Published in final edited form as:

Pediatr Blood Cancer. 2013 March ; 60(3): 508–511. doi:10.1002/pbc.24402.

Assembly of a Cohort of Children Treated for Acute Myeloid Leukemia at Free-Standing Children's Hospitals in the United States Using an Administrative Database

Marko Kavcic, MD^{1,*}, Brian T. Fisher, DO, MSCE^{2,3,4}, Kari Torp, PhD¹, Li Yimei, PhD¹, Yuan-Shung Huang, MS⁴, Alix E. Seif, MD, MPH^{1,4}, Marijana Vujkovic, PhD¹, and Richard Aplenc, MD, PhD^{1,3,4,5}

¹Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

²Division of Infectious Diseases, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

³The Center for Pediatric Clinical Effectiveness, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

⁴Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

⁵Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

Abstract

Pediatric Health Information System data were used to establish a multi-center cohort of 1,686 children treated for newly diagnosed acute myeloid leukemia (AML). The cohort assembly process, which included myeloid leukemia ICD-9 discharge diagnosis codes and manual review of induction chemotherapy, was validated by chart review at a single institution. The use of ICD-9 codes alone resulted in a poor positive predictive value (PPV; 31%). Inclusion of the results from the chemotherapy review improved the PPV to 100% without compromising sensitivity (95.7%). This cohort provides a reliable source for future comparative effectiveness and clinical epidemiology studies in pediatric AML.

Keywords

cohort validation; health administrative data; pediatric acute myeloid leukemia

INTRODUCTION

Survival in children with acute myeloid leukemia (AML) has improved significantly over the past several decades from <20% to over 60%, largely due to increased therapeutic intensity [1–5]. This increased intensity comes with a substantial risk of treatment-related morbidity and mortality [6,7]. Thus, clinical epidemiology and comparative effectiveness studies are becoming increasingly more important to further improve outcomes. However, such studies require large cohorts with comprehensive resource utilization data [8].

Administrative databases represent a viable source of data for assembly of such cohorts. However, standardized approaches for valid cohort assembly are necessary to reduce misclassification [9]. Therefore, we assembled and validated a representative cohort of patients treated for newly diagnosed AML at freestanding children's hospitals across the United States using the Pediatric Health Information System (PHIS) database. PHIS data include 85% of all hospitalizations in free-standing children's hospitals in the United States registered by the Children's Hospitals Association (CHA) [10]. It has been used in over 200 peer-reviewed publications including studies of renal failure in pediatric AML [11], and end of life care in children with cancer [12].

METHODS

Data Source

PHIS is a comparative administrative database including inpatient data from 42 tertiary children's hospitals affiliated with the CHA. PHIS data are derived from the hospital's medical record system, which provides patient identification, demographics, dates of service, discharge disposition, and up to 41 ICD-9 diagnosis and procedure codes. Additionally, the hospital's billing system provides resource utilization data for every patient for every inpatient day of service. These resource utilization data include all billed pharmaceuticals, laboratory tests (without results), imaging procedures (without results), and supplies. Each patient's vital status at the time of the hospital discharge is also recorded. Each patient is assigned a unique identifier in the PHIS database. Therefore, patients can be followed across inpatient admissions. Oversight of PHIS data quality is a joint effort between the CHA (data management center), Thomson Reuters Healthcare (data processing partner), and participating hospitals [13].

Cohort Assembly

Children (<19 years) admitted to one of 43 PHIS-contributing hospitals between January 1999 and March 2010 were considered for inclusion if they had at least one hospitalization that included an International Classification of Diseases-9 (ICD-9) discharge code for myeloid or unspecified leukemia (205.xx–208.xx). Patients with evidence of an alternative malignancy (ICD-9 codes: 140.xx–239.xx, excluding 205.xx–208.xx) or bone marrow transplantation (BMT; ICD-9 codes: v42.81, v42.82, 279.50, or 996.85; procedure code 41.0.x; pharmacy codes for: busulfan, melphalan, and thiotepea) within 60 days of the index admission were excluded.

