SEER Registries: Population-Based Infrastructure to Support Cancer Research

Contract Renewal Proposal

Board of Scientific Advisors
March 29, 2016
Lynne Penberthy, MD, MPH
Associate Director, Surveillance Research Program
Division of Cancer Control and Population Sciences
The SEER Program is a national resource supporting research on the diagnosis, treatment and outcomes of cancer since 1973.

- Currently covers 30% of the US population (450,000+ incident cases reported annually)

- SEER are the only population-based registries
  - With >80% real time e-path reporting (360+ path labs)
  - Intensive visual editing/ quality control processes
  - Integrated with NCI designated cancer centers
  - Capturing a broad set of clinical variables
    - 32 predictive and prognostic biomarkers
Evaluating SEER Program Progress - Usage

Used extensively to support statistical analyses and research

- Most commonly used data to represent trends over time
- > 4,000 downloads of SEER public use file annually
- 17,000 publications using SEER data since 1975
- ~40,000 manuscripts referencing SEER data
- 112 research grants ($87 million) funded in 2011-2012 where SEER data was critical to the grant

National Cancer Institute
Evaluating SEER Program Progress: Quality and Direction

• Ongoing quality studies to improve data collection.
  • Individual registry studies for quality improvement
  • Studies to improve surveillance nationally
    – TNM study
    – Elements of stage study

• SEER*DMS external review
  • Assure optimized SEER-wide software solutions

• Internal review of guidelines to direct data collection
  • Clinically relevant predictive and prognostic markers

• Focus groups to direct new SEER initiatives
  • Natural Language Processing (November 2015)
  • Virtual SEER-Linked Biorepository (VTR) (February 2014)
  • Identifying new data for SEER (September/October 2014)
  • Virtual Pooled Registry (February 2015)
Cancer Surveillance Challenges Facing SEER

• Complexity of cancer care
  o Treatment (new modalities/ ongoing Rx/ multiple cycles)
  o Outcomes other than survival (recurrence/progression/patient reported data)

• Expansion of data characterizing each cancer (precision medicine)
  o Complex molecular and genetic characterization of cancers
  o Need new data sources – require novel linkages and automation (NLP) for capture of relevant results (e.g., Oncotype DX)

• Dispersion of cancer diagnosis and treatment across multiple health care providers/locations (no longer only hospital-based)
  o Requires new methods and processes to assure complete and comprehensive data capture (e.g., for cases diagnosed and treated exclusively in the community)
Cancer Surveillance Challenges Facing SEER

- Resources required for manual collection not sustainable
  - Aging registrar population
  - Multiple new data sources to be accessed by registry personnel
  - Complexity of data interpretation challenging for non-medical personnel

- Changing demographic distribution and aging US population with increasing caseload of cancers to be abstracted
SEER Funding - relatively flat; innovative approaches required to meet challenges

Average increase over 8 years of $480k
Represents 1.3% average annual increase

Fiscal Year

Funding (in Millions)
Strategic Priorities for the SEER Program to address the challenges

1. Represent data in more clinically relevant categories with better representation of special U.S. populations
2. Automate and directly capture data via
   • Linkages
   • Auto-processing of data (Natural Language Processing)
3. Expand outcomes data collection
4. Expand the capacity of SEER to support cancer research
Strategic Priorities for the SEER Program to Address the Challenges

SRP is focusing on efficient central processes where feasible but additional resources are necessary to:

1. Capture data to represent the changing US population
   - Growing and disparate Hispanic and Asian American subgroups with differing cancer risk and outcomes
   - Aging population

2. Support the capture of increasingly complex and important data
   - Develop and sustain new methods for automation and linkages
   - Need for ongoing manual adjudication by registry staff
3. Develop and support an infrastructure to enhance the capacity of SEER to support cancer research

- Virtual SEER Linked Biorepository
- Cohort identification
- Virtual Pooled Registry
  - National effort to include a broad range of central cancer registries
  - De-duplication of incidence
  - More accurate assessment of multiple primary incidence
Proposed Changes to Current SEER Structure

We are requesting resources to enhance SEER’s capacity to meet the surveillance challenges through:

1. The addition of new “core” registries to provide more complete population coverage and representation

2. Expanding the SEER infrastructure to support key cancer research activities via addition of registries to support research

3. Supporting both core and research registries and special projects by leveraging new methods and linkages through central processes (i.e., SEER*DMS)
Pool of all U.S. Central Cancer Registries

