Accuracy of Adverse Event Ascertainment in Clinical Trials for Pediatric Acute Myeloid Leukemia

Tamara P. Miller, Yimei Li, Marko Kavcic, Andrea B. Troxel, Yuan-Shun V. Huang, Lillian Sung, Todd A. Alonzo, Robert Gerbing, Matt Hall, Marla H. Daves, Terzah M. Horton, Michael A. Pulsipher, Jessica A. Pollard, Rochelle Bagatell, Alix E. Seif, Brian T. Fisher, Selina Luger, Alan S. Gamis, Peter C. Adamson, and Richard Aplenc

Tamara P. Miller, Yimei Li, Marko Kavcic, Yuan-Shun V. Huang, Rochelle Bagatell, Alix E. Seif, Brian T. Fisher, Peter C. Adamson, and Richard Aplenc, The Children's Hospital of Philadelphia; Yimei Li, Andrea B. Troxel, Rochelle Bagatell, Alix E. Seif, Brian T. Fisher, Selina Luger, Peter C. Adamson, and Richard Aplenc, University of Pennsylvania School of Medicine, Philadelphia, PA; Lillian Sung, The Hospital for Sick Children, Toronto, Ontario, Canada; Todd A. Alonzo and Michael A. Pulsipher, University of Southern California; Michael A. Pulsipher, Children's Hospital of Los Angeles, Los Angeles; Todd A. Alonzo and Robert Gerbing, Children's Oncology Group, Monrovia, CA; Matt Hall, Children's Hospital Association, Overland Park, Kansas; Marla H. Daves, Children's Healthcare of Atlanta, Atlanta, GA; Terzah M. Horton, Texas Children's Hospital, Houston, TX; Jessica A. Pollard, Seattle Children's Hospital, Seattle, WA; and Alan S. Gamis, Children's Mercy Hospital, Kansas City, MO.

Published online ahead of print at www.jco.org on February 16, 2016.

Supported by the National Institutes of Health Grant No. R01 CA165277 and by Pediatric Pharmacoeconomics Training Grant No. 5T32HD064567-04.

Presented at the 55th Annual ASH Meeting and Exposition, New Orleans, LA, December 7-10, 2013; and the 51st Annual Meeting of the American Society of Clinical Oncology, Chicago, IL, May 29-June 2, 2015.

Authors' disclosures of potential conflicts of interest are found in the article online at www.jco.org. Author contributions are found at the end of this article.

Corresponding author: Tamara P. Miller, MD, The Children's Hospital of Philadelphia, 3501 Civic Center Blvd, CTRB 3022, Philadelphia, PA 19104; e-mail: millertp@e-mail.chop.edu.

© 2016 by American Society of Clinical Oncology

0732-183X/16/3413w-1537w/\$20.00

DOI: 10.1200/JCO.2015.65.5860

ABSTRACT

Purpose

Reporting of adverse events (AEs) in clinical trials is critical to understanding treatment safety, but data on AE accuracy are limited. This study sought to determine the accuracy of AE reporting for pediatric acute myeloid leukemia clinical trials and to test whether an external electronic data source can improve reporting.

Methods

Reported AEs were evaluated on two trials, Children's Oncology Group AAML03P1 and AAML0531 arm B, with identical chemotherapy regimens but with different toxicity reporting requirements. Chart review for 12 AEs for patients enrolled in AAML0531 at 14 hospitals was the gold standard. The sensitivity and positive predictive values (PPV) of the AAML0531 AE report and AEs detected by review of Pediatric Health Information System (PHIS) billing and microbiology data were compared with chart data.

Results

Select AE rates from AAML03P1 and AAML0531 arm B differed significantly and correlated with the targeted toxicities of each trial. Chart abstraction was performed on 204 patients (758 courses) on AAML0531. AE report sensitivity was < 50% for eight AEs, but PPV was > 75% for six AEs. AE reports for viridans group streptococcal bacteremia, a targeted toxicity on AAML0531, had a sensitivity of 78.3% and PPV of 98.1%. PHIS billing data had higher sensitivity (> 50% for nine AEs), but lower PPV (< 75% for 10 AEs). Viridans group streptococcal detection using PHIS microbiology data had high sensitivity (92.3%) and PPV (97.3%).

Conclusion

The current system of AE reporting for cooperative oncology group clinical trials in pediatric acute myeloid leukemia underestimates AE rates. The high sensitivity and PPV of PHIS microbiology data suggest that using external data sources may improve the accuracy of AE reporting.

J Clin Oncol 34:1537-1543. © 2016 by American Society of Clinical Oncology

INTRODUCTION

Clinical trials have dramatically improved outcomes for children with acute myeloid leukemia (AML)^{1,2}; however, AML therapy is intensive and causes substantial treatment-related adverse effects that are captured in adverse event (AE) reports.^{3,4} Clinical trial AE data define the expected toxicities of standard therapy, thus informing patients and clinicians about potential therapy complications. In addition, clinical trials have an ethical imperative to monitor AEs; therefore, accurate assessment of AEs is critical for the trials.

The National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE)

system^{5,6} was developed to standardize AE reporting in oncology clinical trials⁷⁻¹¹; however, even with the use of CTCAE, monitoring of AEs is complex and labor intensive. Intensive therapies cause multiple AEs per patient; 11,000 non-hematological, grade 3 to 5 AEs were reported on 1,022 patients for Children's Oncology Group trial AAML0531.¹ In addition, CTCAE is complex, which makes reproducible, systematic AE capture challenging.^{11,12}

Despite their importance, AEs are likely under-reported in cooperative group AML clinical trials. One study demonstrated that data from the National Cancer Institute Clinical Data Update System were missing in 27% of published

AEs, and 28% of Clinical Data Update System AEs were not in the published articles.¹³ Targeted data collection on cardiac toxicities for patients receiving tyrosine kinase inhibitors revealed higher rates of cardiac AEs than previously reported.¹⁴ There is significant variation and lack of guidance in AE reporting.^{7,11,15} Compliance with the Consolidated Standards of Reporting Trials guidelines is often limited, with a median completeness score of 8 of 14 reporting elements.¹⁶

Whereas AEs are clearly under-reported, no study has rigorously compared cooperative group clinical trial AE reports with the gold standard of chart abstraction to determine the accuracy of AE reports. We hypothesized that comparison with chart data would reveal poor sensitivity and positive predictive value (PPV), indicating global AE under-reporting; that a comparison of two trials with identical chemotherapy regimens would reveal that reporting is increased for targeted toxicities; and that AE reporting can be improved by using external electronic data sources.

METHODS

Study Design

AE reports from two AML trials (AAML03P1 and AAML0531) with identical chemotherapy regimens were compared. AE reports for patients enrolled in AAML0531 at 14 hospitals were compared with AEs identified by chart abstraction for patients in AAML0531. Lastly, billing and laboratory data from an external electronic data source were compared with chart data. Institutional review board approval was obtained at all hospitals.

Clinical Trials for Pediatric AML

Trial AAML03P1 tested the safety of gemtuzumab ozogamicin (GO) with standard chemotherapy.² Trial AAML0531 compared standard chemotherapy with or without GO¹; arm B had chemotherapy identical to that of AAML03P1. All nonhematological, grade 3 to 5 AEs were reported using the CTCAE version 3 except a small number from AAML0531, which used CTCAE version 4 after it was released. All grades of protocol-specified targeted toxicities were reported. AAML03P1 had six targeted toxicities: ALT, AST, hyperbilirubinemia, typhlitis, veno-occlusive disease, and infections. AAML0531 had three targeted toxicities: veno-occlusive disease, cardiac toxicities, and infections. An infection case reporting form (CRF) supplemented CTCAE reports; the AAML0531 CRF was more detailed than that of AAML03P1. Two AAML0531 study members (RA and LS) were notified prospectively of all AAML0531 AEs and reviewed all reported AEs for accuracy. AEs were first reported for the chemotherapy course during which the AE began and for all subsequent courses if the AE persisted. A subset of patients enrolled at each Children's Oncology Group site is audited every 3 years with a focus on study eligibility, informed consent, and ascertainment of primary study end points.

