Center-level variation in accuracy of adverse event reporting in a clinical trial for pediatric acute myeloid leukemia: a report from the Children's Oncology Group

While cure rates for pediatric acute myeloid leukemia (AML) range from 50-60% in Children's Oncology Group Trials (COG)[1,2] and 60%-70% in European cooperative groups,[3,4] AML therapy is very intensive and causes substantial treatment-related morbidity. We recently showed that adverse events (AE) are under-reported on cooperative group clinical trials for *de novo* pediatric AML.[5] Other investigators have also demonstrated heterogeneity in the types of AEs reported.[6,7] Given the limited resources available on cooperative group trials,[8] there is potential for significant variation between hospitals as resources and staffing for AE reporting are often impacted by hospital resources and priorities.[9] Center-level variation in cancer treatment outcomes has been demonstrated based on hospital volume, with improved outcomes at hospitals with greater volumes.[10] This study is an extension of our previously published work[5] and sought to determine if there is center-level variation in AE reporting in order to better understand under-reporting of AEs in pediatric AML clinical trials. We hypothesized that there would be variation in AE reporting between hospitals and across AEs, and that hospitals that treat more children with AML would report AEs more robustly. Using gold standard chart abstraction data collected on 12 AEs at 14 hospitals, this study found that AE reporting sensitivity varied by hospital for anorexia (P<0.001), viridans group streptococcal (VGS) bacteremia (P=0.016), and pain (P=0.001) with a trend toward improved sensitivity at larger centers (anorexia: OR 1.7, P<0.001; pain: OR 1.5, P=0.022). This study demonstrates center-level variation in AE reporting and concludes that the association between hospital volume and reporting accuracy should be evaluated in larger cohorts.

Children's Oncology Group (COG) trial AAML0531 randomized eligible patients to receive standard chemotherapy with or without gemtuzumab ozogamicin.[1] As per standard COG procedures, grades 3 and higher non-hematologic AEs were collected on AAML0531 *via* manual review of the medical record by clinical research associates (CRAs), as noted in our prior publication.[5] For the current study, two pediatric oncologists (T.P.M, M.K.) performed a retrospective medical record abstraction for all patients with available medical records at fourteen hospitals across the United States to evaluate the accuracy of AE reporting.[5] The following grade 3-4 AEs were examined on each inpatient day for all chemotherapy courses: hypertension, hypotension, hypoxia, acute respiratory distress syndrome (ARDS), microbiologically-proven viridans group streptococcal (VGS) bacteremia, microbiologically-proven sterile site invasive fungal disease (IFD), anorexia, typhlitis, disseminated intravascular coagulation (DIC), pain, seizure and acute renal failure. AEs were selected to represent all organ systems and a range of clinically important AEs. Prior to initiation of chart abstraction, AE definitions were developed based on the Common Terminology Criteria for Adverse Events (CTCAE) version 3.

Patient demographic, disease and treatment characteristics were obtained from the COG AAML0531 database. As previously reported, the distributions of patient characteristics of abstracted and non-abstracted patients on AAML0531 were compared using χ^2 or Wilcoxon rank sum tests.[5] Specifically, Compared with all patients in AAML0531, chart abstraction patients were younger (median age, 8.4 years *vs.* 10.0 years; P=0.05), and a greater percentage were black (P=0.02) and had a higher WBC count at presentation (median WBC, 32.8 *vs.* 22.6; P=0.01). All enrolled patients were included in chart abstractions at centers participating in the Pediatric Health Information Systems Database (PHIS). PHIS data are restricted to free-standing pediatric hospitals.[11] Chart data were used as the gold standard to determine sensitivity of trial AE reporting for each AE and by hospital.[5] Fisher's exact test or chi square test compared the sensitivity of AE reports across hospitals; χ^2 test was used for AEs with larger numbers of courses with the AE. Logistic regression was used to determine if sensitivity was associated with AML patient volume at a hospital. All analy-

