SEER: a changing paradigm for cancer surveillance
National Cancer Policy Forum 2019

NIH NATIONAL CANCER INSTITUTE

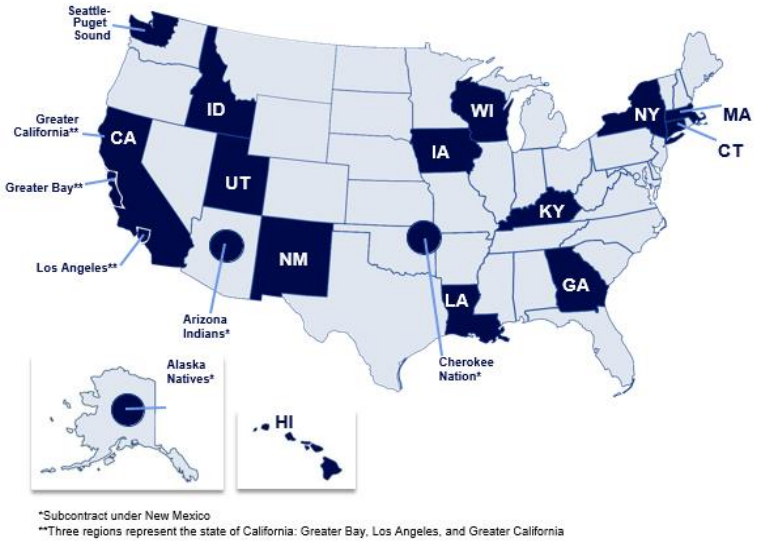
Objectives

Briefly review:

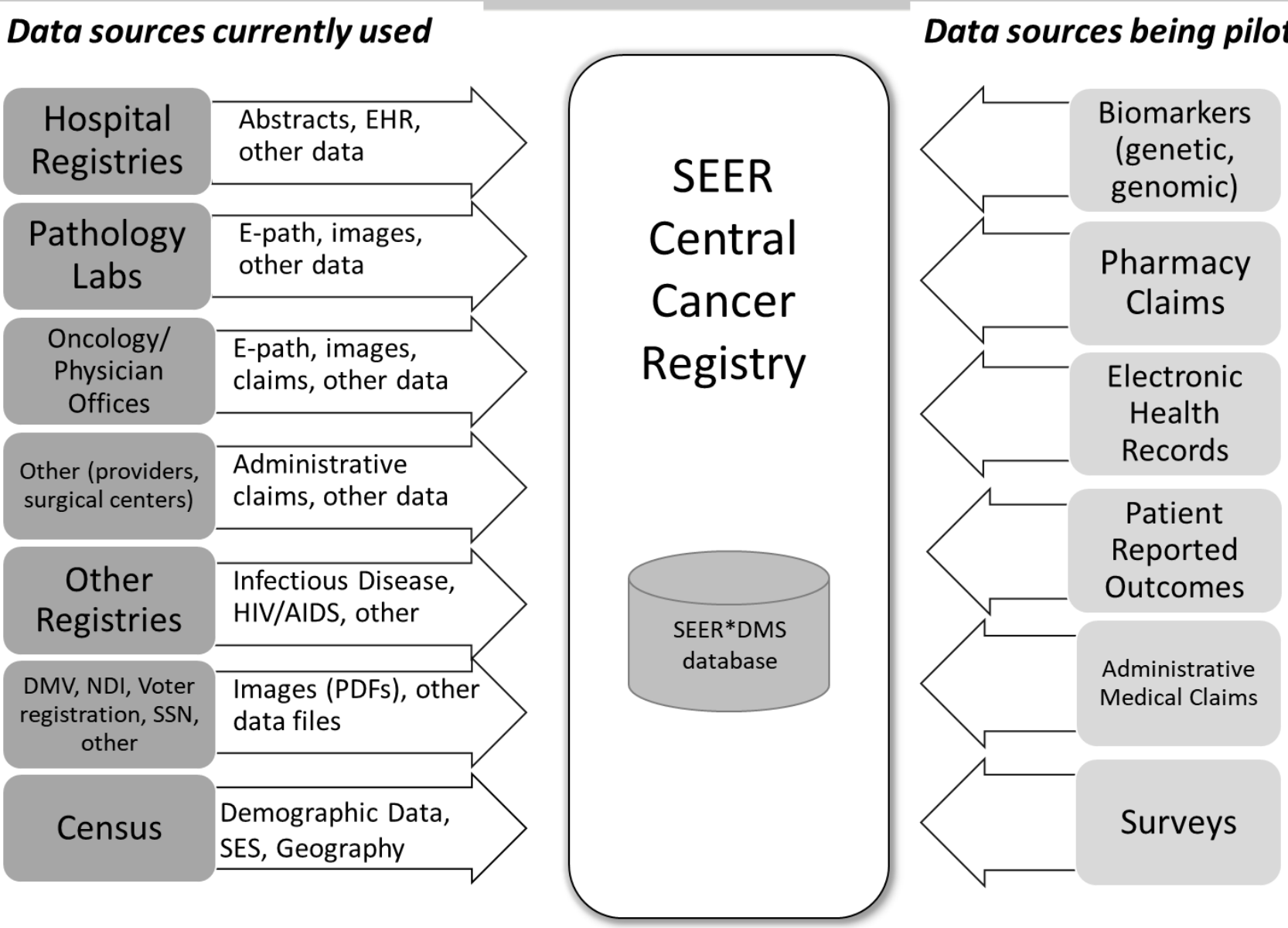
- The SEER Program
- Important areas cancer surveillance must address
- Examples of new initiatives to enhance the data