Chart Abstraction

Two pediatric oncologists (TPM and MK) performed chart abstraction for all children treated in AAML0531 at 14 hospitals with medical charts available as the AE gold standard. All chemotherapy course data were collected, and patients were censored at stem-cell transplantation. Chart abstraction assessed 12 toxicities on each inpatient day: hypertension, hypotension, hypoxia, acute respiratory distress syndrome, microbiologically-proven viridans group streptococcal (VGS) bacteremia, microbiologically-detected sterile-site fungal pathogen, anorexia, typhlitis, disseminated intravascular coagulation, pain, seizure, and acute renal failure. Standardized toxicity identification algorithms were used (Appendix Table A1, online only). AEs were categorized by definition complexity and ascertainment effort on the basis of the experience of the chart abstractors. Ascertainment

effort included difficulty in locating the AE in the chart. Laboratory-based AEs required less effort, and AEs that required review of clinician notes and radiology results were considered greater effort.

External Electronic Data Source

The Pediatric Health Information System (PHIS) is an administrative and billing database of inpatient data from 47 tertiary-care US children's hospitals. PHIS includes demographics, admission and discharge dates, and International Classification of Diseases, Ninth Revision, codes for procedures and diagnoses for each hospitalization as well as daily pharmacy, laboratory, imaging, procedure, and ancillary service billing data. Six PHIS hospitals contribute laboratory, microbiology, and radiology results (PHIS+).^{17,18} Identification of the 12 AEs in PHIS was based on a combination of International Classification of Diseases, Ninth Revision, codes and billing codes (Appendix Table A1). Data on 384 patients in AAML0531 were previously merged with PHIS.¹⁹

Statistical Analyses

Distributions were compared by using χ^2 or Wilcoxon rank sum tests. AE rates per day were obtained from Poisson regressions adjusted for on-protocol therapy time. AE rates were compared as rate ratios (RRs) with 95% CIs.

For each AE, the sensitivity, specificity, PPV, and negative predictive value (NPV) of AE reports from AAML0531 and of PHIS data were calculated by using chart data as the gold standard. The cutoff for AE report data was June 30, 2013. The sensitivity, specificity, PPV, and NPV of PHIS+ data were calculated for VGS compared with chart data. In the primary analyses, each chemotherapy course was considered independent, and exact 95% CIs were calculated. In the secondary analyses, logistic regression models with generalized estimating equations (GEE) were used to obtain robust 95% CIs to separately account for potential within-subject and within-hospital course correlations. Statistical analyses were performed in SAS (SAS/STAT User's Guide, Version 9.3; SAS Institute, Cary, NC) and STATA software version 12 (STATA, College Station, TX; Computing Resource Center, Santa Monica, CA).

RESULTS

Patients and Demographics

Data for 339 patients in trial AAML 03P1 were available. Chart abstraction was completed for 204 of 1,022 eligible patients (20.0%; 758 courses) enrolled in AAML0531. Compared with all patients in AAML0531, chart abstraction patients were younger (median age, 8.4 years ν 10.0 years; $P = .05$), and a greater percentage were black ($P = .02$) and had a higher WBC count at presentation (median WBC, 32.8 ν 22.6; $P = .01$). There were no differences in gender, ethnicity, or percentages treated with GO (Table 1).

Comparison of AEs Between AAML03P1 and AAML0531 Arm B

Figure 1 compares grades 3 to 5 AEs per patient between AAML03P1 and AAML0531 arm B. Bloodstream infections, a targeted toxicity in both studies but with enhanced monitoring in AAML0531, were increased in AAML0531 arm B compared with AAML03P1 (RR, 2.11; 95% CI, 1.90 to 2.34; $P < .01$). Hepatic toxicity, targeted only in AAML03P1, was higher in AAML03P1 compared with AAML0531 arm B (hyperbilirubinemia: RR, 0.64; 95% CI, 0.42 to 0.95; $P = .04$; and ALT/AST: RR, 0.54; 95% CI, 0.44 to 0.66; $P < .01$).

Table 1. Demographics and WBC Count at Presentation and Percentage of Patients Assigned to Treatment With GO in AAML03P1, AAML0531, and for the Abstracted and Nonabstracted Patients in AAML0531

Variable	AAML03P1 All Patients (N = 339)	AAML0531 All Patients (N = 1,022)	AAML0531 Abstracted Patients (n = 204)	AAML0531 Nonabstracted Patients (n = 818)	P (Abstracted v Nonabstracted AAML0531 Patients)
Median age (range), years	9.5 (0.07-21.6)	9.7 (0.01-29.8)	8.4 (0.02-18.7)	10.0 (0.01-29.8)	.05
Median WBC at presentation (range)	19.6 (0.7-495)	24 (0.2-827.2)	32.8 (0.4-610)	22.6 (0.2-827.2)	.01
Female	155 (45.7)	514 (50.3)	112 (54.9)	402 (49.1)	.14
Race					.02
White	232 (68.4)	748 (73.2)	152 (74.5)	596 (72.8)	
Black	46 (13.6)	116 (11.4)	32 (15.7)	84 (10.3)	
Asian	20 (5.9)	50 (4.9)	7 (3.4)	43 (5.3)	
Other	7 (2.1)	6 (0.6)	2 (1.0)	4 (0.5)	
Unknown	34 (10.0)	102 (10.0)	11 (5.4)	91 (11.1)	
Ethnicity					.63
Hispanic	57 (16.8)	189 (18.5)	35 (17.2)	154 (18.8)	
Not Hispanic	266 (78.5)	794 (77.7)	163 (79.9)	631 (77.2)	
Unknown	16 (4.7)	39 (3.8)	6 (2.9)	33 (4.0)	
Treated with GO	339 (100)	511 (50)	104 (51.0)	407 (49.8)	.75

NOTE. All data are given as No. (%) unless otherwise noted.
Abbreviation: GO, gemtuzumab ozogamicin.

Comparison of AAML0531 AE Report With Chart Data

Table 2 shows the sensitivity, specificity, PPV, and NPV of AAML0531 AE report data compared with AAML0531 chart data. Sensitivity ranged from 10.2% to 78.3% for AEs observed in ≥ 10 chemotherapy courses in chart data. VGS, a targeted toxicity in AAML0531, had the highest sensitivity (78.3%; 95% CI, 70.2% to

85.1%); the two highest sensitivities were for AEs with the lowest definition complexity and ascertainment effort (VGS and sterile-site fungal pathogen). There were no other patterns in sensitivity and PPV by AE definition complexity or ascertainment effort (Fig 2). Eight of 12 AEs had sensitivities $< 50\%$. PPV was $\geq 75\%$ for one half of AEs and 45.5% to 98.1% for AEs present in ≥ 10 chart

Fig 1. Pyramid plot of average number of grade 3 to 5 adverse events per patient compared between AAML03P1 and AAML0531 arm B. Red bars indicate targeted toxicities in AAML03P1 and/or AAML0531 arm B. An infection case reporting form was created for AAML0531 and was required to be completed as part of the Common Terminology Criteria for Adverse Events submission for reporting of bloodstream infections on the clinical trial.

