The National Childhood Cancer Registry: Challenges and Solutions

Lynne Penberthy MD, MPH

August 13, 2021

NIH > **NATIONAL CANCER INSTITUTE**

Presentation Objectives

- Briefly Describe the National Childhood Cancer Registry –purpose and goals

- Illustrate examples of specific challenges related to childhood cancer as a rare disease – data access and privacy

- Describe methods and considerations used as the NCCR is being developed

NCCR Purpose

- Leverage and link disparate data from multiple sources to create an infrastructure that can better support research on childhood cancer

- Core data derived from cancer registries- but extended and expanded to include additional relevant information such as

  - Detailed treatment

  - Genomic characterization

  - Trajectory of care from diagnosis throughout life including

    - Multiple primary cancers

    - Recurrent disease

  - Other relevant factors related to risk and outcome (residential history, SDOH etc.)

- Integrate within the CCDI federated data ecosystem
(https://www.cancer.gov/research/areas/childhood/childhood-cancer-data-initiative)

National Childhood Cancer Registry:

- Leverages existing data sources to capture all pediatric and young adult cancers in the US

- Accumulate data through linkages with Cancer Registries who are:
  - *population based* (capture all cancers within a defined geographic area)
  - *maintain PII* with ability to incorporate data on all childhood cancer cases
  - *HIPAA Exempt*
  - State regulations require *ALL* health care providers to report to the state registry

Initial Registry Participation (77% of US childhood cancers from 23 states)

7 NPCR Registries

- Florida
  - Illinois
  - New Jersey
  - Ohio
  - Pennsylvania
  - Tennessee
  - Texas

- 2020- submission of de-identified NAACCR data in NCCR
- 2021 – full registry submission with PII to DMS*Lite as a repository for linkages and submission to the NCCR
- Will support linkages for residential history (Lexis Nexis) and VPR to identify second

SEER Registries

- GA
- Los Angeles
- Greater CA
- Greater Bay area
- Iowa
- CT
- KY
- LA
- Seattle
- ID
- NY
- MA
- SEER will contractually require submission to the NCCR
- Will support linkages for residential history (Lexis Nexis) and VPR to identify second primaries

Goal is to achieve 100% coverage of all pediatric patients over the next few years

The National Childhood Cancer Registry Components

- Routine linkages will be performed centrally via and Honest Broker with external data sources including:
  - Complete abstracts plus text documentation for each case
    - 1995-2019+
    - Text documentation permits NLP/AI – key treatment information
  - **National Death Index** (NDI)
  - **State vital records**
  - **Lexis Nexis** (linkage to be performed centrally not by state)
    - Residential History (routinely biennially) – essential to perform longitudinal linkages
    - Financial Toxicity – provide data to understand the impact of cancer on patients and families
  - **Virtual Pooled Registry (VPR)**
    - Supports linkage across all cancer registries
    - Capture subsequent Cancers (Annual linkage with ALL registry in the US)

The National Childhood Cancer Registry Components: *Planned Central Linkages*

- ***Pharmacy Data***– CVS/Walgreens/Riteaid (Real Time)/ PBM United Health Care
- ***Longitudinal Radiation oncology data*** acquisition
- ***Claims data*** linkages (Treatment/Comorbidity)
    - United Health Care (linkage in process)
    - Medicaid (2021)
- ***Radiology reports + images*** (case finding/ recurrence)
    - Ambra Health (radiology data exchange platform with significant pediatric facility penetration)
    - AIM e path reporting
    - Selected Cancer Centers data
- ***Genomic Data***
    - In discussions with FMI, Caris
    - Individual biomarkers available from pathology report
- ***Birth Records***
    - Capturing parental addresses at birth (residential history supplements
    - Identify issues at birth (APGAR etc.)

