

# Data quality in the Danish National Acute Leukemia Registry: a hematological data resource

Lene Sofie Granfeldt  
Østgård<sup>1,2</sup>

Jan Maxwell Nørgaard<sup>1</sup>

Marianne Tang Severinsen<sup>3</sup>

Henrik Sengeløv<sup>4</sup>

Lone Friis<sup>5</sup>

Morten Krogh Jensen<sup>6</sup>

Ove Juul Nielsen<sup>7</sup>

Mette Nørgaard<sup>2</sup>

<sup>1</sup>Department of Hematology,

<sup>2</sup>Department of Clinical Epidemiology,

Aarhus University Hospital,

Aarhus, Denmark; <sup>3</sup>Department

of Hematology, Aalborg University

Hospital, Aalborg, Denmark;

<sup>4</sup>Department of Hematology,

Rigshospitalet, Copenhagen, Denmark;

<sup>5</sup>Department of Hematology,

Odense University Hospital,

Odense, Denmark; <sup>6</sup>Department

of Hematology, Herlev Hospital,

Herlev, Denmark; <sup>7</sup>Department

of Hematology, Rigshospitalet,

Copenhagen, Denmark

**Background:** The Danish National Acute Leukemia Registry (DNLR) has documented coverage of above 98.5%. Less is known about the quality of the recorded data.

**Objective:** To describe the present coverage of the DNLR, its completeness, and accuracy of individual variables for acute myeloid leukemia (AML). Furthermore, as a second measure of true coverage of the DNLR, to estimate AML incidence in Denmark from DNLR data and compare it to incidence reported from other AML registries.

**Patients and methods:** By the end of December 2011, the DNLR (established January 2000) included detailed data on a large, well-defined, and nonselected Danish population of 2,665 AML patients. We estimated positive predictive values (PPVs) and completeness for 30 variables, which included patient and disease characteristics, treatment, and treatment outcomes. We identified 260 AML patients (10% of all AML patients recorded in the DNLR). We used information from medical records as the gold standard.

**Results:** Using the Danish National Registry of Patients as a reference, the coverage of the DNLR was 99.6%. The PPVs of the individual variables ranged from 89.4% to 100%. The completeness of individual variables varied between 60.7% and 100%. Stratification by time of registration in the DNLR (before 2006 versus 2006 and later) revealed higher PPVs and lower frequencies of missing data from 2006. Sex-adjusted incidence rates were 6.2/100,000 person-years (95% confidence interval 5.8–6.6) in males and 4.9/100,000 person-years (95% confidence interval 4.5–5.4) in females. Yearly incidence rates of AML were higher than the incidence rates reported from Sweden (4.5 and 4.2/100,000) and the US (4.5 and 3.1/100,000 in Caucasians).

**Conclusion:** With few exceptions, there were high values for PPVs and completeness of recorded data. Data accuracy and completeness have improved since the registry was established. The estimated incidence may indicate that the DNLR truly is more complete than other registries. In conclusion, the DNLR is a valuable resource for clinical research of AML.

**Keywords:** acute myeloid leukemia, incidence, registration completeness, validation studies

## Introduction

Acute myeloid leukemia (AML) is the most common type of acute leukemia that affects adults. Between 200 and 250 new cases are diagnosed annually in Denmark.<sup>1</sup> Although rare, AML is the most common cause of cancer death in men younger than 40 years.<sup>2</sup> Successful AML treatment continues to be a major therapeutic challenge throughout the world. Current cure rates in younger adult patients (age <60 years) average approximately 40% and are even lower in the elderly.<sup>1</sup>

A substantial proportion of our current knowledge about AML prognosis is based on data from clinical trials. Characteristics of patients included the general patient

Correspondence: Lene Sofie Granfeldt Østgård  
Department of Clinical Epidemiology,  
Aarhus University Hospital,  
43–45 Oluf Palmes Allé,  
Aarhus N 8200, Denmark  
Tel +45 8716 8248  
Fax +45 8716 7215  
Email lene.oestgaard@ki.au.dk

population (eg, in terms of age, ethnicity, performance status, comorbidity, and intent of treatment).<sup>3–5</sup> Selection of patients into clinical trials may hamper the generalizability of treatment outcomes to the general population of AML patients. Thus, registry- and population-based studies may have an important role in the examination of some aspects of AML prognosis.<sup>2</sup>

Leukemia registries are potentially a valuable source of AML data. The data are readily available and can be utilized at minimal cost compared with data from randomized trials. Furthermore, because these registries collect data for administrative purposes that are unrelated to specific research questions, the risk of certain biases is reduced (eg, recall, nonresponse, and losses to follow-up).<sup>6,7</sup> However, one important disadvantage of registry data is that investigators cannot control data collection and quality. Registry data should be validated before it is used for research.

Danish medical registries are generally known to be complete and accurate.<sup>8</sup> The Danish National Acute Leukemia Registry (DNLR) is a nationwide registry that includes 98.5% of Danish AML patients. However, little is known about the quality of DNLR data.<sup>1</sup> The main objective of this study was, therefore, to quantify DNLR data quality by estimating the completeness and positive predictive values (PPVs) for 30 variables. The second objective was to estimate coverage of the patient population. Coverage is highly dependent on the reference used. As an indirect measure of coverage, we additionally calculated incidence (total and sex-specific) based on the Danish National Acute Leukaemia Database (ALDB) and compared it to incidence estimates from other population-based leukemia registries.

