SEER: A Changing Paradigm for Cancer Surveillance

CTAC July 17, 2019

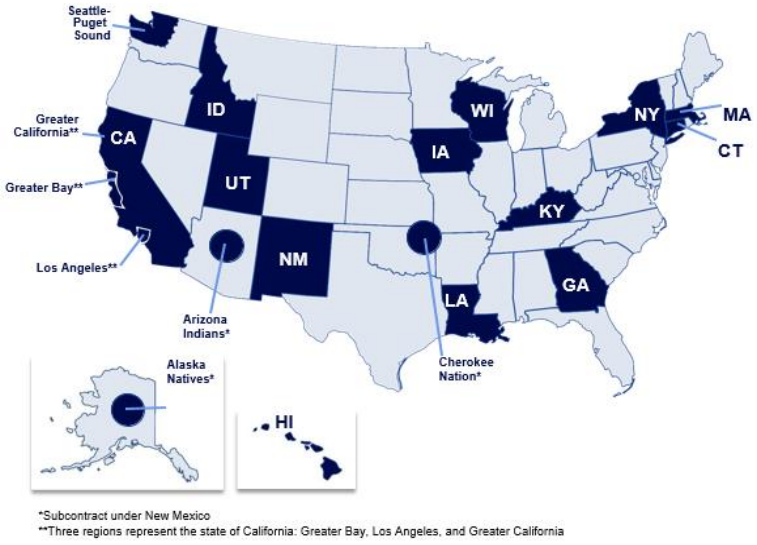
NIH NATIONAL CANCER INSTITUTE

Objectives

- Background on SEER
- Expanding the capacity of SEER to support research
- Examples of new initiatives & results towards enhancing 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
- 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



Cancer Surveillance

- Reminder- reporting to state cancer registries is HIPAA exempt and registries are required to maintain PII for linkages and follow up.
- Registries are legally permitted to collect information from all health care providers on the patient, the cancer, treatment and outcomes

SEER Data Currently Collected

- Data collected routinely includes:
  - Demographics
  - Geospatial data
  - Characterization of the tumor at diagnosis
    - Stage
      - Consolidating data from clinical imaging and pathology
    - Tumor characteristics (including 32 biomarkers)
      - Breast (ER/PR, HER2, Multigene assays (Onco*type* DX and MammaPrint))
      - CRC (CEA, KRAS)
      - Testis (hCG, AFP, LDH)
      - Pharynx (HPV)
      - Liver/ billiary tract (AFP)
      - Ovary (CA-125)
      - Neuroendocrine (Serum Chromogranin, urinary 5-HIAA)
      - Prostate (PSA)
      - Hematologic Malignancies (JAK2)
      - Melanoma (LDH)
  - Treatment (first course)
  - Survival and Cause of death
    - Actively and routinely followed

Value of Surveillance data in the "Real World"

Registries data are valuable for many reasons

- They represent data on all cancer patients in a defined geographic area- not just from a cancer center or hospital system
  - Many real world data sources represent a nonrandom set of patients (from a single center or EMR) which may not reflect what is going on in the general population of cancer patients

- They consolidate information across many sources
  - Typically more than one source of information is used by SEER registries to complete each cancer abstract:
  - Average of 4 records/ case
    - hospital abstracts, physician reports, pathology reports and death certificates
    - Plus additional sources real time data feeds from pharmacies and oncology practices

Value of Surveillance data in the “Real World”

- Active monitoring of patients from diagnosis until death
  - Many data sources lack outcomes to provide context for a dataset (TCGA, Clinical Trials pharma studies etc.)
- Structured data with key clinical information about each patient
  - > 65% of critical information from EMRs is held in unstructured text
  - Extremely costly to pay for structured data collection
- Registry data are curated and adjudicated by trained and experienced personnel
  - While not perfect, the consolidation, manual review and centralization makes the data highly accurate and complete

While SEER Data are very good….there are challenges to capturing clinically meaningful surveillance data

Challenges for cancer surveillance

- Current manual abstraction process of > 215 variables per CASE directly abstracted by registrars- requiring review of many EMR components
- Data elements often complex
  - Staging
    - Registrars need to know how to stage ALL 118 different EOD and/or TNM schemas
    - While clinicians typically specialize on a single organ system with limited diversity in who they stage

Challenges for cancer surveillance

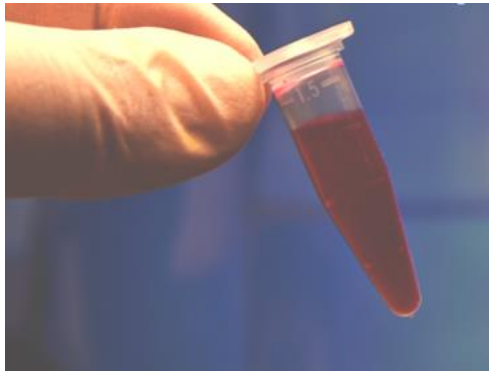
Diversity of health care organizations where patients receive care may provide limited or no access by registrars to these data

Michael Hankins, MD, MPH
http://himss.files.cms-plus.com/2014Conference/handouts/111.pdf

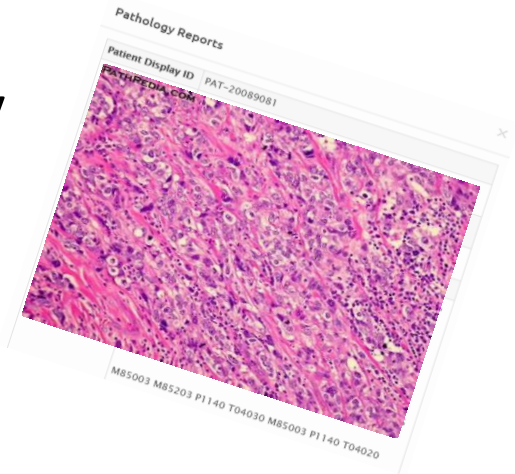
Challenges for cancer surveillance

Rapid pace of change in cancer diagnosis and treatment

- Liquid biopsies
  - Changing the way we diagnose
  - Changing the way we follow patients