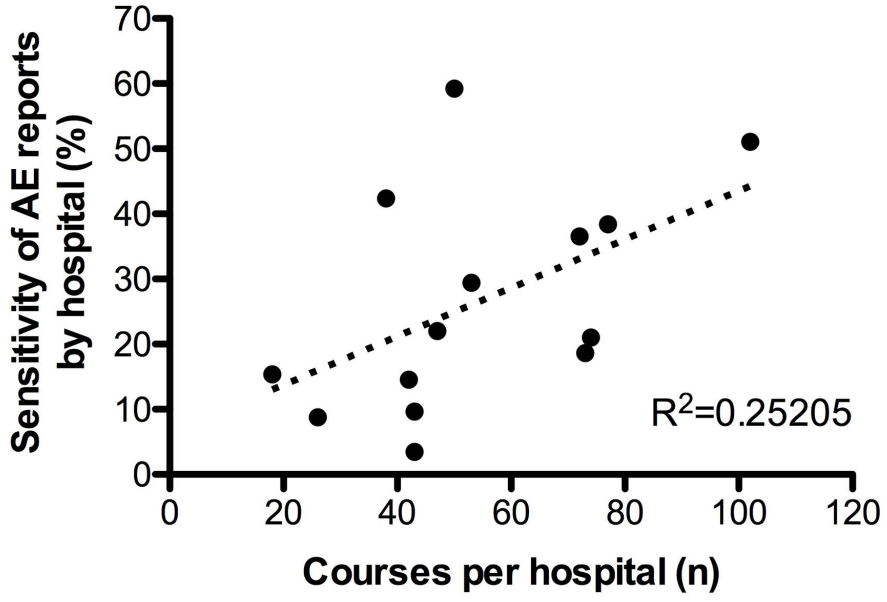

Figure 1. Relationship between the number of courses per hospital and sensitivity of all clinical trial adverse event reports by hospital.

ses used course-level data, and each course was considered a separate event based on prior research that showed no differences when generalized estimating equation models adjusted for within-subject and within-hospital correlations.[5] A two-sided *P*-value of 0.05 was considered statistically significant. Statistical analyses were performed in STATA version 12 (College Station, TX).

As previously described, chart abstraction was performed on 204 patients (758 courses, 20.0% of all patients on AAML0531) at fourteen hospitals.[5] The median number of patients per hospital was 18 (range 5 to 25). The average number of patients per year per hospital had a median of 3 (range 2 to 6). The median number of courses per hospital was 72 (range 18 to 102).

The sensitivity of AE reports varied widely (0% to 78.3%), as did the sensitivity at individual hospitals (0% to 100%) (Table 1). There were statistically significant differences in the sensitivity of clinical trial AE reporting between hospitals for anorexia ($P<0.001$), VGS ($P=0.016$), and pain ($P=0.001$) (Tables 1 and 2). Despite occurring often in chart data at all sites, pain was never reported at three hospitals, anorexia was never reported at four hospitals, and hypoxia was never reported at six hospitals. Some hospitals had higher sensitivity across all AEs, while other hospitals reported some AEs and did not report other AEs at all. For example, one hospital had sensitivity of at least 75% for hypotension and VGS, but less than 25% for hypoxia and pain. Another hospital had sensitivity greater than 0% for only one AE.

Figure 1 shows a trend toward higher sensitivity of trial AE reports at hospitals that administered larger numbers of chemotherapy courses. This is also demonstrated with the trend toward improved sensitivity for the majority of AEs at sites with more chemotherapy courses administered in Table 2. Logistic regression was performed for hypoxia, anorexia, VGS, and pain compared to the number of courses per site as these AEs each had more than 100 AEs in chart data. For every increase in 20 courses of chemotherapy at a hospital, there was a 1.7-fold increased rate of correctly-reported anorexia (OR 1.7, 95% CI 1.290-2.314, $P<0.001$) and 1.5-fold increased rate of correctly-reported pain (OR 1.5, 95% CI 1.062-2.155, $P=0.022$). There was a trend towards improved AE reporting for VGS (OR 1.6, 95% CI 0.986-2.690, $P=0.057$) and no association for hypoxia.