## Materials and methods

The Danish population (~5.5 million people)<sup>9</sup> is guaranteed free access to tax-supported medical care, which is provided by the public health care system. AML patients are treated with curative intent at five highly specialized centers (Copenhagen, Herlev, Odense, Aarhus, and Aalborg). Treatment consists of combination chemotherapy, which is combined with allogeneic stem-cell transplantation (alloHCT) in selected cases. Two hospital departments are accredited to perform alloHCT (Rigshospitalet, Copenhagen and Aarhus University Hospital). Five hematological departments (Viborg, Holstebro, Esbjerg, Roskilde, and Næstved) treat AML patients only with palliative intent. AML patients diagnosed at nonhematological departments are transferred to

hematological departments. No AML treatment in Denmark occurs in private hospitals.

## DNLR

### Design, setting, and participants

All AML patients in Denmark are recorded in the DNLR, which was established by the Danish Society of Hematology in January 2000. Cases of advanced myelodysplastic syndrome (MDS) that are treated using AML protocols are also recorded in the DNLR. Since January 2005, patients diagnosed with acute lymphoblastic leukemia have also been included in the registry. Since 2005, the DNLR has been part of the Danish Common Hematological Database, which consists of separate databases for lymphomas, MDS, chronic lymphocytic leukemia, myeloproliferative disorders, and multiple myeloma. A report based on the ALDB is published annually by the Acute Leukemia Group steering committee.<sup>1</sup> The aims of the hematological clinical databases are to gain insight into the epidemiology of hematological cancers, to evaluate treatment response, and to compare results between the five regions of Denmark. Participating hospital departments must report requested data to the DNLR. Physicians are not required to obtain patient consent. Financial support for the registry depends directly on registration completeness and coverage. By January 2011, the DNLR consisted of data on 2,665 AML patients (including acute promyelocytic leukemia). In August 2012, these patients had a mean follow-up time of 635 days (range 1–4,596) and a combined follow-up time of 4,578.5 years.

### Data collection and variables

Standardized registration forms are used to collect DNLR data. Until 2005, participating departments could submit a registration on paper or computer disk to the DNLR secretary. Since then, the registry has used a web-based reporting system, which allows for programmed validation checks. Up to five registration forms are completed during the disease course of an AML patient. The first form is completed at diagnosis, the second when first-line treatment (in chemotherapy-treated patients) is completed, another in case of relapse, and a last form is completed at follow-up, death, or termination of outpatient follow up. An additional form is completed for alloHCT patients. Until May 2012, clinicians entered the cytogenetic data. To optimize data quality, these data are now entered by the clinical cytogeneticist. Depending on treatment schedule and clinical course, data on more than 150 variables (eg, patient- and leukemia-related, treatment- and outcome-related) can be recorded for each patient (Table 1).

**Table I** Data recorded on five registration forms used by the Danish National Leukemia Registry

Registration form and time of registration	Variables
<b>Registration form</b>	Name, civil registration (CPR) number, and demographic data
At diagnosis	Diagnosis according to WHO (2008)/ICD-10
From 2011 the cytogenetic details are included in a separate registration form	Date of diagnosis and date of first visit to a health care unit
	Prior hematological or solid cancer (specified, if present)
	Prior treatment with radio- or chemotherapy
	WHO Performance Status Score (WHO PS)
	Height and body weight
	Presence of extramedullary leukemia (EML) (specified, if present)
	FAB-type, blast percentage in the blood and bone marrow
	Results of cytogenetic evaluation (conventional cytogenetics and FISH)
	Leucocyte and platelet counts and level of lactate dehydrogenase (LDH)
	Planned chemotherapy (yes/no)
	Date of initiated treatment/decision of "no treatment"
<b>Treatment registration form</b>	Treatment intent (palliative or curative)
After completed first treatment	Date of first initiation of treatment
Only completed in chemotherapy treated patients	Protocol inclusion (yes, no and specified)
	For each cycle of combination chemotherapy, if given:
	Date of initiation, type and dose
	Classification of bone marrow response
	Date of evaluation
<b>Relapse registration form</b>	Date of relapse
At relapse	Treatment intent (palliative or curative)
If chemotherapy is initiated the registration form is filed after completion	For each cycle of combination chemotherapy, if given:
	Date of initiation, type and dose
	Classification of bone marrow response
	Date of evaluation
<b>Transplantation registration form</b>	Date of transplantation
In case of bone marrow transplantation	Type of transplantation (myeloablative, reduced intensity conditioning, or autologous)
	Donor type (sibling/registry)
	Type of transplant (bone marrow/peripheral stem cells)
<b>Follow up registration form</b>	Status (dead/alive)
At death or termination of follow up as an outpatient	Date of death
	Cause of death

**Abbreviations:** ICD, International Classification of Diseases; FISH, fluorescence in situ hybridization; WHO, World Health Organization; FAB, French-American-British.

