Surveillance Epidemiology and End Results (SEER) Program Update for Joint BSA/NCAB Meeting

Lynne Penberthy MD, MPH

June 14th, 2021

NIH NATIONAL CANCER INSTITUTE

Objectives

- Briefly describe the SEER History, research value and evolution
- New activities and enhancements of SEER Data through Linkage and Automation
- SEER Data Access/Patient Privacy
- Other SEER-related activities
  - Virtual Pooled Registry (VPR)
  - National Childhood Cancer Registry (NCCR)

SEER History and Research Use

Mission: Monitor cancer trends and support research on the diagnosis, treatment and outcomes of cancer.

SEER Growth Over Time		
SEER Version	**Year Initiated**	**Coverage**
SEER-9	1975	9.4%
SEER-13	1992	13.4%
SEER-18	2000	27.8%
SEER-21	2017	36.7%
2021	**2021**	**48%**

SEER Utilization Metrics (2020)

Publications to date

17,000+ SEER Data for the Primary Analysis

86,000+ Referencing SEER Data

2,200+ Using SEER-Medicare

Grants 2020 ($55.3 M)

SEER primary focus **(61/ $37.3M)**

SEER as reference data Source **(13/ $4.6M)**

SEER Intramural grant Support **(10/ $12.9M)**

>4,500 Annual Data Downloads

Enhancing SEER to Meet Real World Data Needs

- 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
  - How new diagnostic methods and treatment are used and
  - Their impact 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
- We need ***population level data*** to understand the use and effect of these new treatments in patients outside clinical trials in the real world

Methods for Increasing SEER Relevance

Expanding the data and methods for data capture through

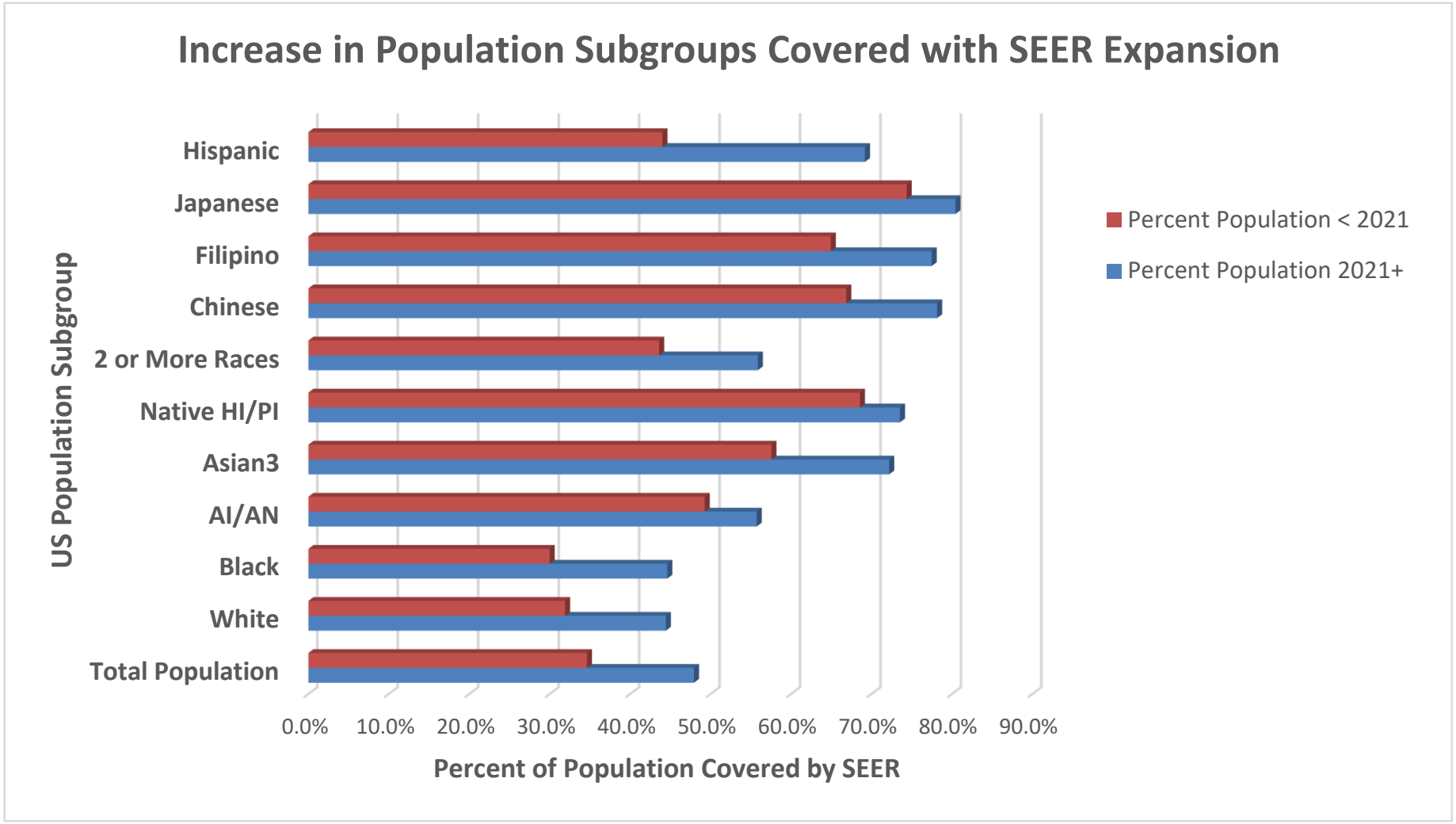
- ***linkages*** with external partners holding key clinical data for understanding cancer care/outcomes
- developing ***automated methods*** for data capture through deep learning and NLP

