



# **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

# Surveillance Epidemiology and End Results (SEER)



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

# 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)

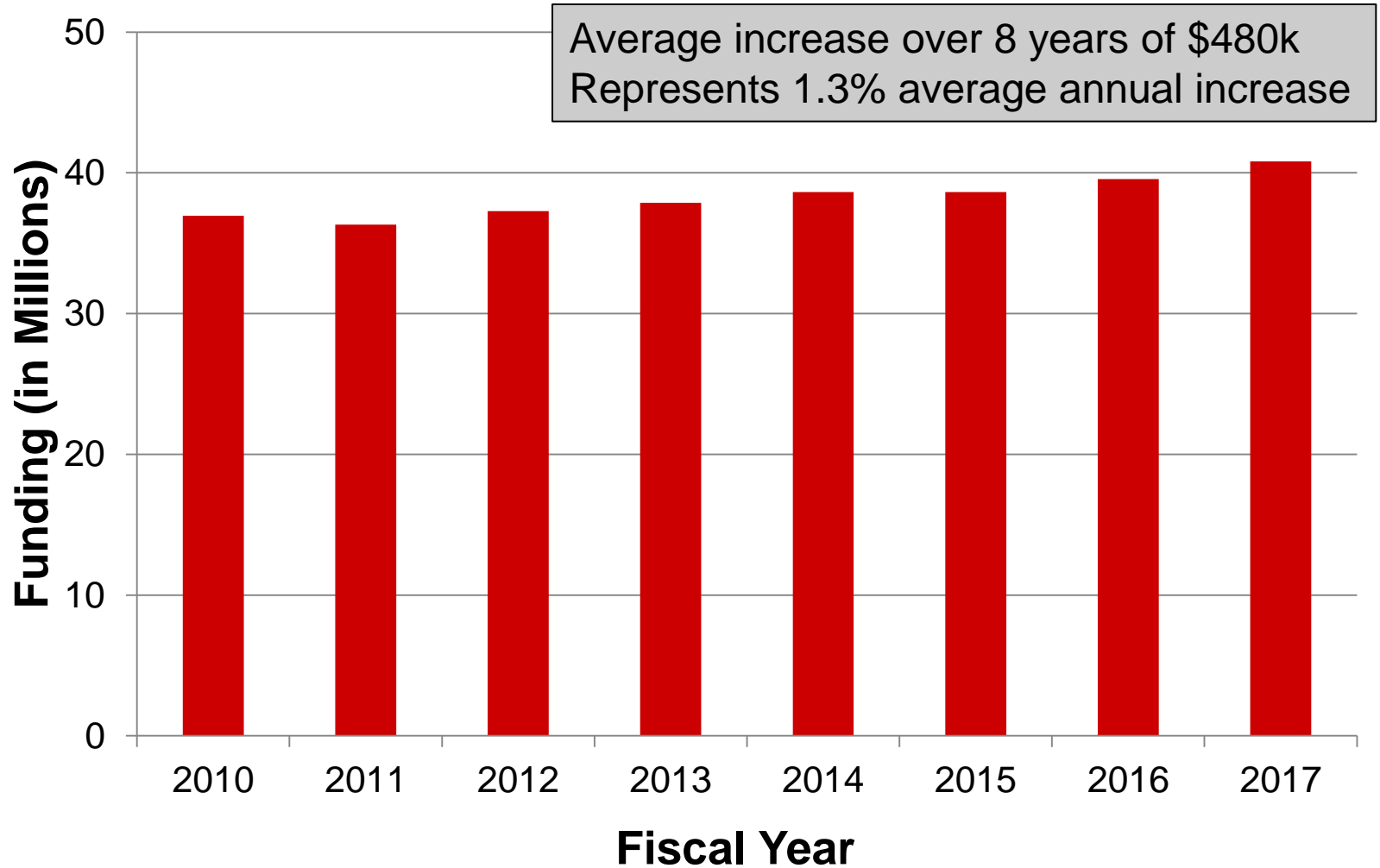
# Cancer Surveillance Challenges Facing SEER