## Central registries of importance to DNLR

### The Civil Registration System

Since 1968, Danish residents have been assigned a unique ten-digit civil registration (CPR) number at birth or immigration, which encodes date of birth and sex. The Civil Registration System (CRS) is administered by the Danish National Board of Health, which tracks data that include information about vital status and residential area. These electronic records are updated daily. The CPR number is recorded when information about, for example, social service and health care is entered into Danish medical or administrative databases.<sup>10,11</sup> The CPR number allows linkage of records between all medical registries.

### The Danish National Registry of Patients (DNRP)

The DNRP, which is administered by the Danish National Board of Health, includes data on all hospital admissions since 1977. Outpatient hospital visits have been included

since 1995. Data include CPR numbers and diagnosis at discharge (International Classification of Diseases [ICD]-10 since 1994).<sup>8</sup> The registry is updated monthly.

### The Danish National Pathology Registry

The Danish National Pathology Registry and its national online registration database, the Danish Pathology Data Bank (Patobank), include detailed nationwide records of all specimens analyzed for pathology since 1997. The registry is almost complete, and can be used for precise and efficient determination of specimen location. To ensure quality, all diagnostic descriptions are approved by a pathologist.<sup>12,13</sup> The data bank is accessible from computerized medical records at most hospitals.

## Coverage (registration completeness)

The reference population, in relation to coverage by the DNLR, was defined as patients recorded with AML in

the DNRP (ICD-10 codes C92.0–C94.9). To ensure high coverage and completeness of forms, the registry is linked to the CRS and the DNRP every 3 months. Reminders are sent to the departments that include lists of patients yet to be recorded (ie, newly diagnosed AML, according to the DNRP), reminders to fill in the treatment registration form (when “yes” is entered for the variable “planned chemotherapy?” in the registration form), and reminders to register follow-up when the CRS categorizes a patient as “dead.”

## Validation of variables and data quality

We included all patients diagnosed between January 1, 2000 and December 31, 2011 and who were recorded in the DNLR with a diagnosis consistent with AML ( $n = 2,665$ ). A computer-generated random sample of 230 patients treated with curative intent (or in which intent of treatment was missing) was drawn to determine the degree of agreement between information in medical records and in the DNLR. To minimize selection bias, patients with missing data (eg, treatment registration form not yet completed) were included. In addition, a sample of 30 palliatively treated patients (diagnosed in the Central and in the Northern Denmark regions) was randomly selected to examine differences in registry data quality for the variable “intent of treatment.” These 260 patients corresponded to 10% of all patients recorded during the 12-year period. For validation, we selected 30 variables that we regarded to be most important for studying prognosis (Table 2). Data on vital status and date of death were not included in the study, because this information is routinely drawn from the CRS and linked to the DNLR.<sup>11</sup>

The variables selected from the DNLR were validated using information from medical records (eg, paper records, electronic records, Patobank, and LABKA [clinical laboratory information system]) as the gold standard (Table 2). During validation, one of four outcomes was recorded: consistent with reference standard, inconsistent with reference standard, missing value (information present in reference), or not relevant (eg, “prior cancer, specified,” when no antecedent solid cancer). All medical records were reviewed by one clinician (LSO). Questions about medical records were discussed with a hematological specialist, and a consensus was reached.

## Statistical methods

Registration completeness was estimated as the number of patients diagnosed before January 2012 and recorded in

the DNLR by the end of September 2012, divided by the number of patients diagnosed with AML (according to the DNRP) before January 2012, after verification of AML diagnoses recorded in the DNRP but not in the DNLR. The PPV for the selected variables was calculated as the number of patients with identical information recorded in the DNLR, and the reference divided by the number of patients for which information was recorded in the medical records.<sup>7</sup> The completeness of the individual variables was calculated as the number of patients with information recorded in the DNLR divided by the total number of patients with information on the variable that were recorded in the reference. To examine whether data quality differed by intent of treatment (curative versus palliative), by department or by year of reporting to the DNLR, we repeated the analyses stratified by these groups. The Jeffreys method was used to estimate 95% confidence intervals (CIs).<sup>14</sup> We obtained information about Danish population size per calendar year from Statistics Denmark (StatBank).<sup>9</sup> The annual AML incidence was calculated by dividing the number of AML patients diagnosed in a given year by the total number of Danish citizens by calendar year (per 100,000). Patients were followed until death or emigration or until August 31, 2012, whichever occurred first. Data were entered into EpiData (EpiData Association, Odense, Denmark). Stata 11 software (StataCorp, College Station, TX, USA) was used for the statistical analyses.

## Ethics

DNLR registration is approved by the National Board of Health and the Danish Data Protection Agency. The Danish Data Protection Agency (2011-41-6456) approved the study, and the National Board of Health (Department of Monitoring and Patient Safety) approved access to medical records.

