

The National Childhood Cancer Registry (NCCR): Leveraging Data Standards for Improving Interoperability

*Wayne Liang, Coy Austin Fitts,
Nikki Wood, David Noyd*

Today's Speakers



Wayne Liang, M.D., M.S.

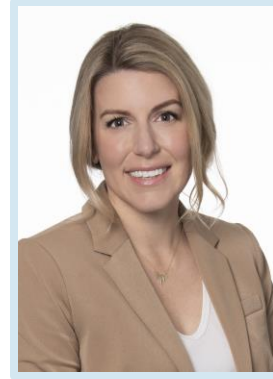
Pediatric Hematology/
Oncology Physician and
Clinical Informaticist at the
Aflac Cancer and Blood
Disorders Center of
Children's Healthcare of
Atlanta

Assistant Professor of
Pediatrics at Emory
University



**Austin Fitts,
Pharm.D.**

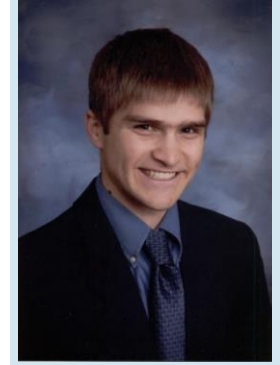
Postdoctoral Fellow,
Surveillance
Research Program at
NCI



Nikki Wood, D.O.

Pediatric Hematology/
Oncology Physician and
Clinical Informaticist at
Children's Mercy Kansas
City

Assistant Professor of
Pediatrics at the
University of Missouri,
Kansas City



David Noyd, M.D., M.P.H.

Pediatric Hematology/
Oncology Physician and
Clinical Informaticist at
Seattle Children's Hospital

Assistant Professor of
Pediatrics at the University
of Washington

Disclosures

- Wayne Liang, Nikki Wood, and David Noyd are Westat consultants to the NCI Surveillance Research Program and Clinical Consultants for Oncology Research and Evaluation Program.
- Wayne Liang is also a consultant and advisor for ChemoMap and Genoplex.ai.

Agenda

1. NCCR Data Platform and Pediatric Data Standards
2. NCCR Data Harmonization Process
3. Implementing Treatment Standards: HemOnc.org
4. Minimal Common Oncology Data Elements (mCODE)

NCCR Data Platform and Pediatric Data Standards

Wayne H. Liang, M.D., M.S.

Central Cancer Registries



NCI SEER (Surveillance, Epidemiology, and End Results):

- 50% of U.S. population
- More detailed data



CDC NPCR (National Program of Cancer Registries):

- 96% of U.S. population
- Less detailed data

Characteristics

- **State-based**
- **Federally mandated** reporting
- **Manually abstracted** from treatment centers
- Generates **national cancer statistics**
- **Lacks comprehensive data** required to understand the full patient journey

NCCR Data Sources and Domains

Data Sources

- Central Registries
- **NCI-Designated Cancer Centers**
- Children's Oncology Group (COG)
- Pediatric Proton/Photon Consortium Registry (PPCR)
- Virtual Pooled Registry (VPR)
- Administrative data sources
- Clinical trials, survivorship studies, etc.

Data Domains

- Longitudinal treatment, comorbid conditions, procedures, outcomes, etc.
- Social determinants of health
- Radiation oncology treatment
- Clinical genomic reports



Source: Canva

NCCR Data Products

NCCR*Explorer

- Open access
- 25 NCCR registries:
 - 70% of U.S. population
 - 1,700,440 reported cancer cases
- Sources: SEER NCCR; NAACCR CiNA submissions

nccrexplorer.ccdi.cancer.gov/application.html

NCCR Data in SEER*Stat

- Registered access
- User queries
- 15 NCCR registries:
 - 49% of U.S. population
 - 973,935 reported cancer cases
- Sources: NAACCR CiNA submissions; transition to NCCR submissions

seer.cancer.gov/seerstat/

datacatalog.ccdi.cancer.gov/data-set/CCDI-NCCR%20SEER*Stat

Coming soon!