- Complexity of cancer care
  - Treatment (new modalities/ ongoing Rx/ multiple cycles)
  - Outcomes other than survival (recurrence/progression/patient reported data)
- Expansion of data characterizing each cancer (precision medicine)
  - Complex molecular and genetic characterization of cancers
  - 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)
  - 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



# 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

# Strategic Priorities for the SEER Program to Address the Challenges

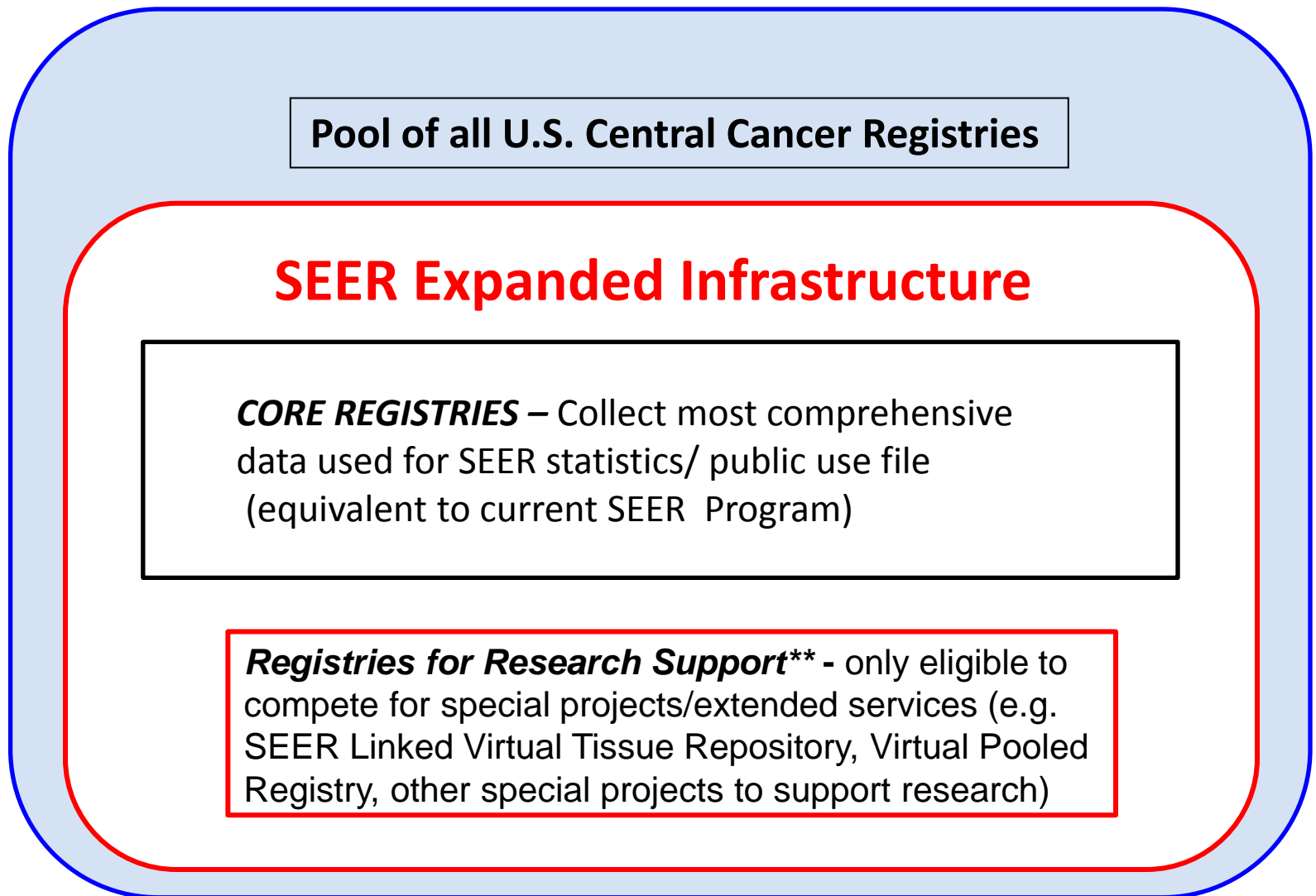
3. Develop and support an infrastructure to enhance the capacity of SEER to support cancer research
  - Virtual SEER Linked Biorepository
  - Cohort identification
  - Virtual Pooled Registry
    - National effort to include a broad range of central cancer registries
    - De-duplication of incidence
    - 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)