The SEER Program

- Funded by NCI ***to support research*** on the diagnosis, treatment and outcomes of cancer since 1973
- 16 population-based registries ***now*** covering **35%** of the US population
  - RFP for expansion in process
- With new registries −550,000 incident cases received annually
  - Approximately 85% of cases with real time electronic pathology (e-path) reporting
  - Facilitates rapid case identification supporting research
- All registries will be on a common data platform (SEER DMS) that permits
  - central linkages with external partners
  - facilitates scaling of new initiatives across all registries simultaneously



SEER Data Sources- current and in testing



While Surveillance data are very good….we must enhance what and how we collect the data to be more clinically relevant and meet the needs of cancer research

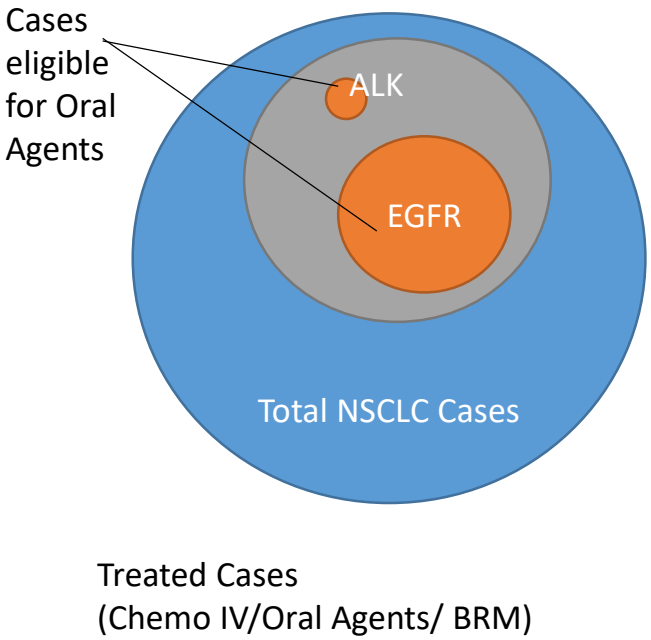
Why do we need detailed treatment?

- Real world treatment data from registries would permit studying key questions around
  - Affordability
    - E.G. New Immunotherapies today:
      - Cost: $1.01 Million
      - Out of pocket: ~$200K*
  - Adherence and compliance
  - Disparities in who receives the treatments
  - Understanding outcomes in non clinical trial patients (>95% of all cancer patients)

*Kantarjian et al. High Cancer Drug Prices in the United States: Reasons and Proposed SolutionsDOI: 10.1200/JOP.2013.001351 Journal of Oncology Practice 10, no. 4 (July 1 2014) e208-e211.

Why do we need genomic data?

- How patients are treated is changing based on targeted mutations.
  - Represents current standards of care and quality of care
- These mutations may represent a small subset of the cancer population or there may be a population subgroup where a variant is significant.
- Use case: NSC Lung Cancer - EGFR & ALK. BRAF. Her2NEu etc



Lung Biomarker	Prevalence	Treatment Example
EGFR	19-41% (varies by location and ethnicity, many factors) median overall prevalence=33.1% Is this impacting increased survival and decreased mortality in lung cancer patients?	Erlotinib
ALK	2% in the population (varies from 2-7%, updated articles are toward the lower estimate) Kwak EL, Bang YJ, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non–small-cell lung cancer. *N Engl J Med*. 2010;363:1693-1703.	Crizotinib

Why do we need registries to represent "Real World Data"?

Because Randomized Control Trials cannot test all permutations of patient situations.

Use Case- Orally administered targeted therapy (Larotrectinib).