NCCR Data Platform

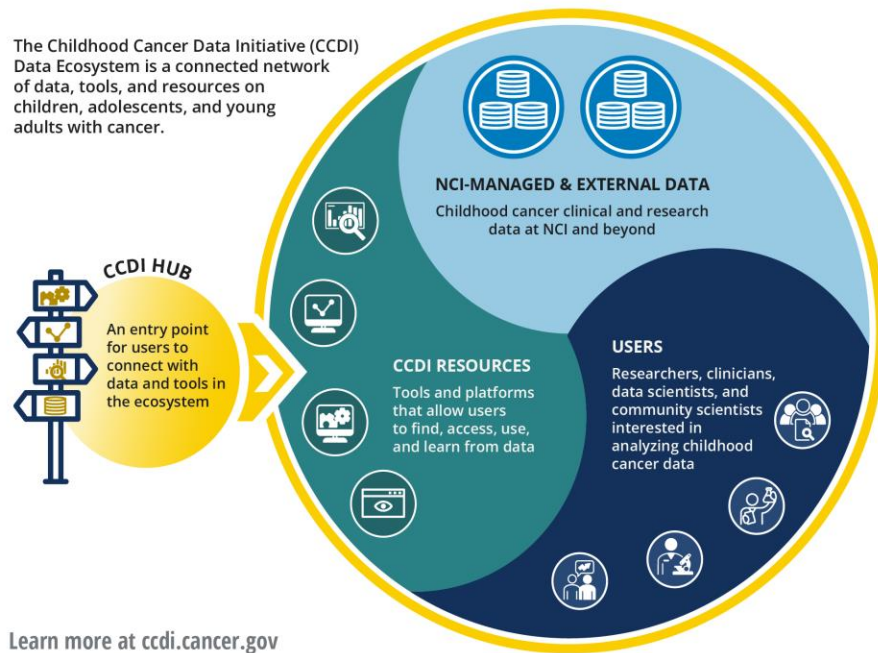
- Registered and controlled access
- User queries
- Cloud-based or downloaded custom data sets (approved, individual-level)
- 18 SEER core and 4 SEER RSR registries:
 - 57% of U.S. population
 - 1,543,269 reported cancer cases
- Sources: Multiple linked and harmonized data sources (central registries, NCCR submissions, other)

*Dx @ ages 0-19 for some registries

Data Standards: Cross-Domain Collaboration

- **Data standards** are critical for **interoperability** across the CCDI Data Ecosystem
- **CCDI data standards** are being developed in collaboration with the:
 - NCI Semantics Infrastructure Team
 - Pediatric Cancer Data Commons (University of Chicago)
 - CCDI Molecular Characterization Initiative Disease Committee
 - HemOnc.org
 - mCODE

NATIONAL CANCER INSTITUTE THE CCDI DATA ECOSYSTEM



NCI's Semantics Infrastructure: Promote Interoperability

NCI Thesaurus (NCIt):

- Controlled vocabulary of biomedical concepts

Cancer Data Standards Registry and Repository (caDSR):

- Structured repository for clinical and research metadata

Common Data Elements (CDEs):

- Standardized metadata elements for clinical and research use
- Incorporates terms from NCIt
- Managed within caDSR



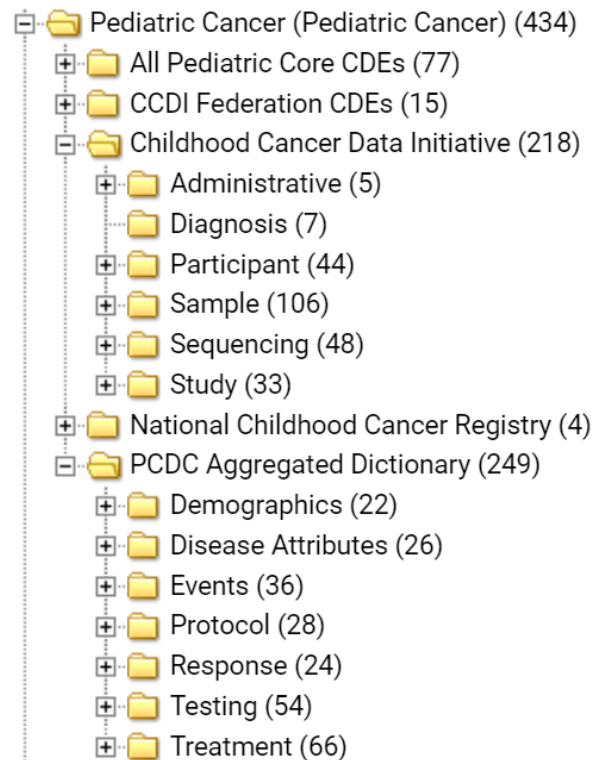
Source: Shutterstock

Pediatric Cancer CDEs

- 434 CDEs (as of July 2024)
- Multiple clinical and research domains:
 - Administrative, Demographics, Diagnosis, Treatment, Specimen, Molecular, and Imaging, etc.
- Linked to reference data standards or cross-walked to other CDEs