Table 2. Chart Abstraction Data Compared With Clinical Trial Adverse Event Report for Each of the 12 Grade 3 to 5 Toxicities

Toxicity	Chart Abstraction, No. (%) [*]	Adverse Event Report				
		No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
Hypertension	28 (3.7)	9 (1.2)	21.4 (8.3 to 41.0)	99.6 (98.8 to 99.9)	66.7 (29.9 to 92.5)	97.1 (95.6 to 98.2)
Hypotension	46 (6.1)	35 (4.6)	56.5 (41.1 to 71.7)	98.7 (97.6 to 99.4)	74.3 (56.7 to 87.5)	97.2 (95.8 to 98.3)
Hypoxia	167 (22.0)	30 (4.0)	17.4 (12.0 to 24.0)	99.8 (99.1 to 100)	96.7 (82.8 to 99.9)	81.0 (78.0 to 83.8)
ARDS	13 (1.7)	11 (1.5)	38.5 (13.9 to 68.4)	99.2 (98.3 to 99.7)	45.5 (16.8 to 76.6)	98.9 (97.9 to 99.5)
Anorexia	307 (40.5)	100 (13.2)	30.6 (25.5 to 36.1)	98.7 (97.1 to 99.5)	94.0 (87.4 to 97.8)	67.6 (63.9 to 71.2)
Typhilitis	27 (3.6)	11 (1.5)	37.0 (19.4 to 57.6)	99.9 (99.2 to 100)	90.9 (58.7 to 99.8)	97.7 (96.4 to 98.7)
DIC	59 (7.8)	7 (0.9)	10.2 (3.8 to 20.8)	99.9 (99.2 to 100)	85.7 (42.1 to 99.6)	92.9 (90.9 to 94.7)
VGS	129 (17.0)	103 (13.6)	78.3 (70.2 to 85.1)	99.7 (98.9 to 100)	98.1 (93.2 to 99.8)	95.7 (93.9 to 97.1)
IFI	10 (1.3)	10 (1.3)	60.0 (26.2 to 87.8)	99.5 (98.6 to 99.9)	60.0 (26.2 to 87.8)	99.5 (98.6 to 99.9)
Pain	324 (42.7)	56 (7.4)	15.7 (12.0 to 20.2)	98.9 (97.3 to 99.6)	91.1 (80.4 to 97.0)	61.1 (57.4 to 64.7)
Seizure	5 (0.7)	2 (0.3)	0 (0.0 to 52.2)	99.7 (99.0 to 100)	0 (0.0 to 84.2)	99.3 (98.5 to 99.8)
Renal failure	6 (0.8)	4 (0.5)	50.0 (11.8 to 88.2)	99.9 (99.3 to 100)	75.0 (19.4 to 99.4)	99.6 (98.9 to 99.9)

NOTE. All data are for patients enrolled in clinical trial AAML0531 for whom chart abstraction was performed.

Abbreviations: ARDS, adult respiratory distress syndrome; DIC, disseminated intravascular coagulation; IFI, invasive fungal infection; NPV, negative predictive value; PPV, positive predictive value; VGS, viridans group streptococcus.

^{*}Chart abstraction data are the gold standard.

courses. Specificity was $\geq 98\%$ for all AEs, and the NPV was $\geq 80\%$ for all AEs except anorexia and pain. Results were similar when using GEE models adjusting for within-subject and within-hospital correlations (Appendix Tables A2 and A3, online only). Whereas accuracy of AE reporting varied substantially between hospitals (Appendix Table A4, online only), no clear patterns of variation were discernable.

Fig 2. Adverse event (AE) definition complexity and work required to ascertain AEs. Pink shading indicates AEs that could be electronically captured; yellow shading indicates AEs that could be electronically captured, but may need manual review to verify grading; blue shading indicates AEs that can be triggered for review electronically, but would need manual review for ascertainment and grading; italics indicate AEs that were not included in chart abstraction, but were commonly reported in cooperative group acute myeloid leukemia clinical trials as shown in Figure 1. AE definition complexity is based on a combination of the ease of interpretation of the Common Terminology Criteria for Adverse Events (CTCAE) definition and the clinical complexity of the AE. AKI, acute kidney injury; ARDS, adult respiratory distress syndrome; DIC, disseminated intravascular coagulation; IFI, invasive fungal infection; VGS, viridans group streptococci. (*)CTCAE version 4, which is currently used, does not include acute renal failure, which is a clinically defined toxicity on the basis of use of dialysis, but instead includes acute kidney injury, which is a laboratory-based toxicity defined by creatinine level, except for grade 4, which requires the patient to receive dialysis.

Comparison of PHIS Toxicities With Chart Data

Table 3 shows the sensitivity, specificity, PPV, and NPV of PHIS data compared with AAML0531 chart data. Sensitivity was 10.9% to 99.1% for AEs present in ≥ 10 chart courses. The sensitivity was $\geq 50\%$ for nine of 12 AEs and overall was higher for PHIS data than for AE report data. Infectious AEs had similar or lower sensitivity for PHIS data than for AE report data. The PPV of PHIS AEs was $\geq 75\%$ for two AEs and overall was lower (range, 35.0% to 98.7%) than AE report data. Specificity of PHIS AEs was $\geq 95\%$ for all but two AEs (hypoxia and pain), and the NPV was $\geq 86\%$ for all AEs. Results were similar when using GEE models (Appendix Tables A5 and A6, online only).

Comparison of PHIS + VGS Data With Chart Data

Three hospitals (53 patients; 202 chemotherapy courses) with chart data contributed microbiology data to PHIS+. The sensitivity of VGS in PHIS+ data was 92.3% (95% CI, 79.1% to 98.4%), specificity was 99.4% (95% CI, 96.6% to 100%), PPV was 97.3% (95% CI, 85.8% to 99.9%), and NPV was 98.2% (95% CI, 94.8% to 99.6%).

DISCUSSION

To our knowledge, this is the first study to evaluate the accuracy of AE reports in cooperative group clinical trials using patient chart abstraction as the gold standard. We show that AE reporting is substantially impacted by the specification of targeted toxicities, and that the current AE reporting processes underestimate toxicity in complex clinical trials. Given the intensity of AML chemotherapy, the intrinsic complexity of current AE reporting, and the limited resources to support this reporting, these results are not unexpected.

The comparison of AE rates reported in AAML03P1 and AAML0531 arm B highlights that protocol-defined, targeted toxicities are preferentially reported. Despite identical chemotherapy

Table 3. Chart Abstraction Data Compared With Resource Use (PHIS) Data for Each of the 12 Grade 3 to 5 Toxicities

Toxicity	Chart Abstraction, No. (%) [*]	PHIS Data					PHIS Data Source
		No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	
Hypertension	28 (3.7)	44 (5.8)	64.3 (44.1 to 81.4)	96.4 (94.8 to 97.7)	40.9 (26.3 to 56.8)	98.6 (97.4 to 99.3)	Pharmacy
Hypotension	46 (6.1)	11 (1.5)	10.9 (3.6 to 23.6)	99.2 (98.2 to 99.7)	45.5 (16.8 to 76.6)	94.5 (92.6 to 96.0)	Pharmacy
Hypoxia	167 (22.0)	201 (26.5)	73.7 (66.3 to 80.2)	86.8 (83.8 to 89.4)	61.2 (54.1 to 68.0)	92.1 (89.5 to 94.2)	ICD-9, procedure, clinical
ARDS	13 (1.7)	11 (1.5)	30.8 (9.1 to 61.4)	99.1 (98.1 to 99.6)	36.4 (10.9 to 69.2)	98.8 (97.7 to 99.5)	ICD-9
Anorexia	307 (40.5)	237 (31.3)	76.2 (71.1 to 80.9)	99.3 (98.1 to 99.9)	98.7 (96.4 to 99.7)	86.0 (82.7 to 88.9)	Pharmacy
Typhlitis	27 (3.6)	26 (3.4)	66.7 (46.0 to 83.5)	98.9 (97.9 to 99.5)	69.2 (48.2 to 85.7)	98.8 (97.7 to 99.4)	ICD-9
DIC	59 (7.8)	63 (8.3)	79.7 (67.2 to 89.0)	97.7 (96.3 to 98.7)	74.6 (62.1 to 84.7)	98.3 (97.0 to 99.1)	Pharmacy, ICD-9
VGS	129 (17.0)	53 (7.0)	30.2 (22.5 to 38.9)	97.8 (96.3 to 98.8)	73.6 (59.7 to 84.7)	87.2 (84.5 to 89.6)	ICD-9
IFI	11 (1.5)	20 (2.6)	70.0 (34.8 to 93.3)	98.3 (97.1 to 99.1)	35.0 (15.4 to 59.2)	99.6 (98.8 to 99.9)	ICD-9
Pain	324 (42.7)	511 (67.4)	99.1 (97.3 to 99.8)	56.2 (51.4 to 61.0)	62.8 (58.5 to 67.0)	98.8 (96.5 to 99.8)	Pharmacy
Seizure	5 (0.7)	19 (2.5)	80.0 (28.4 to 99.5)	98.0 (96.7 to 98.9)	21.1 (6.1 to 45.6)	99.9 (99.3 to 100)	ICD-9
Renal failure	6 (0.8)	6 (0.8)	83.3 (35.9 to 99.6)	99.9 (99.3 to 100)	83.3 (35.9 to 99.6)	99.9 (99.3 to 100)	Clinical

NOTE. All data are for patients enrolled in clinical trial AAML0531 for whom chart abstraction was performed.

Abbreviations: ARDS, adult respiratory distress syndrome; DIC, disseminated intravascular coagulation; ICD-9, International Classification of Diseases, Ninth Revision; IFI, invasive fungal infection; NPV, negative predictive value; PHIS, Pediatric Health Information System; PPV, positive predictive value; VGS, viridans group streptococcus.

^{*}Chart abstraction data are the gold standard.

regimen, the rates of reported hepatic toxicities were significantly higher in AAML03P1 than in AAML0531. Conversely, the bacteremia rate was nearly two-fold higher in AAML0531 than in AAML03P1. Of note, the additional reporting guidance in AAML0531 resulted in a high PPV of 98.1% for VGS bloodstream infections. These results also demonstrate that infection-specific CRFs and prospective monitoring, as were used in AAML0531, are critical for accurate reporting of bloodstream infections.

Sensitivity of the clinical trial AE report was $\leq 50\%$ for eight of 12 complex toxicities, with no discernable pattern in AE reporting across organ systems. However, the PPV was generally high, indicating that the AE reports were typically true positives. As hypothesized, the sensitivity of AE capture was higher for PHIS compared with clinical trial AE ascertainment; however, PHIS AE sensitivity varied substantially by AE (10.9% to 99.1%) and PPV was consistently $< 75\%$, which is lower than that for clinical trial AE reports. These data expand on previous work by including daily resource utilization data and manual chart review to confirm that use of administrative and billing data alone is unlikely to improve AE monitoring.²⁰

Of note, microbiology data in PHIS+ had both high sensitivity and PPV for VGS data (92.3% and 97.3%, respectively). These results strongly suggest that external electronic collection of primary data, rather than billing data, may improve the accuracy of AE reporting. An algorithm using administrative, pharmacy, and electronic data to capture AEs in women treated for breast cancer had an overall sensitivity of 89%, but ranged from 0% to 100% for specific AEs.²¹ Whereas this study did not report PPV or compare results with clinical trial data,²¹ the high overall sensitivity indicates that automated electronic capture of primary patient data may be effective. Our concordant results suggest this approach could improve AE reporting on clinical trials.

The observed modest sensitivity, generally high PPV, and high NPV indicate that individual AE reports are typically accurate, but overall AE rates are grossly underestimated. This likely stems

from multifactorial difficulties in ascertainment. The limited time available for AE reporting, the complexity of certain CTCAE definitions, and the clinical complexity of some AEs are obvious potential etiologies. Cooperative oncology groups have substantially fewer resources for AE reporting than industry-sponsored trials,²² and clinical research assistants perform manual AE reporting as one of many duties. One study found clinical research assistants spend 18 minutes per day reporting AEs.²³ Given AML chemotherapy intensity and the complexity of many chemotherapy-associated AEs, this time is likely inadequate for comprehensive AE reporting. Moreover, such time constraints favor selective reporting of targeted toxicities. These time constraints are heightened by the complexity of CTCAE version 4, which encompasses 790 toxicities that do not correspond to specific medical interventions.⁶ For example, total parenteral nutrition is listed in 45 CTCAE toxicities, including eight pertaining to mucositis-related anorexia and weight loss.⁶

Figure 2 summarizes the grading and phenotypic complexity of frequently reported AEs as well as the varying ascertainment efforts based on our chart abstraction experience. Laboratory-based AEs, such as VGS bacteremia, are the most straightforward to capture. Some AEs, like pain, require a modest amount of work to identify, but the CTCAE definition is complex. Other AEs, like hypotension, require significant work and have moderately complex definitions. A final group of AEs require substantial work to capture and have complex definitions.

There are at least four potential strategies for addressing these reporting challenges. First, the marked increase in bloodstream infection reports in AAML0531 demonstrates the effectiveness of well-designed, specified AE-targeted reporting. Careful selection of targeted toxicities is crucial as the preferential reporting of targeted toxicities likely decreases other AE reporting. For example, the focused reporting of grades 1 and 2 hepatic toxicities in AAML03P1 was not clinically beneficial and may have decreased reporting of other important toxicities (Fig 1). Indeed, Mahoney

et al²⁴ reported that the majority of non-life-threatening AEs in 26 oncology trials were clinically unimportant. As a first step in initiating a discussion of AML-specific targeted toxicities, Appendix Table A7 (online only) provides a list of potential targeted toxicities.

Second, for large trials, Kaiser et al²⁵ suggested subsampling patients to provide accurate estimates of toxicities without necessitating AE reporting on all patients. Whereas subsampling may benefit trials enrolling at least 400 patients, it may miss rare but serious AEs and will be underpowered on smaller trials. However, this risk could be partially mitigated by requiring reporting of predetermined serious AEs on all patients. Furthermore, careful a priori determination of a subsampling algorithm will be particularly important for pediatric clinical trials involving multiple small centers that enroll limited numbers of patients.

Third, automated electronic capture and grading of primary clinical data may improve AE reporting, but will require methodologies specific to each AE being captured. The PHIS+ results for VGS bacteremia concur with other data demonstrating that laboratory-based AEs can be ascertained electronically.²¹ Whereas moderately complex AEs may be identified electronically, grading will likely need to be performed manually. For example, pain may be identified by analgesics documented in the medication administration record, but grading will require a manual assessment of the impact of pain on the activities of daily living. AEs with high phenotypic complexity may also be identified electronically, but will require manual review for confirmation and grading. For example, acute respiratory distress syndrome could be identified through searching chest x-ray results for bilateral infiltrates with subsequent manual review of temporally associated clinical data. Data managers reviewing high-complexity AEs would need training and guidelines for the evaluation and grading of each AE. With remote access to electronic medical records, such grading could be performed remotely by a central data review group. Furthermore, these three strategies could be combined, and selected chart review of important targeted AEs could be performed for a subset of patients.