Next, chemotherapy during the first 120 days was reviewed for each patient and compared against established AML induction regimens [2,3,14]. Induction regimens were categorized as: ADE (cytarabine, daunorubicin, and etoposide) with or without gemtuzumab ozogamicin, intensively timed DCTER (two cycles of dexamethasone, cytarabine, thioguanine, etoposide, and rubidomycin [daunomycin] or idarubicin given 10 days apart), DAT (daunorubicin, cytarabine, thioguanine), 7 + 3 (cytarabine and daunorubicin), and other: standard-timed DCTER and MAE (mitoxantrone, cytarabine, and etoposide). Patients who did not receive chemotherapy consistent with described regimens were excluded.

Cohort Validation

The positive predictive value (PPV) was calculated by reviewing the charts of PHIS identified patients admitted to the Children's Hospital of Philadelphia (CHOP) from 2004 (when CHOP joined PHIS) through 2009. Members of the study team reviewed bone marrow pathology results, attending oncologists' progress notes, and medication records to determine the actual malignancy diagnosis and induction regimen administered. In order to calculate the sensitivity of PHIS cohort assembly, patients identified in PHIS were compared

against the CHOP AML registry for the same time period. Chart abstractionists were blinded to PHIS pharmacy billing data at the time of chart review. The negative predictive value ($TN/[FN + TN]$) was not calculated as the result of such a calculation would be predictably high for two reasons. First, the PHIS database contains millions of non-AML admissions and second, AML is a relatively rare disease. The former would result in a high number of true-negative patients and the latter into a relatively low number of false negative patients. Likewise, in this scenario the specificity ($TN/[TN + FP]$) would be predictably high as again there would be millions of patients that are TN and relatively few that are FP. It is important to note that the actual number of false negatives (e.g., those patients with new onset AML that were not included in the final cohort) could not be quantified by the data available.

Induction case fatality rate was determined for the subgroup of patients treated with ADE, the current standard induction regimen. The induction period was defined from the start of the first to the initiation of the third course of chemotherapy. Demographic characteristics (age, gender, and race) were compared with those reported by the National Cancer Institutes' SEER Program from 1992 to 2004 using Pearson's χ^2 test. Analyses were performed using Microsoft Excel, SAS version 9.2 (Cary, NC) and STATA statistical software version 10.0 (College Station, TX).

RESULTS

A total of 5,226 patients with an ICD-9 code consistent with myeloid leukemia were identified in the PHIS database. Patients without chemotherapy exposure during the first 120 days ($n = 1,326$), with presumed alternative malignancy ($n = 534$), or with bone marrow transplantation ($n = 627$) within 60 days of index admission were excluded. Finally, 1,053 patients were excluded due to chemotherapy patterns not consistent with AML therapy. This process yielded a cohort of 1,686 pediatric patients with presumed newly diagnosed AML. Most children (75%) were treated with ADE ($n = 931$) or intensively timed DCTER ($n = 333$) and the remaining received DAT ($n = 243$), 7 + 3 ($n = 84$), or other ($n = 95$) induction regimens. Of 44 patients included in the final cohort who were treated at CHOP, 35 were treated with ADE, 6 with 7 + 3 and 3 with DAT induction regimens.

At CHOP, 147 children identified in the PHIS database had ICD-9 discharge diagnosis codes consistent with myeloid leukemia during the validation period (2004–2009). Of these, 46 patients truly had newly diagnosed AML by chart review. The sensitivity of using ICD-9 codes alone was 100%, but the PPV was only 31%. After manual chemotherapy review, the PPV dramatically improved to 100% without significantly compromising sensitivity (only two false negatives). Similar findings were observed when the cohort was limited to patients receiving ADE (Table I).

Characteristics of patients who received ADE induction ($n = 931$) were comparable to those reported from the National Institute of Cancer's SEER Program for 1992–2004 (Table II). Induction fatality rate in that subgroup was 3.8%.

DISCUSSION

This study presents the first validated cohort of children with AML identified from a health care administrative database. Importantly, the observed induction case fatality rate in the homogenous subgroup of patients treated with ADE is similar to that reported by cooperative group clinical trials [3,4]. Although there were statistically significant differences in the comparison of our cohort's race to SEER data, the actual proportions were concordant. These statistical differences are likely the result of large sample sizes rather than clinically meaningful differences.