Reference Data Standards

HL7 FHIR mCODE, HemOnc.org, OMOP, PRISMM, OncoTree, UBERON, OBO, GA4GH, HGVS, PCORnet, HPO, SNOMED CT, LOINC, ASTRO, etc.



cadsr.cancer.gov/onedata/Home.jsp

NCCR Data Harmonization Process

C. Austin Fitts, Pharm.D.

NCCR Cancer Center Data Harmonization Clinical Team



Austin Fitts, Pharm.D.
National Cancer Institute



Elaine Fan, M.D.
University of Kansas,
Wichita



David Noyd, M.D., M.P.H.
University of Washington
and Seattle Children's
Hospital



Nikki Wood, D.O.
Children's Mercy Kansas
City and University of
Missouri, Kansas City



Wayne Liang, M.D., M.S.
Children's Healthcare of
Atlanta and Emory
University



Jeremy Warner, M.D., M.S.
Rhode Island Hospital and
Brown University

HemOnc.org Editorial Board
Pediatric Team



NCCR Cancer Center
Data Harmonization
(2021)

* **CC-DIRECT** (Childhood Cancer—
Data Integration for Research,
Education, Care, and Clinical
Trials)

Standards Task Team

Methods for Harmonization of Complex Data Elements

Patient data submitted by cancer centers



Identify standardized data elements



Validation of interoperability and accuracy



Transform data to common standards



Release harmonized data set to researchers

Harmonization Example: Transplant Type

■ Cancer Center Permissible Values

- Matched Unrelated
- Autologous
- Matched Related
- Unknown
- Syngeneic
- Haploidentical Related
- Other

Donor Relationship Type (13363450)

▲ 1 Permissible Value
Biological Parent
Biological Relative
Biologically Unrelated
Not Reported
Unknown

Transplant Protocol Type (13376562)

▲ 1 Permissible Value
Allogeneic
Autologous
Not Reported
Unknown

HLA Status (13376564)