Larotrectinib efficacy established

- Based on 3 clinical trials
- Population: 55 pediatric and adult patients
- Biomarker: identified neurotrophic receptor tyrosine kinase (NTRK) gene fusion
  - metastatic or where surgical resection not reasonable
- A total of 12 cancer types were represented:
- 75 percent overall response rate (ORR) across different types of solid tumors

Orphan Drug with accelerated approval to fill an unmet medical need (November 2018)

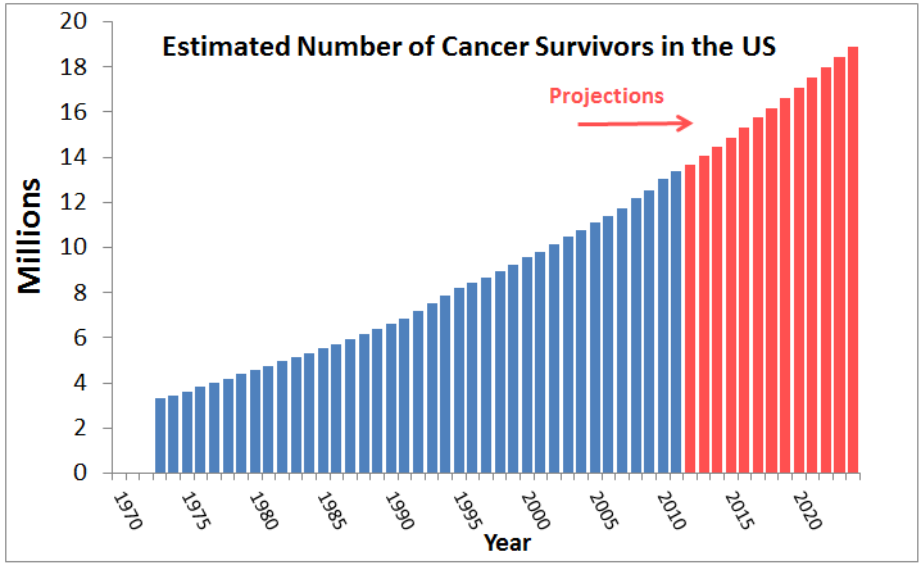
Near real time data feeds from CVS and Walgreens permits:

- monitoring the dissemination of new agents and
- complement the info captured in the RCTs
  - new population subgroups
  - ages
  - pts with comorbidity

Why do we need to capture recurrence

- With nearly 17 million cancer survivors in the US alone (nearly 5% of the population) lack of recurrence information is no longer acceptable- cancer is a chronic disease.
- Many clinical trials are now focused on recurrent disease and our most intransigent cancers with the highest mortality are likely to manifest with recurrence/metastatic disease

There are likely survival differences among population subgroups related to differential treatment, genetics or other factors.



[1] DeSantis C, Chunchieh L, Mariotto AB, et al. (2014). Cancer Treatment and Survivorship Statistics, 2014. CA: A Cancer Journal for Clinicians. In press.

Approaches to Enhancing SEER: A prototype for new surveillance processes

Main Goals in Enhancing SEER

- Create a system representing ***population level*** real world data to supplement clinical trials and understand effectiveness of oncology care for the 95% of patients outside the clinical trial setting through
  - Linkages to capture current and new data items
  - Developing tools for automation (NLP/machine learning) – DOE partnership
  - Leveraging these activities through collaborations with external partners both commercial and public
    - Pharmacies: CVS, Walgreens, Riteaid, PBMs
    - Claims data: United Health Care, Aetna, Unlmited Systems, Statewide APAC
    - Genomic/Genetic testing labs: GHI, Castle, Myriad, Ambra etc
    - Data Integration Partners: CLQ, Syapse, Tempus, Varian, Elekta