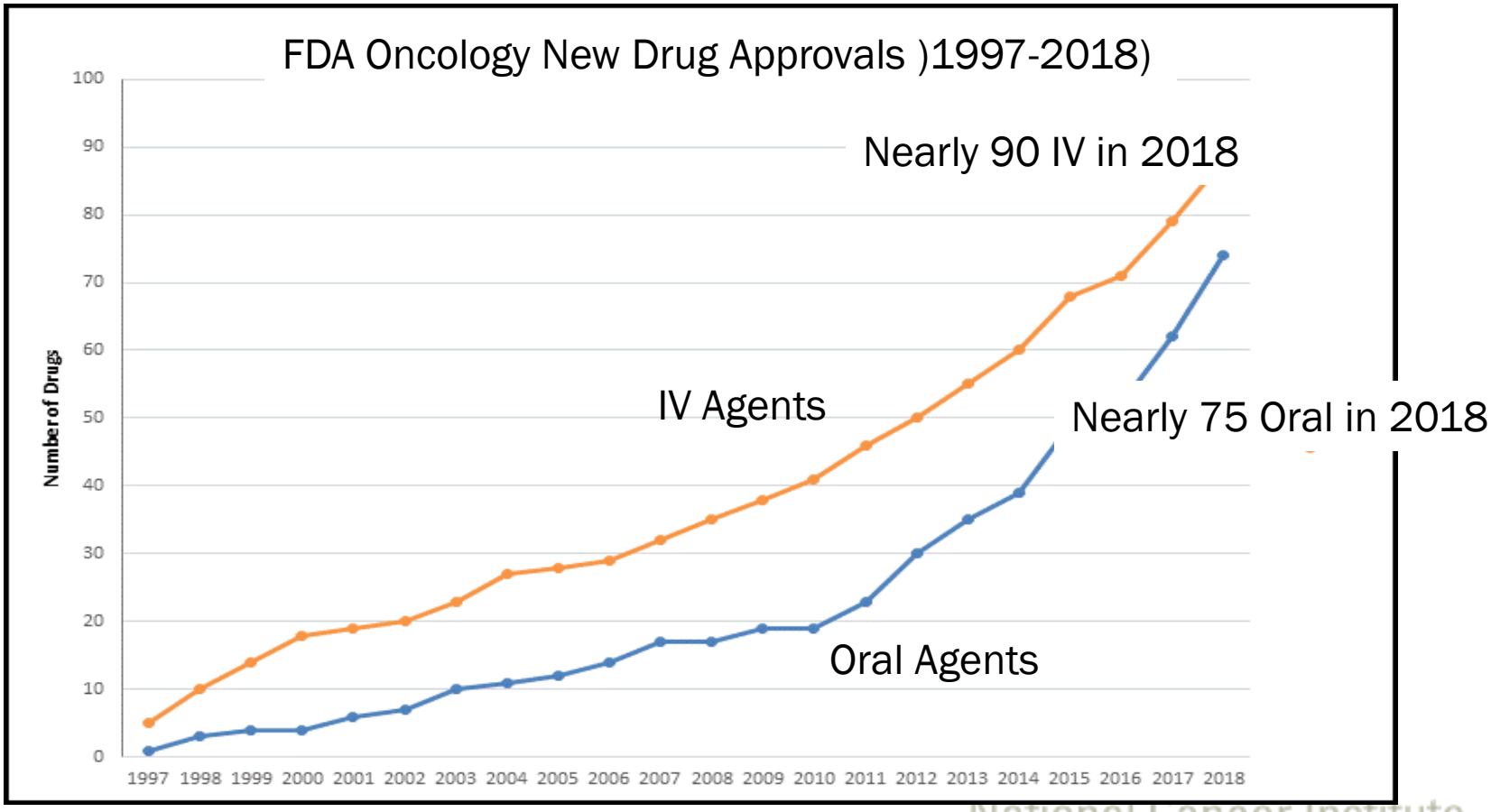


- Digital imaging for path and radiology
  - Features from images not well captured in report documents (TILs)



Challenges for cancer surveillance

Increasing pace of new therapies being approved



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

Approval of new therapies are often based on small samples of selected patients.

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

Challenges for cancer surveillance

- Current manual abstraction is slow-
  - data not timely and
  - with the rapid pace of change the 2 year delay reduces the value of the data

- Registrars may not have access to the appropriate information
  - Outpatient delivered chemotherapy and testing
  - Pharmacy delivered oral antineoplastic therapy
  - Multiple courses of therapy over years

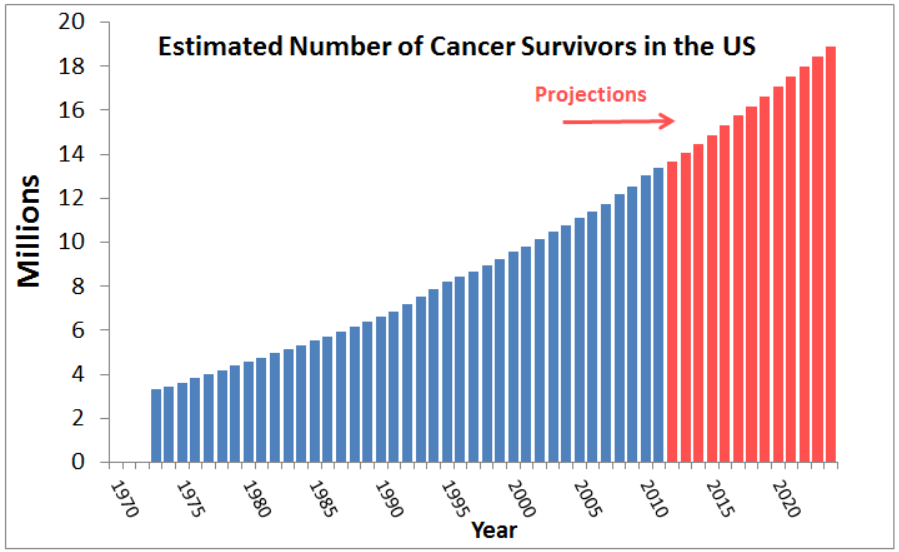
Challenges for Cancer Surveillance

Capturing outcomes other than survival -recurrence

- Cancer is a chronic disease requiring
  - long term measures of outcome (recurrence)
  - Subsequent courses of therapy
  - Comorbid conditions impacting therapy and resulting from therapy

