National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

SEER Update
February 24, 2017

Population Health Assessment in Cancer Center Cachement Areas Meeting
Columbus, Ohio

Objectives

- Describe Current SEER structure & potential value for research
- Describe
  - gaps in surveillance (SEER and others)
  - How these gaps are being filled by SEER – and potentially other regsitries

Surveillance Epidemiology and End Results (SEER)

The SEER Program is a national resource:

- Funded by NCI ***to support research*** on the diagnosis, treatment and outcomes of cancer since 1973
- Population-based registries covering 30% of the US population
  - Representing racial and ethnic minorities
  - Various geographic subgroups
- –450,000+ incident (newly diagnosed) cases reported annually
- Undergoing contract recompete- full and open for the first time in 44 years!
  - Expanding the coverage (increase in registries)
  - Developing two categories of registries under the SEER Program
    - Registries to support research
    - From these- the Core Data Reporting Registries will be selected

SEER: Data Currently Collected

- Demographics
- Detailed histopathologic characterization of the tumor at diagnosis
  - Histology
  - Molecular characterization (limited currently)
- Stage at Dx
  - Anatomic (tumor size, nodal involvement, location of mets)
  - Non-anatomic factors (ER, PR, Her2, PSA, Gleason, grade, Oncotype Dx etc)
- Initial course of therapy
  - Surgery
  - Radiation
  - Chemotherapy (under-reported and currently limited to yes/no/unknown)
- Survival
  - Routine follow up
  - Date of death
  - Cause of death

SEER Program

- Only population-based system in the US that includes a broad set of clinical variables
- Variable selection driven by guidelines and standards
  - Current - 32 predictive & prognostic biomarkers
- In Process –
  - Guideline review to identify relevant new variables to be collected for example:
    - EGFR/ALK lung cancer
    - BRAF/MSI Colon cancer,
    - BRAF Melanoma
  - Capture of detailed longitudinal treatment information
    - Claims based
    - Pharmacy linkages for orally administered agents
  - Developing methods for capturing outcomes other than survival
    - Recurrence/progression
    - Patient generated data

FILLING THE GAPS TO MEET CURRENT & FUTURE RESEARCH NEEDS

6

Summary of the Strategic Vision for SEER

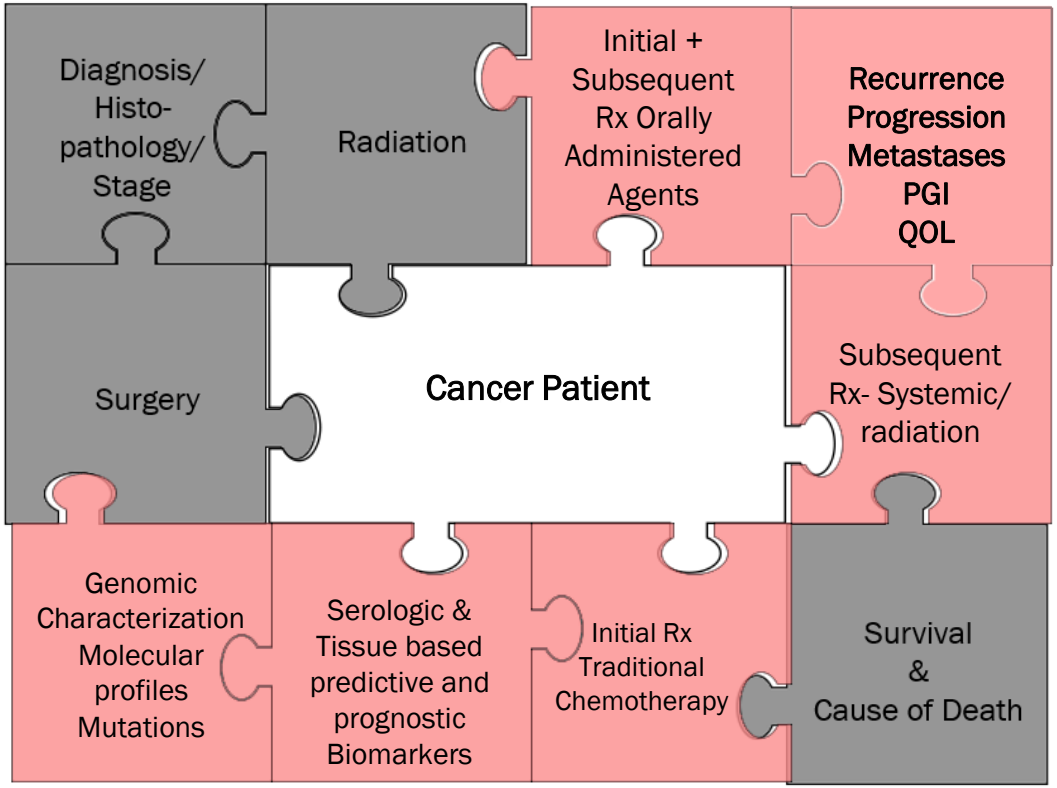
Enhancing SEER as a national resource to support cancer research