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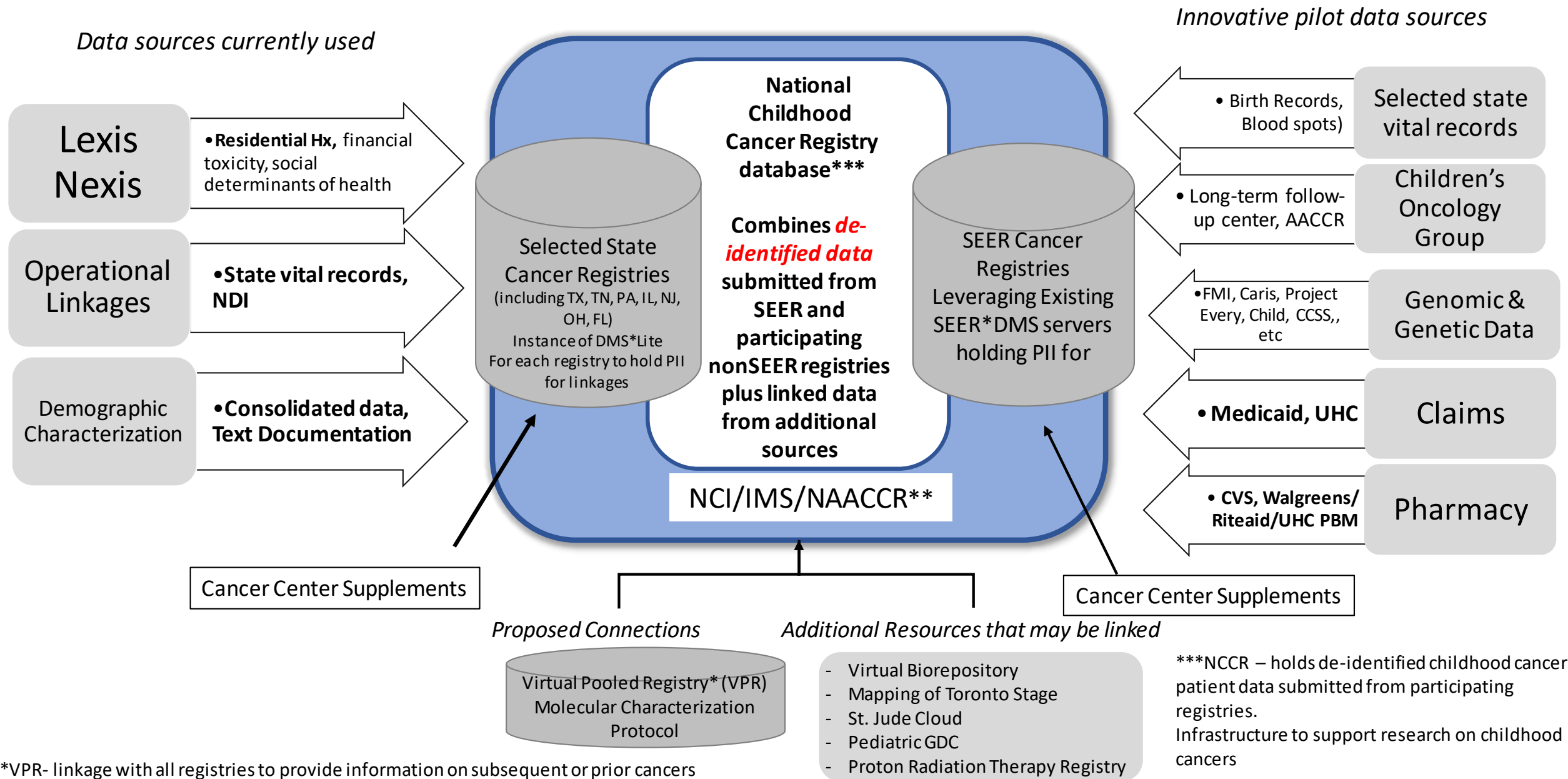
Use of a CDK 4/6 inhibitor (Palbociclib) in 4,302 Patients by Cancer Site (2013-2020)