▲ 1 Permissible Value
Match
Non-Match
Not Reported
Unknown

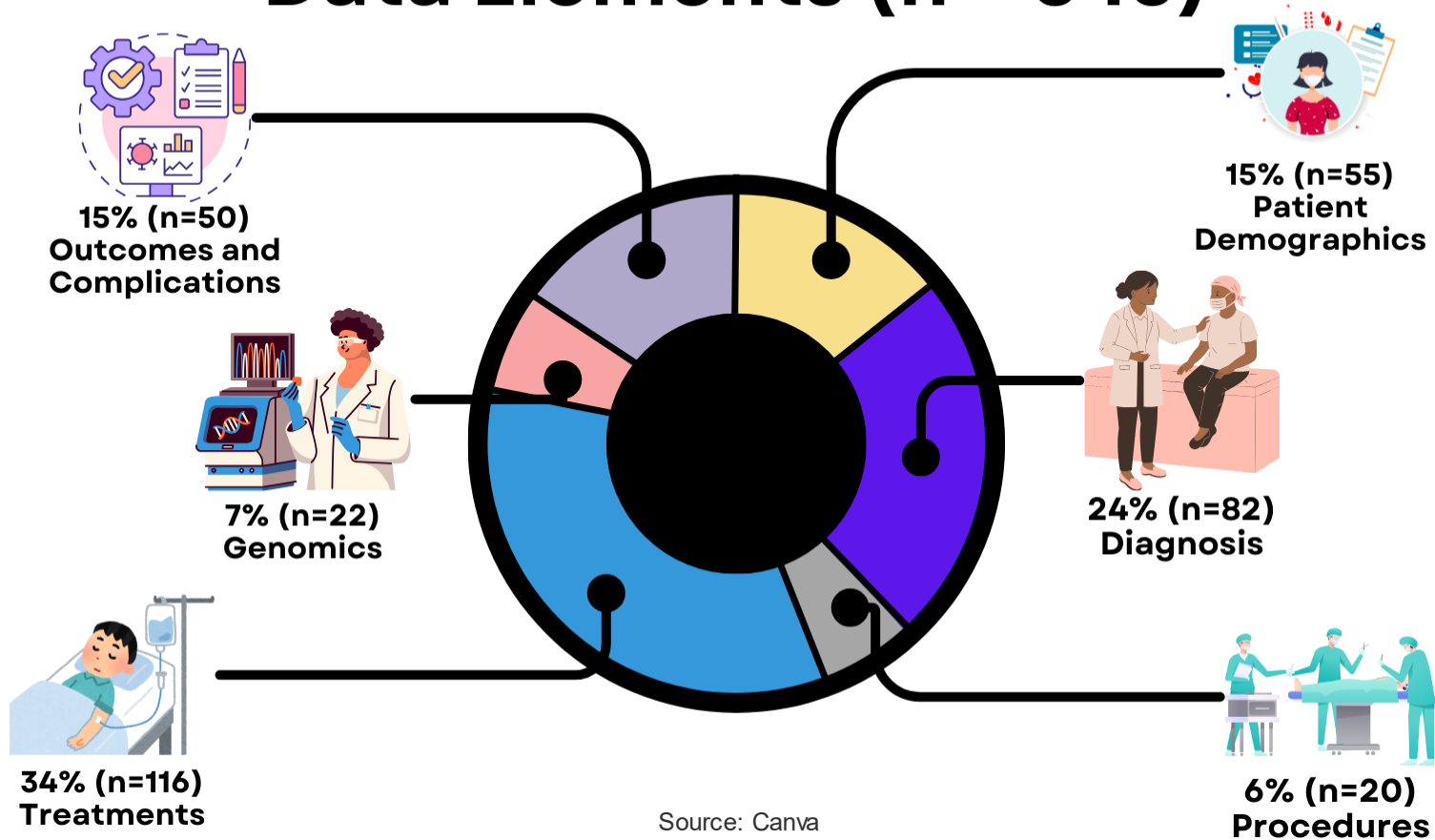
Harmonization Example: Transplant Type

Cancer Center Permissible Values	Donor Relationship Type (13383450)	Transplant Protocol Type (13376562)	HLA Status (13376564)
Autologous	Not Applicable	Autologous (C16039)	Not Applicable
Matched Unrelated	Biologically Unrelated (C130053)	Allogenic (C46089)	Match (C129972)
Haploidentical related	Biological Relative (C71384)	Allogenic (C46089)	Non-Match (C126298)
Syngeneic	Biological Sibling (C100809)	Allogenic (C46089)	Match (C129972)

Cancer Center Supplements: Data Harmonization

- 12 Cancer Centers
 - Submitted 6,500 columns
 - Representing over 25,000 patients
 - Across 23 million rows of patient data
- General Stats
 - Over **3.5 billion** non-null **data points** to harmonize
 - Average of **77,000** data points submitted **per patient**
 - Average **data completeness** was **96%** (% of non-null data)

Breakdown of Harmonized Data Elements (n = 345)



Examples of Cancer Center Data Elements Augmenting

Patient Demographics	Diagnosis	Treatments	Outcomes
Insurance	Performance Status	Chemotherapy Received	Lab Results
Education Level	Morphology	Radiation Received	Complications and Adverse Events
Past Medical History	Tumor Genomics	Stem Cell Transplants	Hospital Admissions



Cloud-Native Data Sharing through the NCCR Data Platform

Authentication and Authorization

- Federated identities for authentication
- Approve users using SEER Data Access Request process
- Continually validate appropriate data use and manage authorization

Data Browser

- Learn about data sources linked to registry data using PII or PPRL

Cohort Discovery

- Search for a defined cohort of patients and view aggregate statistics and counts

Governance

- Request access to individual-level data and support reviews of requests
- Create custom data sets for a user based on the data access request

Data Release and Analysis

- Release custom data sets to researchers for analysis with statistical software programs like R Studio, Jupyter Notebooks, SAS

CCDI Data Ecosystem

- Enable data sharing with other CCDI data resources (e.g., CRDC) at NCI and other institutions

Implementing Treatment Standards: HemOnc.org

Nikki Wood, D.O.