# Figure 1. New Proposed SEER Structure



\*\* 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:
  - Expansion of Core registries
  - 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**



State	% of US Pop.	% of Cancer Cases
California	12.2	10.3
Texas	8.6	6.5
Florida	6.3	6.8
New York	6.2	6.8
Illinois	4	4.2
Pennsylvania	4	4.9
Ohio	3.6	3.9
Georgia	3.2	3.0
North Carolina	3.1	3.2
Michigan	3.1	3.4
New Jersey	2.8	3.1
Virginia	2.6	2.3
Washington	2.2	2.2
Arizona	2.1	1.8
Massachusetts	2.1	2.3
Indiana	2.1	2.1
Tennessee	2.1	2.2
Missouri	1.9	2.1
Maryland	1.9	1.8
Wisconsin	1.8	1.9
Minnesota	1.7	1.8
Colorado	1.7	1.4
South Carolina	1.5	1.6
Alabama	1.5	1.6
Louisiana	1.5	1.6

State	% of US Pop.	% of Cancer Cases
Kentucky	1.4	1.7
Oregon	1.3	1.3
Oklahoma	1.2	1.2
Connecticut	1.1	1.3
Iowa	1	1.1
Utah	0.9	0.6
Mississippi	0.9	1.0
Arkansas	0.9	1.0
Kansas	0.9	0.9
Nevada	0.9	
New Mexico	0.7	0.6
Nebraska	0.6	0.6
West Virginia	0.6	0.7
Idaho	0.5	0.5
Hawaii	0.5	0.4
New Hampshire	0.4	0.5
Maine	0.4	0.5
Rhode Island	0.3	0.4
Montana	0.3	0.3
Delaware	0.3	0.3
South Dakota	0.3	0.3
North Dakota	0.2	0.2
Alaska	0.2	0.2
District of Columbia	0.2	0.2
Vermont	0.2	0.2
Washington	0.2	0.2

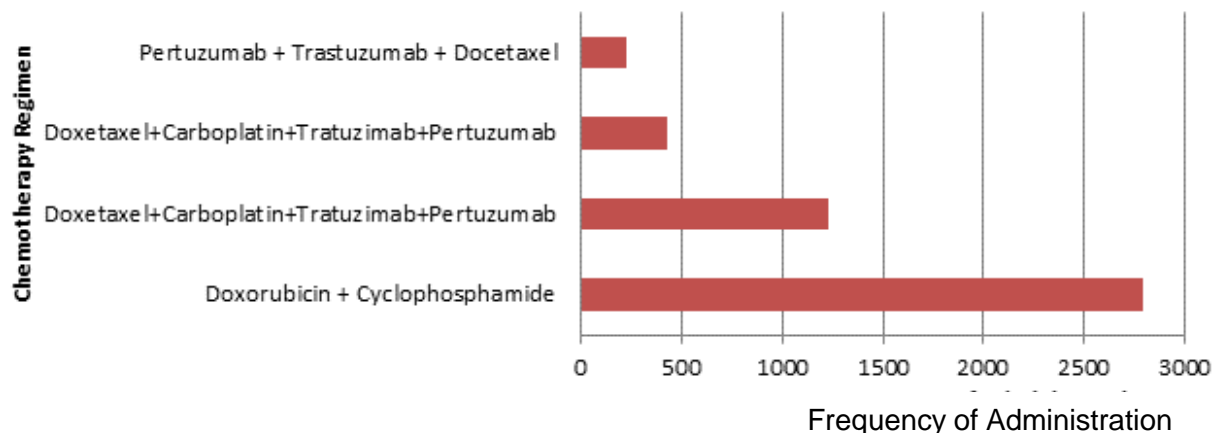
# INITIAL TREATMENT RESULTS

# Infusion Chemotherapy from Claims

- Working with single large central oncology practice claims processor
  - Potential to acquire 25-50% of oncology practice data in 7 current registries
  - Pilot in GA with 6 practices sending 3 years retrospective and prospective data stream
  - 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