This study demonstrates substantial variation between hospitals in sensitivity of AE reporting on a pediatric AML clinical trial. For the frequently identified AEs of pain, anorexia, and VGS, there were statistically significant differences in AE reporting sensitivity between hospitals. This indicates that not only are AEs under-reported on clinical trials,[5] but also that reporting varies substantially between treating institutions.

The variation identified in AE reporting by institution is likely multifactorial and has concerning implications. Although data collection and analysis beyond the scope of this letter would be needed to identify these factors, several possibilities include the following: ambiguity in CTCAE definitions; variation in research staff availability, training, experience and workload compared to number of open trials; and variable familiarity with experimental agents. Since prior work has shown improved outcomes at hospitals treating more cancer patients,[10] larger patient volume may also improve AE reporting accuracy. Given the large number of COG centers, it seems likely that this variation would result in inaccuracies and overall AE under-reporting rather than differential reporting by study arm, but further work is needed to confirm this assessment.

Of note, Good Clinical Practice (GCP) training at each site is a shared responsibility between the site and COG. Specifically, each site is responsible for ensuring that all GCP training requirements are met for each individual

Table 1. Variation in sensitivity of reporting adverse events by hospital.

AE	Courses with AE in chart data across all hospitals n (%)	Courses with AE in clinical trial AE report across all hospitals n (%)	Sensitivity of clinical trial AE reports across all hospitals (95% CI)	Range of sensitivity at hospitals	*P* comparing sensitivity between hospitals
Hypertension	28 (3.7)	9 (1.2)	21.4 (8.3-41.0)	14.3-100%	0.074*
Hypotension	46 (6.1)	35 (4.6)	56.5 (41.1-71.7)	0-100%	0.794*
Hypoxia	167 (22.0)	30 (4.0)	17.4 (12.0-24.0)	0-100%	0.075*
ARDS	13 (1.7)	11 (1.5)	38.5 (13.9-68.4)	0-100%	0.054*
Anorexia	307 (40.5)	100 (13.2)	30.6 (25.5-36.1)	0-76.3%	<0.001**
Typhlitis	27 (3.6)	11 (1.5)	37.0 (19.4-57.6)	0-100%	0.268*
DIC	59 (7.8)	7 (0.9)	10.2 (3.8-20.8)	0-50%	0.599*
VGS	129 (17.0)	103 (13.6)	78.3 (70.2-85.1)	33.3-100%	0.016*
IFD	10 (1.3)	10 (1.3)	60.0 (26.2-87.8)	0-100%	0.924*
Pain	324 (42.7)	56 (7.4)	15.7 (12.0-20.2)	0-35.6%	0.001**
Seizure	5 (0.7)	2 (0.3)	0 (0.0-52.2)	0%	N/A
Renal Failure	6 (0.8)	4 (0.5)	50.0 (11.8-88.2)	0-100%	0.400*

CI: confidence interval. *Using Fisher's exact test or **using χ^2 test depending on number of AEs. N/A: unable to be evaluated due to small number of gold standard AEs in chart data; ARDS: adult respiratory distress syndrome; DIC: disseminated intravascular coagulation; VGS: viridans group streptococcal bacteremia; IFD: invasive fungal disease.

Table 2. Adverse event sensitivity by number of courses per hospital.