Comparing Cancer Treatment Using Real World Data



Challenges

Central registry data lacks detailed treatment data

Treatment as Prescribed vs Treatment Received—Dose modifications, missed doses, missing data



Needs

Medication administration record and prescription filling data

Match drugs/treatments to cancer regimens

A standardized vocabulary of regimens

Challenge of Paper Chemotherapy Roadmaps

Name: [redacted] MR [redacted] DOB: [redacted]
 Initial ht: cm Initial wt: kg BSA: m² → 142cm, 43.5kg, 1.29 m²
 Diagnosis: Hemophagocytic Lymphohistiocytosis
 Modified treatment as per HLH-1994 Weekly; + Vinblastine from COG ANHL0131
 Primary Attending [redacted]

→ Antitoxin due to
 delay in use of
 drugs while
 awaiting review
 post 3/29/17

week	day	date due	date given	Chemotherapy	VP-16	Vinblastine	Studies	IVIg	Pentam
1	1	3/19/17	3/19/17	Dexamethasone 1 8mg BID (started 3/16/17 PM x1 dose)	165mg (75% dose) due to c/cet	5.6mg	1,2	22g	3/20/17
	4	3/22/17	3/22/17		157mg (75% dose)	5.6mg	1 (given on 3/24/17)		
2	1	3/26/17	3/26/17	Dexamethasone 8mg BID	165mg (75% dose)	5.6mg	1,2,3		
	4	3/29/17			160mg (75%)	5.6mg	1		
3	1	4/1/17		Dexamethasone 2 4mg BID (moved up)	160mg (75%)	5.6mg	1,2		Delayed (H)
4	1	4/7/17		Dexamethasone 2 4mg BID (moved up)	160mg (75% dose due to c/cet still)	5.6mg	1,2,3,4	Delayed	
5	1	4/14/17	4/20/17	Dexamethasone 3 2mg BID	160mg	5.6mg	1,2	22g	4/30/17
6		5/5/17	5/5/17	Dexamethasone 3 2mg BID	210mg (100%)	5.6mg	1,2,3		
7		5/12/17	5/12/17	Dexamethasone 4 1mg BID	210mg	5.6mg	1,2		
8		5/20/17		Dexamethasone taper	210mg	5.6mg	1,2,3,4		

Delayed due to HHT 4/28/17

Labs/Studies

- CBC, differential, LFTs, Coags, lytes
- Ferritin, Triglycerides, fibrinogen, LDH
- Soluble IL-2 receptors (CD25)
(Draw M-Th after 10am, needs to be in lab by 3pm)
- NK activity (if WBC low, discuss with team mL needed)
(Draw M-Th after 10am, needs to be in lab by 3pm)
- BMT referral at week 4 if no improvement

* CRIS disease treated with intrathecal chemo on (H) Week 1, Day 1 and week 2 Day 4

* Delayed after 4 HHT due to chronic perforation + surgical repair at outside hospital.

- Results from 3/29/17 LP was negative for malignant cells -> since HHT to EBV/HT.

- Chemotherapy regimens are complex
- Paper roadmaps are the mainstay for documentation
- Efforts to create machine-interpretable electronic roadmaps will facilitate standardized, discrete, and transferable chemotherapy regimen metadata

Wyatt KD, Noyd DH, Wood NM, et al. Data standards in pediatric oncology: Past, present, and future. *Pediatric Blood & Cancer*. 2023 Feb. doi: 10.1002/pbc.30128.