## Results

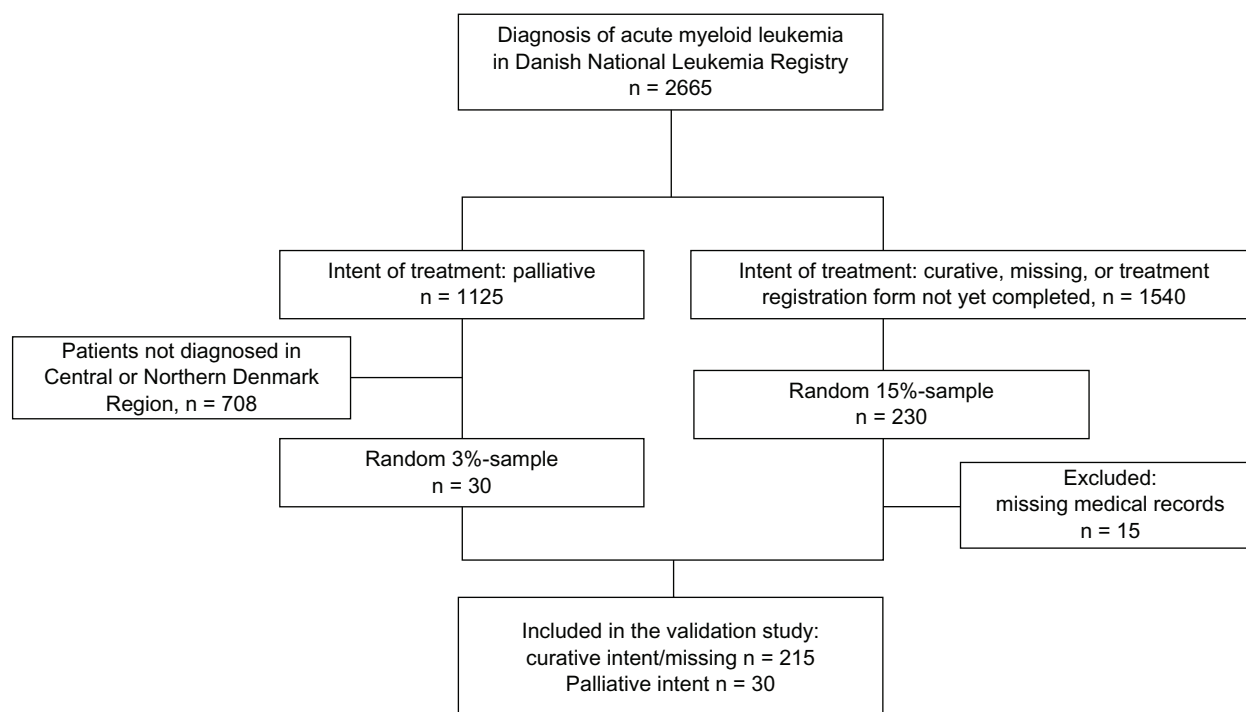
We were unable to retrieve 15 of 260 patient medical records (twelve at Rigshospitalet, two at Odense University Hospital, and one at Aarhus University Hospital) from the DNLR. These patients were excluded from the study. All patients were confirmed to have AML by bone marrow findings recorded in Patobank (Figure 1). If one or more variables from a patient could not be evaluated because of missing information (eg, single pages missing), the patient was not included when computing PPV for this specific variable. A patient can be treated for as long as a year. Therefore, patients who were diagnosed in 2011 but did not have a

**Table 2** Accepted values and gold standards for the validation of 30 variables from the Danish National Leukemia Registry

Variable	Correct specification of registered data	Primary gold standard
Diagnosis	AML unspecified or specification of sub diagnosis <sup>15</sup>	Patobank
Time of diagnosis	Date of diagnostic bone marrow sample ( $\pm 24$ hours)	Patobank
Cytogenetic result	Normal, abnormal, or not done	Medical records including photocopy of diagnostic cytogenetic results
Cytogenetic result, specified	Major specific changes correctly specified Specified clonal changes lead to correct grouping according to Grimwade's criteria <sup>16</sup>	Medical journals including copy of diagnostic cytogenetic results
Prior hematological disease	Yes/no/uncertain	Medical records and Patobank
Prior hematological disease, specified	MDS, CMML, ET, PV, MF, other (specified) Disease must be diagnosed more than three month prior to AML	Medical records and Patobank
Prior cancer, other than hematological	Yes/no/uncertain (lack of registration of basal cell carcinoma was accepted)	Medical records and Patobank
Prior cancer, specified	ICD 10 diagnosis (lack of registration of basal cell carcinoma was accepted)	Medical journals and Patobank
Prior chemotherapy	Yes/no/uncertain	Medical journals (paper and electronic)
Prior radiotherapy	Yes/no/uncertain	Medical journals (paper and electronic)
WHO performance status	Exact status, or description in accordance with 0, 1, 2, 3, 4, or 5	Medical journals (paper and electronic)
Extramedullary leukemia	Defined as a leukemic tumor mass or infiltrate at an anatomical site other than the blood and bone marrow Interpreted as "not present" when both references contained no information regarding EML	Medical journals and Patobank
Extramedullary leukemia, specified	Skin, oral, CNS, liver, spleen, lymph nodes, testis, other	LABKA
White blood cell count $10^9/L$	On the day of diagnosis. If not monitored, values within 2 days of diagnosis (before chemotherapy initiated) were accepted	LABKA
Platelet count $10^9/L$	At the day of diagnosis. If not monitored, values within 2 days of diagnosis (before chemotherapy initiated) were accepted	LABKA
Lactate dehydrogenase (U/L)	At the day of diagnosis. If not monitored, values within 2 days of diagnosis (before chemotherapy initiated) were accepted	LABKA
Weight	Weight in kg ( $\pm 1$ kg)	Medical journals (including chemotherapy treatment plan)
Height	Height in cm ( $\pm 1$ cm)	Medical journals (including chemotherapy treatment plan)
Curative intent?	Yes/no	Medical journals
Protocol participation?	Included for patients diagnosed in 2006 or later Defined as participation in clinical trial or protocol Yes/no, as well as specified	Medical journals (including chemotherapy treatment plan)
Time of initiated treatment	Exact date	Medical journals (including chemotherapy treatment plan)
1st course of induction chemotherapy, specified	Specified combination regimens (categorical)	Medical journals (including chemotherapy treatment plan)
Dose of 1st course of induction chemotherapy, specified	100%, 75%, 50%, 25%	Medical journals (including chemotherapy treatment plan)
Bone marrow response after 1st course of chemotherapy	Complete remission/partial remission/stable disease/progressive disease/not evaluated <sup>17</sup>	Medical journals and Patobank
2nd course of induction chemotherapy, specified	Specified combination regimens (categorical)	Medical journals (including chemotherapy treatment plan)
Dose of 2nd course of induction chemotherapy, specified	100%, 75%, 50%, 25%	Medical journals (including chemotherapy treatment plan)
Bone marrow response after 2nd course of chemotherapy	Complete remission/partial remission/stable disease/progressive disease/not evaluated <sup>17</sup>	Patobank and medical journals
Relapse?	If yes, date specified ( $\pm 48$ hours)	Patobank and medical journals
Extramedullary leukemia at relapse?	Yes/no, and specified if present	Patobank and medical journals
Cause of death	Within 1 week of induction chemotherapy/more than 1 week after induction chemotherapy/progressive disease/treatment-related death in complete remission, other cause (specified)	Medical journals (paper and electronic) and Patobank (autopsy results)