Expanding the breadth of patients covered in SEER

Reasons for SEER Expansion 2021

In order to represent real world data at the population level

- Increase data coverage to:
  - enable reporting of trends in more refined, clinical categories such as
    - histologic subtype
    - biomarkers status
    - treatment categories
  - and by important population subgroups
- As of June 1, 2021 SEER now covers ~50% of the US population
  - Represents >850,000 incident cancers reported annually



Percent Increase for US Population Subgroups with SEER Expansion May 2021										
Total Populatoin	White	Black	AI/AN	Asian3	Native HI/PI	2 or More Races	Chinese	Filipino	Japanese	Hispanic
13.3%	12.5%	14.6%	6.5%	14.6%	5.0%	12.3%	11.3%	12.5%	6.1%	25.2%

SEER Expansion

Two categories of Registries added to SEER

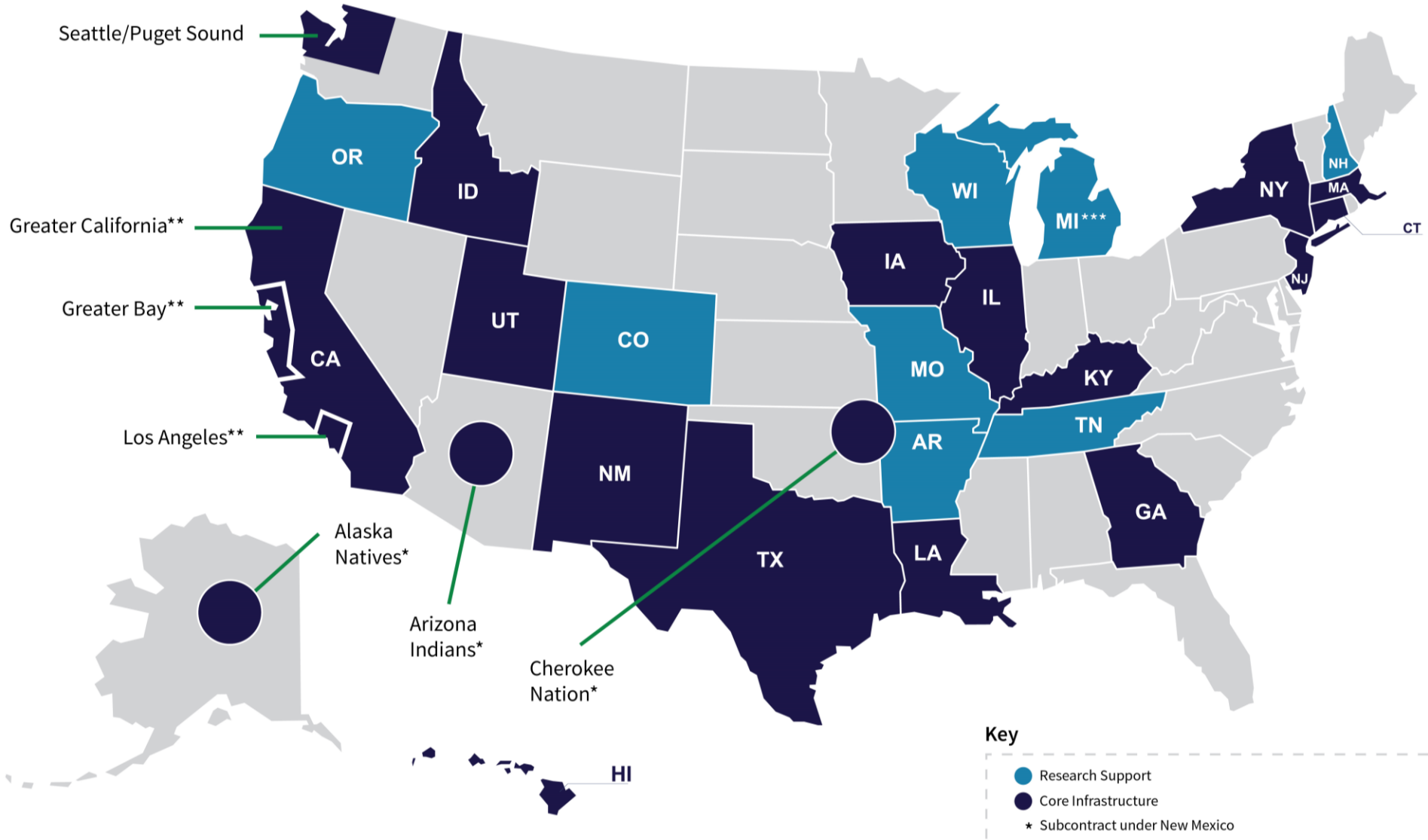
- ***Core Registries-*** these submit data annually to the SEER program and data accessible to the research community (~50% of US Population)

- ***Research Support Registries*** – 9 additional Registries eligible to participate in specific activities such as
  - Virtual Pooled Registry (VPR)
  - Virtual SEER-Linked Biorepository
  - National Childhood Cancer Registry (NCCR)
  - Special Research Projects and Pilots
  - Participation in linkages with external data partners (Pharmacies, genomic laboratories, claims data etc.)

Map of SEER Program June 1, 2021

