter/>. Accessed April 30, 2021.
32. 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. doi:10.1182/blood-2005-10-4149
33. Platzbecker U, Fenaux P, Adès 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. doi:10.1182/blood-2018-06-857102
34. Broe MO, Jensen PB, Mattsson TO, Pottegård A. Validity of anti-neoplastic procedure codes in the Danish national patient registry: the case of colorectal cancer. *Epidemiology*. 2020;31(4):599–603. doi:10.1097/EDE.0000000000001208
35. Østgård LSG, Lund JL, Nørgaard JM, et al. Impact of allogeneic stem cell transplantation in first complete remission in acute myeloid leukemia: a National Population-Based Cohort Study. *Biol Blood Marrow Transplant*. 2018;24(2):314–323. doi:10.1016/j.bbmt.2017.10.019

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventive initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress