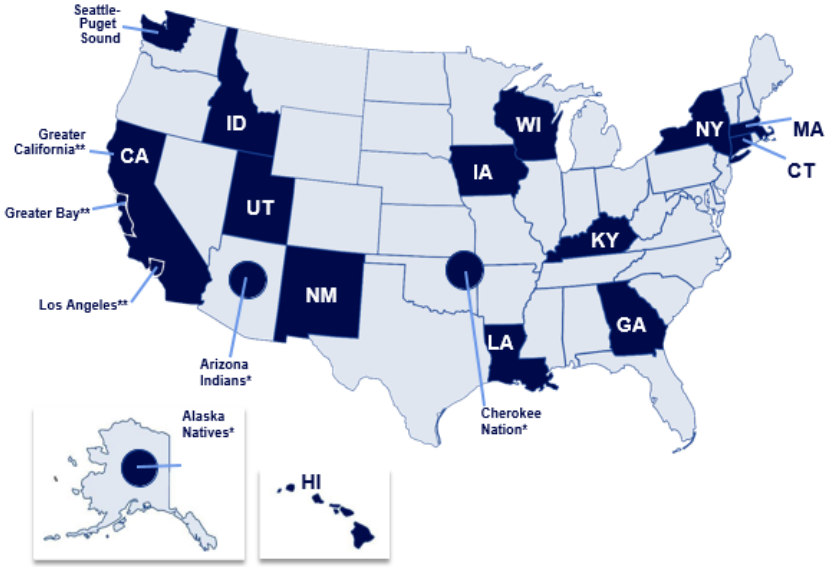
SEER DMS Face to Face Meeting
September 26, 2018

NIH NATIONAL CANCER INSTITUTE

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 **34%** of the US population
  - Registries represent racial and ethnic minorities
  - Various geographic areas



*Subcontract under New Mexico
**Three regions represent the state of California: Greater Bay, Los Angeles, and Greater California

The SEER Program

- With new registries −550,000 incident cases received annually
  - Approximately 85% of cases with real time electronic pathology (e-path) reporting
    - Jan 2019 mandated e path reporting from all pathology labs in CA
  - Goal of increased e-path coverage to facilitate
    - automate data capture of key data elements
      - Beginning with Site, Histology, Behavior, Grade and Laterality
    - support real time case ascertainment for clinical trials and other studies
    - provide capacity to perform early incidence reporting

Why have a focused DMS meeting?

- SEER DMS is central to much of what we are doing in enhancing SEER
- Needed for all the many key daily activities that you perform
- Essential to have a system that allows us to
  - Link data from various external partners
  - Consolidate data from multiple sources
  - Integrate new methods for quality
  - Build new tools for automation (i.e. NLP, Automation Algorithms)
- Plus we like to see you all in person!

Challenges for DMS- Many and heterogeneous data sources

- Data consolidated centrally from multiple sources by trained personnel to capture most accurate data from:
  - Hospitals
  - Path labs
  - Physician practices
  - Linked to original sources
    - Commercial entities (e.g. Genomic Health Incorporated)
    - Radiology facilities (reports)
    - Vital records, NDI, DMV, SSA
    - Accurint/Lexus-Nexus

Challenges for DMS- Many and heterogeneous data sources

- Central review and consolidation of multiple sources permits validation of key elements
  - Typically more than one source of information is used by SEER registries to complete each cancer abstract:
    - Average of 3.6 records/ case including hospital abstracts, physician reports, pathology reports and death certificates
    - Average of 1.7 path reports/ case
    - Additional sources include real time data feeds from pharmacies, meaningful use reports from oncology practices, claims from oncology practices

Challenges for DMS- Need to support new and complex infrastructures

- Virtual Biorepository
- Virtual Pooled Registry
- Real Time data integration and automated data extraction (DOE pilot)

Challenges to DMS: Supporting new data and new data types

- Genomics
  - Multigene panels
    - Multiple cancers(Foundation, Caris etc.)
    - Individual cancers (GHI, Myriad, GenomceDx)
  - Genetic panels (BRCa Linkage)
- Longitudinal treatment data
  - Commercial insurers
  - Central claims processers
  - Pharmacy data
  - Automated processing of abstract text documentation for chemo
- Outcomes
  - Recurrence
  - Patient generated data

Use Case Examples of Patient Trajectories – Linked data from multiple sources

	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 10/16**	Biopsy/ Wide excision (9/15)	No systemic chemo)	Dabrafenib/ Tretinitinibt (11/16 –present)	Vital Status Alive 2/18

Time since Diagnosis

Challenges for DMS

In summary DMS needs to:

- deal with different and varied data types
  - Longitudinal treatment (claims)
- integrate across and combine differing data types
  - Pharmacy data and claims and possibly NLP extracted data for treatment
  - Path digital images
  - Large volumes of path reports and radiology reports
- support and integrate with different data systems in new ways
  - How do we link with complex genomic data? Integrate with GDC
  - Provide integration with the cloud for differing applications?
  - VTR Central Repository
  - Virtual Pooled Registry

What are we doing?

Working to develop methods that ensure a stable environment that :

- Takes an incremental approach to adding new components and usability testing in small sets to assure ability to scale
- Flexible to allow new data and new types of data
- Supports some registry specific and project needs without excessive customization
- Enables integration with important key infrastructures (VTR, VPR etc)

What is our approach?

- Focusing on key areas of short term benefit.
- Priorities from the NCI perspective include:
  - Migrate all SEER Funded Registries onto SEER*DMS
  - Develop system supported edit checks
  - Integrate “Uncertainty Quantification” (UQ) to reduce manual review
  - Update the system through expert usability evaluations to enhance efficiency and quality
  - Using DMS to support efficient use of QIEs leveraging their expertise in particular areas that will direct system builds

In summary

- We hope to use this annual meeting to help us
  - Gather feedback about where we have been going
  - Share with you our proposed next steps
  - Help define best practices for operational processes in SEER*DMS
  - Get input from the broader community to help us prioritize next steps and future development of the SEER*DMS product

Thank you

NIH NATIONAL CANCER INSTITUTE