Challenges for Cancer Surveillance

- With nearly 17 million cancer survivors in the US alone (nearly 5% of the population) lack of recurrence information is no longer acceptable
- 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
  - Pancreas
  - Ovarian
  - Melanoma
  - GBM



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

Challenges to capturing recurrence

- Diagnostic methods for recurrence differ by cancer site and provider including one or more of the following:
  - Biopsy
  - Imaging
  - Serologic tests (clinical laboratory tests)
  - Signs and symptoms
- Differential time from diagnosis to recurrence and risk of recurrence for different cancer sites
  - Colorectal vs. breast

Background: challenges to capturing recurrence

- Registrars are unlikely to have access to the heterogeneous data sources from which evidence of recurrence should be derived
  - Path reports
  - Radiologic reports or images
  - Longitudinal serologic lab tests (PSA, CEA etc)
  - Clinical notes from physicians
  - Claims data

- As technology advances (e.g. circulating tumor DNA), when and how “recurrence” is defined will change

Approaches to Enhancing SEER

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

- We are taking an incremental approach using small demonstration pilots to enable us to:
  - Test methods using cost efficient pilots prior to scaling
  - Understand and address barriers and challenges
  - Then scale to all of SEER to create a longitudinal picture of each cancer patient's trajectory from diagnosis to death

Solutions in process at SEER

- Efficiently enhance completeness and expand the clinical data collected through:
  - Linkages to capture current and new data items
    - Cost efficient
    - Increased accuracy and timeliness (real time data feeds often possible)
    - Ability to incorporate data not available manually (e.g. genomic panels)
  - Developing tools for automation (NLP/machine learning) – DOE partnership
    - Reducing manual abstraction
    - Increasing consistency and accuracy above human curation
    - Opportunity to provide real time data to support cancer research
  - Leveraging these activities through collaborations with external partners both commercial and public (CVS, Walgreens, Tempus, Caris, Myriad, etc.)

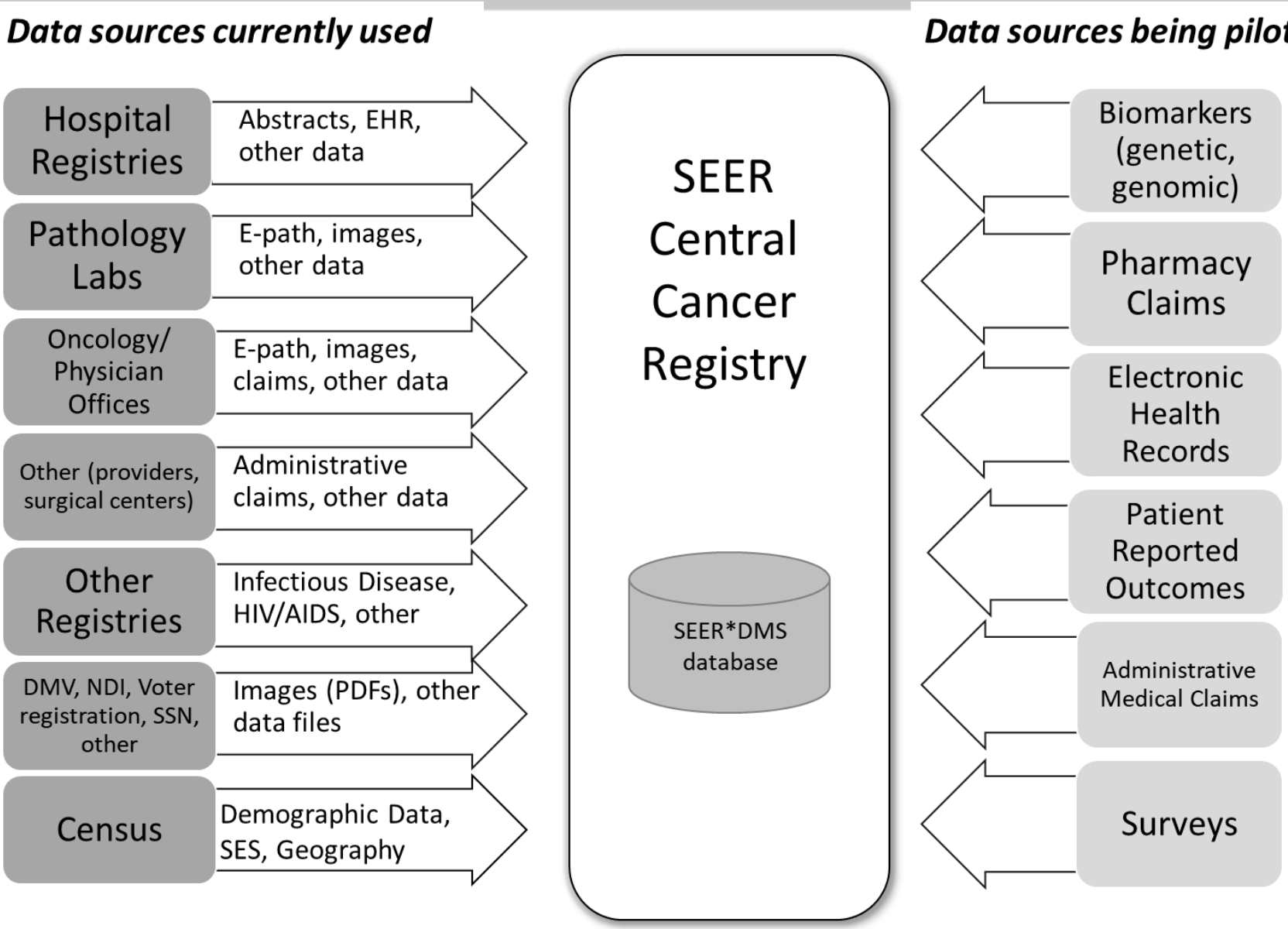
Specific gaps in current surveillance data being addressed with new initiatives

