Deadline Feb 28th 11:59 PST
3000 characters including spaces: currently 2,984
1 table/figure as jpg

Tamara P. Miller, MD, MSCE
Vikki Nolan, DSc, MPH
Edward Krause, MS
Yang Xie, MBBS, MS
Gregory Joseph, MS
Jun Gun Jo
Yimei Li, PhD
M. Monica Gramatges, MD, PhD
Jennifer J. Wilkes, MD, MSCE
Gregory T. Armstrong, MD, MSCE
Richard Aplenc, MD, PhD

Title: Automating Ascertainment of Chemotherapy Exposure for the Childhood Cancer Survivor Study with the ExtractEHR+ Toolkit

Background/Purpose
Since the National Cancer Institute established a pediatric cancer survivor cohort in 1971, manual chart abstraction has identified chemotherapy exposures. While many teams, including the Childhood Cancer Survivor Cohort (CCSS), have performed transformative research using manual abstraction, it is time intensive. The ExtractEHR+ toolkit, an R-based software package that includes ExtractEHR and MedCleanEHR, automatically extracts and curates electronic health record (EHR) data. We sought to test concordance between manually and ExtractEHR+ ascertained chemotherapy data for medications in the CCSS medical record abstraction form (MRAF) at 3 sites.

Methods
MRAF specified medications, including novel therapies and supportive care, were manually abstracted by trained staff using CCSS abstraction guidance for 50 patients across 14 diagnoses at Children's Healthcare of Atlanta, Children's Hospital of Philadelphia, and Texas Children's Hospital. Included patients were aged 0-20 years at diagnosis, English or Spanish speaking, and diagnosed with *de novo* cancer after EHR implementation at each site. ExtractEHR extracted all medication orders and administrations, including outpatient prescriptions. MedCleanEHR centrally cleaned and identified unique MRAF exposures. For 146 medications on the MRAF, presence (yes) or absence (no) in the medical record was ascertained manually and by ExtractEHR+. Concordance between manual and ExtractEHR+ MRAF medication exposure (yes/no) was tabulated and discrepancies were evaluated. Ascertainment time was estimated for both methods.

Results
Of 12,752 identified survivors, 11,818 had EHR data available. A total of 27,709,224 MRAF medication actions were extracted by ExtractEHR; after MedCleanEHR processing, there were 8,523,729 unique medication administrations. The 150 patients with manual chart abstraction data had 21,900 possible MRAF defined medication exposures (150*146). Overall concordance for yes/no exposure between manual and ExtractEHR+ was high at 98.3% with 20,633 no/no (94.1%) and 902 yes/yes (4.1%) concordance. 278 exposures were identified only by ExtractEHR (1.3%) and 87 were identified only in manual review (0.4%). Of 1,267 total MRAF exposures, 21.9% (278) were not identified by manual

abstraction and 6.9% (87) were not identified by ExtractEHR+. The average time required for manual exposure ascertainment was 3 hours per patient while MedCleanEHR ascertainment required 0.001 seconds per patient.

Conclusions

ExtractEHR+ accurately and efficiently ascertains chemotherapy exposures for the CCSS MRAF. Manual abstraction had larger medication identification errors and ExtractEHR+ provides a 15% absolute gain in identification of chemotherapy exposures among patients with an exposure. Work is ongoing to complete planned analyses at 5 centers, understand intra/inter center level manual abstraction variability, analyze concordance of cumulative dose data, and complete a cost analysis.