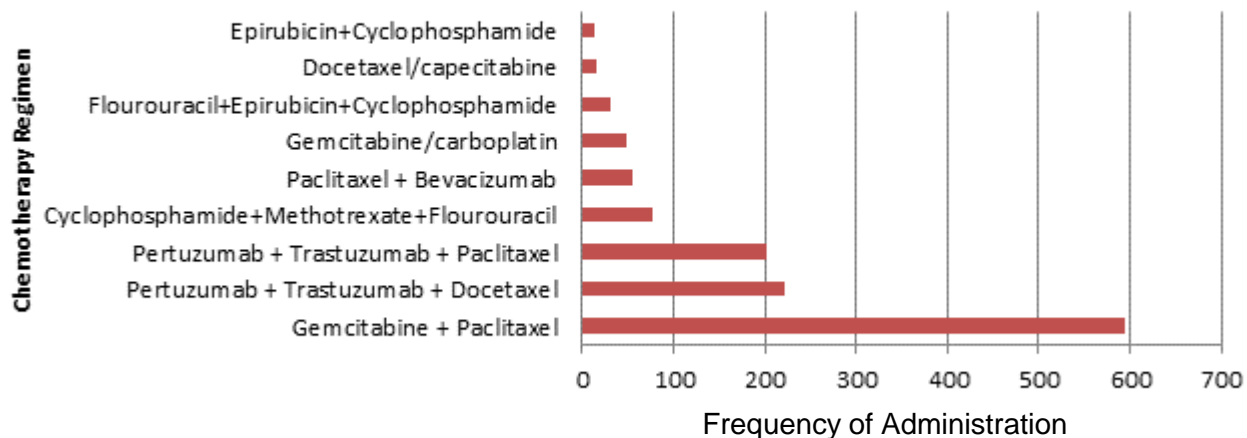
Administration frequency for chemotherapy regimens commonly used for initial breast cancer treatment (6 months of data)



Common Regimens  
Initial treatment of  
Breast Cancer  
4 regimens - 4,676  
administrations

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

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
  - Two large pharmacy switchers (central claims processors)
  - 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)

<b>Capecitabine</b>			
	Lousiana	Seattle	California***
Change Healthcare*	230	153	867
Medicare A/B**	136	127	205
<b>Imatinib</b>			
	Lousiana	Seattle	California***
Change Healthcare*	129	52	336
Medicare D**	37	18	15
<b>Lenalidomide</b>			
	Lousiana	Seattle	California***
Change Healthcare*	212	176	874
Medicare D**	95	72	132
<b>Tamoxifen</b>			
	Lousiana	Seattle	California***
Change Healthcare*	741	541	1302
Medicare D**	318	354	319

