New Informatics Tools and Methods to Enhance U.S. Cancer Surveillance Research

Angela Mariotto and Nadia Howlader
Surveillance Research Program

NIH NATIONAL CANCER INSTITUTE

ITCR
November 4th, 2016
Presentation Objectives

- Introduction to SEER and its mission
- Discuss the challenges to cancer surveillance in the current setting for cancer care
- Program Announcement:
  - New Informatics Tools and Methods to Enhance U.S. Cancer Surveillance Research (UG3/UH3)
Surveillance Epidemiology and End Results (SEER)

- SEER’s mission is to support research
  - Diagnosis, treatment, and outcomes of cancer since 1973
- Population-based registries representing 30% of the U.S. population
  - Are charged to collect every reportable cancer case in their catchment area
- Over 450,000 incident cases reported annually
- Approximately 80% cases electronic path (E-path) reports received in near real time by registries
  - Over 1 million E-path reports processed manually per year
  - Three year lag between actual diagnosis and reporting
SEER and the CDC National Program of Cancer Registries (NPCR)
Key categories of data in SEER

- Demographics
- Tumor Characterization
- Initial Course of Therapy
- Biomarkers
- Outcomes (Survival/Prevalence)
Challenges to the SEER Mission

- More diversity on the delivery of cancer care
  - Expand data collection beyond Hospitals

- Outcomes other than survival
  - Recurrence/metastases/progression
  - Longitudinal information: subsequent treatment, Quality of Life (QOL)

- Increased number of cancer cases (aging population) to be abstracted with static funding
  - Current processes are mainly manual

- Increasing complexity of cancer care with expanded treatment protocols and new biomarkers
Strategic Vision for SEER

- Improve Registry Data Collection Methods
  - Automation and direct capture/processing of data from diverse data sources
- To expand, enhance and improve the quality of SEER data
- Expand methods to analyze, interpret, report and visualize SEER and related databases
- Expand the capacity of SEER to support cancer research across a broader continuum
New Informatics Tools and Methods to Enhance U.S. Cancer Surveillance Research (UG3/UH3) PAR-16-349
Purpose of the Program Announcement

- To advance the science of cancer surveillance.
- To improve the collection and integration cancer registry data.
- To expand the data items collected and the usefulness of registry-collected data to support high-quality cancer research.
- To enhance the registry core infrastructure.
  - Applications must be built on partnership with U.S. population-based central cancer registries.
Main Areas of Research: Application may cover one or more

- **Area 1**: Methods for automatic/unsupervised/minimally supervised extraction and integration of discrete cancer-related data from free-text medical reports.

- **Area 2**: Approaches for optimized integration of registry with various types of data.

- **Area 3**: Approaches to obtain prospective data stream to be part of the registry data flow.
Area 1: Methods for data extraction and integration

- Extraction of discrete cancer-related data from unstructured free-text medical reports (e.g. pathology reports, radiology reports, diagnostic imaging interpretations, laboratory results, EMR);

- Machine assisted human review methods that combine both NLP and manual review; information is automatically extracted and stored when there is certainty that the correct information is being retrieved and recommended for manual review when the probability of a correct response is lower.
Area 2: Approaches for optimized integration of registry with various types of data.

- Linkages with
  - Claims from health insurance plans, large managed care organizations, providers to obtain detailed treatment information (agents, dose, frequency);
  - Commercial pharmacy databases to obtain detailed oral cancer drug information;
  - Linkages with surveys to obtain information on risk factors
Area 3: Approaches to obtain prospective data stream to be part of the registry data flow

- Continuous automatic feed from oncology practice EHRs;
- Develop approaches for continuous collection of genetic or genomic test information via linkage with industry providers;
- Development of tools/methods to receive data streams from laboratories on an ongoing basis.
UG3/UH3 Phase Innovation Awards Cooperative Agreement

UG3 Phase (Up to 2 years)

Investigator → U.S. Registry A → Pilot methods in Registry A → Transition from UG3 to UH3: After NCI administrative review

No more than $300K per year (direct)

UH3 Phase (Up to 3 Year)

U.S. Registry B + → Scale up in Registry B or more

No more than $500K per year (direct)
Eligibility and Key Dates

- Eligibility: U.S. Institutions only
- Partnering with a minimum of two U.S. population-based central cancer registries (funded by the SEER or NPCR programs)
  - Contact information:
- 2 reviews per year for 3 years:
  - Open Date: September 14, 2016
  - Application Due Date(s): October 14, 2016; April 14, 2017; November 30, 2017; April 16, 2018; November 30, 2018; April 16, 2019
Acknowledgments

SRP
- Angela Mariotto
- Paul Fearn
- Kathy Cronin
- Eric Feuer
- Rose Fredua
- Carol Kosary
- Zaria Tatalovich
- Lynne Penberthy

DCCPS
- Shobha Srinivasan
- David Chambers
- Mark Alexander
Questions?

Thanks

Angela.Mariotto@nih.gov