**Abbreviations:** MDS, myelodysplastic syndrome; CMML, chronic myelomonocytic leukemia; ET, essential thrombocythemia; PV, polycythemia vera; MF, myelofibrosis; ICD, International Classification of Diseases; EML, extramedullary leukemia; CNS, central nervous system; LABKA, clinical laboratory information system; WHO, World Health Organization; AML, acute myeloid leukemia.





**Figure 1** Sampling strategy for the validated population. Patients treated with curative intent were validated on a national level. To examine differences in registration quality with regards to treatment intent, a small sample of palliatively treated patients were included. To minimize selection bias patients with missing information for intent of treatment were included in the study.

completed treatment registration form at the time of data retrieval were excluded from validation for the variables on the treatment and relapse registration forms. The frequency of “missing” values in the DNLR for these cases and for the 15 patients with missing medical records did not differ from the remaining patients.

DNLR coverage was 99.6% (95% CI 99.3–99.8), which corresponded to ten missing cases of verified AML that were recorded in the registry. Additionally, ten AML cases (0.4%) in the DNRP were found to be misclassified. The correct diagnoses were MDS and myeloproliferative neoplasia. For our sample from the DNLR, only one patient was found to be misclassified as AML (natural killer-cell leukemia treated as AML), yielding an AML diagnosis PPV of 99.6% (95% CI 98.1–100).

Table 3 shows the overall completeness and PPV for each of the 30 variables included in this study. The completeness of entered data ranged from 60.7% (95% CI 42.3%–77.0%) for “prior cancer, specified” to 100% for diagnosis, time of diagnosis, prior chemotherapy, prior radiation therapy, and World Health Organization (WHO) performance status score. Completeness was greater than 90% for 23 out of 30 variables. PPVs ranged from 89.4% (95% CI 85.1%–92.8%) for date of diagnosis ( $\pm 24$  hours) to 100% (95% CI 98.8%–100%) for body weight. PPV was greater than 90% for 29 of 30 variables.

The variable “prior solid cancer” had a high frequency of “missing/uncertain” entered in the database. For 54 of the 58 patients with a missing value for prior solid cancer, the patient had no prior cancer. Four patients (7%) had a confirmed antecedent cancer diagnosis. Completeness increased to 100% (95% CI 99.0%–100%) if all missing/uncertain entries were categorized as “no prior cancers,” and PPV decreased slightly from 97.3% (95% CI 94.2%–99.0%) to 96.7% (95% CI 93.9%–98.4%). In almost 10% of the cases, the date recorded was the date the pathologist confirmed the diagnosis instead of the defined date of diagnosis, ie, the date of diagnostic bone marrow examination. If dates of  $\pm 6$  days from the exact date of diagnosis were accepted as correct, PPV increased from 89.4% (95% CI 85.1%–92.8%) to 96.3% (95% CI 93.4%–98.2%).