\*Change HealthCare represents approx. 20% of pharmacy data

# **LINKAGES: ONCOTYPE DX EXAMPLE**

# Clinical data: Example of linkages with commercial partners

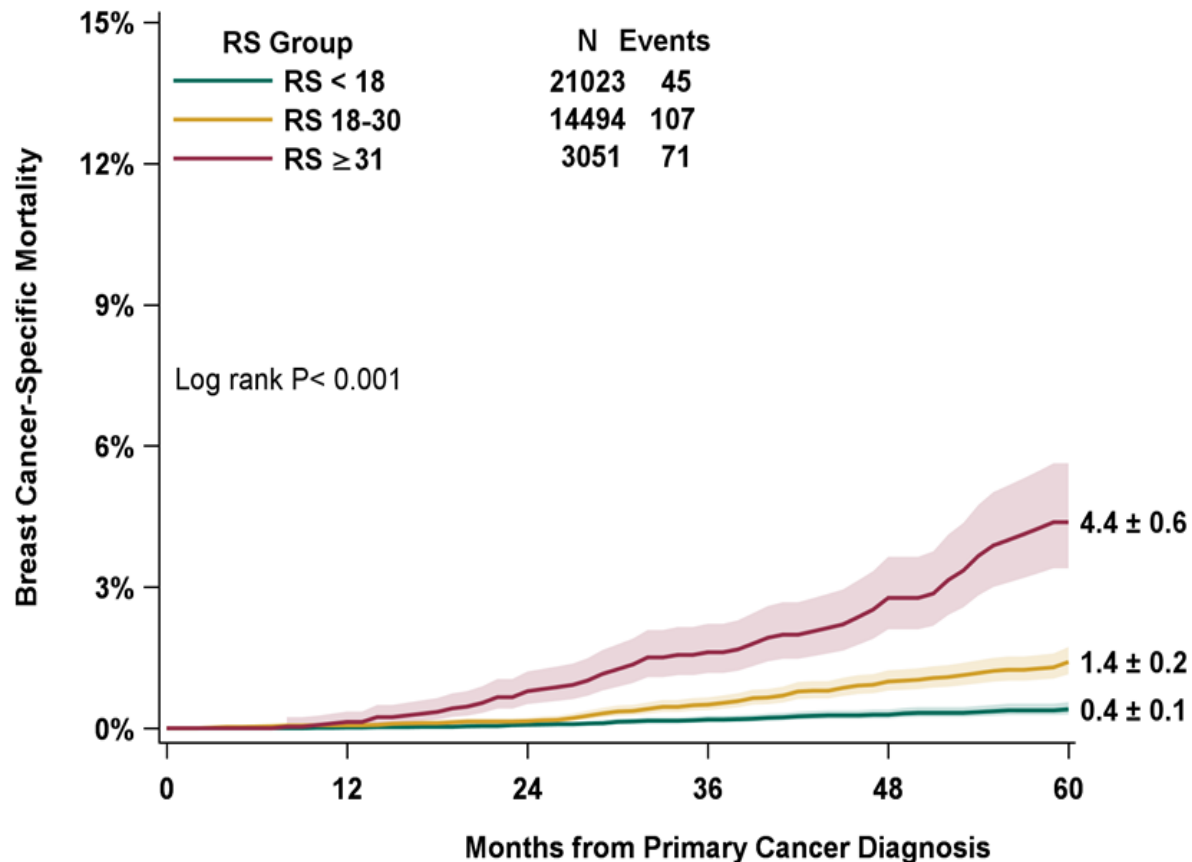
## Oncotype DX: Linkage with GHI data 2004-2013

- Added 40% of test results to existing data from hospital (only) reported results
- Largely test results sent directly to physician practices