SEER Expanded Infrastructure

**CORE REGISTRIES** – Collect most comprehensive data used for SEER statistics/public use file (equivalent to current SEER Program)

**Registries for Research Support** – only eligible to compete for special projects/extended services (e.g. SEER Linked Virtual Tissue Repository, Virtual Pooled Registry, other special projects to support research)

**These registries will transition to SEER*DMS to enable consistent and enhanced data collection/quality through central linkages and automation. These registries will not be funded for core data collection.**
Budget for the SEER renewal

• Budget request for a 10 year contract period

• Requested increase of 10% for a total budget request of $46.2 million per year to support:
  o Expansion of Core registries
  o Inclusion of registries to support research (small contracts initially with option to compete for larger contracts)
Conclusion

We are proposing to enhance the existing infrastructure

• through central processes (e.g. linkages, NLP, and automated processes) and
• through innovative expansion of participating registries to optimize the SEER program

to meet research needs given the changes in cancer care in the United States.
QUESTIONS?
STATE POPULATIONS AND DISTRIBUTION OF CANCER CASES
<table>
<thead>
<tr>
<th>State</th>
<th>% of US Pop.</th>
<th>% of Cancer Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>12.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Texas</td>
<td>8.6</td>
<td>6.5</td>
</tr>
<tr>
<td>Florida</td>
<td>6.3</td>
<td>6.8</td>
</tr>
<tr>
<td>New York</td>
<td>6.2</td>
<td>6.8</td>
</tr>
<tr>
<td>Illinois</td>
<td>4</td>
<td>4.2</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>4</td>
<td>4.9</td>
</tr>
<tr>
<td>Ohio</td>
<td>3.6</td>
<td>3.9</td>
</tr>
<tr>
<td>Georgia</td>
<td>3.2</td>
<td>3.0</td>
</tr>
<tr>
<td>North Carolina</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td>Michigan</td>
<td>3.1</td>
<td>3.4</td>
</tr>
<tr>
<td>New Jersey</td>
<td>2.8</td>
<td>3.1</td>
</tr>
<tr>
<td>Virginia</td>
<td>2.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Washington</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Arizona</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>2.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Indiana</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Tennessee</td>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Missouri</td>
<td>1.9</td>
<td>2.1</td>
</tr>
<tr>
<td>Maryland</td>
<td>1.9</td>
<td>1.8</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>1.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Minnesota</td>
<td>1.7</td>
<td>1.8</td>
</tr>
<tr>
<td>Colorado</td>
<td>1.7</td>
<td>1.4</td>
</tr>
<tr>
<td>South Carolina</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Alabama</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Louisiana</td>
<td>1.5</td>
<td>1.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State</th>
<th>% of US Pop.</th>
<th>% of Cancer Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kentucky</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>Oregon</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Connecticut</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>Iowa</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Utah</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Mississippi</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Arkansas</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Kansas</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Nevada</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>Nebraska</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>West Virginia</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Idaho</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Hawaii</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Maine</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Montana</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Delaware</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>South Dakota</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>North Dakota</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Alaska</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Vermont</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Wyoming</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>
INITIAL TREATMENT RESULTS
Infusion Chemotherapy from Claims

• Working with single large central oncology practice claims processor
  o Potential to acquire 25-50% of oncology practice data in 7 current registries
  o Pilot in GA with 6 practices sending 3 years retrospective and prospective data stream
  o Next year implement in
    • New Mexico
    • Louisiana
    • Kentucky
    • New Jersey
    • Utah
Preliminary Data from 6 Months Claims in 4 Georgia Oncology Practices: Common Regimens for Treatment of Initial and Recurrent Breast Cancer

Common Regimens
- Treatment of Initial Breast Cancer
  - 4 regimens - 4,676 administrations
- Treatment of Recurrent or Metastatic Breast Cancer
  - 9 regimens - 1,262 administrations

Administration frequency for chemotherapy regimens commonly used for initial breast cancer treatment (6 months of data)

Administration frequency for chemotherapy regimens commonly used for treatment of recurrent or metastatic breast cancer (6 months of data)
Oral Treatment

• Discussions with two large pharmacy chains
  o Two large pharmacy switchers (central claims processors)
  o If agreed to provide data will capture 75-80% of antineoplastic prescriptions in SEER areas
## Preliminary data for Specific Drug Prescriptions from Medicare Part A/B or D compared with Change HealthCare* (Switcher)