We found no systematic difference in data quality after stratification by each of the variables relevant for both groups for intent of treatment (data not shown). Stratification by department revealed that two departments had low completeness for date of relapse: 63.3% (95% CI 49.3%–75.7%) and 70.7% (95% CI 52.9%–83.2%) compared with 86.4%–100% for the other six departments. This difference was due to lower frequency of registration-form submission from these two departments. There was no significant difference between departments for information on treatment, but six

**Table 3** Completeness and positive predictive value (PPV) of 30 variables from the Danish National Leukemia Registry

Variable	Number of correctly coded records/number of relevant records reviewed	Completeness (%) (95% CI)	PPV (%) (95% CI)
Diagnosis	244/245	100 (99.0; 100)	99.6 (98.1; 100)
Time of diagnosis ( $\pm 24$ hours)	219/245	100 (99.0; 100)	89.4 (85.1; 92.8)
Time of diagnosis ( $\pm 6$ days)	236/245	100 (99.0; 100)	96.3 (93.4; 98.2)
Cytogenetic result	220/231	99.6 (98.0; 99.9)	95.2 (91.9; 97.4)
Cytogenetic result, specified	224/231	99.1 (97.3; 99.8)	97.0 (94.4; 98.6)
Prior hematological disease	231/240	98.0 (96.6; 99.2)	96.3 (93.2; 98.1)
Prior hematological disease, specified	42/44	83.0 (71.3; 91.3)	95.5 (86.2; 99.0)
Prior cancer, other than hematological	182/186	75.9 (70.3; 80.1)	97.8 (95.0; 99.3)
Prior cancer (missing/uncertain categorized as "no")	238/245	100 (99.0; 100)	97.1 (94.5; 98.7)
Prior cancer, specified	16/17	60.7 (42.3; 77.0)	94.1 (75.7; 99.4)
Prior chemotherapy	234/244	100 (99.0; 100)	95.9 (92.9; 97.9)
Prior radiotherapy	239/244	100 (99.0; 100)	98.0 (95.6; 99.2)
WHO performance status	236/241	100 (99.0; 100)	97.9 (95.5; 99.2)
Extramedullary leukemia	217/230	96.2 (93.2; 98.1)	94.3 (90.8; 96.8)
Extramedullary leukemia, specified	27/29	72.5 (57.5; 84.4)	93.1 (79.7; 98.6)
White blood cell count ( $1 \times 10^9/L$ )	218/232	97.9 (95.4; 99.2)	94.0 (90.3; 96.5)
Platelet count ( $1 \times 10^9/L$ )	211/231	97.5 (94.9; 98.9)	91.3 (87.2; 94.5)
Lactate dehydrogenase (U/L)	157/161	67.9 (61.8; 73.6)	97.5 (94.2; 99.2)
Weight	216/216	90.4 (86.2; 93.6)	100 (98.8; 100)
Height	214/215	90.0 (85.7; 93.2)	99.5 (97.8; 99.9)
Curative intent?	229/233	97.5 (94.9; 98.9)	98.3 (96.0; 99.4)
Protocol participation?	116/119	94.4 (89.4; 97.5)	97.5 (93.4; 99.3)
Time of initiated treatment	205/206	98.6 (96.2; 99.6)	99.5 (97.8; 99.8)
1st course of induction chemotherapy, specified	194/204	97.1 (94.2; 98.8)	95.1 (91.5; 97.5)
Dose of 1st course of induction chemotherapy, specified	203/204	97.1 (94.2; 98.8)	99.5 (97.7; 99.9)
Bone marrow response, after 1st course of chemotherapy	202/205	97.1 (94.2; 98.8)	98.5 (96.1; 99.6)
2nd course of induction chemotherapy, specified	195/202	96.2 (92.9; 98.2)	96.5 (93.3; 98.4)
Dose of 2nd course of induction chemotherapy, specified	201/203	96.7 (93.6; 98.5)	99.0 (96.9; 99.8)
Bone marrow response, after 2nd course of chemotherapy	197/200	95.2 (91.7; 97.5)	98.5 (96.1; 99.6)
Relapse? (if yes, date specified)	157/161	94.7 (90.6; 97.3)	97.5 (94.2; 99.2)
Extramedullary leukemia at relapse?	53/55	88.7 (79.1; 94.8)	96.4 (88.9; 99.2)
Cause of death	136/146	81.1 (74.9; 86.3)	93.2 (88.2; 96.4)

**Abbreviation:** CI, confidence interval; WHO, World Health Organization.

patients without timely submission of treatment registration forms were treated at the same department. There was no difference between departments for accuracy of reporting. Stratification by year of registration in the DNLR (before 2006 versus 2006 and later) revealed that in the latter period there was an improvement in completeness and PPV for six variables. PPV of date of diagnosis improved from 86.7% (95% CI 75.4%–90.0%) to 93.2% (95% CI 88.3%–96.4%), and for cytogenetic result PPV increased from 90.5% (95% CI 83.4%–95.2%) to 98.5% (95% CI 95.4%–99.7%). The completeness of prior solid cancer improved from 40.8% (95% CI 31.5%–50.7%) to 100% (95% CI 97.5%–100%), and the completeness of lactate dehydrogenase improved from 33.0% (95% CI 24.1%–42.9%) to 90.1% (95% CI 85.4%–94.8%).

The overall incidence of AML was 5.4/100,000 person-years (95% CI 4.99–5.74) in persons 15 years or older.

The sex-adjusted incidence rate was 6.2/100,000 person-years (95% CI 5.8–6.6) for male patients and 4.9/100,000 person-years (95% CI 4.5–5.4) for female patients. The sex-specific incidence rates varied over the years, but the combined incidence remained stable. The yearly incidence rates of AML found in our study are higher than the incidences reported from Sweden (4.5 and 4.2/100,000 in men and women, respectively).<sup>18</sup> Incidence rates in Caucasians in the US of 4.5 and 3.1/100,000 in men and women, respectively, are even lower.<sup>19</sup>

## Discussion

Our study showed that the DNLR is highly complete and valid. The DNLR includes almost 100% of AML patients in Denmark, and the PPV of an AML diagnosis in the registry is 99.6%. For specific variables, completeness ranged from 60.7% to 100%, and PPVs ranged from 89.4% to 100%.