Specific gaps in current surveillance data being addressed with new initiatives

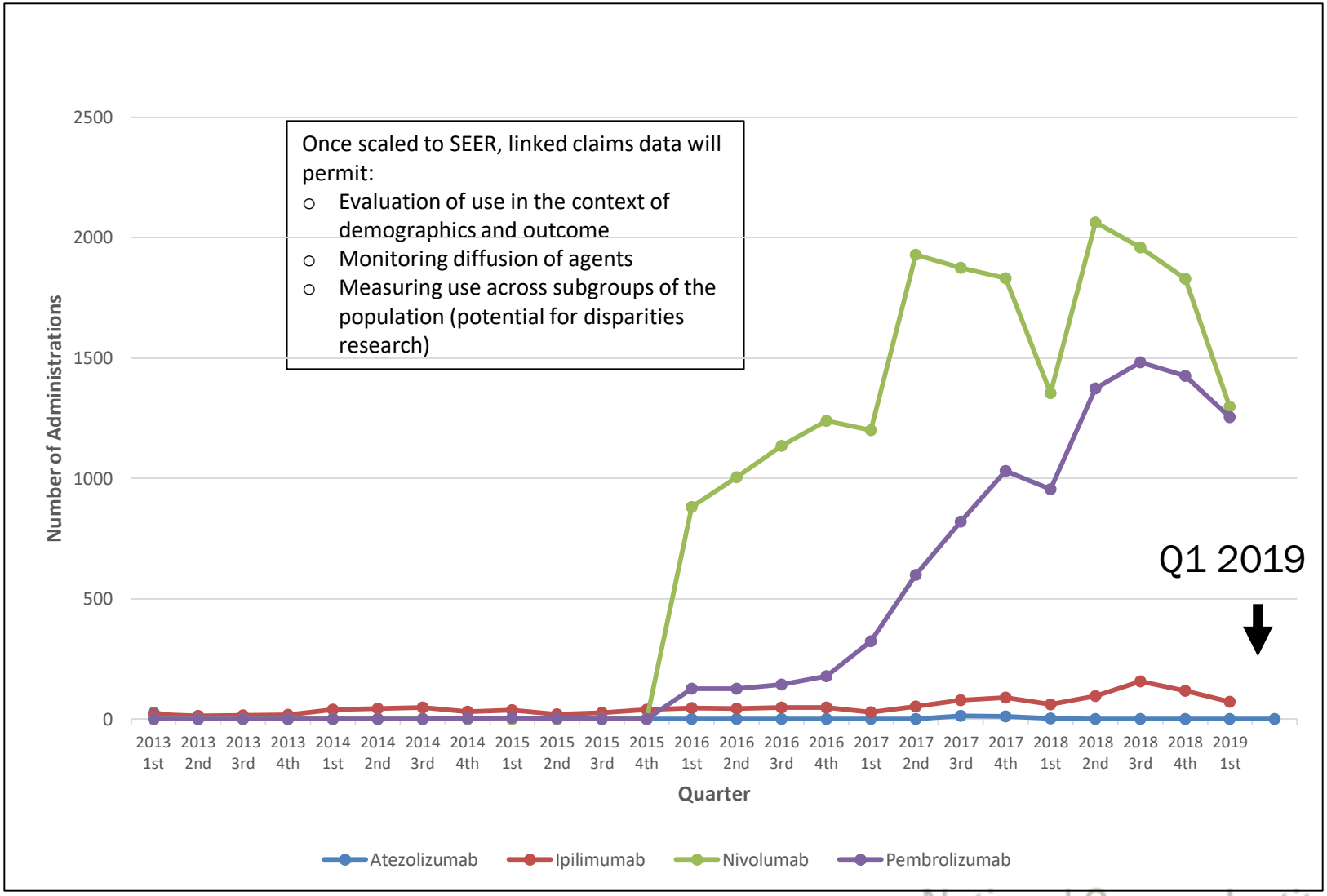
- Data Capture
  - Detailed longitudinal treatment data
  - Comprehensive genomic data characterizing the cancer
  - Outcomes other than survival and cause of death (recurrence)
  - Comorbidity to provide context for therapies and outcomes
- Developing infrastructure to support cancer research
  - SEER wide mechanisms for Rapid Case Ascertainment for patient eligibility assessment for RCTs and other studies (including patient contact studies)
  - Virtual Pooled Registry (VPR)
  - Virtual SEER Linked Biorepository (VTR)

Current pilot study results

The changing paradigm for surveillance: Examples of what we can do

- We are beginning to collect data that will permit
  - Tracking and monitoring dissemination of specific treatments over time
  - Evaluation of standards of care in oncology practice over time
  - Corroboration of clinical trial results in the real world
  - Representation of trends by more clinically relevant categories
  - Doing so where feasible using automated deep learning data extraction

Example: Post marketing surveillance- Tracking the dissemination of checkpoint inhibitor use in oncology practice claims (2013-2019) –claims linkages



National Cancer Institute

*Represents 12-35% of oncologists in 6 SEER registries and approximately 10,000 administrations

Cancer Site	Total Unique patients receiving at least one admininstraion of a cehckpoint inhibitor			
	Nivolumab	Pembrolizumab	Ipilimumab	Combined
All	1178	735	237	2150
Tongue	12	13		25
Oral Cavity	26	25	1	52
Esophagus	12	17	2	31
Stomach	7	19	1	27
Colon	15	18	4	37
Rectum	3	14	3	20
Anus, Anal Canal and Anorectum	10	5	2	17
Liver	31	1	1	33
Intrahepatic Bile Duct/GB/Other Biliary	3	4	1	8
Pancreas	11	4	5	20
Other Digestive Organs	1	5		6
Larynx	4	13		17
Lung and Bronchus	573	354	26	953
Melanoma of the Skin	136	78	137	351
Other Non-Epithelial Skin	2	2	1	5
Breast	18	15	2	35
Cervix Uteri	2	7		9
Corpus Uteri	5	15	1	21
Ovary	10	1	1	12
Prostate	19	23	2	44
Urinary Bladder	20	36	2	58
Kidney and Renal Pelvis	190	8	30	228
Ureter	2	7		9
Thyroid	2	8		10
Hodgkins	10	3		13
Non-Hodgkins	4	4	1	9
Mesothelioma	8	8		16