1. Automation and direct capture/processing of data to expand, enhance and improve the quality of SEER data to better support research by filling gaps in the current data collection
   - Linkages
   - Auto-processing of data
   - Quality Assurance/ Quality Improvement

2. Expand the capacity of SEER & SRP to support cancer research across a broader continuum
   - SEER linked Virtual Bio-Repository (VTR)
   - Virtual Pooled Repository (VPR)
   - Generate FOAs to support surveillance methods research

Putting the puzzle together for each cancer patient: Data Collection

Diagnosis → Death



Automate and directly capture data via linkages – treatment

Problem: Lack of complete and detailed treatment

Solutions:

- Link with existing data for pharmacy-provided oral drugs
- Capture and process standardized claims for infusion therapy
- Receive data directly from medical oncology and radiation oncology vendors for enhanced completeness and detail on treatment and outcomes

Treatment Linkages: targeting high volume/ high relevance sources - status

Pharmacy data

- In process for pilot testing receipt of CVS data in GA
- Walgreens- in discussion re: DUA using CVS template

Treatment Linkages: Pilots with high volume oncology "vendors" - status

- Claims – Unlimited Systems (>2000 oncologists across US- practice based)
  - 3 years retrospective data received and real time data flow developed for 6 practices
  - **Expanding to 6 other registries this year**
  - Coverage to include 15-45% of oncology practices in 7 SEER registries
  - **Potential for implementation of mandatory claims reporting from ALL oncology practices in a state** (modeling on FL)

- **Planning claims workshop 2017**
  - **Purpose: assess scalability of existing work at Fred Hutch and U KY to SEER wide implementation of linking claims data to supplement treatment reporting and case finding**

Treatment Linkages: Pilots with high volume oncology "vendors" - status

- Radiation Oncology/EMRs – Metrik/Mosaic (45%) & Varian (50%)
  - Pilots in development/ implementation in KY

- Oncology Practice EMR data – CancerLinQ
  - MOU in development for use of SEER data for QC in practices
  - Pilot in development for direct data exchange of practice data in GA, Iowa and Utah

Automate and directly capture data via linkages – genomics data

Problem: Need to expand and automate collection of genomic characterization of tumors- not collectible via manual abstraction

Solutions:

- Link with existing external labs performing these tests
- Change the mandate for “reportability” to cover capture of panels

Genomics Data Linkages: targeting high volume/ high relevance sources - status

- GHI
  - Oncotype DX 21 and 16 gene assay completed and repeating annually
- Syapse
  - Genomic data acquisition specialist (analogous to Unlimited Claims vendor)
  - Receive and store data from >15 genomic testing labs for clinical client access
  - Developing pilot in GA for linkage with selected genomic lab data
- Foundation Medicine
  - Pilot in discussion with Foundation
- BRCA Breast and Ovarian
  - Pilot linkage in GA/ CA completed for 2013-15 data
  - Processes for linkage confirmation/data access in development
  - Goal – link data from Myriad et al across all SEER

Expand Outcomes Data Collection

Focus on better understanding the course of disease among > 15 million cancer survivors

Capturing Disease Progression/Recurrence

- Complex diagnostic patterns require multiple approaches varying by cancer site (e.g., NLP and serologic biomarkers)

Collecting Patient-Generated Health Information

- Working with partners to test solutions, e.g., patient portals, direct patient reporting, and patient-generated data sources
  - MBCA/ACS/NCI funded activities for focus groups on patient reporting to registries
  - NCI funded registry study in 5 registries to explore methods for patients to provide data to registries

