Workshop #4

Innovative Models and Data Management

Session 3

Workshop #4 Session 3

Facilitators

Jaime GUIDRY AUVIL - US National Cancer Institute

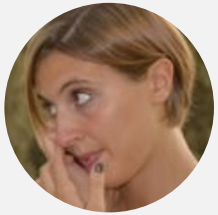
Gijs GELEIJNSE - Netherlands Comprehensive Cancer Organisation

Subthemes

- Would a hybrid model (a combination of options discussed at subthemes 1 and 2 be a feasible option? What types/categories of data could be released via each of these tiers?
- Are there different types of research questions/analysis that can be answered and cancer statistics that can be calculated using the different mechanisms?

Discussants

Tamara MILLER
USA
Emory University School of Medicine
Pediatric Oncologist



Laura BOTTA
Italy
Istituto Nazionale dei Tumori Foundation
Senior biostatistician – Cancer epidemiology

Paris Conference for an International Childhood Cancer Data Partnership
Newcap Event Center (Paris, France), November 7-8, 2023

Leveraging Automated Methods to Ascertain and Use Cancer Data

Tamara P. Miller, MD, MSCE

Emory University/Children's Healthcare of Atlanta, United States

- Cancer cohorts and clinical trials currently rely on manual collection of data
- Inherent limitations in manual data ascertainment
  - Time- and labor-intensive
  - Typically only include a select number of targeted data elements
  - Missed and incorrect data

66% of key Adverse Events missed
25% of reported Adverse Events were incorrect

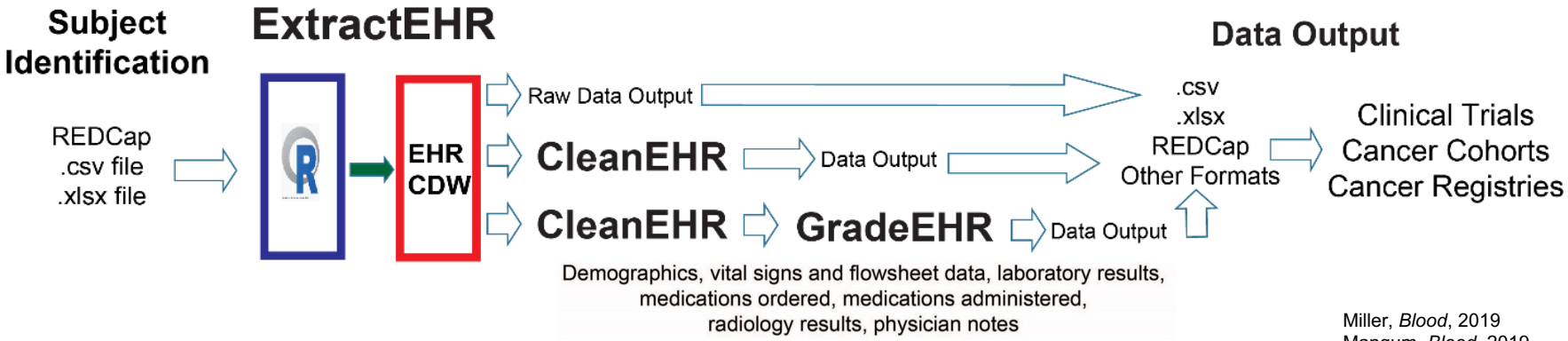
JOURNAL OF CLINICAL ONCOLOGY | ORIGINAL REPORT

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

Miller, *JCO*, 2016

- Developed ExtractEHR as a potential solution
- R package that extracts data from electronic health record (EHR) data warehouse
  - Series of post-extraction packages process extracted EHR data to provide clinical context for cancer cohorts



Miller, *Blood*, 2019
Mangum, *Blood*, 2019
Myers, *Blood*, 2019
Yi, *Haematologica*, 2022
Miller, *Lancet Haem*, 2022