Dark Blue represents ***Core Registries*** (Reporting data)

Light Blue represents ***Research Support Registries-*** participate in special projects



SEER Data Enhancements

SEER Traditional and Enhanced (New) Data

Data traditionally collected

- Demographics
- Geospatial data
- Characterization of the tumor at diagnosis
  - Stage
  - Tumor characteristics (including 32 biomarkers)
- Treatment (first course)
  - Surgery, radiation, Chemo/hormonal (y/n)
- Survival and Cause of death

New Data being integrated into SEER

- Detailed & longitudinal Treatment
  - infusion and Oral (Chemo/Hormone/Immune)
  - Radiation
  - Both Initial & subsequent Rx courses
- Expanded tumor characteristics
  - Genomics/Biomarkers (EGFR, ALK, BRAF MSI etc.)
  - Multigene panels for specific cancers (Oncotype, Decipher, Castle LS)
- Capturing Metastatic Recurrence
  - Leveraging multiple data sources (pathology, radiology, claims, hospital reports)
- Longitudinal Residential History

Examples of SEER-Linked* Pharmacy Data (2013-2020)

Sample of Tyrosine Kinase Inhibitor Use - 2013-2020

Drug Name	Patients	Filled Prescriptions
TARCEVA	2,129	17,423
SPRYCEL	1,934	27,729
IMATINIB MESYLATE	1,929	31,326
		8,035
		20,208
		4,580
		8,718
TAGRISSO	1,247	13,936
SUTENT	1,235	8,189
TASIGNA	1,020	15,830
CABOMETYX	791	4,939
INLYTA	744	4,884
LENVIMA	544	2,569
TYKERB	490	2,496
XALKORI	488	4,020