- 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)

Partnerships and linkages to enhance SEER

- Partnerships with organizations to acquire source data
  - Genomic/Genetic testing companies (GHI, FMI, Caris LS, Myriad, etc.)
  - Claims sources
    - Unlimited Systems (oncology claims processor)
    - Large insurers (United Health Care)
    - All Payer All Claims (6 SEER registries have state wide APAC)
    - Pharmacy (CVS and Walgreens)
      - Working to scale across all registries beyond GA
- Partnerships with technology companies aggregating and using clinical data
  - CancerLinQ, Syapse, Tempus
  - Varian/Elekta

SEER Data Sources- current and in testing



Current examples of pilot: Leveraging real world data for clinical utility

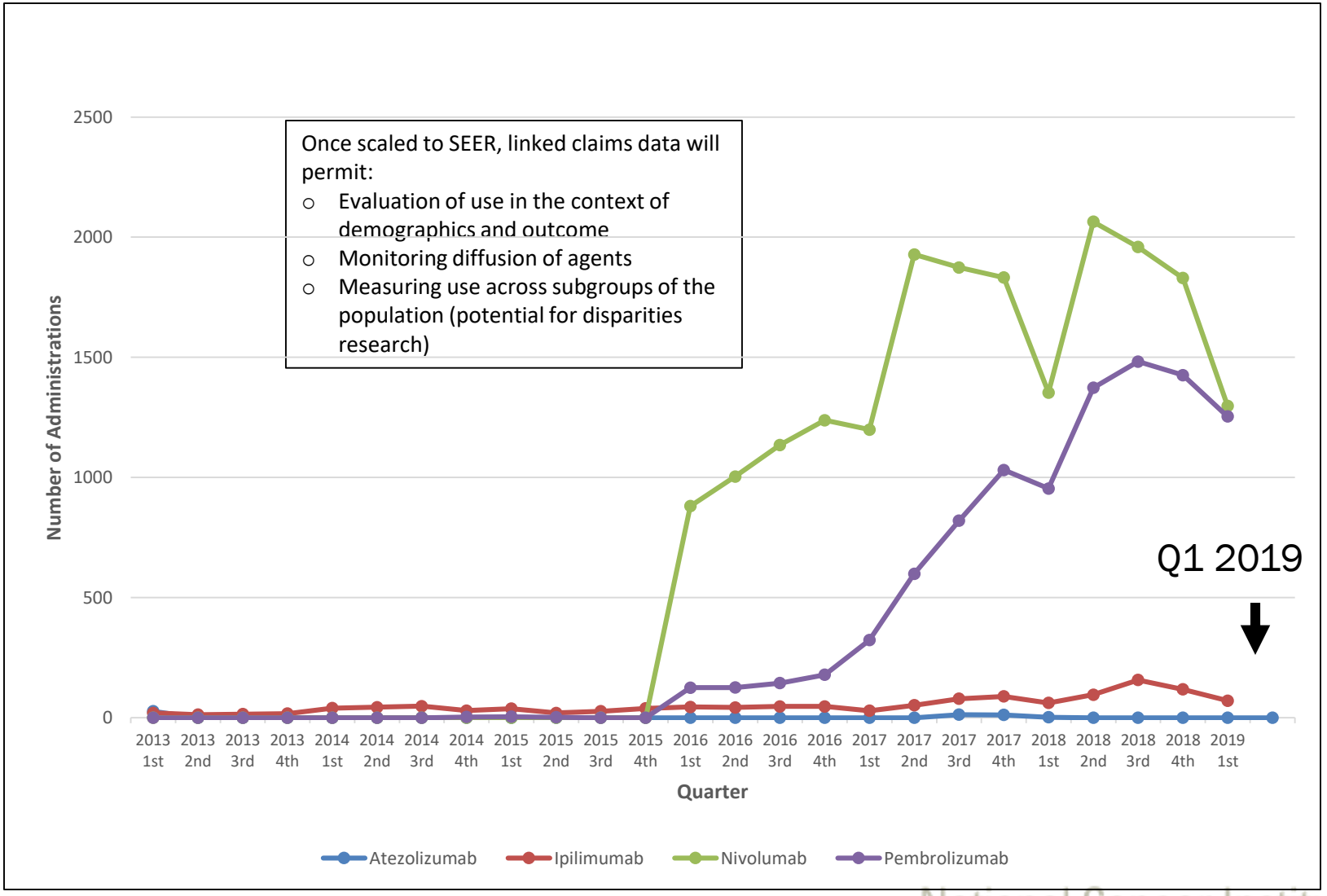
The changing paradigm for surveillance: Examples of what we can do

We are beginning to collect data that will permit

- Tracking and monitoring dissemination of specific treatments over time – beyond the clinical trial setting
- Evaluation of standards of care in oncology practice
- Corroboration of clinical trial results in the real world
- Representing trends by more clinically relevant categories
- Developing tools to support automation
  - CanMed
  - Automated extraction from unstructured text documents

Complimenting Clinical Trial Results with “Real World” Data

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



*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, AnalCanal andAnorectum	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 DigestiveOrgans	1	5		6
Larynx	4	13		17
Lung andBronchus	573	354	26	953
Melanoma ofthe Skin	136	78	137	351
OtherNon-EpithelialSkin	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 andRenal 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: Understanding approved and off label use of Checkpoint Inhibitors by cancer site - (2013- March 31, 2019)

Leveraging Radiation Oncology EMRs: Capturing detailed, real time information from Varian/Elekta EMR products.