The 205.xx–208.xx series of codes was predictably sensitive but likely captured patients with AML at various stages such as *de novo*, secondary, or relapsed disease. Additionally, other types of myeloid leukemia such as chronic myeloid leukemia may be identified by these codes. Finally, these ICD-9 codes may also have resulted in the inclusion of patients that were incorrectly assigned a myeloid leukemia code because of initial diagnostic uncertainty. Therefore, although the resultant sensitivity was 100%, the extremely low PPV confirmed ICD-9 codes alone as unreliable for defining an AML cohort from administrative data. Reviewing daily pharmacy billing data to identify predefined chemotherapy patterns was a necessary step in cohort assembly, as evidenced by improved PPV and preserved sensitivity. Although such work is time-consuming, these data and previous work from our group [13] indicate that a similar approach is necessary to establish a cohort of patients with malignancy from an administrative database.

PHIS administrative data have the advantage of daily billing data to determine chemotherapy utilization but also have several limitations. First, since chemotherapy patterns were used as inclusion criteria, patients who died before the initiation of chemotherapy were not detected with our method. However, such events are rare (<1%) [15], and the impact of missing these patients is predicted to be minimal. Second, due to limited resources, we were only able to validate our cohort assembly at one institution. Since PHIS data capture may vary by center, a multi-institutional validation could alter estimates of sensitivity and PPV. Finally, we acknowledge that our process of establishing the AML cohort from the PHIS database had the potential to result in missing patients that had AML (e.g., false negatives). Such patients may have been erroneously excluded from cohort inclusion at multiple steps including inaccurate assignment of ICD-9 discharge diagnosis codes (e.g., a patient with AML was assigned an ICD-9 code consistent with ALL), inaccurate assignment of an ICD-9 code for an alternative malignancy, or because of inaccuracies in the chemotherapy billing data. Based on the data available from PHIS it is not possible to accurately quantify the number of false negative cases. However, validation of the cohort assembly process at our institution yielded a sensitivity of 100% for myeloid leukemia ICD-9 codes alone and the subsequent steps resulted in only 2 false negative cases. Therefore, we suspect that the total number of missed AML patients across all PHIS institutions was low.

In summary, the use of ICD-9 discharge diagnosis codes coupled with the systematic review of chemotherapy billing data contained in the PHIS database resulted in a large multi-center cohort of pediatric patients with newly diagnosed AML. Based on the excellent sensitivity and PPV estimates from the single center validation, our cohort represents a reliable source for future clinical epidemiology and comparative effectiveness analyses. Additionally, the multi-center nature of the cohort offers the opportunity to investigate practice variation among participating institutions. Finally, since the PHIS database contains day-by-day resource utilization billing data, it has the potential to serve as an efficient source of data to monitor treatment-related toxicities, and to determine the relative difference in healthcare expenditures between treatment arms in future pediatric AML randomized trials.

Acknowledgments

NIH; Grant number: R01CA165277; Grant sponsor: The Slovenian Ministry of Education, Science, Culture, and Sport; Grant number: J3-4220.

This study was supported by NIH (grant R01CA165277 (R.A.)) and by The Slovenian Ministry of Education, Science, Culture, and Sport (grant J3-4220 (M.K.)).