Finally, as a fourth strategy, consideration should be given to aligning CTCAE more closely with the operational realities of AE reporting and monitoring.^{26,27} Currently, CTCAE aims both to identify toxicity signals and to report specific AEs in granular detail. Separation of these two aims may improve the reliability of CTCAE reporting. Specifically, the first step of toxicity signal detection could be on the basis of automated laboratory data analysis. The second step of detailed AE reporting could be directed by identified toxicity signals. Detailed reporting could be performed centrally, and anchoring some CTCAE definitions in specific medical resources may further streamline detailed AE reporting.

The primary limitation of this study is that chart abstraction was only performed for 12 AEs. Whereas these represent all organ systems, the reporting of other key toxicities, such as gram-negative rod bacteremia and left-ventricular systolic dysfunction, was not

assessed. Given the similarity in reporting processes for VGS and gram-negative rod bacteremia, VGS results will likely generalize to all microbiologically determined bacteremia. Work is currently ongoing to determine the accuracy of other laboratory-based AEs and left-ventricular systolic dysfunction. Because pediatric AML therapy has high AE rates, the sensitivity of these 12 AEs may be higher for less-intensive chemotherapy regimens with fewer adverse effects. Thus, this study should be replicated in other malignancies, those with both high AE rates and lower anticipated rates.

This study used chart abstraction to demonstrate the challenges and limitations of current AE reporting in pediatric AML trials. Pediatric cooperative oncology group clinical trials have dramatically improved outcomes for children with cancer despite steadily decreasing per-patient reimbursements.²² However, therapy remains intensive, and, therefore, accurate AE reporting is crucial. Because a substantial increase in resources for cooperative group clinical trials is unlikely, and increased resources alone may not fully address AE reporting difficulties, alternative approaches for more accurate AE ascertainment are needed. Given the importance and scope of AE reporting difficulties, these approaches will likely require substantial changes in reporting expectations, reporting processes, and AE definitions. Though daunting, these challenges may be overcome by using more efficient electronic AE ascertainment processes and the collaborative efforts of investigators, cooperative groups, federal agencies, patients, and families.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at www.jco.org.

AUTHOR CONTRIBUTIONS

Conception and design: Tamara P. Miller, Marko Kavcic, Lillian Sung, Rochelle Bagatell, Alix E. Seif, Brian T. Fisher, Richard Aplenc

Financial support: Richard Aplenc

Administrative support: Richard Aplenc

Provision of study materials or patients: Marla H. Daves, Terzah M. Horton, Michael A. Pulsipher, Jessica A. Pollard, Alan S. Gamis, Richard Aplenc

Collection and assembly of data: Tamara P. Miller, Marko Kavcic, Yuan-Shun V. Huang, Todd A. Alonzo, Robert Gerbing, Matt Hall, Marla H. Daves, Terzah M. Horton, Michael A. Pulsipher, Jessica A. Pollard, Alan S. Gamis, Richard Aplenc

Data analysis and interpretation: Tamara P. Miller, Yimei Li, Andrea B. Troxel, Yuan-Shun V. Huang, Todd A. Alonzo, Marla H. Daves, Michael A. Pulsipher, Rochelle Bagatell, Alix E. Seif, Brian T. Fisher, Selina Luger, Peter C. Adamson, Richard Aplenc

Manuscript writing: All authors

Final approval of manuscript: All authors

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* 32:3021-3032, 2014

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* 118:761-769, 2012

3. 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* 111:1044-1053, 2008

4. Smith FO, Alonzo TA, Gerbing RB, et al: Long-term results of children with acute myeloid leukemia: A report of three consecutive Phase III trials by the Children's Cancer Group: CCG 251, CCG 213 and CCG 2891. *Leukemia* 19:2054-2062, 2005

5. National Cancer Institute: Common Terminology Criteria for Adverse Events (CTCAE), Version 3.0. http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf

6. National Cancer Institute: Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0. http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf

7. Trotti A, Bentzen SM: The need for adverse effects reporting standards in oncology clinical trials. *J Clin Oncol* 22:19-22, 2004

8. Trotti A, Byhardt R, Stetz J, et al: Common toxicity criteria: Version 2.0. An improved reference for grading the acute effects of cancer treatment: impact on radiotherapy. *Int J Radiat Oncol Biol Phys* 47:13-47, 2000

9. Trotti A, Colevas AD, Setser A, et al: CTCAE v3.0: Development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol* 13:176-181, 2003

10. Williams J, Chen Y, Rubin P, et al: The biological basis of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol* 13:182-188, 2003

11. Gwede CK, Johnson DJ, Daniels SS, et al: Assessment of toxicity in cooperative oncology clinical trials: The long and short of it. *J Oncol Manag* 11:15-21, 2002

12. Huynh-Le MP, Zhang Z, Tran PT, et al: Low interrater reliability in grading of rectal bleeding using National Cancer Institute Common Toxicity Criteria

and Radiation Therapy Oncology Group Toxicity scales: A survey of radiation oncologists. *Int J Radiat Oncol Biol Phys* 90:1076-1082, 2014

13. Scharf O, Colevas AD: Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J Clin Oncol* 24:3933-3938, 2006

14. Schmidinger M, Zielinski CC, Vogl UM, et al: Cardiac toxicity of sunitinib and sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol* 26:5204-5212, 2008

15. Fromme EK, Eilers KM, Mori M, et al: How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol* 22:3485-3490, 2004

16. Sivendran S, Latif A, McBride RB, et al: Adverse event reporting in cancer clinical trial publications. *J Clin Oncol* 32:83-89, 2014

17. Gouripeddi R, Warner PB, Mo P, et al: Federating clinical data from six pediatric hospitals: Process and initial results for microbiology from the PHIS+ consortium. *AMIA Annu Symp Proc* 2012:281-290, 2012

18. Narus SP, Srivastava R, Gouripeddi R, et al: Federating clinical data from six pediatric hospitals: Process and initial results from the PHIS+ consortium. *AMIA Annu Symp Proc* 2011:994-1003, 2011

19. 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* 21:37-43, 2012 (suppl 2)

20. Lamont EB, Herndon JE, II, Weeks JC, et al: Measuring clinically significant chemotherapy-related toxicities using Medicare claims from Cancer and Leukemia Group B (CALGB) trial participants. *Med Care* 46:303-308, 2008

21. Mandelblatt JS, Huang K, Makgoeng SB, et al: Preliminary development and evaluation of an algorithm to identify breast cancer chemotherapy toxicities using electronic medical records and administrative data. *J Oncol Pract* 10:1200/JOP.2013.001288 [epub ahead of print on August 26, 2014]

22. Nass SJ, Moses HL, Mendelsohn J: A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program. Washington, DC, National Academies Press, 2010, pp 163

23. Roche K, Paul N, Smuck B, et al: Factors affecting workload of cancer clinical trials: Results of a multicenter study of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol* 20:545-556, 2002

24. Mahoney MR, Sargent DJ, O'Connell MJ, et al: Dealing with a deluge of data: An assessment of adverse event data on North Central Cancer Treatment Group trials. *J Clin Oncol* 23:9275-9281, 2005

25. Kaiser LD, Melemed AS, Preston AJ, et al: Optimizing collection of adverse event data in cancer clinical trials supporting supplemental indications. *J Clin Oncol* 28:5046-5053, 2010

26. Witteles RM, Telli M: Underestimating cardiac toxicity in cancer trials: Lessons learned? *J Clin Oncol* 30:1916-1918, 2012

27. Thanarajasingam G, Hubbard JM, Sloan JA, et al: The imperative for a new approach to toxicity analysis in oncology clinical trials. *J Natl Cancer Inst* 107:1-4, 2015

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Accuracy of Adverse Event Ascertainment in Clinical Trials for Pediatric Acute Myeloid Leukemia

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or jco.ascopubs.org/site/ifc.