Count	Radiation Site	Technique	Modality
182	Vaginal Cuff	Brachy-Intracavitary	Ir-192
167	Breast @ Left	3D/conformal	6X
111	Prostate/SV/Nodes	IMRT	6-X
99	Breast _R	3D-Conformal	Mixed-X
92	Breast @ Right	3D/conformal	6X
92	Breast _R	3D-Conformal	6-X
90	Prostate/SV	IMRT	6-X
85	Prostate	BRACHY-Interstitial	Pd-103
77	Whole Brain	Opposed Laterals	6X
71	Breast _L	3D-Conformal	6-X
70	Breast _L	3D-Conformal	Mixed-X

Example of detailed data automatically captured from a single practice- 2017
Working collaboratively with ASTRO to coordinate with their MDS

Example- Capturing Oral Anti-neoplastics:10 Classes of Medications from Pharmacy Claims (Georgia 2013-2017)

		CVS		Walgreens	
Major Class	Minor Class	Number of Patients	Number of Fills	Number of Patients	Number of Fills
Aromatase inhibitor		11204	133707	3665	37872
Selective Estrogen Receptor Modulator (SERM)		6641	79112	1953	22338
Antiandrogen	non-steriodal	1967	13309	794	4591
Antimetabolite	Pyrimidine Analog	1128	5834	1089	4759
Miscellaneous agent	Antimetabolite/organ ooxygen	730	7459	323	2389
Immunomodulator	Thalidomide analog	687	8333	179	1357
Antiandrogen		674	5218	7	25
Tyrosine kinase inhibitor	BCR-ABL	447	5935	207	1775
Antimetabolite	Purine analog	423	4427	219	1908
Antimetabolite	Folic Acid Analog	417	4488	831	9006
Cyclin dependent kinase inhibitor	CDK 4/6	340	2896	207	1087
Antiandrogenic	CYP17 inhibitor	273	2325	142	886

Initial pilot in GA

✓ once data assessed will scale to entire SEER program

These types of real world data will permit:

- Trend Analyses
- Monitoring of patient adherence and compliance
- Assessing clinical outcomes and disparities

Tracking oral anti-neoplastics through pharmacy data linkages. Example: TKI Use by Cancer Site and Target in GA (2013-2017)

Represents >1,700 patients and >20,000 fills

These types of real world data will permit:

- Trend Analyses
- Monitoring of patient adherence and compliance

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

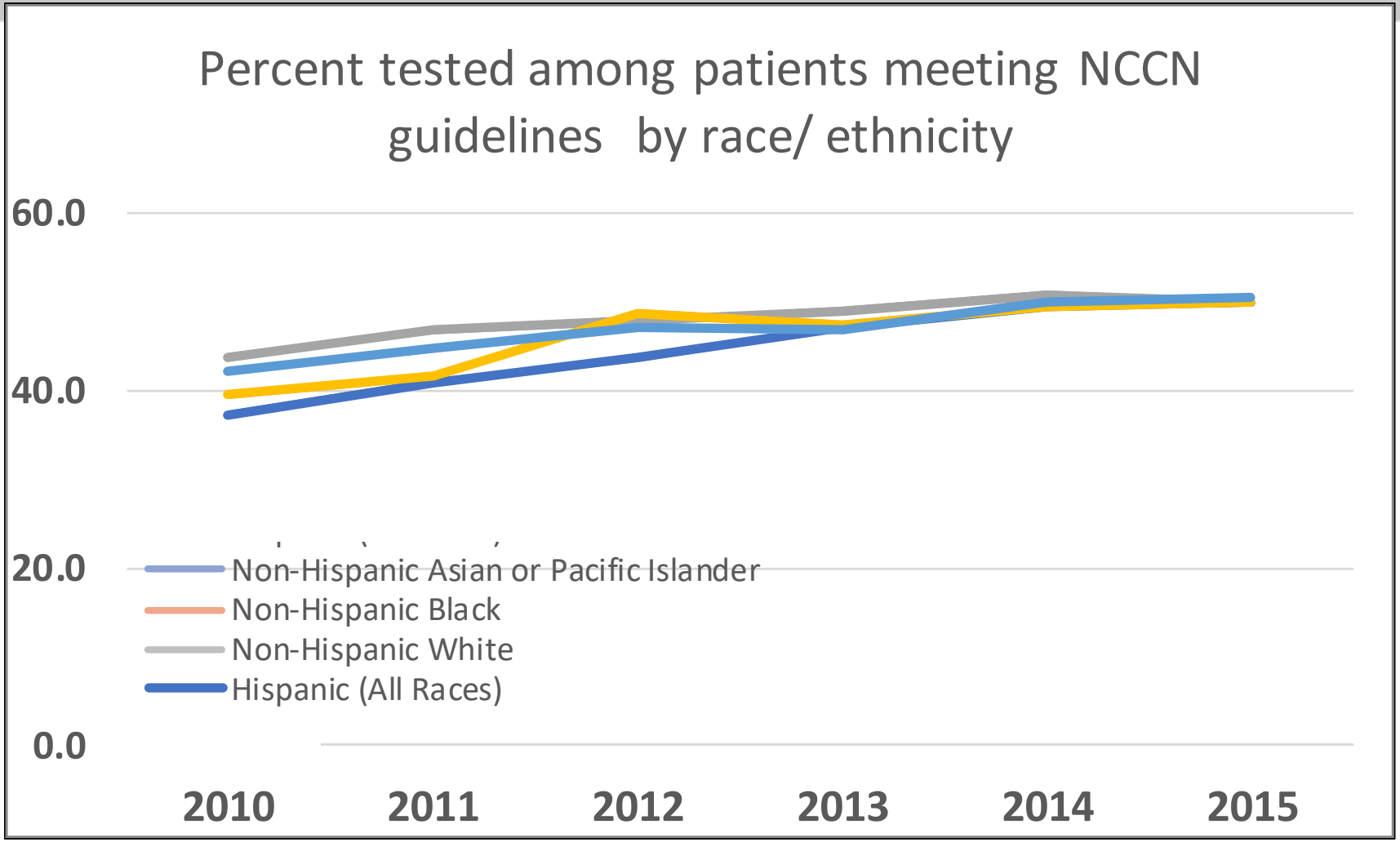
Leveraging SEER for Monitoring Standards of Care