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/

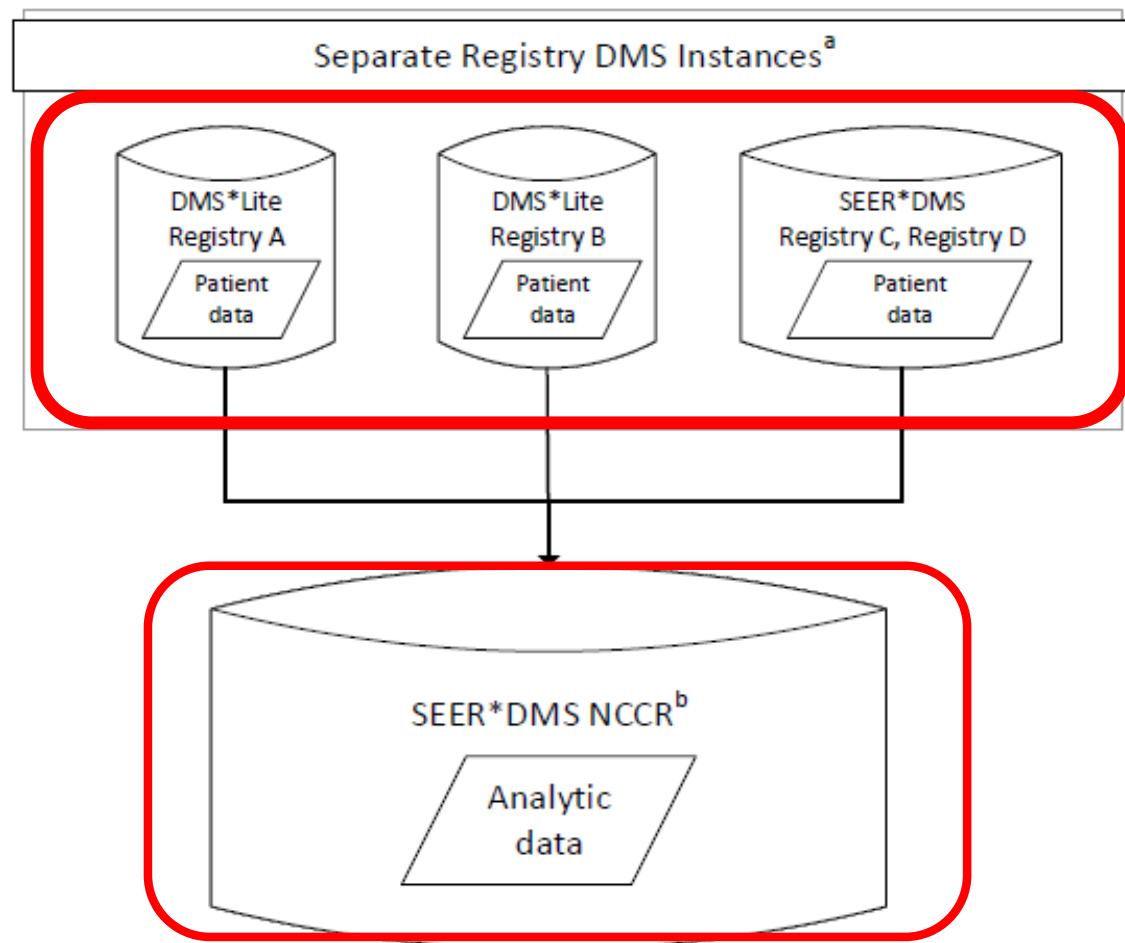# NCCR Workflow Process and Data Platform

Conceptual Framework: National Childhood Cancer Registry

Data sources currently used

Innovative pilot data sources

Lexis Nexis

•**Residential Hx,** financial toxicity, social determinants of health

Operational Linkages

•**State vital records, NDI**

Demographic Characterization

•**Consolidated data, Text Documentation**

Selected State Cancer Registries
(including TX, TN, PA, IL, NJ, OH, FL)
Instance of DMS*Lite
For each registry to hold PII for linkages

National Childhood Cancer Registry database***

Combines *de-identified data* submitted from SEER and participating nonSEER registries plus linked data from additional sources

NCI/IMS/NAACCR**

SEER Cancer Registries Leveraging Existing SEER*DMS servers holding PII for

• Birth Records, Blood spots)