References

1. Kaspers GJL, Creutzig U. Pediatric acute myeloid leukemia: International progress and future directions. *Leukemia*. 2005; 19:2025–2029. [PubMed: 16304569]
2. Lange BJ, Smith FO, Feusner J, et al. Outcomes in CCG-2961, a children's oncology group phase 3 trial for untreated pediatric acute myeloid leukemia: A report from the children's oncology group. *Blood*. 2008; 111:1044. [PubMed: 18000167]
3. Cooper TM, Franklin J, Gerbing RB, et al. AAML03P1, a pilot study of the safety of gemtuzumab ozogamicin in combination with chemotherapy for newly diagnosed childhood acute myeloid leukemia: A report from the Children's Oncology Group. *Cancer*. 2012; 118:761–7769. [PubMed: 21766293]
4. Gibson BES, Webb DKH, Howman AJ, et al. Results of a randomized trial in children with Acute Myeloid Leukaemia: Medical Research Council AML12 trial. *Br J Haematol*. 2011; 155:366–376. [PubMed: 21902686]
5. Rubnitz JE, Inaba H, Dahl G, et al. Minimal residual disease-directed therapy for childhood acute myeloid leukaemia: Results of the AML02 multicentre trial. *Lancet Oncol*. 2010; 11:543–552. [PubMed: 20451454]
6. Sung L, Gamis A, Alonzo TA, et al. Infections and association with different intensity of chemotherapy in children with acute myeloid leukemia. *Cancer*. 2009; 115:1100–1108. [PubMed: 19156894]
7. Ethier M-C, Blanco E, Lehnbecher T, et al. Lack of clarity in the definition of treatment-related mortality: Pediatric acute leukemia and adult acute promyelocytic leukemia as examples. *Blood*. 2011; 118:5080–5083. [PubMed: 21937689]
8. Juliusson G, Lazarevic V, Hörstedt A-S, et al. Acute myeloid leukemia in the real world: Why population-based registries are needed. *Blood*. 2012; 119:3890–3899. [PubMed: 22383796]
9. Benchimol EI, Manuel DG, To T, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol*. 2011; 64:821–829. [PubMed: 21194889]
10. Bertoch, D. Personal communication. 2011.
11. Fisher BT, Zaoutis TE, Leckerman KH, et al. Risk factors for renal failure in pediatric patients with acute myeloid leukemia: A retrospective cohort study. *Pediatr Blood Cancer*. 2010; 55:655–661. [PubMed: 20533519]
12. Orsey AD, Belasco JB, Ellenberg JH, et al. Variation in receipt of opioids by pediatric oncology patients who died in children's hospitals. *Pediatr Blood Cancer*. 2009; 52:761–766. [PubMed: 18989880]
13. Fisher, BT.; Harris, T.; Torp, K., et al. [Accessed June 5, 2012] Establishment of an 11-year cohort of 8733 pediatric patients hospitalized at United States free-standing children's hospitals with de novo acute lymphoblastic leukemia from health care administrative data. *Med Care*. 2012. [Epub ahead of print] Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22410405>
14. Hann IM, Stevens RF, Goldstone AH, et al. Randomized comparison of DAT versus ADE as induction chemotherapy in children and younger adults with acute myeloid leukemia. Results of the Medical Research Council's 10th AML trial (MRC AML10). Adult and childhood leukaemia working parties of the Medical Research Council. *Blood*. 1997; 89:2311–2318. [PubMed: 9116274]
15. Creutzig U. Early deaths and treatment-related mortality in children undergoing therapy for acute myeloid leukemia: Analysis of the multicenter clinical trials AML-BFM 93 and AML-BFM 98. *Journal of Clinical Oncology*. 2004; 22:4384–4393. [PubMed: 15514380]

Estimates of Sensitivity and Predictive Value Using AML ICD-9 Codes Alone and in Combination With Manual Chemotherapy Review

TABLE I

	True positives	False positives	False negatives	Sensitivity (%)	Positive predictive value (%)
Step 1: ICD-9 discharge series of codes 205.xx-208.xx	46	101	0	100	31.3
Step 2a: Step1, lack of an ICD-9 diagnosis, procedure code or pharmacy billing codes consistent with an alternative malignancy or bone marrow transplant and induction chemotherapy review	44	0	2	95.7	100
Step 2b: Same as step 2a, but includes only patients with ADE induction chemotherapy	35	0	2	94.6	100

ADE, cytarabine, daunorubicin, and etoposide with or without gentuzumab ozogamicin.

Characteristics of Patients From PHIS-Assembled AML Cohort and National Cancer Institute^a SEER Program AML Data

TABLE II

	PHIS-assembled, ADE induction regimen (N = 931)		SEER 1992-2004 (N = 1,085)		P-value
	No.	%	No.	%	
Age category ^a					
<1	95	10.20	109	10.05	0.0841 ^a
1 and <5	244	26.21	279	25.71	
5 and <10	147	15.79	169	14.93	
10 and <15	267	28.68	231	21.29	
15 and <19	178	19.12	—	—	
15 and <20	—	—	297	27.37	
Gender					
Male	496	53.28	579	53.36	0.9685
Female	435	46.72	506	46.64	
Race					
White	643	69.07	808	74.47	0.0235
Black	124	13.32	125	11.52	
Other ^b	164	17.62	152	14.01	

PHIS, pediatric health information systems; ADE, cytarabine, daunorubicin, and etoposide with or without gemtuzumab ozogamicin; SEER, surveillance epidemiology and end results.

^aSEER cohort includes patients <20 years of age, PHIS-assembled cohort is limited to children less than 19 years of age. Thus, only age groups for patients <15 years old are considered in the statistical comparison;

^b“Other” include Asian/Pacific Island, American Indian, other and unknown.