- Multi-site implementation: ExtractEHR implemented at 4 hospitals, 4 additional in process
  - Successful at hospitals using Epic and Cerner EHR vendors
  - Once installed, ExtractEHR can extract data repeatedly
- Customizable inclusion of EHR components and post-extraction packages based on use case
  - ExtractEHR can extract specified or all laboratory results data
  - CleanEHR processes and cleans laboratory data, removing false positive results
  - GradeEHR grades adverse events (AEs) per NCI Common Terminology Criteria for Adverse Events definitions
- More accurate than manually ascertained data (Miller, *BJH, 2017)*
  - Manually collected trial data: 85% of laboratory AEs missed, 50% incorrect
  - Automated data collection: 0.2% of laboratory AEs missed, 0.5% incorrect

- Provision of data to NCI Surveillance, Epidemiology, and End Results (SEER) Program
  - Initial pilot: use ExtractEHR to move data from Children's Healthcare of Atlanta to Georgia (GA) Cancer Registry
    - Included patients diagnosed in 2019 or 2020 with 5 cancer types to establish feasibility
    - Extraction of wide range of EHR data elements, selected due to clinical relevance
      - No post-extraction processing
    - Transfer data securely to GA Cancer Registry
  - Expanding to include all cancer diagnoses and 3 additional hospital-registry pairs in 2023-2024

- 306 patients, 2,241,963 extracted data elements

	Unique Patients With Results	Number of Data Elements Extracted
Addresses	306	914
Demographics	306	306
Inpatient visits	295	2262
Clinic visits	306	33891
Laboratory test results	304	831614
Microbiology	276	56348
Pathology	302	2536
Medication orders	305	280867
Medication administrations	305	574462
Procedures	306	344548
Height	305	16691
Weight	306	53521
Radiology result reports	305	7062
Oncology notes	300	34756
Genomics	147	787

- Automatically ascertained EHR data can be formatted for sharing across institutions/cohorts
- Leukemia Electronic Abstraction of Records Network (LEARN): observational cohort of pediatric patients
  - Merged ExtractEHR data from 4 hospitals to create granular dataset for clinical epidemiology research
    - De-identify protected health information post-extraction (e.g. names, medical record numbers, dates)
    - Per data use agreement, de-identified data transferred and stored centrally
  - Parallel formatting between sites due to coding embedded in ExtractEHR
    - Post-extraction central processing harmonizes differences in data elements between sites
      - e.g. Different nomenclature for results in EHR data (“White Blood Cell” vs. “WBC”)
- Children’s Oncology Group/Pediatric Brain Tumor Consortium trial PEPN21EHR/PBTC-N15 (NCT05020951)
  - ExtractEHR or locally developed packages extract and format laboratory results
  - Pre-specified formatting permits direct upload into electronic data capture system

Emory/Children's Healthcare of Atlanta

- Nicholas DeGroote, MS
- Judy Lee
- Anjali Khanna, MBBS, MPH

Children's Hospital of Philadelphia

- **Richard Aplenc, MD, PhD**
- Evanette Burrows, MPH
- Brian Fisher, DO, MSCE
- Kelly Getz, PhD
- Robert Grundmeier, MD
- Allison Heath, PhD
- Yun Gun Jo
- Edward Krause, MS
- Yimei Li, PhD
- Mark Ramos
- Adam Resnick, PhD

Texas Children's Hospital

- M. Monica Gramatges, MD, PhD
- Philip Lupo, PhD
- Karen Rabin, MD, PhD
- Michael Scheurer, PhD

Seattle Children's Hospital

- Jennifer Wilkes, MD, MSCE

COG

- Peter Adamson, MD
- Todd Alonzo, PhD
- Douglas Hawkins, MD

NIH

- Malcolm Smith, MD, PhD
- Lynne Penberthy, MD, MPH
- Johanna Goderre, MPH

Funding

- 3P30CA138292-14S2
- K07 CA211956
- National Cancer Institute/CCDI Support
- Damon Runyon-Sohn Pediatric Fellowship
- Alex's Lemonade Stand Foundation
- CHOP Philanthropy

Paris Conference for an International Childhood Cancer Data Partnership
Newcap Event Center (Paris, France), November 7-8, 2023

Experience and challenges in clinical and population-based cancer registry for analysing rare cancers data

Laura Botta

Evaluative Epidemiology Unit, Department of Epidemiology and Data Science, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

laura.botta@istitutotumori.mi.it

Where we are?