Example longitudinal claims from oncology practices (Unlimited Systems):
Understanding approved and off label use of Checkpoint Inhibitors by cancer site - (2013-March 31, 2019)

Tracking oral anti-neoplastics through pharmacy data. Example: TKI Use by Cancer Site and Target in GA (2013- Dec 2018)

Overall >65,000 patients in GA registry with ≥ 1 antineoplastic therapy prescription (>500k fills)

This table represents >2800 patients and >20,000 fills
These types of real world data will permit:

- Trend Analyses
- Monitoring of patient adherence and compliance
- Disparities in receipt

Cancer Site	Target	Generic Drug Name	# Unique Patients with Anti-neplastic Prescriptions	
			CVS	Walgreens
NSCLC	ALK	alectinib, ceritinib,crizotinib	42	13
NSCLC	EGFR	afatinib, erlotinib, osimertinib, Gefinitib	229	174
CML	BCR-ABL	bosutinib, dasatinib, Imatinib, nilotinib, ponatinib	675	300
RCC/Thyroid	VEGF	cabozantinib	100	41
RCC	VEGFR	axitinib	47	
RCC	VEGF, FLT, PDGFR, Kit, RET, CSF	sunitinib	118	72
RCC	VEGF FGF, PDGFR, Kit, RET, CRAF, BRAF	sorafenib	138	122
RCC	VEGF, FGF, PDGFR, Kit, Lck, FMS	pazopanib	143	167
CRC/ HCC	VEGF, FGF, PDGFR, Kit, RET, TIE2....	regorafenib	115	69
BC	HER2, EGFR	lapatinib, neratinib	100	41
