SEER: a changing paradigm for cancer surveillance
National Cancer Policy Forum 2019
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
  o RFP for expansion in process

• With new registries –550,000 incident cases received annually
  o Approximately 85% of cases with real time electronic pathology (e-path) reporting
  o Facilitates rapid case identification supporting research

• All registries will be on a common data platform (SEER DMS) that permits
  o central linkages with external partners
  o facilitates scaling of new initiatives across all registries simultaneously
SEER Data Sources - current and in testing

**Data sources currently used**

- **Hospital Registries**: Abstracts, EHR, other data
- **Pathology Labs**: E-path, images, other data
- **Oncology/Physician Offices**: E-path, images, claims, other data
- **Other (providers, surgical centers)**: Administrative claims, other data
- **Other Registries**: Infectious Disease, HIV/AIDS, other
- **DMV, NDI, Voter registration, SSN, other**: Images (PDFs), other data files
- **Census**: Demographic Data, SES, Geography

**Data sources being piloted**

- **Biomarkers (genetic, genomic)**
- **Pharmacy Claims**
- **Electronic Health Records**
- **Patient Reported Outcomes**
- **Administrative Medical Claims**
- **Surveys**
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)

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

<table>
<thead>
<tr>
<th>Lung Biomarker</th>
<th>Prevalence</th>
<th>Treatment Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR</td>
<td>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?</td>
<td>Erlotinib</td>
</tr>
</tbody>
</table>
**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

---

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

---

*Orphan Drug with accelerated approval to fill an unmet medical need (November 2018)*
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.
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, Unlimited 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
  o Detailed longitudinal treatment data
  o Comprehensive genomic data characterizing the cancer
  o Outcomes other than survival and cause of death (recurrence)
  o Comorbidity to provide context for therapies and outcomes

• Developing infrastructure to support cancer research
  o SEER wide mechanisms for Rapid Case Ascertainment for patient eligibility assessment for RCTs and other studies (including patient contact studies)
  o Virtual Pooled Registry (VPR)
  o 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
  o Tracking and monitoring dissemination of specific treatments over time
  o Evaluation of standards of care in oncology practice over time
  o Corroboration of clinical trial results in the real world
  o Representation of trends by more clinically relevant categories
  o 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

Once scaled to SEER, linked claims data will permit:
- Evaluation of use in the context of demographics and outcome
- Monitoring diffusion of agents
- Measuring use across subgroups of the population (potential for disparities research)

*Represents 12-35% of oncologists in 6 SEER registries and approximately 10,000 administrations
<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>Total Unique patients receiving at least one administration of a checkpoint inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nivolumab</td>
</tr>
<tr>
<td>All</td>
<td>1178</td>
</tr>
<tr>
<td>Tongue</td>
<td>12</td>
</tr>
<tr>
<td>Oral Cavity</td>
<td>26</td>
</tr>
<tr>
<td>Esophagus</td>
<td>12</td>
</tr>
<tr>
<td>Stomach</td>
<td>7</td>
</tr>
<tr>
<td>Colon</td>
<td>15</td>
</tr>
<tr>
<td>Rectum</td>
<td>3</td>
</tr>
<tr>
<td>Anus, Anal Canal and Anorectum</td>
<td>10</td>
</tr>
<tr>
<td>Liver</td>
<td>31</td>
</tr>
<tr>
<td>Intrahepatic</td>
<td></td>
</tr>
<tr>
<td>Bile Duct/GB/Other Biliary</td>
<td>3</td>
</tr>
<tr>
<td>Pancreas</td>
<td>11</td>
</tr>
<tr>
<td>Other Digestive Organs</td>
<td>1</td>
</tr>
<tr>
<td>Larynx</td>
<td>4</td>
</tr>
<tr>
<td>Lung and Bronchus</td>
<td>573</td>
</tr>
<tr>
<td>Melanoma of the Skin</td>
<td>136</td>
</tr>
<tr>
<td>Other Non-EPithelial Skin</td>
<td>2</td>
</tr>
<tr>
<td>Breast</td>
<td>18</td>
</tr>
<tr>
<td>Cervix Uteri</td>
<td>2</td>
</tr>
<tr>
<td>Corpus Uteri</td>
<td>5</td>
</tr>
<tr>
<td>Ovary</td>
<td>10</td>
</tr>
<tr>
<td>Prostate</td>
<td>19</td>
</tr>
<tr>
<td>Urinary Bladder</td>
<td>20</td>
</tr>
<tr>
<td>Kidney and Renal Pelvis</td>
<td>190</td>
</tr>
<tr>
<td>Ureter</td>
<td>2</td>
</tr>
<tr>
<td>Thyroid</td>
<td>2</td>
</tr>
<tr>
<td>Hodgkins</td>
<td>10</td>
</tr>
<tr>
<td>Non-Hodgkins</td>
<td>4</td>
</tr>
<tr>
<td>Mesothelioma</td>
<td>8</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>Target</th>
<th>Generic Drug Name</th>
<th># Unique Patients with Anti-neoplastic Prescriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CVS</td>
<td>Walgreens</td>
</tr>
<tr>
<td>NSCLC</td>
<td>ALK</td>
<td>42</td>
<td>13</td>
</tr>
<tr>
<td>NSCLC</td>
<td>EGFR</td>
<td>229</td>
<td>174</td>
</tr>
<tr>
<td>CML</td>
<td>BCR-ABL</td>
<td>675</td>
<td>300</td>
</tr>
<tr>
<td>RCC/Thyroid</td>
<td>VEGF</td>
<td>100</td>
<td>41</td>
</tr>
<tr>
<td>RCC</td>
<td>VEGFR</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>RCC</td>
<td>VEGF, FLT, PDGFR, Kit, RET, CSF</td>
<td>118</td>
<td>72</td>
</tr>
<tr>
<td>RCC</td>
<td>VEGF FGF, PDGFR, Kit, RET, CRAF, BRAF</td>
<td>138</td>
<td>122</td>
</tr>
<tr>
<td>RCC</td>
<td>VEGF, FGF, PDGFR, Kit, Lck, FMS</td>
<td>143</td>
<td>167</td>
</tr>
<tr>
<td>CRC/ HCC</td>
<td>VEGF, FGF, PDGFR, Kit, RET, TIE2....</td>
<td>115</td>
<td>69</td>
</tr>
<tr>
<td>BC</td>
<td>HER2, EGFR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma/NSCLC</td>
<td>BRAF V600</td>
<td>30</td>
<td>29</td>
</tr>
</tbody>
</table>
Example: Evaluating standards of care- BRCa testing among patients with ovarian (and breast) cancer - CA & GA (2013-2015) *

