Enabling 21st Century Applications for Cancer Surveillance through Enhanced Registries and Beyond
Session 1.

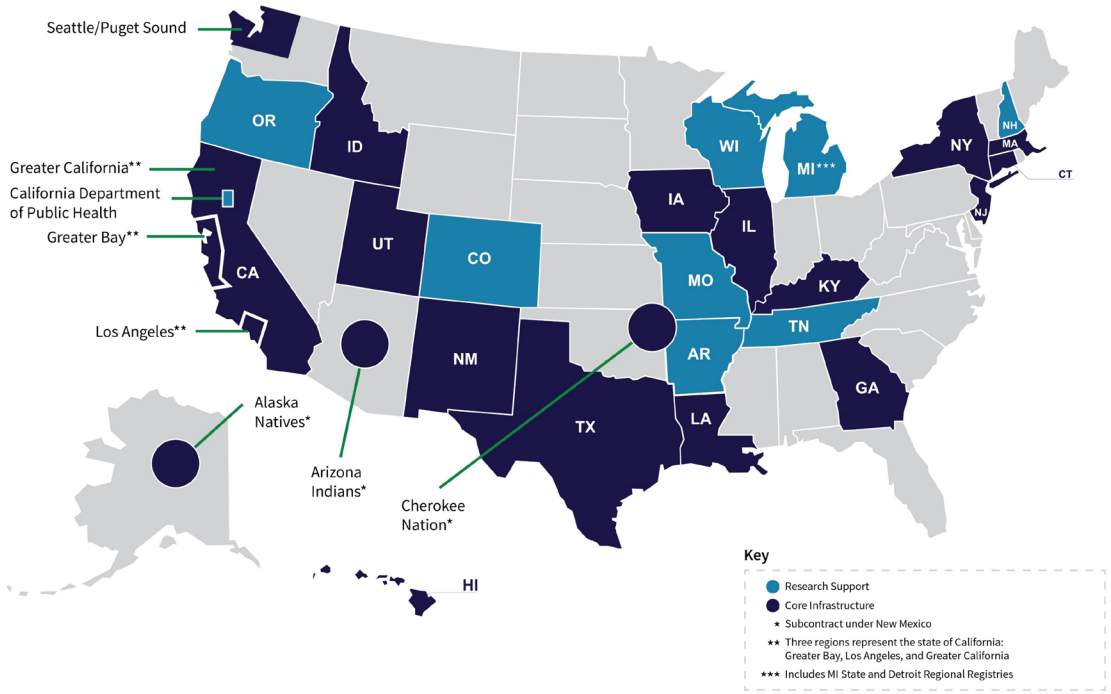
Understanding the complexity of cancer surveillance operations

July 29, 2024

Lynne Penberthy

The SEER Program:

- Funded by NCI ***to support research*** on the diagnosis, treatment and outcomes of cancer since 1973
  - Data collected under public health reporting regulations in each state (reporting HIPAA exempt)
  - Requires reporting by ***all cancer care providers***
  - De-identified (limited dataset) submitted to NCI (and CDC)
- 18 population-based central registries that submit data annually covering **48%** of the US population (dark blue)
- ~900,000 incident cases received annually
- Use of a common data platform in a central IMS enclave permits centralized efficient linkages through an Honest Broker (IMS)



Additional Detail Backup Slides 21-22

Cancer Surveillance Challenges facing Population Based Registries (SEER & NPCR)

- How cancer patients are diagnosed and managed is changing so rapidly it is challenging even for clinicians to keep up to date

- There are no data *outside clinical trials* that provide information on
  - dissemination and general use of new diagnostic methods and treatments
  - the impact of new treatments on outcomes in the general cancer population

- Guidelines for treatment are based on clinical trials but…
  - these capture <5% of the cancer population
  - are non-representative - largely white, younger and no comorbidities

- Therefore we need rapid, efficient collection of ***population level data*** to understand the use and effects of these new treatments in patients outside clinical trials in the real world

Strengths and Limitations of Registries

Registries generally have a broad mandate enabling targeted increments in data collected as clinical care evolves (e.g. genomics, oral antineoplastics, social determinants of health)

However:

- Data still often manually collected
- Registry accession/reporting is slower than desired
- Registries do not have comprehensive access to the entire patient medical record
  - Multiple EHRs across specialty providers
  - ***Many*** other data sources used for cancer surveillance limiting capture of the longitudinal patient record (recurrence, subsequent therapies)

Data sources leveraged for cancer surveillance

While EHR data are important – Many more data sources (up to 35) are essential for creating a complete cancer case across SEER

- Hospital Abstracts
- Physician office reporting
- E path reports
- Pathology Images
- Radiology reports (case identification and recurrence/mets)
- Genomic testing data from labs/specialty labs
  - Exact Sciences
  - Decipher
  - Foundation
  - Castle Life Sciences
- Medical Claims based treatment information from physician practices initial and subsequent infusion treatments (oncology, radiation oncology)
  - United Health Group (All registries)
  - Unlimited Systems
  - State Medicaid
  - Humana
  - Anthem
  - Blue Cross
  - Varian
  - Dermatology Clinical reports
- Genetic testing data from labs (germline mutations for actionable mutations) Myriad, Invitae, Ambry
- Pharmacy data from CVS, Walgreens, RiteAid, UHC PBM (longitudinal prescriptions anti-neoplastics)
- Treatment information from inpatient-based claims (surgery, radiation, infusion-based codes)
- Clinical Trials/Research Studies (COG, National Breast & Cervical Cancer Early Detection Network, Multi-Ethnic Cohort)
- All Payer Claims data
- Vital Records
- NDI
- Motor Vehicle Data
- Voter Registration
- Interstate Data Exchange
- Veteran's administration data exchange
- Lexis Nexis (res hx, measures of financial status)
- SSA
- CMS
- Indian Health Service Linkage for confirmation of native American Status

A use case example for surveillance complexity: Capturing recurrence

Identification of recurrent metastatic disease is a critical outcome for both patients and providers

- There are no population level data on risk of recurrent metastatic disease
- Challenge for surveillance as mets are diagnosed
  - across inpatient & outpatient settings and
  - by multiple specialties (no single source of "truth")
  - Via a variety of diagnostic modalities (pathologic confirmation, imaging, biochemical analysis)

Potential solutions require multiple methods/approaches and data sources

- Leveraging DOE E path API (for recurrence)
- Imaging report often initial modality of diagnosis (10s of thousands of radiology facilities)
- Algorithm using ICD DX codes from inpatient and outpatient claims "Secondary malignancy"
- Longitudinal treatment algorithm
- Hospital reported (low sensitivity and high validity)
- Linking across all registries to capture subsequent malignancies (and data consolidate) VPR (46 states now participate)
- Potential policy solution:

Requirement of CMS mandating reporting disease status for each outpatient visit linked to reimbursement (disease free, progression, recurrence)

Thank you for your attention.