Selected state vital records

• Long-term follow-up center, AACCR

Children's Oncology Group

•FMI, Caris, Project Every, Child, CCSS,, etc

Genomic & Genetic Data

• **Medicaid, UHC**

Claims

• **CVS, Walgreens/ Riteaid/UHC PBM**

Pharmacy

Cancer Center Supplements

Cancer Center Supplements

Proposed Connections

Additional Resources that may be linked

Virtual Pooled Registry* (VPR)
Molecular Characterization Protocol

- Virtual Biorepository
- Mapping of Toronto Stage
- St. Jude Cloud
- Pediatric GDC
- Proton Radiation Therapy Registry

***NCCR – holds de-identified childhood cancer patient data submitted from participating registries.
Infrastructure to support research on childhood cancers

*VPR- linkage with all registries to provide information on subsequent or prior cancers
** NAACCR is the coordinating center – does not hold or access data.

Flow Diagram for SEER and NPCR data submission to NCCR

NCCR Data Platform

- The NCCR will require a specialized data platform over which the data products and data access process can be overlain.
  - Needs to support:
    - cohort discovery,
    - simple linkages,
    - protect privacy and
    - provide a governance structure for accessing various components of the data and data products
- A Request For Information (RFI) for a commercial off-the-shelf data platform was issued
  - Purpose: identify potential applicants, and generate ideas regarding the data platform
- Currently developing the RFP for release in the next 3 months

Data Access and Release Process

NATIONAL CANCER INSTITUTE

NCCR Data Release: Underlying goals

SEER Data Acquisition	SEER Data Evaluation	SEER Data Release

- Policies and processes for sharing cancer surveillance data with researchers

  - part of NCCR/SEER/NCI mission

  - within the NIH data sharing policy framework

  - protecting patient confidentiality/privacy

  - Reducing risk of re-identifiability

  - minimizing risks from inappropriate use of data

    - including analytically inappropriate use

Special Considerations for the NCCR

High risk of re-identifiability because

- Rare Tumors with increased breadth and depth of data on each patient
- Computational capability available to individuals
- Open websites for possible linkage (e.g. Go Fund Me) & re-identification

Solutions

- Tiered system for data release – with potential for IRB review
  - Central IRB now available for SEER/NCCR
- Data release system linked to central authentication and authorization process (eraCommons)
- External consultant assessing risk of re-identifiability and advising steps for risk mitigation
- Data Use Agreement

Data Access & Release Goals

Internal access goals:

- Sharing of data between central cancer registries
  - De-duplication
  - Identification of Multiple Primacy
- Quality control
- Central registry access to and use of linked data

External access goals:

- Applicability of Common Rule and Public Health Reporting for Surveillance
- Enable multiple modes of data access to maximize patient privacy while promoting data utilization and research
- Establish user authentication and authorization system
- Identify access and release restrictions for linked data sources
- Develop criteria to evaluate "*fitness for use*" of linked data

SEER/NCCR Data Access

- To meet the goals - 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)
    - Requires internal review of brief proposals for release
  - Tier 4 – Limited Dataset with longitudinal treatment, dates etc., (May require IRB review) (in dev}
  - Each tier has a Data Use agreement targeted to the level of the data released

Developing Data Products

- Incremental and tiered system with 5 levels of data included
  - Counts and Indexing
  - Ready statistics/Interactive Tools (PEDS*Explorer) *https://seer.cancer.gov/statistics/nccr/*
  - "Canned Analysis" (SEER Stat) increased flexibility but with limitations
  - Cloud Analysis- SAS/SPSS etc. Not downloadable
  - Downloading data – likely requiring IRB approval