Individual pseudonymised data
Population-Based Cancer Registry (PBCR) Experience

Information Network on Rare Cancers RARECARENet

EUROCARE study
Survival of cancer patients in Europe

About 100 PBCRs involved, 30 countries.
In future: privacy assessments between PBCRs and the Joint research center and INT/ISS or hybrid analysis (individual and grouped data).

BENCHISTA project
International benchmarking of childhood cancer survival by stage

About 70 PBCRs involved. To achieve research collaboration: **18** months to finalize the privacy assessment.
In future: all this work will have to be redone.

References:
Long-term survival and cure fraction estimates for childhood cancer in Europe (EUROCARE-6): results from a population-based study. Botta et al. LO 2022
International benchmarking of childhood cancer survival by stage at diagnosis: The BENCHISTA project protocol. Botta et al. PLOS ONE 2022
Cancer data quality and harmonization in Europe: the experience of the BENCHISTA Project. Lopez-Cortes et al. Frontiers 2023

Federated learning approach

Population based cancer registry data:

RARECARENET Asia

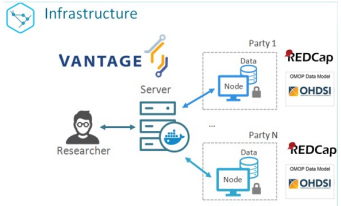
PBCR Head and neck data analyzed using VANTEGE6 An open-source infrastructure for privacy preserving analysis.

Clinical Cancer Registry:

STARTER

STARTER PROJECT
STarting an Adult Rare Tumour European Registry

Hospital based cancer registry collecting rare head and neck cancer data. VANTEGE6. Legal framework and DPIA in place that will last "forever".



References:
Head and neck cancers survival in Europe, Taiwan, and Japan: results from RARECAREnet Asia based on a privacy-preserving federated infrastructure. Botta et al. Frontiers Oncology 2023
The observational clinical registry of the ERN on Rare Adult Solid Cancers: The protocol for the rare head and neck cancers. Trama et al. PLOS ONE 2022

Features	Centralised vs federated
Latency of computation	CENTRALIZED reduced reliance on external systems; in the federated analysis the speed is set to the slower machine involved
Data management/Data analysis	CENTRALIZED Individual level data quality checks; all type of analysis are feasible; Possibility to aggregate countries to overcome rarity issue
Lightness of technical implementation	CENTRALIZED IT infrastructure needed is easier
Data updated	BOTH
Data availability	BOTH FEDERATED data are always accessible when needed but the CENTRALIZED relies less on external sources
Privacy assessment	FEDERATED is more privacy preserving
Security / Data breach	FEDERATED Reduced amount of data in case of breach
Privacy-by-design principles	FEDERATED Avoids creating additional copies of data, stored in the original source system and does not have to be communicated or transfer
Expanding trust	FEDERATED Possibility to opt in/out; all analyses and requests are tracked.

Where are we going?

- Regardless the type of data (population based /clinical data) and whether it is a study or a registry: Privacy assessment is part of research life and we can't ignore it

- Peculiarities of rare cancers

Although the IT infrastructure required is complex, the FEDERATED LEARNING APPROACH is evolving rapidly.
It is THE FUTURE, but it takes time.

PBCRs will be the first to benefit from the federated approach because they are dedicated to research (technical readiness, standardized data collections).

IN THE MEANTIME:

- Standardized the dataset as much as possible across countries and projects

- Using an Hybrid model (individual data + grouped data) if possible. Ok for some statistical analysis (descriptive analysis, univariate models) but difficult for others such as multivariate analysis and Propensity score definition. Difficult for data research/exploration.

Many thanks to you and to all the people who collaborated with me on these projects














laura.botta@istitutotumori.mi.it