To our knowledge, there are only two other large population-based AML registries in existence. The largest, the US Surveillance, Epidemiology, and End Results (SEER) cancer registry, has recorded more than 20,000 patients. However, this registry has several limitations that include lack of validation of the diagnosis and limited clinical data.<sup>19–21</sup> The Swedish Leukemia Registry contains data on more than 3,300 patients that were recorded between 1997 and 2006. The accuracy of the values for the variables that have been entered into these two registries is not known.<sup>2,22</sup> The sex-specific incidences found in this study were higher than reported incidences based on data from the Swedish Leukemia Registry and from the SEER registry. AML incidence may vary by geographic location, but we expected that the differences would be small. The coverage of the referenced data sources may explain the observed differences in incidence. The lower incidence reported from the SEER registry can be explained by a general and systematic underreporting of myeloid leukemias. Craig et al estimated a SEER coverage of 30%–50%. In addition, most secondary leukemias are not recorded, because prior to 2009 more than one cancer diagnosis could not be entered for the same patient.<sup>21</sup> The Swedish Leukemia Registry has a reported coverage of 98%, which is only slightly lower than the coverage found in our study (99.6%).<sup>2</sup> We used the DNRP as a reference in relation to coverage by the DNLR, because the DNRP is considered more complete for AML data than the Danish Cancer Registry, and has fewer false positive diagnoses than the Danish National Pathology Registry.<sup>1</sup> In contrast, the Swedish Leukemia Registry uses the Swedish Cancer Registry as a reference.<sup>2</sup> In a validation study, Aström et al found that the Swedish Cancer Registry underreported 15.8% of AML cases. After combining data from the Swedish Cancer Registry and the Swedish Registry of Death, the total incidence increased to 5.4/100,000 person-years (males 5.9, females 4.9), which is almost identical to the incidence reported in our study.<sup>23</sup> This result could indicate that the Danish Leukemia Registry is more complete than other registries.

In this study, the sensitivity of the AML diagnosis could not be directly estimated, but the completeness is an estimate of sensitivity. Specificity could not be computed, because we did not know the true incidence of AML in the general population. However, specificity is very likely close to 100%, since few persons in the background population will remain undiagnosed because of the severity of the disease. Furthermore, the disease is rare and the background population is large. The specificity of AML diagnosis in the DNLR is therefore assumed to be close to 100%.<sup>7</sup>

The variable “prior solid cancer” had a high frequency of “missing/uncertain” cases. In the majority of missing cases, the patient did not have a prior cancer diagnosis. By categorizing “missing” as “no prior cancer,” the completeness increased to nearly 100% without a substantial decrease in the PPV. The variable “prior cancer, specified” had the lowest completeness of the study variables (60.7%), which was able to be improved by supplementing it with data from the Danish Cancer Registry. The accuracy of almost all of the 30 variables selected from the DNLR was very high. Only date of diagnosis had a PPV that was slightly less than 90%. This variable was affected by a systematic recording error, which was not a factor in the later years.

Strengths of this study include extensive review of medical records and extensive use of detailed clinical data, pathology, and laboratory results. We validated a high number of variables that are regarded to be important for the study of AML prognosis. The study included 12 of 13 years of database information and represented 10% of the patients in the DNLR registry. Medical records (paper and electronic records, including Patobank and LABKA) were chosen as the gold standard, because these data sources represent primary data collection and are accessible to clinicians when patients are recorded in the DNLR. Only a small sample of the included patients were treated with palliative intent. We did not find any difference in data quality between patients treated with curative and palliative intent. The study included palliatively treated patients from only two of the five Danish regions, but we do not expect coding errors or registration completeness for these patients to be different from patients from other regions.

The diagnostic process for hematological diseases can be complex, and the borderline between possible precursor conditions, such as MDS and chronic myelomonocytic leukemia, is somewhat arbitrary. Furthermore, some patients that do not have frank leukemia are treated with AML protocols. This issue may lead to misclassification of these diseases as AML.<sup>24</sup> One reason for the impressive coverage and completeness of the DNLR is that reminder lists are sent from the registry to the clinical departments that report CPR numbers of newly recorded (in the DNRP) AML patients. Patient information is validated by clinicians during registration in the DNLR. If a patient is found to be incorrectly recorded as AML in the DNRP, the DNRP is notified and the diagnosis is corrected. Other strengths of the DNLR include the population-based design and the size of the registry, especially considering that AML is a rare disease. Data are easily accessible, and the individual form computerized data format facilitates linkage to other registries. The registry has only been in operation for 13 years; however, AML has a



high mortality, so the follow-up period was complete for a large proportion of the patients (77.3% of patients were dead or had emigrated at the time of data withdrawal). The mean follow-up time for all patients was 635 days.