Automation – NLP/ Machine learning – DOE Pilot

Dept of Energy Partnership for joint funded 3 year pilot

- Three focus areas
  - NLP/machine learning (Aim 1)
    - Leveraging data pipeline for targeted variables form existing unstructured text documents (e.g. path reports, radiology reports)
  - Linking in additional data from alternative sources for optimizing info in registry (Aim 2)
    - Leveraging claims, pharmacy data, genomic data etc.
  - Modeling of patient trajectory over time including all these diverse data sets (Aim 3)
    - Initial use cases focusing on recurrence or disease progression

EXPANDING THE CAPACITY FOR SEER TO SUPPORT RESEARCH

17

SEER-Linked Virtual Bio-Repository

What is it?

- A ***virtual*** repository of SEER-based tissue with annotation
- Tool for researchers to search de-identified abstracts linked to electronic path reports to select a set of relevant specimens
- Ultimate aims
  - Annotation and search capacity of abstracts + e path reports for all SEER cases with tissue
  - Centralization of requests for specimens and custom annotation
  - Capacity for investigators to custom select relevant cases for their research

SEER-Linked Virtual Bio-Repository: Benefits

- Population based – permitting comparison of subsets
- Available across a broad spectrum of health care facilities/pathology labs (not just academic centers)
- Access to rare cancers and exceptional outcomes
- Linked long term outcomes
- Existing annotation with clinical and demographic data
- Potential for custom annotation
- Renewable with > 400,000 incident cases annually

SEER-Linked Virtual Bio-Repository Pilot

Purpose

- Assess best practices across multiple SEER registries
  - Many registries are performing tissue acquisition but as single entities
- Estimate costs of supporting a scaled SEER-wide system
- Assess availability of specimens
- Understand human subjects/consent as requirements vary by registry and prepare for common rule changes
- Using a test case: provide a publicly available clinical and genomic data set for investigators to access

7 registries funded for pilot of pancreas and breast 9/15

- Focus on “exceptional” survivors
  - 431 early stage node negative breast cancer (< 2 yr survival)
  - 224 pancreatic adenoca long term survivors (> 5 yr survival)

Virtual Pooled Registry with NAACCR

What is it?

- A ***virtual*** national cancer registry
- Tool for researchers to automatically link patients with ***all US cancer registries***
- Ultimate aims
  - Automated linkage via Honest Broker
  - Centralized IRB
  - Return of patient information on cancers, survival, cause of death, treatment etc.

Virtual Pooled Registry

Who would benefit?

NCI with potential cost savings and enhanced efficiency of current linkage processes

- Cohort studies
- Follow up for Clinical Trials

FDA

- Post-marketing surveillance

Cancer registries

- De-duplication of cases
- Accurate assessment of multiple primary incidence

Expanding SEER Capacity for Research Virtual Pooled Registry (VPR)

- ***Two major cohorts linked***
  - Camp Lejeune (approx 500,000) – 45 central registries (17,412 matches)
    - Manual review on subset provided these as high quality matches
  - Radiation Technician Cohort (DCEG) (approx 150,000) 42 central registries
    - Doubled number of matches from survey based cancer identification
    - 22 of 42 registries will accept NCI SSIRB (central IRB approval)

- ***Surveillance Research Program supported "minimal risk" Central IRB in development***
  - Contract for commercially performed CIRB in process for bid

- Supporting development of a new linkage software targeted for use by registries

- Third component of VPR underway
  - ***Hashed matching across all registries***
  - Purpose
    - Deduplication
    - More accurate assesment of multiple primary incidence

THANK YOU FOR YOUR ATTENTION

24

SEER Wide Quality Assurance/ Improvement

- Using epath and abstract text documentation for targeted review
  - Focusing on targeted variables (melanoma depth, HPV, CES etc)
  - NLP hire (Glenn Abastilles) developing queries targeting known issues)melanoma depth, HPV expression, tumor size etc)

- PSA first use case – reviewed and corrected back to 2004
  - Leveraged text documentation and NLP to identify potential errors for efficiency

- SEER Wide QC/QA SOPs
  - Contract with Dilts & Assoc. to develop a ***SEER wide data quality audit plan***
  - Plan to expand Dilts work to broader SEER wide QA QI plan and SOPs
  - Plan to hire a QA person in 2017 to oversee SEER wide processes