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

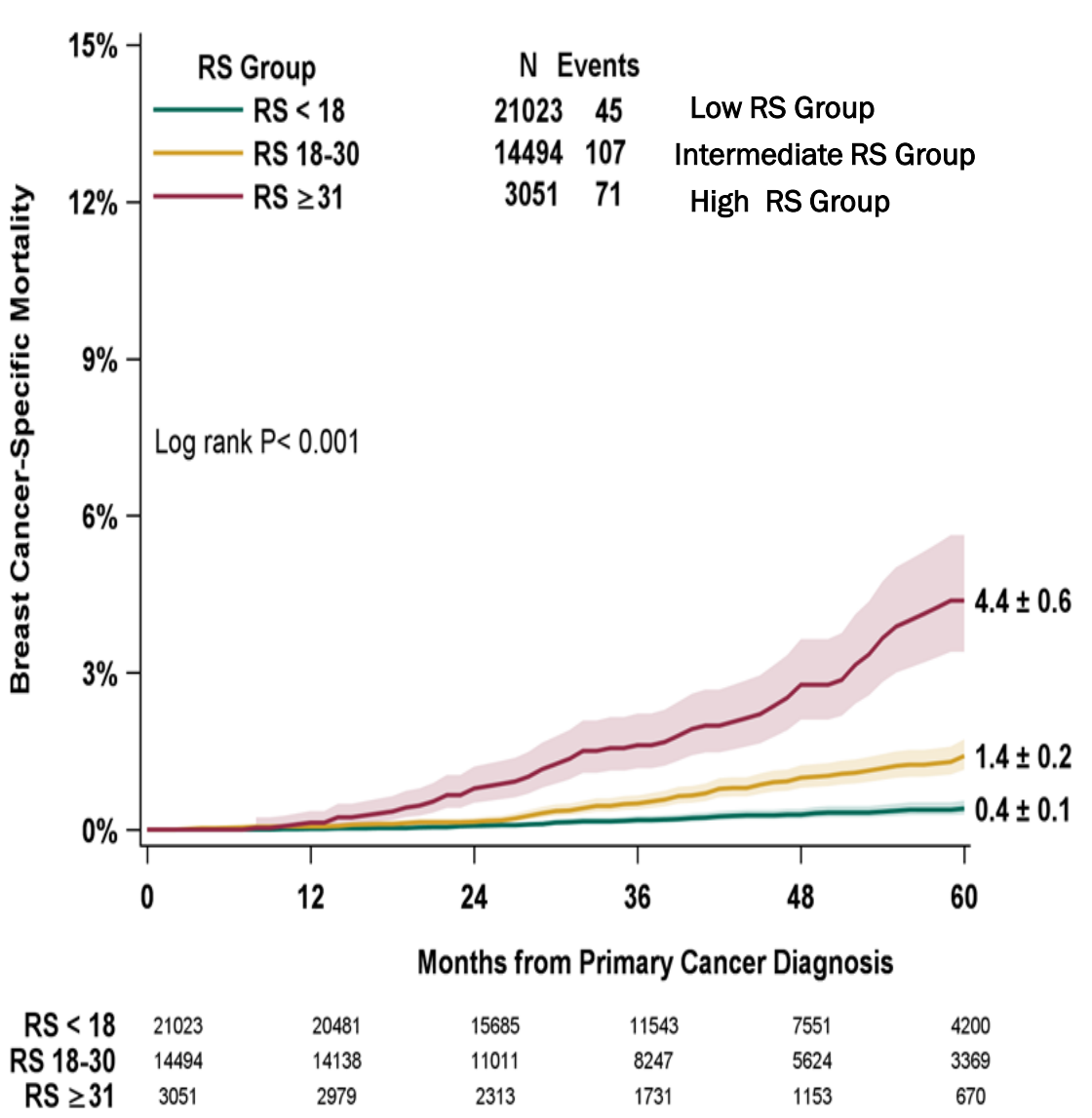


During the initial years (2010-2012), there was some evidence of differential testing by race and ethnicity dependent on age.- recent data suggests disparities are disappearing.

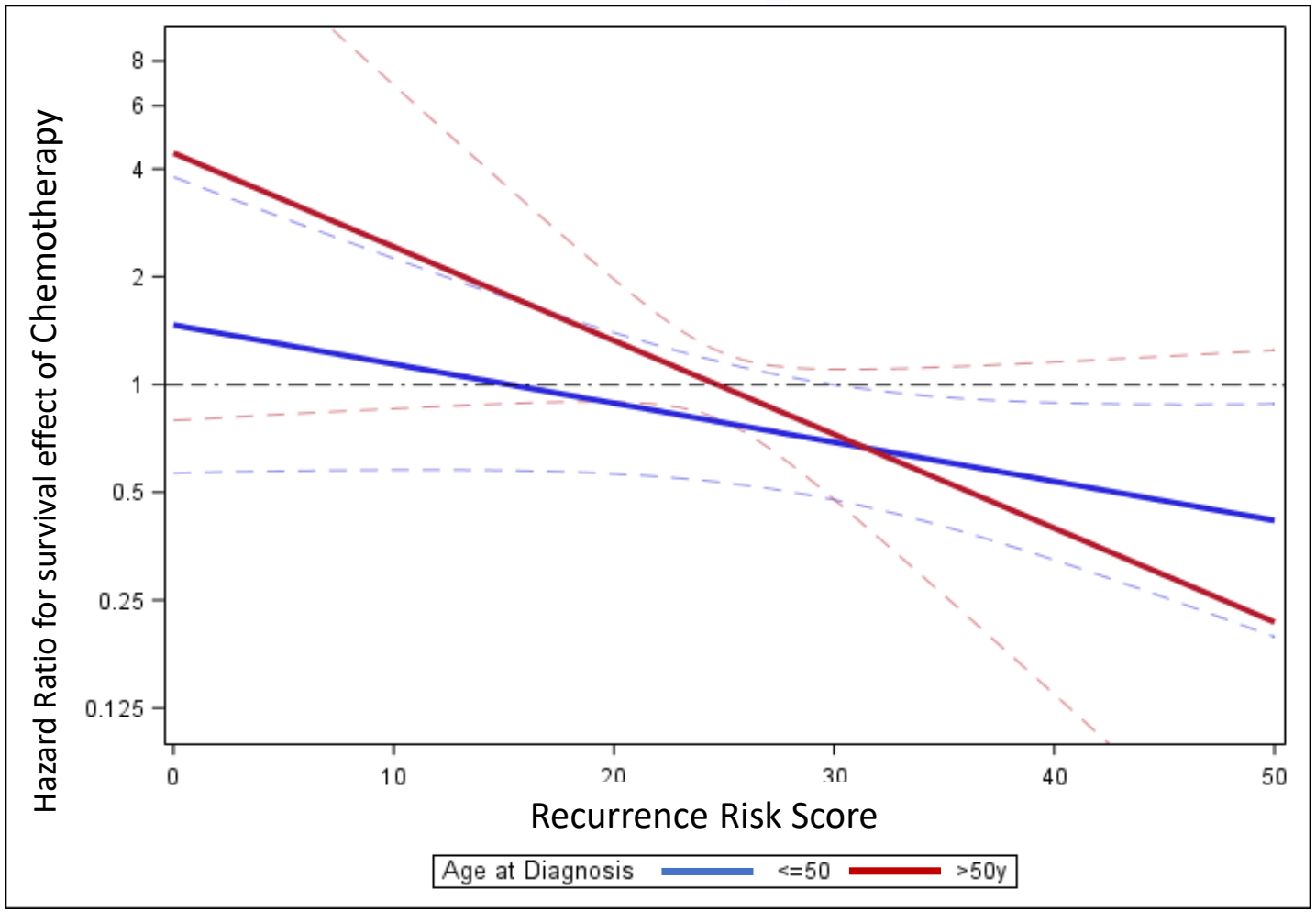
Corroborating Clinical Trial Results in the General Population