Hospital	Number of courses	Sensitivity											
		Hypertension	Hypotension	Hypoxia	ARDS	Anorexia	Typhlitis	DIC	VGS	IFI	Pain	Seizure	Renal Failure
1	18	0	0	0	0	0	100	0	100	0	0	0	0
2	26	0	50	0	0	0	0	0	33.3	0	14.3	0	0
3	38	0	100	33.3	0	64.3	100	20	80	50	21.1	0	0
4	42	0	0	0	0	23.5	0	0	100	0	6.3	0	0
5	43	0	0	0	0	0	0	0	33.3	0	0	0	0
6	43	0	50	16.7	0	0	0	0	100	100	0	0	0
7	47	100	33.3	12.5	0	11.1	25	0	66.7	0	12.5	0	0
8	50	0	66.7	40	100	65.6	50	0	100	0	33.3	0	100
9	53	0	50	0	0	25	100	50	66.7	0	21.4	0	0
10	72	60	75	25	16.7	42.3	0	14.3	81.3	0	18.0	0	0
11	73	0	66.7	9.5	0	2.4	0	0	73.7	50	16.3	0	0
12	74	0	0	0	0	16.7	0	0	76.9	100	0	0	0
13	77	14.3	66.7	31.3	100	33.3	66.7	28.6	100	50	4.8	0	0
14	102	100	71.4	20.8	0	76.3	50	14.3	87.5	0	35. 6	0	100

participating in the COG clinical trial. In addition, COG provides formalized training for clinical research associates (CRAs) in adverse event reporting. This training includes both general and trial specific components. Large phase III trial such as AAML0531 are typically not monitored by the study sponsor. However, there is regular and standardized auditing of data at each COG site. Notably, VGS was the only AE for which all hospitals had AE reports when an event had occurred in chart data. This indicates that the additional guidance provided for this targeted toxicity helped improve sensitivity across hospitals. Due to limited resources and limited research staff time available to devote to AE reporting,[8,12] this result is not unexpected.[5] Judicious use of targeted AE reporting may minimize the impact of variation in center-level reporting at least for specific AEs of particular clinical interest.

In conclusion, patterns of AE under-reporting must be identified so that interventions can be implemented to standardize AE reporting. Standardized systems and guidance on AE ascertainment combined with improvement of AE definitions could lead to more reproducible AE reporting and, ultimately, improve the accuracy of this important component of clinical trials.

Tamara P. Miller,[1,2,3] Yimei Li,[1,4] Marko Kavcic,[1] Kelly D. Getz,[1,2] Yuan-Shun V. Huang,[2] Lillian Sung,[5] Todd A. Alonzo,[6,7] Robert Gerbing,[7] Marla H. Daves,[8] Terzah M. Horton,[8] Michael A. Pulsipher,[9] Jessica Pollard,[10] Rochelle Bagatell,[1,11] Alix E. Seif,[1,11] Brian T. Fisher,[2,4,11,12] Selina Luger,[13] Alan S. Gamis,[14] Peter C. Adamson,[1,11] and Richard Aplenc[1,2,3,11]

[1]*Divisions of Oncology, The Children's Hospital of Philadelphia, PA, USA;* [2]*Center for Pediatric Clinical Effectiveness, The Children's Hospital of Philadelphia, PA, USA;* [3]*Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, PA, USA;* [4]*Center for Clinical Epidemiology and Biostatistics Departments of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA;* [5]*Department of Haematology/Oncology, The Hospital for Sick Children, Toronto, Canada;* [6]*Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA;* [7]*Children's Oncology Group, Monrovia, CA, USA;* [8]*Department of Pediatrics, Section of Hematology-Oncology, Texas Children's Hospital, Houston, TX, USA;* [9]*Division of Hematology, Oncology and BMT, Children's Hospital of Los Angeles, USC Keck School of Medicine, CA, USA;* [10]*Maine Children's Cancer Program, Maine Medical Center, Scarborough, ME, USA;* [11]*Departments of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA;* [12]*Infectious Diseases, The Children's Hospital of Philadelphia, PA, USA;* [13]*Hematology/Bone Marrow Transplant, University of Pennsylvania School of Medicine, Philadelphia, PA, USA and* [14]*Pediatric Hematology/Oncology, Children's Mercy Hospital, Kansas City, MO, USA*

Fundind: this work is supported by the National Institutes of Health R01 CA165277, K12 CA076931, Statistics and Data Center Grant U10CA098413, Chair's Grant U10CA098543, National Clinical Trials Network (NCTN) Operations Center Grant U10CA180886, NCTN Statistics & Data Center U10CA180899, the Pediatric Pharmacoepidemiology Training Grant 5T32HD064567-04, the Damon Runyon Cancer Research Foundation, and the St. Baldrick's Foundation. Staci Arnold, MD, MBA, Jessica Bokland, MD, James Feusner, MD, Samir Kahwash, MD, Michael Kelly, MD, PhD, Matthew Kutny, MD, William Roberts, MD, and Naomi Winick, MD also contributed data for this manuscript.