Tamara P. Miller

No relationship to disclose

Yimei Li

No relationship to disclose

Marko Kavcic

No relationship to disclose

Andrea B. Troxel

Leadership: VAL Health

Consulting or Advisory Role: Asubio Pharmaceuticals, VAL Health

Research Funding: The Vitality Group (Inst), Weight Watchers (Inst)

Yuan-Shun V. Huang

No relationship to disclose

Lillian Sung

No relationship to disclose

Todd A. Alonzo

No relationship to disclose

Robert Gerbing

No relationship to disclose

Matt Hall

No relationship to disclose

Marla H. Daves

No relationship to disclose

Terzah M. Horton

Research Funding: Takeda Pharmaceuticals, Millennium Pharmaceuticals

Michael A. Pulsipher

Consulting or Advisory Role: Chimerix, Novartis, Jazz Pharmaceuticals

Jessica A. Pollard

Consulting or Advisory Role: Celgene

Rochelle Bagatell

No relationship to disclose

Alix E. Seif

No relationship to disclose

Brian T. Fisher

Research Funding: Pfizer (Inst), Merck (Inst)

Selina Luger

Consulting or Advisory Role: Pfizer, Novartis, Karyopharm, Sigma Tau

Research Funding: Amgen (Inst), Celgene (Inst), Cyclacel (Inst)

Alan S. Gamis

Consulting or Advisory Role: Pfizer

Peter C. Adamson

Stock or Other Ownership: Johnson & Johnson, Merck, Pfizer, Gilead Sciences

Travel, Accommodations, Expenses: Genentech, Celgene, Pfizer, Nektar

Richard Aplenc

Honoraria: Sigma-Tau

Travel, Accommodations, Expenses: Sigma-Tau

Acknowledgment

We thank Malcolm Smith for his careful review of this manuscript and support. Staci Arnold, Jessica Bokland, James Feusner, Samir Kahwash, Michael Kelly, Matthew Kutny, William Roberts, and Naomi Winick also contributed data for this manuscript.

Appendix

Table A1. AE Definitions for Chart Abstraction and Identification of AEs Using PHIS Data

AE	Chart Abstraction Definition	PHIS Definition
Hypertension	Hypertension in progress note > 1 antihypertensive in MAR and on nursing flow sheet or IV nicardipine, nitroprusside or labetalol	Pharmacy: Amlodipine (133201), nifedipine (133231), hydralazine (133141), or clonidine (134125) for ≥ 2 consecutive days
Hypotension	Hypotension in progress note Dopamine, dobutamine, norepinephrine, or epinephrine administered for > 24 hours, or dopamine > 5 $\mu\text{g/kg/h}$, confirmed on nursing flow sheet and in MAR	Pharmacy: Dobutamine (131305.20), norepinephrine (131351.20), epinephrine (131321.20), or dopamine (131311.20) for ≥ 2 consecutive days
Hypoxia	Respiratory problem indicated in progress note Nasal canula, blow-by, CPAP, BiPAP, or intubation on respiratory care flowsheet and/or nursing flow sheet	ICD-9 or procedure or clinical code: Supersaturated oxygen therapy (00.49), other oxygen enrichment, including oxygen therapy (93.96), oxygen therapy, including oxygen delivery by cannula, mask, tent, or t-tube (521171), high frequency ventilation (521161), CPAP (521162), BiPAP (521164), mechanical ventilation (521166), or other specific ventilator assistance (521169); Or respiratory arrest (799.1), asphyxia and hypoxemia (799.0), hypoxemia (799.02), acute respiratory failure (518.81), respiratory insufficiency (786.09), ventilator management (96.7), or noninvasive mechanical ventilation (93.9)
ARDS	ARDS in attending progress note	ICD-9: Other pulmonary insufficiency, not otherwise classified (518.82), or pulmonary insufficiency after trauma or surgery (518.5x)
Anorexia	Malnutrition or weight loss in progress note Nasogastric feeds or TPN on nursing flow sheet	Pharmacy: Hyperalministration (146040), dextrose and amino acids (146041), dextrose, amino acids, and fat emulsion (146045), travasol (146011), travasol with electrolytes (146015), TPN electrolyte concentrate (146431), TPN HBC (146019), TPN hepatic (146021), TPN renal (146017), or intralipid (146070)
Typhlitis	Typhlitis in attending progress note	ICD-9: Appendicitis, unqualified (541)
DIC	DIC in progress note Fresh frozen plasma, cryoprecipitate, or factor VII on blood bank transfusion record and on nursing flow sheet	Pharmacy and ICD-9: Fresh frozen plasma or cryoprecipitate procedure (99.07), or ICD9 codes (354041,354020), or DIC (286.6)
VGS	VGS in progress note Positive culture for VGS in laboratory results report	ICD-9: Sepsis (038.00)
IFI	IFI in progress note Positive culture for IFI in a sterile site in laboratory results report	ICD-9: Disseminated candidiasis (112.5), aspergillosis (117.3), or other and unspecified mycoses (117.9)
Opiate and PCA	Pain in progress note IV or PCA morphine, hydromorphone, fentanyl, Demerol, or nubain in MAR and on nursing flow sheet	Pharmacy: Morphine (112131), hydromorphone (112117), fentanyl (112115), or nalbuphine (112163)
Seizure	Seizure in progress note Antiepileptic medication in MAR and on nursing flow sheet	ICD-9: Epilepsy or recurrent seizures (345.xx), seizure NOS, or convulsive disorder NOS (780.39)
Renal failure	Renal failure in progress note Renal dialysis on nursing and/or dialysis flow sheet	Clinical: Hemodialysis (525201), peritoneal dialysis (525205), hemoperfusion (525215), or continuous arteriovenous hemofiltration (525221)

Abbreviations: AE, adverse event; ARDS, adult respiratory distress syndrome; BiPAP, bilevel positive airway pressure; CPAP, continuous positive airway pressure; DIC, disseminated intravascular coagulation; HBC, HBC total parenteral nutrition; ICD-9, International Classification of Diseases, Ninth Revision; IFI, invasive fungal infection; IV, intravenous; MAR, medication administration record; NOS, not otherwise specified; PCA, patient-controlled anesthesia; PHIS, Pediatric Health Information System; TPN, total parenteral nutrition; VGS, viridans group streptococcus.