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Breast Cancer</th>
<th>Ovarian Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Cases</td>
<td>Tested* Cases</td>
</tr>
<tr>
<td>State and year of diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>California§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>30,367</td>
<td>7,314</td>
</tr>
<tr>
<td>2014</td>
<td>30,012</td>
<td>6,951</td>
</tr>
<tr>
<td>2013-2014</td>
<td>60,379</td>
<td>14,265</td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>8,296</td>
<td>2,066</td>
</tr>
<tr>
<td>2014</td>
<td>8,410</td>
<td>2,270</td>
</tr>
<tr>
<td>2013-2014</td>
<td>16,706</td>
<td>4,336</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic (NH) White</td>
<td>48,063</td>
<td>11,635</td>
</tr>
<tr>
<td>NH Black</td>
<td>9,039</td>
<td>2,095</td>
</tr>
<tr>
<td>NH American Indian</td>
<td>207</td>
<td>51</td>
</tr>
<tr>
<td>NH Asian</td>
<td>9,061</td>
<td>2,034</td>
</tr>
<tr>
<td>Hispanic</td>
<td>10,715</td>
<td>2,786</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>HR+/HER2- Breast</th>
<th>SEER Diagnostic Data</th>
<th>SEER Surgery/Rad Rx Data</th>
<th>Treatment Claims Data</th>
<th>Treatment Pharmacy Data</th>
<th>Outcome SEER</th>
</tr>
</thead>
<tbody>
<tr>
<td>49 YO</td>
<td>70 YO</td>
<td>83 YO F</td>
<td>23 YO M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage IA ductal</td>
<td>Stage IA invasive</td>
<td>Stage IIB adeno EGFR+</td>
<td>Stage IIIC Melanoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oncotype Score=36</td>
<td>Invasive breast</td>
<td>Exxon19 ALK -</td>
<td>BRAF V600E/V600K mutation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beam Radiation</td>
<td>Beam Radiation</td>
<td>No Rad</td>
<td>Ipilimumab 12/15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HR+/HER2- 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

**Stage IA invasive breast**

- Lumpectomy (1/15)
- Beam Radiation
- Trastuzumab (3/15-3/16)
- Docetaxel/Carbo (3/15-3/16)
- Letrizole 10/15-present 4/18
- Vital Status Alive- 5/18

**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 IIIC Melanoma**

- Biopsy/Wide excision/ (9/15)
- Ipilimumab 12/15
- Dabrafenib/Tretinitinib Begun 11/16
- Vital Status Alive 4/19

**Groin Mets- Node dissection 10/16**
Thank you
Question and Answers?
How/When can researchers access the data

• Currently evaluating quality and completeness
  o Many not population based yet
  o 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
  o “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
  o 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
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.