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

Oncotype Risk Score Category predicted breast cancer specific mortality



Corroboration of TAILORx findings: Chemotherapy Benefit as a function of Oncotype Dx Risk Score and Age in SEER data (N=70,087)



1.0 = No CT Benefit

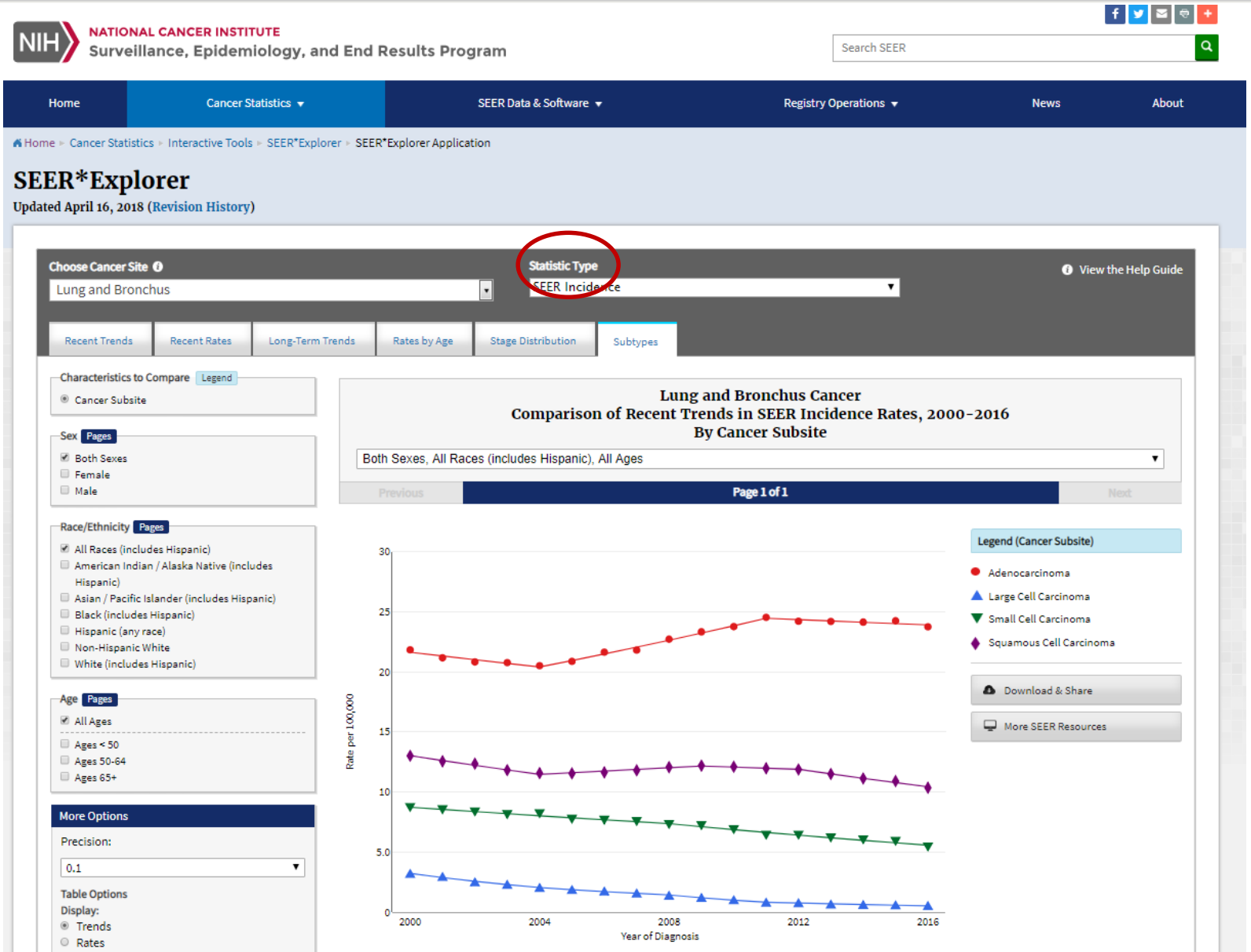
Increasing Chemo Benefit

Replicated TAILORx findings showing increased chemotherapy benefit with increasing RS for younger and older women (HR<1 = protective effect)

Benefit of chemo in younger women starts at a lower risk score (~16)

Reporting Data in Clinically Relevant Categories

Example: representing trends in clinically relevant categories: Lung Cancer Incidence by Histologic Subtype (SEER*Explorer)