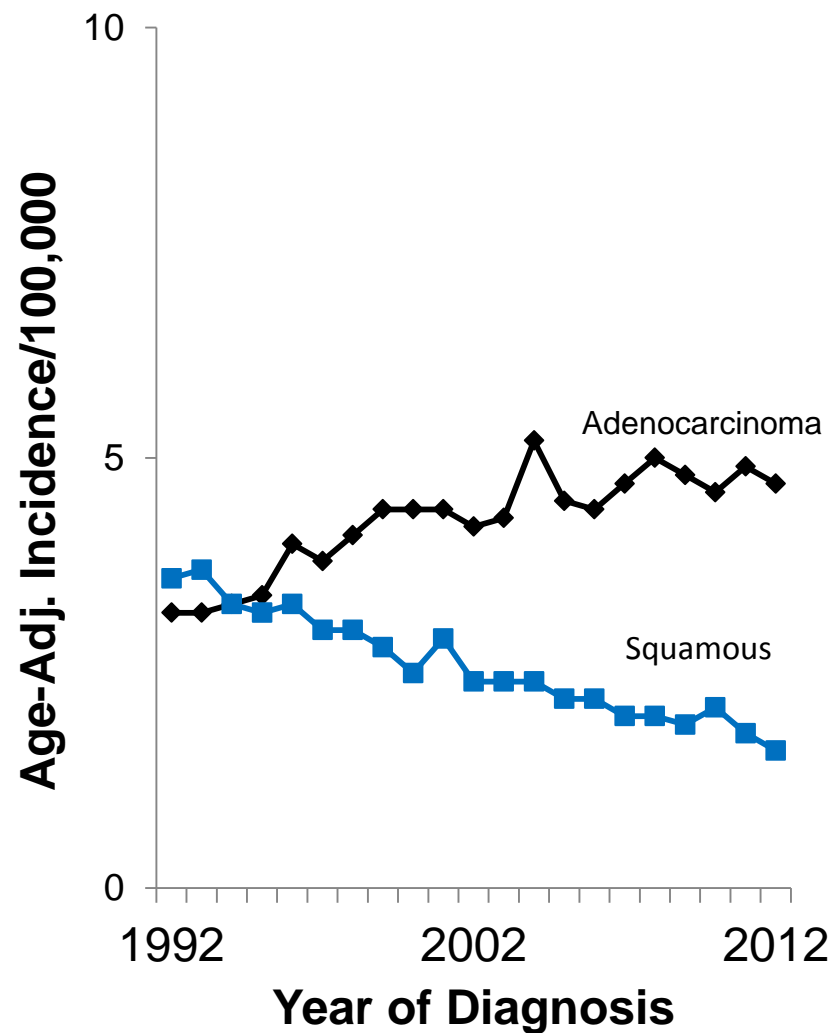
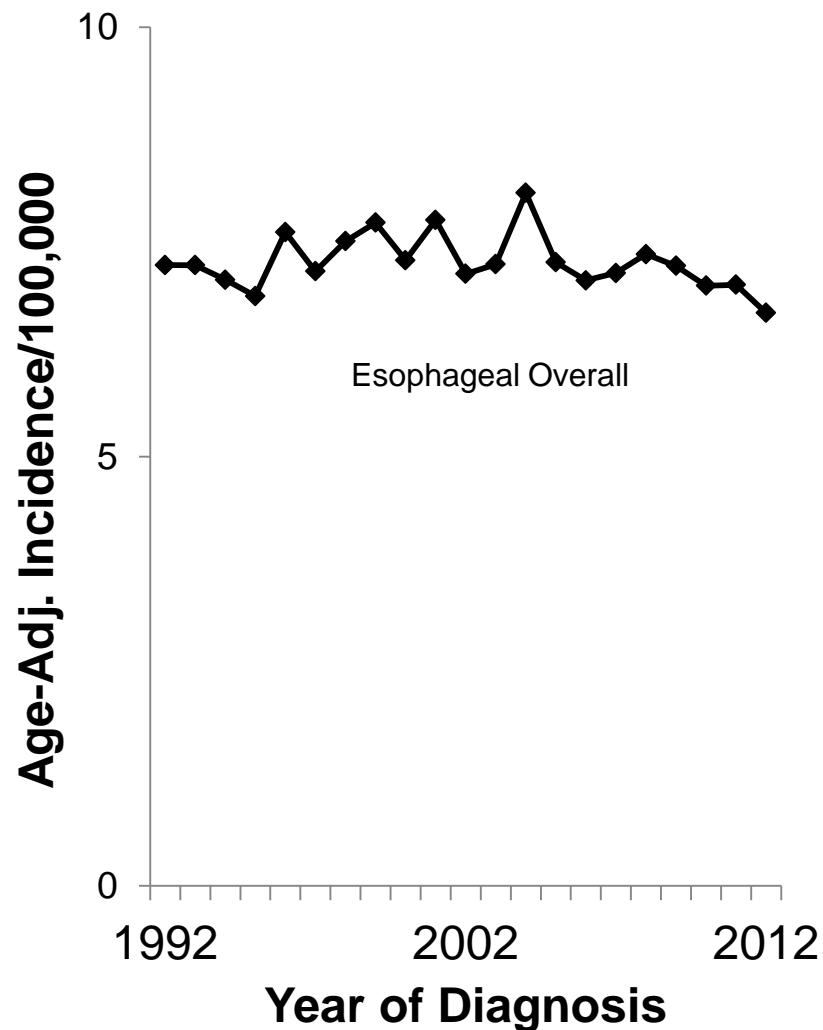
# Pre-specified BCSM Analysis of N0, HR+, HER2- Patients Age 40-84: Results (n=38,568)

- 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)



RS < 18	21023	20481	15685	11543	7551	4200
RS 18-30	14494	14138	11011	8247	5624	3369
RS ≥ 31	3051	2979	2313	1731	1153	670

# Example: Reporting data in more clinically relevant categories - Esophageal Cancer in men - overall and by histologic subtype