The values for the variables included in the DNLR are, with few exceptions, very accurate compared with other Danish clinical databases.<sup>25,26</sup> PPV and completeness of AML diagnosis were considerably higher compared with results from a validation study of hematological diagnosis recorded in the DNRP from 1994 to 1999 (completeness 89.0%, 95% CI 80.4%–94.1%; PPV 67.6%, 95% CI 58.3%–75.7%).<sup>24</sup>

Observational studies cannot replace clinical protocols or trials, but they do have a well-accepted role in medical research, especially for the study of risk factors, diagnosis, and prognosis.<sup>6</sup> The large amount of validated clinical and research-relevant data recorded for unselected patients in the DNLR makes it a valuable tool for research of the prognosis and clinical course of AML. Additional parameters of interest (eg, social status, comorbidity) can also be linked to the DNLR from other Danish registries.

## Conclusion

The coverage of the DNLR was remarkably high. With few exceptions, high values were obtained for PPV and completeness of recorded data. For the few parameters with lower completeness, data quality may be improved by including data from other Danish national registries. This study supports the importance of the DNLR as a resource for future clinical AML research.

## Acknowledgments

This work was supported by grants from the Danish Cancer Society, the University of Aarhus Faculty of Health, and the Danish Acute Leukemia Group.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. The Danish Acute Leukemia Group. [Annual Leukemia Report]. Århus: ALG; 2012. Danish.
2. Juliusson G, Lazarevic V, Hörstedt AS, Hagberg O, Höglund M. Acute myeloid leukemia in the real world: why population-based registries are needed. *Blood*. 2012;119(17):3890–3899.
3. Estey E. Do commonly used clinical trial designs reflect clinical reality? *Haematologica*. 2009;94(10):1435–1439.
4. Tsimberidou AM, Estey E. Relevance of clinical trials in acute myeloid leukaemia. *Hematol Oncol*. 2008;26(3):182–183.
5. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med*. 1999;341(27):2061–2067.
6. Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology*. 2006;44(5):1075–1082.
7. Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*. 1996;25(2):435–442.
8. Andersen TF, Madsen M, Jørgensen J, Mellemkjær L, Olsen JH. The Danish National Hospital Register. A valuable source of data for modern health sciences. *Dan Med Bull*. 1999;46(3):263–268.
9. Statistics Denmark. Population and elections. Available from: <http://www.statbank.dk/statbank5a/default.asp?w=1366>. Accessed June 18, 2013.
10. Frank L. Epidemiology. When an entire country is a cohort. *Science*. 2000;287(5462):2398–2399.
11. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health*. 2011;39(Suppl 7):22–25.
12. Erichsen R, Lash TL, Hamilton-Dutoit SJ, Bjerregaard B, Vyberg M, Pedersen L. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. *Clin Epidemiol*. 2010;2:51–56.
13. Bjerregaard B, Larsen OB. The Danish Pathology Register. *Scand J Public Health*. 2011;39(Suppl 7):72–74.
14. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Stat Sci*. 2001;16(2):101–133.
15. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*. 2011;117(19):5019–5032.
16. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116(3):354–365.
17. Cheson BD, Bennett JM, Kopecky KJ, et al. Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J Clin Oncol*. 2003;21(24):4642–4649.
18. Derolf AR, Kristinsson SY, Andersson TM, Landgren O, Dickman PW, Björkholm M. Improved patient survival for acute myeloid leukemia: a population-based study of 9729 patients diagnosed in Sweden between 1973 and 2005. *Blood*. 2009;113(16):3666–3672.
19. Surveillance Epidemiology and End Results [website on the Internet]. Available from: <http://seer.cancer.gov>. Accessed June 18, 2013.
20. Dore GM, Devesa SS, Curtis RE, Linet MS, Morton LM. Acute leukemia incidence and patient survival among children and adults in the United States, 2001–2007. *Blood*. 2012;119(1):34–43.
21. Craig BM, Rollison DE, List AF, Cogle CR. Underreporting of myeloid malignancies by United States cancer registries. *Cancer Epidemiol Biomarkers Prev*. 2012;21(3):474–481.
22. Lehmann S, Lazarevic V, Hörstedt A, et al. Poor outcome in secondary acute myeloid leukemia (AML): A first report from the population-based Swedish acute leukemia registry. *ASH Ann Meet Abstr*. 2012;120:130.
23. Åström M, Bodin L, Tidefelt U. Adjustment of incidence rates after an estimate of completeness and accuracy in registration of acute leukemias in a Swedish population. *Leuk Lymphoma*. 2001;41(5–6):559–570.
24. Nørgaard M, Skriver MV, Gregersen H, Pedersen G, Schønheyder HC, Sørensen HT. The data quality of haematological malignancy ICD-10 diagnoses in a population-based hospital discharge registry. *Eur J Cancer Prev*. 2005;14(3):201–206.
25. Lamberg AL, Cronin-Fenton D, Olesen AB. Registration in the Danish Regional Non-melanoma Skin Cancer Dermatology Database: completeness of registration and accuracy of key variables. *Clin Epidemiol*. 2010;2:123–136.
26. Pedersen CG, Gradus JL, Johnsen SP, Mainz J. Challenges in validating quality of care data in a schizophrenia registry: experience from the Danish National Indicator Project. *Clin Epidemiol*. 2012;4:201–207.

**Clinical Epidemiology****Dovepress****Publish your work in this journal**

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

reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, 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.

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