Table A2. Chart Abstraction Data Compared With Clinical Trial Adverse Event Report Data Adjusting for Within-Subject Course Correlation for each of the 12 Grade 3 to 5 Toxicities

Toxicity	Chart Abstraction, No. (%) [*]	Adverse Event Report				
		No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
Hypertension	28 (3.7)	9 (1.2)	21.7 (10.3 to 40.3)	99.6 (98.7 to 99.9)	63.7 (29.8 to 87.9)	96.7 (97.9 to 94.9)
Hypotension	46 (6.1)	35 (4.6)	56.5 (41.1 to 71.7) [†]	98.7 (97.6 to 99.3)	74.3 (56.7 to 87.5) [†]	97.3 (95.8 to 98.2)
Hypoxia	167 (22.0)	30 (4.0)	17.3 (12.0 to 24.4)	99.8 (98.8 to 100)	96.6 (79.6 to 99.5)	80.9 (84.0 to 77.3)
ARDS	13 (1.7)	11 (1.5)	38.5 (17.0 to 65.6)	99.2 (98.2 to 99.6)	45.5 (20.3 to 73.2)	98.9 (97.9 to 99.5)
Anorexia	307 (40.5)	100 (13.2)	31.1 (25.0 to 37.8)	98.7 (97.1 to 99.4)	94.0 (87.6 to 97.2)	67.0 (61.9 to 71.7)
Typhlitis	27 (3.6)	11 (1.5)	36.1 (21.1 to 54.4)	99.9 (99.0 to 100)	90.9 (56.1 to 98.7)	97.7 (96.2 to 98.6)
DIC	59 (7.8)	7 (0.9)	10.4 (4.7 to 21.3)	99.9 (99.0 to 100)	85.7 (41.9 to 98.0)	92.8 (90.6 to 94.5)
VGS	129 (17.0)	103 (13.6)	77.9 (69.8 to 84.4)	99.7 (98.7 to 99.9)	98.0 (92.7 to 99.5)	95.7 (93.9 to 97.0)
IFI	10 (1.3)	10 (1.3)	60.0 (29.7 to 84.2)	99.5 (98.6 to 99.8)	60.0 (29.7 to 84.2)	99.5 (98.6 to 99.8)
Pain	324 (42.7)	56 (7.4)	14.7 (10.6 to 20.0)	98.8 (97.3 to 99.5)	90.8 (79.6 to 96.1)	60.5 (55.9 to 64.8)
Seizure	5 (0.7)	2 (0.3)	0 (0.0 to 52.2) [†]	99.7 (98.9 to 99.9)	0 (0.0 to 84.2) [†]	99.3 (98.4 to 99.7)
Renal failure	6 (0.8)	4 (0.5)	69.3 (16.8 to 83.2)	99.9 (99.1 to 100)	75.0 (23.8 to 96.6)	99.6 (98.8 to 99.9)

NOTE. All data are for patients enrolled in clinical trial AAML0531 for whom chart abstraction was performed.

Abbreviations: ARDS, adult respiratory distress syndrome; DIC, disseminated intravascular coagulation; IFI, invasive fungal infection; NPV, negative predictive value; PPV, positive predictive value; VGS, viridans group streptococcus.

^{*}Chart abstraction data are the gold standard.

[†]Could not perform generalized estimating equation; % and 95% CI are from unadjusted analyses.

Table A3. Chart Abstraction Data Compared With Clinical Trial Adverse Event Report Data Adjusting for Within-Hospital Course Correlation for Each of the 12 Grade 3 to 5 Toxicities

Toxicity	Chart Abstraction, No. (%) [*]	Adverse Event Report				
		No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
Hypertension	28 (3.7)	9 (1.2)	23.5 (9.3 to 48.0)	99.6 (98.9 to 99.8)	72.3 (53.1 to 85.7)	96.6 (93.2 to 98.3)
Hypotension	46 (6.1)	35 (4.6)	56.5 (41.1 to 71.7) [†]	98.7 (97.8 to 99.3)	76.3 (66.7 to 83.9)	97.2 (95.6 to 98.2)
Hypoxia	167 (22.0)	30 (4.0)	15.9 (9.7 to 25.0)	99.8 (98.9 to 100)	96.7 (80.6 to 99.5)	81.4 (76.7 to 85.3)
ARDS	13 (1.7)	11 (1.5)	38.5 (13.9 to 68.4) [†]	99.2 (98.3 to 99.6)	47.4 (23.8 to 72.2)	99.0 (96.9 to 99.7)
Anorexia	307 (40.5)	100 (13.2)	26.3 (14.1 to 43.8)	98.6 (96.7 to 99.4)	93.8 (90.7 to 95.9)	66.2 (56.8 to 74.5)
Typhlitis	27 (3.6)	11 (1.5)	37.4 (20.8 to 57.6)	99.9 (99.1 to 100)	90.4 (54.5 to 98.7)	97.7 (96.3 to 98.6)
DIC	59 (7.8)	7 (0.9)	10.2 (5.1 to 19.4)	99.9 (99.1 to 100)	85.1 (40.5 to 97.9)	92.9 (91.4 to 94.2)
VGS	129 (17.0)	103 (13.6)	77.6 (67.3 to 85.3)	99.7 (98.9 to 99.7)	98.1 (93.4 to 99.5)	95.6 (93.1 to 97.2)
IFI	10 (1.3)	10 (1.3)	61.3 (39.9 to 79.1)	99.5 (98.3 to 99.9)	60.0 (29.7 to 84.2)	99.4 (98.8 to 99.7)
Pain	324 (42.7)	56 (7.4)	14.1 (8.7 to 22.1)	99.1 (95.7 to 99.8)	93.5 (78.7 to 98.2)	60.9 (66.9 to 54.5)
Seizure	5 (0.7)	2 (0.3)	0 (0.0 to 52.2) [†]	99.7 (99.0 to 99.9)	0 (0.0 to 84.2) [†]	99.4 (98.6 to 99.7)
Renal failure	6 (0.8)	4 (0.5)	50.0 (16.8 to 83.2)	99.9 (99.1 to 100)	75.0 (23.8 to 96.6)	99.6 (98.8 to 99.9)

NOTE. All data are for patients enrolled in clinical trial AAML0531 for whom chart abstraction was performed.

Abbreviations: ARDS, adult respiratory distress syndrome; DIC, disseminated intravascular coagulation; IFI, invasive fungal infection; NPV, negative predictive value; PPV, positive predictive value; VGS, viridans group streptococcus.

^{*}Chart abstraction data are the gold standard.

[†]Could not perform generalized estimating equation; % and 95% CI are from unadjusted analyses.

Accuracy of Adverse Event Reporting in Pediatric AML

Table A4. Variation in AE Reporting Accuracy by Hospital

AE	AE in Chart Data Across All Sites, No. (%)	Sensitivity of COG AE Reports Across All Sites, No. (%)	Range of Accurate AEs Reported at Sites, %	<i>P</i> (comparing accuracy across sites)*
Hypertension	28 (3.7)	21.4 (8.3-41.0)	14.3-100	.074
Hypotension	46 (6.1)	56.5 (41.1-71.7)	0-100	.794
Hypoxia	167 (22.0)	17.4 (12.0-24.0)	0-100	.075
ARDS	13 (1.7)	38.5 (13.9-68.4)	0-100	.054
Anorexia	307 (40.5)	30.6 (25.5-36.1)	0-76.3	< .001
Typhlitis	27 (3.6)	37.0 (19.4-57.6)	0-100	.268
DIC	59 (7.8)	10.2 (3.8-20.8)	0-50	.599
VGS	129 (17.0)	78.3 (70.2-85.1)	33.3-100	.016
IFI	10 (1.3)	60.0 (26.2-87.8)	0-100	.924
Pain	324 (42.7)	15.7 (12.0-20.2)	0-35.6	.001
Seizure	5 (0.7)	0 (0.0-52.2)	0	N/A
Renal failure	6 (0.8)	50.0 (11.8-88.2)	0-100	.400

NOTE. All data are for patients enrolled in clinical trial AAML0531 for whom chart abstraction was performed. Chart abstraction data are the gold standard. Abbreviations: AE, adverse event; ARDS, adult respiratory distress syndrome; COG, Children's Oncology Group; DIC, disseminated intravascular coagulation; IFI, invasive fungal infection; N/A, unable to be evaluated as a result of a small number of gold standard AEs in chart data. VGS, viridans group streptococcus. *Using Fisher's exact test or χ^2 test depending on the number of AEs.