TKIs- 34 agents
188,000 Fills for >20,000 Patients

PARP Inhibitor** Use in 1,095 Patinets by Cancer Site from SEER-Linked Pharmacy Data (2017-2020)

Cancer Site	N Patients
Ovary	504
Breast	229
Other Female Genital Organs	132
Prostate	58
Peritoneum, Omentum and Mesentery	39
	31
	23
	13
Melanoma of the Skin	13
Thyroid	12
Brain	8
Urinary Bladder	7
Lung and Bronchus	6
Acute Myeloid Leukemia	4
Cervix Uteri	4
Other Biliary	4
Other cancer sites*	50

PARP Inhibitors 3 agents
7,000 Fills for 1,095 Pts.

* Sites with < 4 patients receiving agents

** Olaparib (approved 2014), rucaparib (approved Oct 2018), talazaporib (approved Oct 2018)

Use of a CDK 4/6 inhibitor (Palbociclib) in 4,302 Patients by Cancer Site (2013-2020)

Cancer Site	N Patients
Breast	4100
Corpus Uteri	28
Melanoma of the Skin	25
Thyroid	23
Retroperitoneum	22
	18
	17
	16
	13
Other Site	10
Kidney and Renal Pelvis	8
Non-Hodgkin Lymphoma	8
Ovary	7
Prostate	7

CDK 4/6 Inhibitors
Palbociclib
45,000 Fills for 4,302 Pts.

- Includes 11 of 20 SEER Registries- Pharmacy data from CVS/Walgreens/

Automation in SEER: DOE-NCI Collaboration

- API for auto-extraction of key elements (site, histology, behavior, laterality, grade) developed and under refinement
  - Implementing to optimize work flow in 6 SEER registries
  - Automated process **18,000X** faster than manual extraction
  - Full automated extraction of 20% of **all** path reports + 50% partial auto-extraction
  - Initial test in GA registry- automated 24,000 report backlog in <2 minutes!
  - Developing Privacy Preserving version for sharing with CDC, VA and others.
- ***Ultimate Goal: support near real time incidence reporting***

- API for ***reportability*** of cancer pathology reports developed
  - Selects cancer related path reports for case finding
  - modifying for radiology reports (to capture missed cases- CNS tumors and HCC)

- API for ***recurrent metastatic disease*** developed for pathology reports
  - modifying for radiology reports

SEER Data Access

- SEER data collected by registries under state public health reporting authority
  - States require all health care providers to report
  - Reporting is HIPAA exempt
  - SEER Receives only a limited data set for dissemination

- SEER traditionally broadly available (>4,500 downloads per year)
  - Increasingly detailed data increases risk of re-identifiability of individuals

- Therefore a new Multi-tiered Authentication and Authorization Process is in development with increasing requirements by tier
  - Tier 1- De-identified – no dates or geographic variables- available to all with minimal Data Use Agreement (DUA) (live)
  - Tier 2 – Limited Data set with minimal detailed characterization variables (live)
  - Tier 3 – Limited Data set with special variables (biomarkers, multi-gene panels etc.) (live)
  - Tier 4 – Limited Dataset with longitudinal treatment, dates etc., (May require IRB review) (in dev}

Other Significant Activities

Virtual Pooled Registry (VPR)

- Centralized process for linking cohorts/other studies to ***all registries*** in US to
  - Capture information on cancers and outcomes (Survival)
  - Currently 45 US registries are participating (87% US pop)
- Central website for data submission and linkages
- Templated IRB and DUA forms
  - Accepted by ~50% of participating registries to date
- cIRB contract for SRP May 2021
  - Enables central IRB submission and approval through reliance agreements with registries
  - Available to support other data access processes in SRP for SEER data (Tier 4 requests)

The National Childhood Cancer Registry (NCCR): a Component of the CCDI Ecosystem

Centralized data system from 23 registries across the US (77% of all childhood cancers to date)

Base Data

- Complete abstracts from participating state registries (1995+)
- National Death Index (NDI)/State vital records
- Lexis Nexis - Residential History
- Virtual Pooled Registry (VPR) linkage – de-duplication/subsequent cancers