<table>
<thead>
<tr>
<th>Drug</th>
<th>Louisiana</th>
<th>Seattle</th>
<th>California***</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capecitabine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change Healthcare*</td>
<td>230</td>
<td>153</td>
<td>867</td>
</tr>
<tr>
<td>Medicare A/B**</td>
<td>136</td>
<td>127</td>
<td>205</td>
</tr>
<tr>
<td><strong>Imatinib</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change Healthcare*</td>
<td>129</td>
<td>52</td>
<td>336</td>
</tr>
<tr>
<td>Medicare D**</td>
<td>37</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td><strong>Lenalidomide</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change Healthcare*</td>
<td>212</td>
<td>176</td>
<td>874</td>
</tr>
<tr>
<td>Medicare D**</td>
<td>95</td>
<td>72</td>
<td>132</td>
</tr>
<tr>
<td><strong>Tamoxifen</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change Healthcare*</td>
<td>741</td>
<td>541</td>
<td>1302</td>
</tr>
<tr>
<td>Medicare D**</td>
<td>318</td>
<td>354</td>
<td>319</td>
</tr>
</tbody>
</table>

*Change HealthCare represents approx. 20% of pharmacy data
LINKAGES: ONCOTYPE DX EXAMPLE
Clinical data: Example of linkages with commercial partners

Oncotype DX: Linkage with GHI data 2004-2013

• Added 40% of test results to existing data from hospital (only) reported results

• Largely test results sent directly to physician practices
Pre-specified BCSM Analysis of N0, HR+, HER2− Patients Age 40-84: Results (n=38,568)

- Risk Score predicted BCSM in univariate analysis unadjusted for treatment or covariates
  - Known Chemo use
    - 7% of RS <18
    - 34% of RS 18-30,
    - 69% of RS ≥31
- Chemotherapy known to be under-reported in SEER
- No significant association of RS with non-breast cancer mortality (p=0.66)

![Graph showing breast cancer-specific mortality by risk score group](image-url)
Example: Reporting data in more clinically relevant categories - Esophageal Cancer in men - overall and by histologic subtype

- Esophageal Overall
  - Age-Adj. Incidence/100,000

- Adenocarcinoma
  - Age-Adj. Incidence/100,000

- Squamous
  - Age-Adj. Incidence/100,000
Expanding the capacity of SEER to support cancer research
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 and linked e path reports to select a set of relevant specimens
• Ultimate aims
  o Annotation and search capacity of abstracts + e path reports for all SEER cases with tissue
  o Centralization of requests for specimens and custom annotation
  o 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

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 adenocarcinoma long term survivors (> 5 yr survival)

- Purpose
  - Assess best practices across multiple registries
  - Estimate costs of supporting a SEER wide system
  - Assess availability of specimens
  - Understand human subjects/consent as requirements vary by registry and prepare for common rule changes
SEER Biospecimen Repository Proposed Workflow

Central Processing
- Central Website
  - User registration
  - Query de-ID’ed e-Path reports
  - Request submission and status
  - Peer review/approval protocol
  - Honest Broker process

Central Coordinator
- Work with Honest Broker
- Abstract/Annotation
- Linkage - data/specimens
- Interaction with Path Labs & Investigators

Path Lab
- Inventory & processing
- Residual & other specimens
- QC

Investigator
- Study design: funding, protocols, hypothesis

Virtual Tissue Repository

Residual Tissue Repository

SEER Registry
- Work with Honest Broker
- Abstract/Annotation
- Linkage - data/specimens
- Interaction with Path Labs & Investigators

Path lab may ship specimen directly to investigator through registry processes
Virtual Pooled Registry

What is it?

- A **virtual** national cancer registry being developed in collaboration with the North American Association of Central Cancer Registries and CDC
- 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

• Current Status
  o Two large cohorts to pilot linkage (Camp LeJeune and Radiation Technician Cohorts)
  o 47 registries already signed on
  o Funding large linkage across all 47 to
    • De-duplicate
    • Provide more accurate estimate of multiple primacy cancer incidence

• Who would benefit?
  o NCI with potential cost savings and enhanced efficiency of current linkage processes
    • Cohort studies
    • Follow up for Clinical Trials
  o FDA
    • Post-marketing surveillance