# **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
  - 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

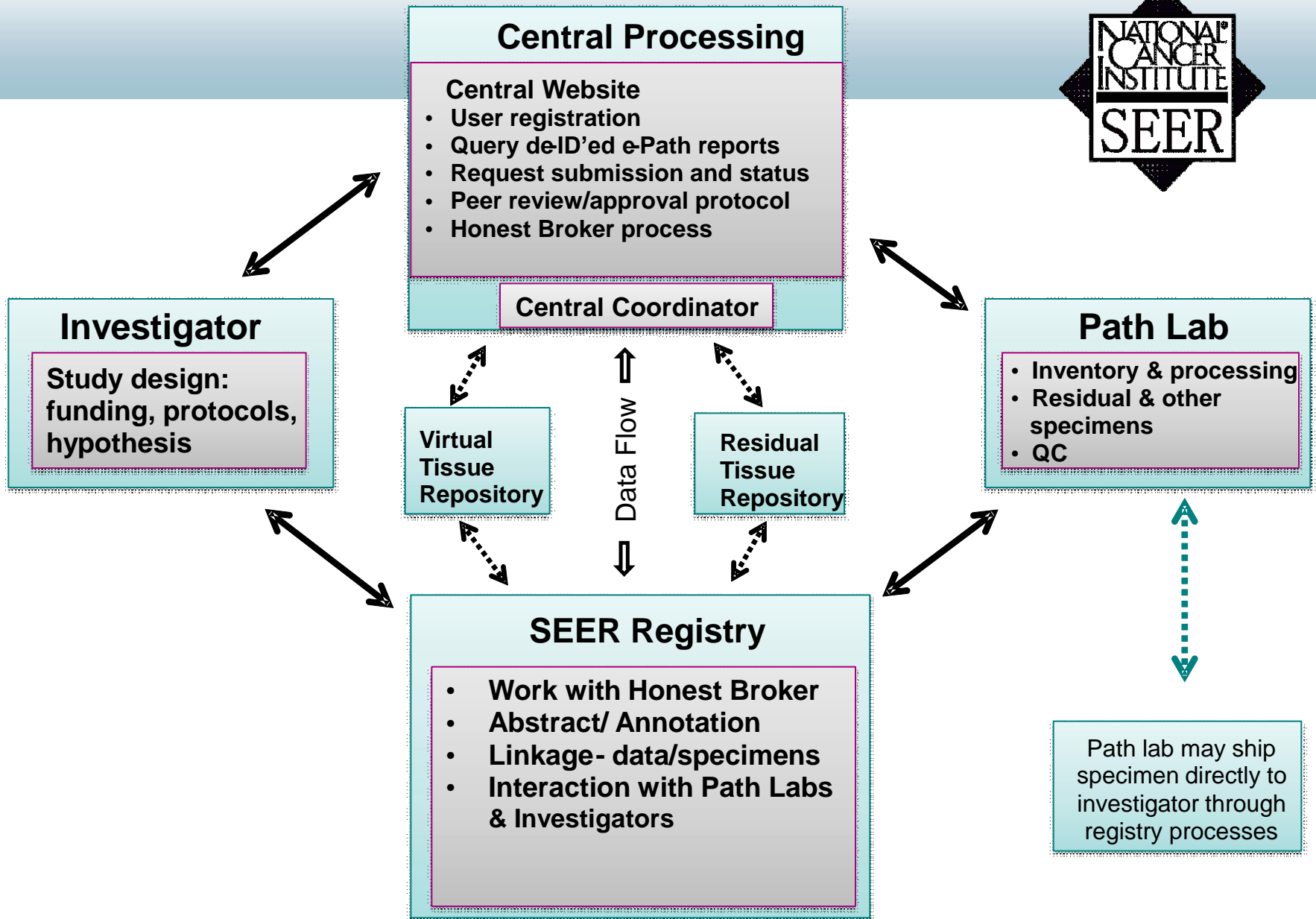
- Population based – permitting comparison of subsets
- Available across a broad spectrum of health care facilities/pathology labs (not just academic centers)
- Access to rare cancers and exceptional outcomes
- Linked long term outcomes
- Existing annotation with clinical and demographic data
- Potential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

# SEER-Linked Virtual Bio-Repository Pilot

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)
- 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



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