Planned Linkages

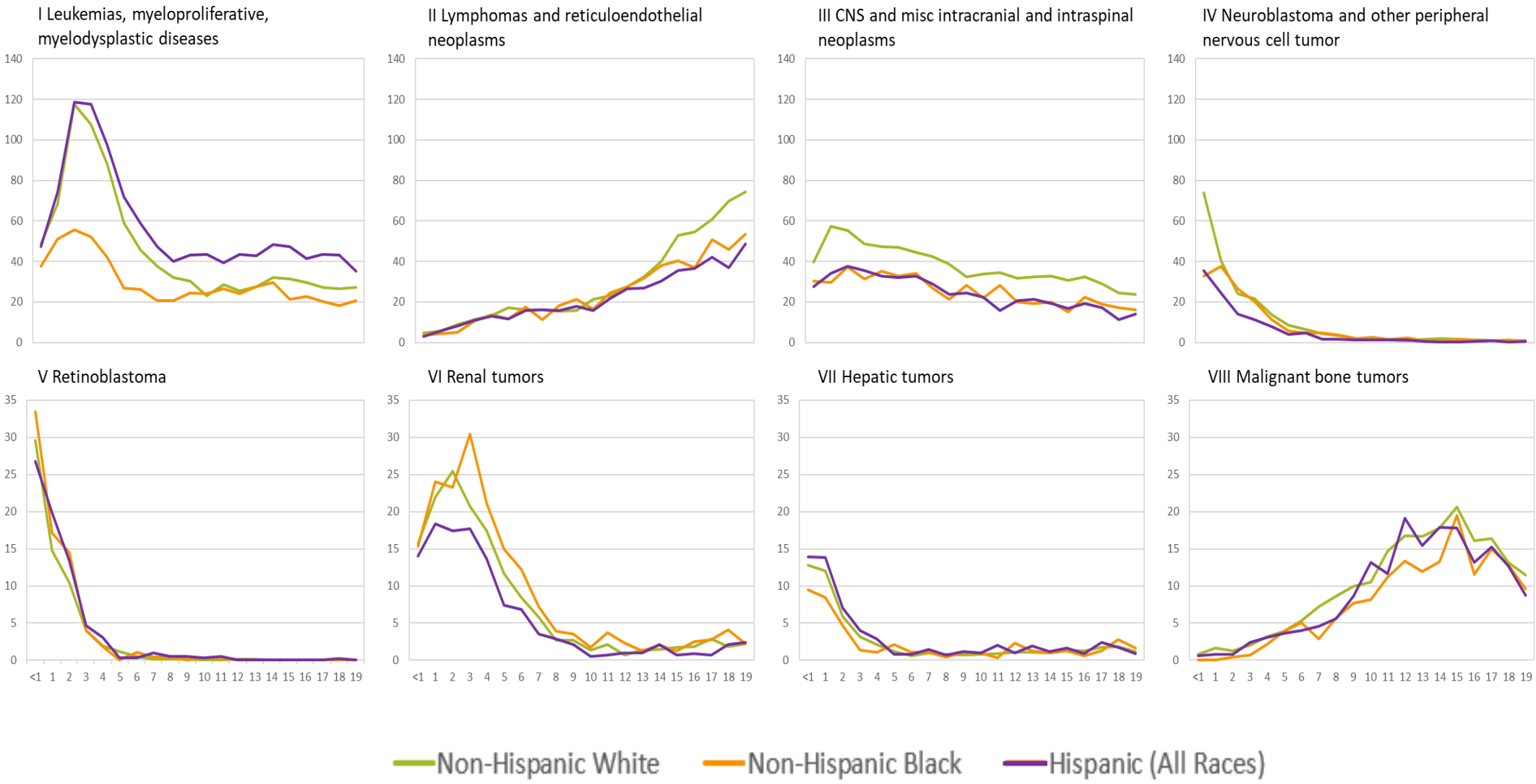
- Pharmacy Data (Oral Therapies)
- Longitudinal Radiation oncology data
- Claims data (Comorbidity and treatment)
- Radiology reports + images (case finding/ recurrence)
- Genomic Data
- COG Clinical Trials and Project Every Child DB – evaluate coverage and under representation for COG

Data Access

- Creation of Data platform and data products to enable controlled access by investigators

*Data Products- Example from PEDS*Explorer*

Age-specific rates by AGE for MAJOR CANCER SITES AND BY RACE (2008-2017)



Thank you