Table A5. Chart Abstraction Data Compared With PHIS Resource Use Data Adjusting for Within-Subject Course Correlation for Each of the 12 Grade 3 to 5 Toxicities

Toxicity	Chart Abstraction, No. (%)*	PHIS Data					PHIS Data Source
		No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	
Hypertension	28 (3.7)	44 (5.8)	62.4 (41.9 to 79.2)	95.9 (93.6 to 97.3)	40.1 (25.6 to 56.6)	98.5 (99.2 to 97.1)	Pharmacy
Hypotension	46 (6.1)	11 (1.5)	11.1 (4.7 to 24.1)	99.2 (97.9 to 99.7)	50.0 (22.5 to 77.5)	94.4 (92.4 to 95.9)	Pharmacy
Hypoxia	167 (22.0)	201 (26.5)	74.8 (66.7 to 81.5)	85.4 (81.2 to 88.9)	63.6 (55.7 to 70.8)	91.9 (88.7 to 94.2)	ICD-9, procedure, clinical
ARDS	13 (1.7)	11 (1.5)	30.8 (12.0 to 59.1)	99.1 (97.7 to 99.6)	36.4 (14.3 to 66.1)	98.8 (97.7 to 99.4)	ICD-9
Anorexia	307 (40.5)	237 (31.3)	81.1 (74.8 to 87.1)	99.3 (97.9 to 99.8)	98.7 (96.2 to 99.5)	86.6 (81.3 to 90.5)	Pharmacy
Typhlitis	27 (3.6)	26 (3.4)	67.4 (49.3 to 81.5)	98.7 (97.4 to 99.4)	69.2 (49.5 to 83.8)	98.8 (97.7 to 99.4)	ICD-9
DIC	59 (7.8)	63 (8.3)	79.2 (67.7 to 87.3)	97.7 (96.2 to 98.6)	75.0 (62.6 to 84.3)	98.3 (97.0 to 99.0)	Pharmacy, ICD-9
VGS	129 (17.0)	53 (7.0)	30.3 (22.5 to 39.3)	97.5 (95.7 to 98.6)	72.2 (59.9 to 81.9)	87.3 (84.5 to 89.7)	ICD-9
IFI	11 (1.5)	20 (2.6)	70.0 (37.6 to 90.0)	98.1 (96.2 to 99.1)	35.0 (17.7 to 57.4)	99.6 (98.8 to 99.9)	ICD-9
Pain	324 (42.7)	511 (67.4)	99.1 (97.2 to 99.7)	54.7 (49.4 to 60.0)	61.7 (56.6 to 66.4)	98.8 (96.4 to 99.6)	Pharmacy
Seizure	5 (0.7)	19 (2.5)	80.0 (30.9 to 97.3)	97.4 (95.2 to 98.7)	19.6 (8.2 to 40.0)	99.9 (99.0 to 100)	ICD-9
Renal failure	6 (0.8)	6 (0.8)	83.3 (36.9 to 97.7)	99.9 (99.1 to 100)	83.3 (36.9 to 97.7)	99.9 (99.1 to 100)	Clinical

NOTE. All data are for patients enrolled in clinical trial AAML0531 for whom chart abstraction was performed. Abbreviations: ARDS, adult respiratory distress syndrome; DIC, disseminated intravascular coagulation; ICD-9, International Classification of Diseases, Ninth Revision; IFI, invasive fungal infection; NPV, negative predictive value; PHIS, Pediatric Health Information System; PPV, positive predictive value; VGS, viridans group streptococcus.

*Chart abstraction data are the gold standard.

Table A6. Chart Abstraction Data Compared With PHIS Resource Use Data Adjusting for Within-Hospital Course Correlation for Each of the 12 Grade 3 to 5 Toxicities

Toxicity	Chart Abstraction, No. (%) [*]	PHIS Data					PHIS Data Source
		No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	
Hypertension	28 (3.7)	44 (5.8)	64.8 (44.4 to 81.0)	96.3 (93.0 to 98.1)	40.0 (28.8 to 52.3)	98.4 (96.2 to 99.3)	Pharmacy
Hypotension	46 (6.1)	11 (1.5)	11.5 (6.0 to 21.0)	99.2 (97.9 to 99.7)	46.4 (19.8 to 75.2)	94.5 (92.1 to 96.2)	Pharmacy
Hypoxia	167 (22.0)	201 (26.5)	74.2 (60.3 to 84.6)	89.7 (72.5 to 96.7)	69.8 (50.7 to 83.8)	92.9 (86.8 to 96.3)	ICD-9, procedure, clinical
ARDS	13 (1.7)	11 (1.5)	33.5 (11.8 to 65.4)	99.0 (97.6 to 99.6)	36.4 (14.3 to 66.1)	98.9 (96.9 to 99.6)	ICD-9
Anorexia	307 (40.5)	237 (31.3)	79.0 (64.2 to 88.7)	99.3 (98.3 to 99.7)	98.5 (96.8 to 99.3)	86.9 (76.9 to 93.0)	Pharmacy
Typhlitis	27 (3.6)	26 (3.4)	65.9 (51.2 to 78.0)	98.9 (97.7 to 99.5)	67.9 (52.9 to 80.0)	98.8 (97.9 to 99.3)	ICD-9
DIC	59 (7.8)	63 (8.3)	79.8 (73.3 to 84.9)	97.7 (96.8 to 98.3)	75.1 (64.0 to 83.7)	98.4 (97.7 to 98.8)	Pharmacy, ICD-9
VGS	129 (17.0)	53 (7.0)	30.2 (22.2 to 39.6)	97.8 (96.2 to 98.7)	73.8 (61.1 to 83.5)	87.3 (84.4 to 89.8)	ICD-9
IFI	11 (1.5)	20 (2.6)	68.7 (35.5 to 89.8)	98.2 (96.8 to 99.0)	34.5 (27.5 to 42.3)	99.5 (98.9 to 99.8)	ICD-9
Pain	324 (42.7)	511 (67.4)	99.0 (97.6 to 99.6)	56.5 (47.1 to 65.4)	62.9 (55.5 to 69.7)	98.1 (97.1 to 98.7)	Pharmacy
Seizure	5 (0.7)	19 (2.5)	80.0 (30.9 to 97.3)	97.7 (95.3 to 97.7)	16.2 (7.7 to 31.0)	99.9 (99.1 to 100)	ICD-9
Renal failure	6 (0.8)	6 (0.8)	83.3 (36.9 to 97.7)	99.9 (99.1 to 100)	83.3 (36.9 to 97.7)	99.9 (99.1 to 100)	Clinical

NOTE. All data are for patients enrolled in clinical trial AAML0531 for whom chart abstraction was performed.

Abbreviations: ARDS, adult respiratory distress syndrome; DIC, disseminated intravascular coagulation; ICD-9, International Classification of Diseases, Ninth Revision; IFI, invasive fungal infection; NPV, negative predictive value; PHIS, Pediatric Health Information System; PPV, positive predictive value; VGS, viridans group streptococcus.

^{*}Chart abstraction data are the gold standard.

Table A7. Proposed Targeted Acute Toxicities for Pediatric Acute Myeloid Leukemia Clinical Trials

Organ System	Toxicity
Cardiac	Hypertension
	Hypotension
	Left-ventricular systolic dysfunction
Pulmonary	Adult respiratory distress syndrome
	Hypoxia
Gastroenterology	Anorexia
	Typhlitis
	Veno-occlusive disease
Hematology	Disseminated intravascular coagulation
	Hyperbilirubinemia
	Time to count recovery
Infectious disease	Blood stream infection
	Invasive fungal infection
Neurology	Pain
	Seizure
Renal	Renal failure