NCCR Static Report Example of Initial Data Product: Figure 2.2: Incidence Rates* by Cancer Site, 2008-2017#
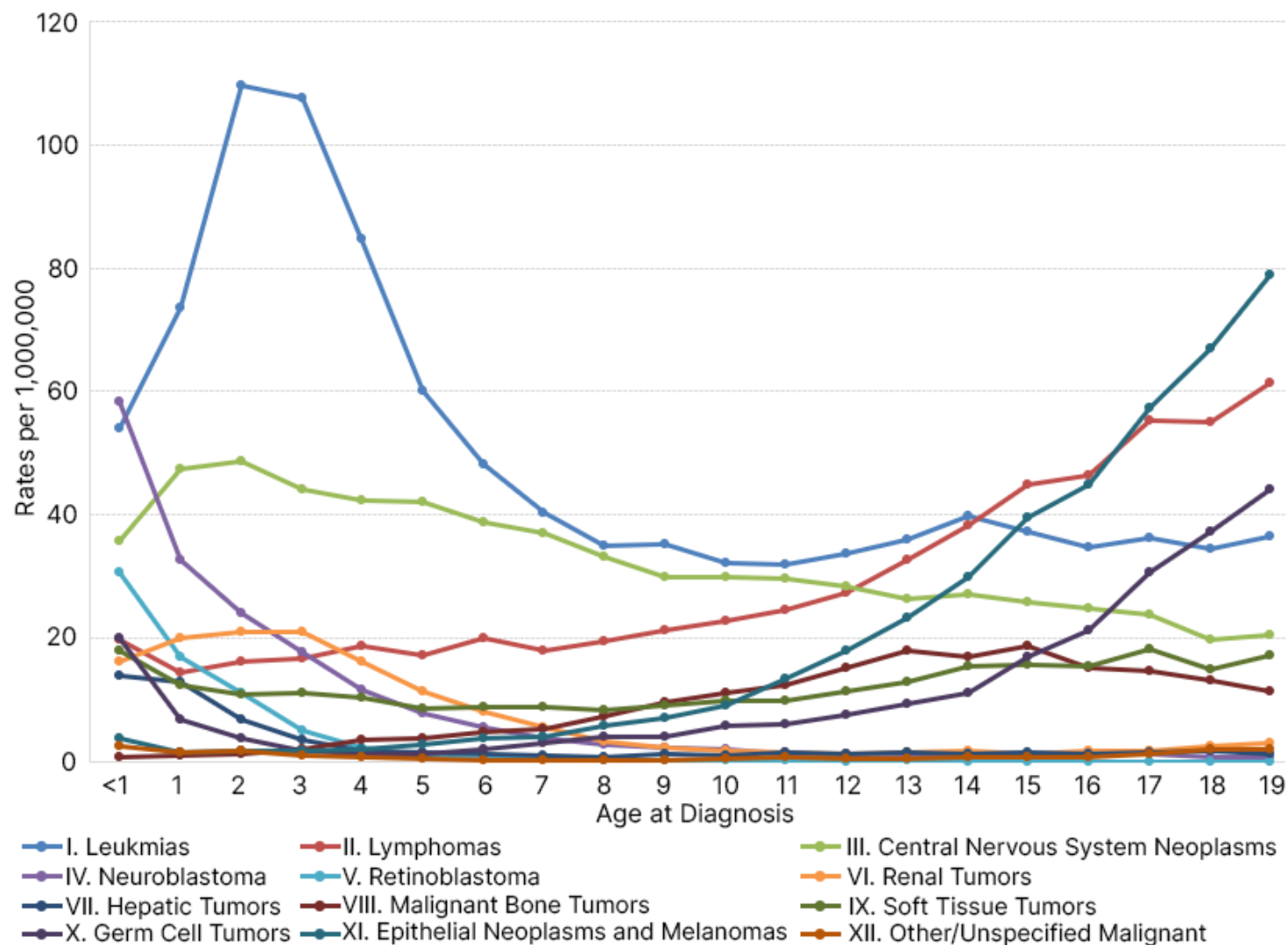
Other Considerations for registries

Special Considerations related to State Health Departments Reporting

1. Because each state is slightly different it has been essential for the registries to be engaged with the state legislature
2. Often at the state DOH or at a university.
3. Common practice to work with the state legislature annually to create or modify reporting requirements
4. Variation from state to state and even from year to year within a state

Privacy Preserving Patient Linkages

- Typically use PII hashing and tokenization

- Many companies who have this capability

- NCI has performed a landscape analysis (27 companies reviewed)
  - Report available

- 4 companies under evaluation for formal assessment of accuracy, etc. against a set of gold standard manually validated datasets

- In summary:
  - Linking using PII always optimal
  - Linkage results vary depending on completeness, quality and type of PII available in both datasets being linked
  - Customized variations can improve the accuracy of linkage using P3RL products

Thank you