National Cancer Policy Forum Workshop: Session 3 Interoperable Electronic Health Records in Cancer Surveillance

Pete Stetson, MD, MA, FAMIA Chief Health Informatics Officer

Enabling 21st Century Applications for Cancer Surveillance through Enhanced Registries and Beyond Challenges and Opportunities: Data Collection Methodologies and Technological Advances in Cancer Surveillance National



Memorial Sloan Kettering Cancer Center

Problem: High-quality, sharable data

- We want to predict cancer onset, tumor evolution, treatment response, toxicity burden so we can improve outcomes and value
- We want to tailor therapies to the whole patient
- Can't do these without <u>good</u>
 <u>data still a constraint on Al</u>





Approach:

Tackling real-world problems through quality data

- Deep phenotyping of genotyped cohort
- Standardizing data using Core Clinical Data Elements (CCDE)
 - n=2000 concepts; ~200 typically really hard to get
- Current Use Cases (sample)
 - Curation pipeline for researchers
 - Cancer registry automation pipeline
 - Clinical trial cohort management
 - Predicting tumor evolution
 - Fusion models exploration



Outputs **Concept-based** phenotypes (and **ePROs** phenotype-lets)

MSK Extract Pipeline Effort led by John Philip, MS Neil Shah, MBBS

Extract CCDE Curation Summary Dashboard | Treatment & Outcomes (筆)

Description: A summary and exploration tool for a subset of the 150+ clinical data elements curated for over 25k patients as part of the ongoing Extract CCDE Curation initiative. See the 'i ' at top right for more info and a data release timeline.

Instructions; Make selections using dropdown filters to query across data domains. Use the 'Make Selections' button below to make additional filter selections, or the 'View List' button to see a deidentified cohort list and individual patient journeys. See the '? ' at top right for more info.



? 6/3/2024 10:02:50 PM

Dataline

Data Updated * :

*Hover right for more info

Benefits Impact on AI in Oncology; Learning Health Systems











Smith MA et al Nature Methods

Relieve the data quality constraint for AI Provide proving grounds for AI models Data density further sped by AIenabled curation Computable phenotypes (like eMerge) and journeys

Combined Tumor/Clinical Response predictions



Future State:

Possible roadmap

- Automated cancer registry at national scale
- Federated tuning of <u>AI-enabled curation</u> (generalized model assurance)
- Industry/Academic partnerships for a <u>computable phenotype ecosystem</u>
 - Major EHR and major cloud compute companies enable sharing/publishing of pragmatic computable oncology phenotype (PCOPs) definitions into local EHR instances (Like TurboBooster for research order sets)
 - Your data stays local
 - AI tools know what they're analyzing
 - Obtain semantic interoperability, not just data interoperability
- <u>Cross-institution computable journeys</u> (care started at hospital A, continued at hospital B, failed 1st-3rd line Rx now on trial, toxicity at Hospital C, etc)