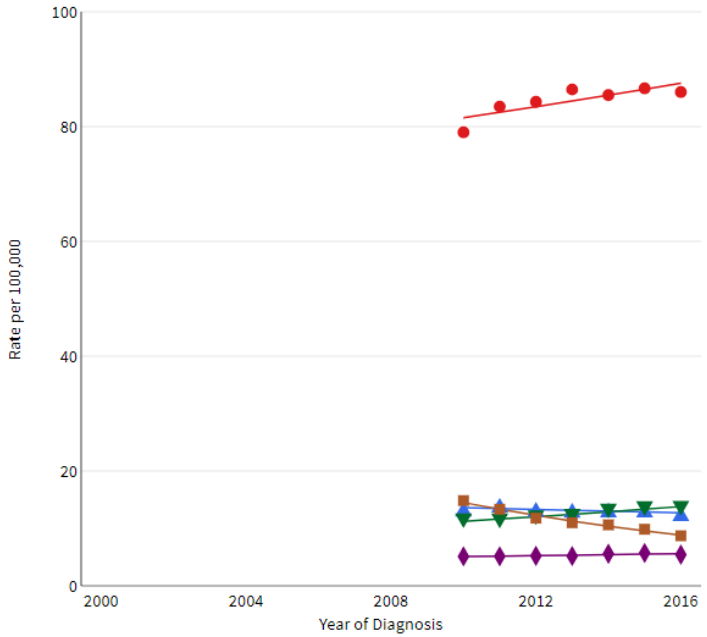
Example: representing trends in clinically relevant categories: breast cancer incidence by molecular subtype

Cancer Subsite	Annual Percent Change			
	Year Range	Estimate (%)	P-Value	Direction
HR+/HER2- (Luminal A)	2010-2016	1.2	0.0	↑
R2- (Triple Negative)	2010-2016	-1.2	0.0	↓
:R2+ (Luminal B)	2010-2016	3.5	0.0	↑
R2+ (HER2-enriched)	2010-2016	1.6	0.0	↑
vn subtype	2010-2016	-8.0	0.0	↓

Breast Cancer
Comparison of Recent Trends in SEER Incidence Rates, 2000-2016
By Cancer Subsite

Female, All Races (includes Hispanic), All Ages

Previous | Page 1 of 1 | Next



Legend (Cancer Subsite)

- HR+/HER2- (Luminal A) [a]
- HR-/HER2- (Triple Negative) [a]
- HR+/HER2+ (Luminal B) [a]
- HR-/HER2+ (HER2-enriched) [a]
- Unknown subtype [a]

[a] Estimates based on less than 16 cases are suppressed and not shown. Trend lines can not be calculated when a rate in the series is suppressed.

Download & Share

More SEER Resources

National Cancer Institute

Developing Tools to Support Automated, Real-time Data Capture

Observational Research in Oncology Toolbox

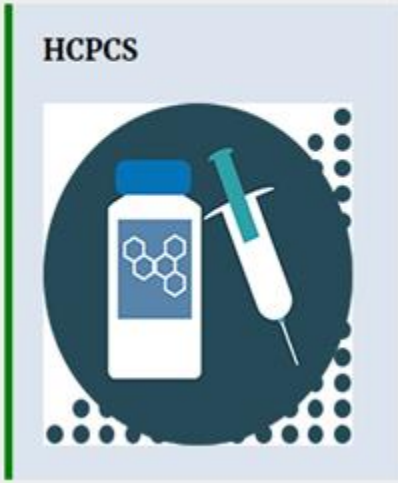
A resource to standardize mapping of relevant oncology treatment codes for automated systems, manual abstraction, and research analyses in cancer surveillance and pharmacoepidemiology

- Used in SEER to automatically categorize and structure data consistently
- Available for download

Part I

CanMED: Cancer Medications Enquiry Database

The Cancer Medications Enquiry Database (CanMED) is a two-part resource for cancer drug treatment related studies. It is intended to facilitate cancer surveillance, epidemiology, and pharmacoepidemiology research that uses the National Drug Code (NDC) and Healthcare Common Procedure Coding Systems (HCPCS) nomenclatures.



National Drug Codes (NDC)

Example: Cyclophosphamide 10019-0945-01

Healthcare Common Procedure Coding System (HCPCS)

Example: Bevacizumab, C9214

Part II: Development ongoing for CPT and ICD9/10

Website address: https://seer.cancer.gov/oncologytoolbox/canmed

API to automatically extract in real time 5 key data items

- Being developed via the DOE partnership – NLP algorithms for real time data extraction
- Path screening task. Currently a registrar manually codes site, histology, behavior, grade, and laterality in this task
  - Mean time to manually complete a path screening task - **0.93 minutes (55 sec) per report** based on 2.2 million manual tasks in 10 SEER registries; 2015+
  - Mean time for the automated algorithm to process 614,230 path screening tasks in 2018 - **12 milliseconds per report**

- Testing the algorithm across 11 SEER registries (3.3 million path reports): **43% percent of path reports had all 5 data elements coded correctly**

DOE partnership – NLP algorithms for real time data extraction– early results

- Estimated time savings for one year based on 43% of 616,230 path reports for which all 5 data elements were correctly extracted
  - **4,048 hours for manual process** for 1 year and 11 registries in the study
  - **0.88 hours (53 minutes) for automated process**
- Caveat: the api runs against all path reports including resection, FNA, Biomarker etc
  - Developing a mechanism to pre-screen path reports as to utility (e.g. surgical resection, molecular test result, lymph node etc.)
- Purpose of the API:
  - Real time incidence reporting
  - Real time identification of patients eligibility in trials and other studies
- Next steps
  - Production implementation of the API in SEER
  - Capturing recurrence (3 studies and an 2 algorithms in development)
  - Capturing biomarkers (currently developing the infrastructure to support automation for selected molecular tests)

Our Goal: Provide a detailed longitudinal picture of treatment and outcomes for each cancer patient

Our goal: to have linked data from multiple sources representing each patient's trajectory over their 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 11/18

Time since Diagnosis

Thank you

NIH NATIONAL CANCER INSTITUTE

Questions for discussion:

1. Given the described enhancements- what do you see as additional priorities for data integration into the SEER program
2. Are there additional data sources that we should consider (COG, NCTN etc.?)
3. Do you have suggestions for how to integrate the real time data capture in SEER to support the CTs system?
    - Ideas for pilots that would be useful to develop and test such a system?