Correspondence: aplenc@email.chop.edu
doi:10.3324/haematol.2017.168815

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Gamis AS, Alonzo TA, Meshinchi S, et al. Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: results from the randomized phase III Children's Oncology Group trial

AAML0531. J Clin Oncol. 2014;32(27):3021-3032.
2. 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(3):761-769.
3. Lohmann DJ, Abrahamsson J, Ha SY, et al. Effect of age and body weight on toxicity and survival in pediatric acute myeloid leukemia: results from NOPHO-AML 2004. Haematologica. 2016; 101(11):1359-1367.
4. Creutzig U, Zimmermann M, Bourquin JP, et al. Randomized trial comparing liposomal daunorubicin with idarubicin as induction for pediatric acute myeloid leukemia: results from Study AML-BFM 2004. Blood. 2013;122(1):37-43.
5. Miller TP, Li Y, Kavcic M, et al. Accuracy of adverse event ascertainment in clinical trials for pediatric acute myeloid leukemia. J Clin Oncol. 2016;34(13):1537-1543.
6. Sivendran S, Latif A, McBride RB, et al. Adverse event reporting in cancer clinical trial publications. J Clin Oncol. 2014;32(2):83-89.
7. Atkinson TM, Li Y, Coffey CW, et al. Reliability of adverse symptom event reporting by clinicians. Qual Life Res. 2012;21(7):1159-1164.
8. Nass SJ, Moses HL, Mendelsohn J. A national cancer clinical trials system for the 21st century: reinvigorating the NCI cooperative group program: The National Academies Press, 2010.
9. Gwede CK, Johnson DJ, Daniels SS, Trotti A. Assessment of toxicity in cooperative oncology clinical trials: the long and short of it. J Oncol Manag. 2002;11(2):15-21.
10. Hillner BES, T.J.; Desch, C.E. Hospital and physician volume or specialization and outcomes in cancer treatment: Importance in quality of cancer care. J Clin Oncol. 2000;18(11):2327-2340.
11. Aplenc R, Fisher BT, Huang YS, et al. Merging of the National Cancer Institute-funded cooperative oncology group data with an administrative data source to develop a more effective platform for clinical trial analysis and comparative effectiveness research: a report from the Children's Oncology Group. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 2:37-43.
12. Roche K, Paul N, Smuck B, et al. Factors affecting workload of cancer clinical trials: results of a multicenter study of the National Cancer Insititue of Canada clinical trials group. J Clin Oncol. 2002;20(2):545-556. Novel mouse hemostasis model for real-time determination of bleeding time and hemostatic plug composition. J Thromb Haemost. 2014;13(3):417-425.
12. Andre P, Delaney SM, LaRocca T, et al. P2Y12 regulates platelet adhesion/activation, thrombus growth, and thrombus stability in injured arteries. J Clin Invest. 2003;112(3):398-406.
13. Gros A, Syvannarath V, Lamrani L, et al. Single platelets seal neutrophil-induced vascular breaches via GPVI during immune-complex-mediated inflammation in mice. Blood. 2015;126(8):1017-1026.
14. Nonne C, Lenain N, Hechler B, et al. Importance of platelet phospholipase Cgamma2 signaling in arterial thrombosis as a function of lesion severity. Arterioscler Thromb Vasc Biol. 2005;25(6):1293-1298.
15. Schafer AI. Effects of nonsteroidal antiinflammatory drugs on platelet function and systemic hemostasis. J Clin Pharmacol. 1995;35(3):209-219.