Melanoma/ NSCLC	BRAF V600	vemurafenib, dabrafenib, trametinib	30	29

Example: Evaluating standards of care- BRCa testing among patients with ovarian (and breast) cancer - CA & GA (2013-2015) *

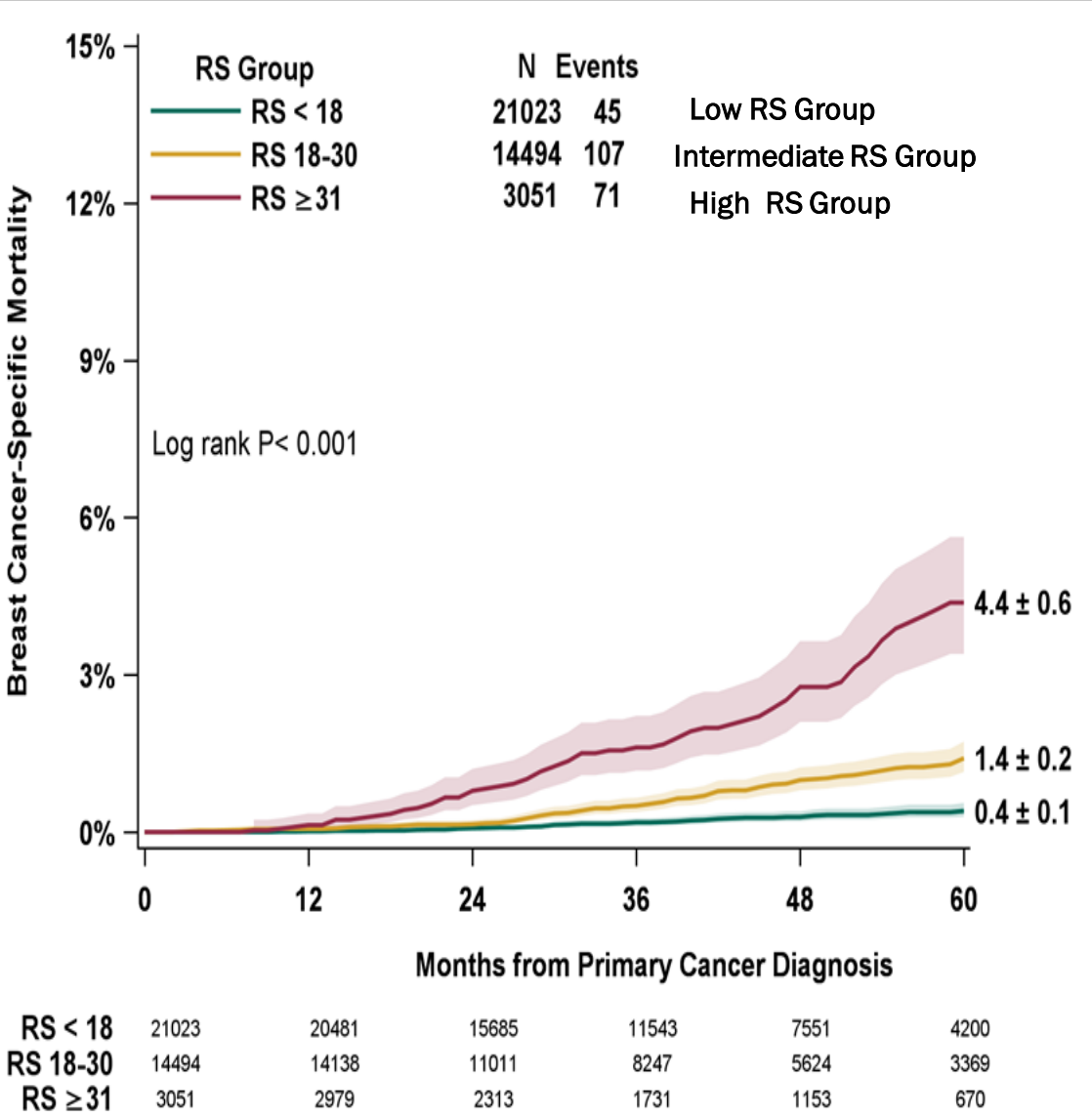
	Breast Cancer			Ovarian Cancer		
Characteristics	**Total Cases**	**Tested* Cases**	**Proportion Tested* % (95% CI)**	**Total Cases**	**Tested* Cases**	**Proportion Tested* % (95% CI)**
State and year of diagnosis						
California§						
2013	30,367	7,314	24.1 (23.6-24.6)	2,388	707	29.6 (27.8-31.5)
2014	30,012	6,951	23.2 (22.7-23.6)	2,390	732	30.6 (28.8-32.5)
2013-2014	60,379	14,265	23.6 (23.3-24.0)	4,778	1,439	30.1 (28.8-31.4)
Georgia						
2013	8,296	2,066	24.9 (24.0-25.9)	618	206	33.3 (29.6-37.2)
2014	8,410	2,270	27.0 (26.0-28.0)	605	209	34.5 (30.8-38.5)
2013-2014	16,706	4,336	26.0 (25.3-26.6)	1,223	415	33.9 (31.3-36.7)
Race/Ethnicity						
Non-Hispanic (NH) White	48,063	11,635	24.2 (23.8-24.6)	3,701	1,251	33.8 (32.3-35.3)
NH Black	9,039	2,095	23.2 (22.3-24.1)	523	113	21.6 (18.1-25.4)
NH American Indian	207	51	24.6 (18.9-31.1)	19	5	26.3 (9.1-51.2)
NH Asian	9,061	2,034	22.5 (21.5-23.3)	728	229	31.5 (28.1-35.0)
Hispanic	10,715	2,786	26.0 (25.2-26.8)	1,030	256	24.9 (22.2-27.6)

Overall testing (2013-2015) **24% breast cancers and 31% ovarian** cancers.
Substantial variation for ovarian cancer testing ranging from 22% in Black women to 34%in white women
* Kurian et al. JCO April 9, 2019

Example: OncotypeDx Population-based results corroborating CTs in a real world setting (n=38,568)

Oncotype Risk Score Category predicted breast cancer specific mortality

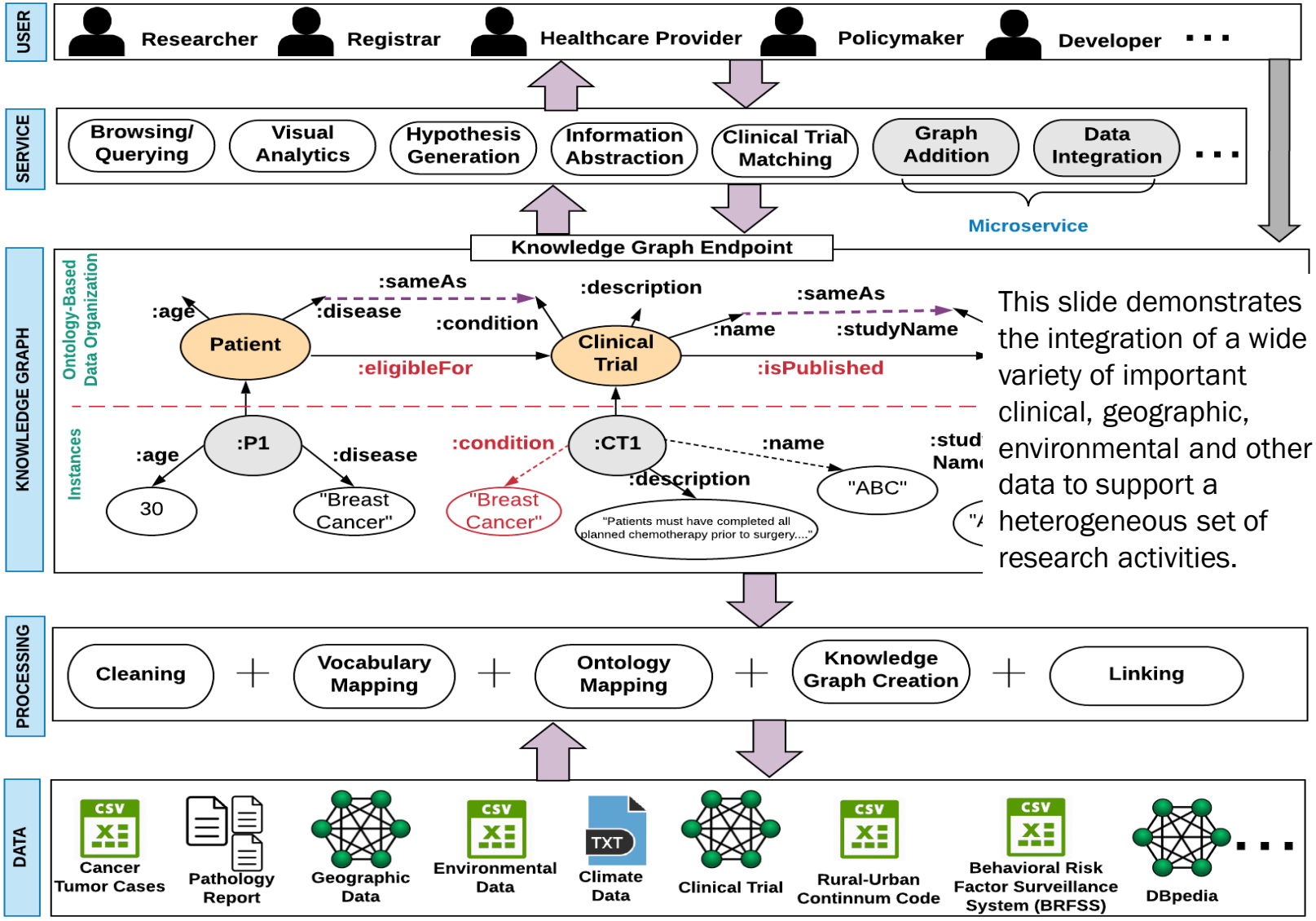
- These data support analysis overall as well as by racial and ethnic subgroup
- Populations NOT captured well in RCTs



Added information to SEER: Residential History and Social Determinants of Health

- Lexis Nexis initial linkage to obtain complete PII and residential history
  - Completed for 15/16 registries back to 2009 diagnosis years (2.9 million cancer cases).
  - Capture of residential history will be performed annually to enable improved linkages and to support work on exposure estimations in the appropriate latency period
  - Residential history critical to:
    - Provide longitudinal address information for linkages
    - Support research looking at exposures and cancer
- Working on bringing in a set of Social Determinants of Health for a subset of registries (approved funding for DOE project)

Leveraging SEER Data: Creating a Knowledge Graph to support research, including clinical trial enrollment



This slide demonstrates the integration of a wide variety of important clinical, geographic, environmental and other data to support a heterogeneous set of research activities.

Use Case – Linked data from multiple sources representing patient trajectory over the disease course

	SEER Diagnostic Data	SEER Surgery/ Rad Rx Data	Treatment Claims Data	Treatment Pharmacy Data	Outcome SEER
HR+/HER 2- Breast	49 YO Stage IA ductal Oncotype Score=36	Lumpectomy (7/15) Beam Radiation	Docetaxel, Cyclo-Phosphamide (OCT NOV 2015)	Anastrozole 1 prescription 4/18	Vital Status Alive- 4/18
ER+/HER2+ Breast	70 YO Stage IA Invasive breast	Lumpectomy (1/15) Beam Radiation	Trastuzumab (3/15-3/16) Docetaxal/Carbo (3/15-3/16)	Letrizole 10/15- present 4/18	Vital Status Alive- 5/18
Lung	83 YO F Stage IIB adeno EGFR + Exxon19 ALK -	No Surg No Rad	No systemic chemo)	Gefitinib Nov 2016-Jan 2017 Erlotinib (Feb 2017)	Vital Status Dead 6/17
Stage III Melanoma	23 YO M Stage IIIC Melanoma BRAF V600E/V600K mutation **Groin Mets- Node dissection 10/16**	Biopsy/ Wide excision/ (9/15)	Ipilumimab 12/15	Dabrafenib/ Tretinitinibt Begun 11/16	Vital Status Alive 4/19

Time since Diagnosis

Thank you

NIH NATIONAL CANCER INSTITUTE

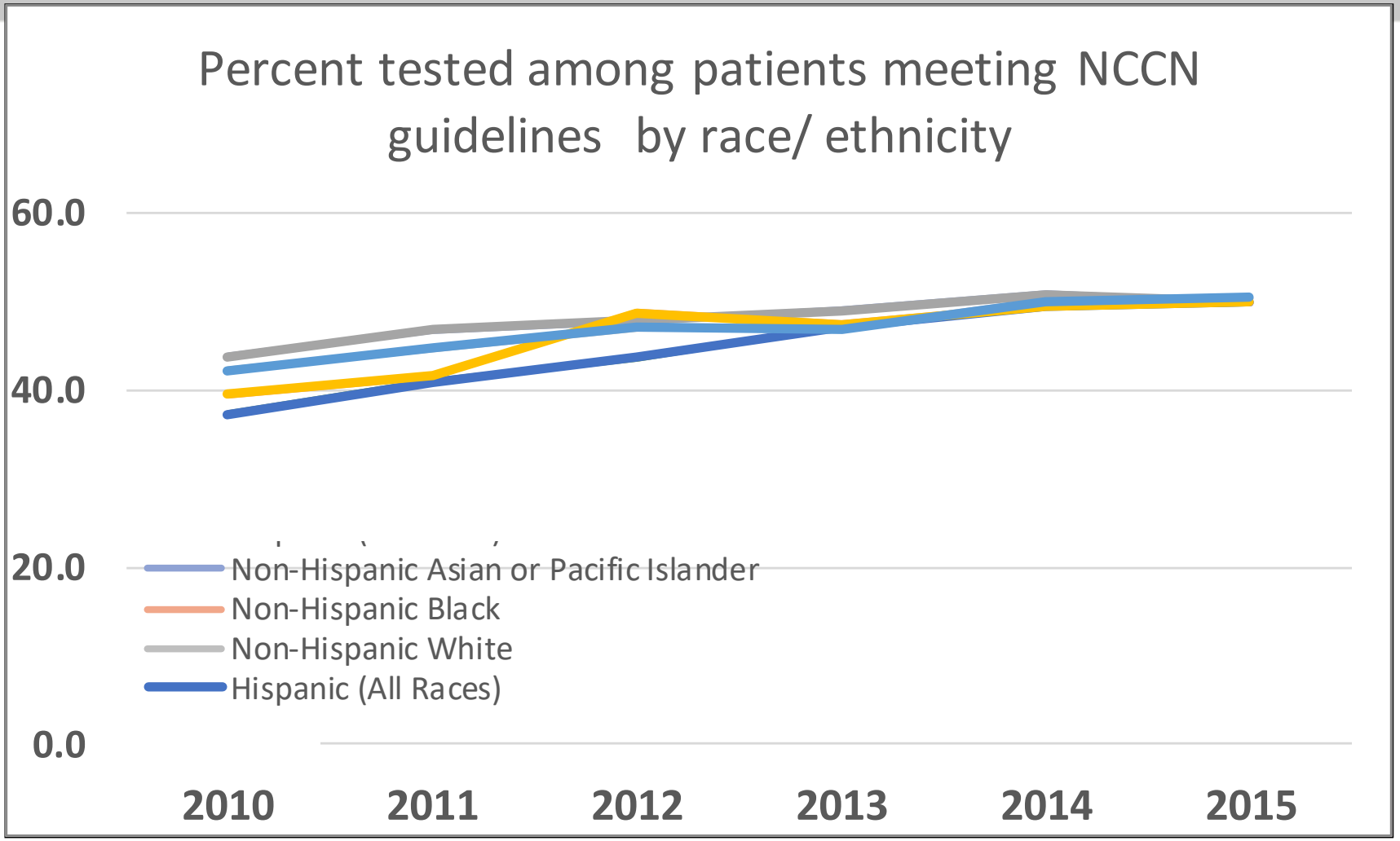
Question and Answers?

NIH NATIONAL CANCER INSTITUTE

How/When can researchers access the data

- Currently evaluating quality and completeness
  - Many not population based yet
  - Sandbox environment being developed to
    - Allow researchers with specific questions and/or expertise to perform analysis in collaboration with NCI staff
- Working towards a different model of data access
  - "SEER Data Commons"
    - Linked data from registries, genomics, longitudinal treatment in the cloud
    - Differing levels of controlled access in the cloud (HIPAA Limited Datasets)
    - Ability to perform analysis in the cloud
    - Downloadable ONLY in special circumstances
  - New method for access will enable
    - Better security (non downloadable) for increasingly detailed data
    - Differing levels of access (e.g. at the most detailed level will likely require IRB (minimal risk study) with different SEER product lines
    - NCI SRP developing a cIRB that will be available and linked
    - Will support access to more detailed and more refined data not currently available

Example: Evaluating trends in standards of care- disparities in Oncotype DX testing rates



During the initial years (2010-2012), there was some evidence of differential testing by race and ethnicity dependent on age.- recent data suggests disparities are disappearing.