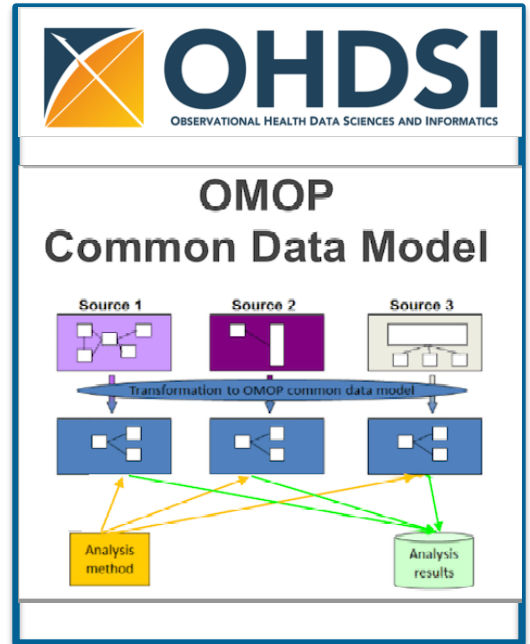
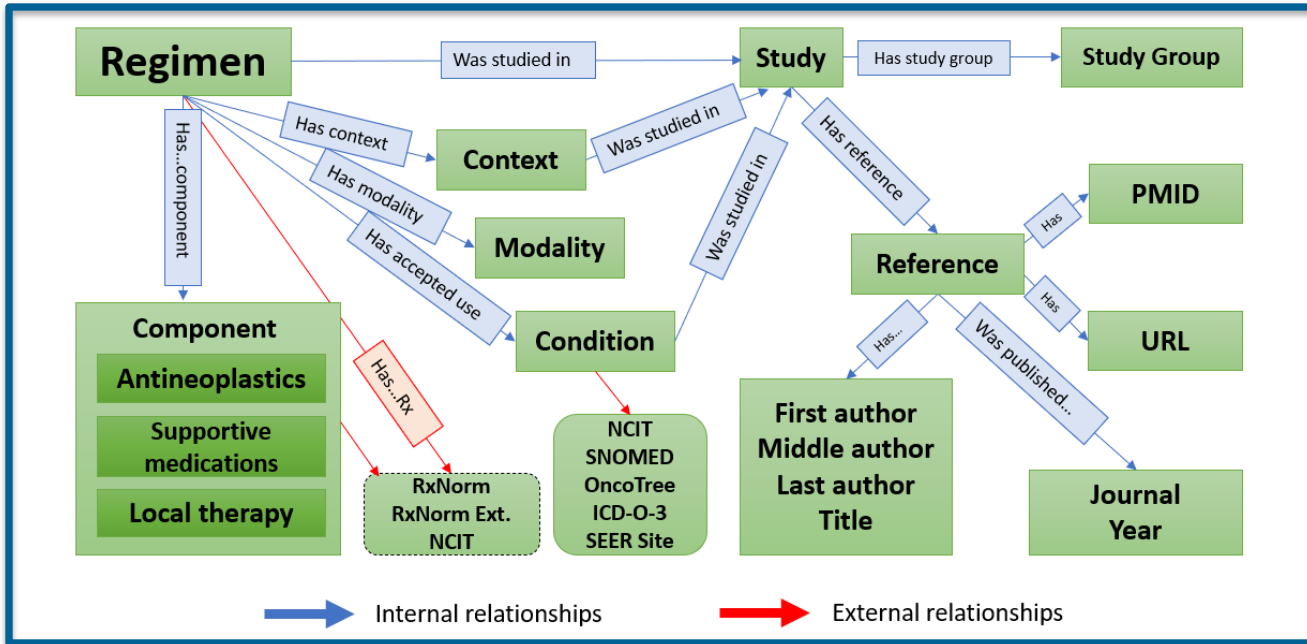
HemOnc.org Vocabulary of Cancer Regimens

- Largest freely available, computable knowledge resource of cancer regimens
- **Machine readable standard vocabulary** for chemotherapy regimen representation in the OMOP common data model
- Derived from HemOnc.org website
- **Free for non-commercial use**
- Available within **OHDSI ATHENA** vocabularies repository: athena.odhsi.org
- Available on **Harvard Dataverse** for download: dataverse.harvard.edu/dataverse/HemOnc

“An ontology is a FORMAL, EXPLICIT, specification of a SHARED conceptualization.”

– Rudi Studer (1998)

The HemOnc Regimen Model



Warner JL, Dymshyts D, Reich CG, et al. HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. *Journal of Biomedical Informatics*. 2019 Aug. doi: 10.1016/j.jbi.2019.103239.

HemOnc.org Example Regimen: COG AHOD0031

COG AHOD0031

Rapid Early Responders (Standard Arm)

ABVE-PC x 2 Cycles

ABVE-PC: [A](#)driamycin (Doxorubicin), [B](#)leomycin, [V](#)incristine, [E](#)toposide, [P](#)rednisone, [C](#)yclophosphamide

[back to top](#)

Study	Evidence
Friedman et al. 2014 (COG AHOD0031) #P	Non-randomized portion of phase 3 RCT

Chemotherapy

- Doxorubicin (Adriamycin) 25 mg/m² IV over 10 to 30 minutes once per day on days 1, 2
- Bleomycin (Blenoxane) 5 units/m² IV over 10 to 20 minutes or SQ once on day 1
- Bleomycin (Blenoxane) 10 units/m² IV over 10 to 20 minutes or SQ once on day 8
- Vincristine (Oncovin) 1.4 mg/m² (maximum dose of 2.8 mg) IV once per day on days 1, 8
- Etoposide (Vepesid) 125 mg/m² IV over 1 hour once per day on days 1, 2, 3
- Prednisone (Sterapred) 40 mg/m²/day PO divided BID or TID on days 1 to 7
- Cyclophosphamide (Cytosol) 800 mg/m² IV over 1 hour once on day 1

21-day cycle for 2 cycles

References

1. COG AHOD0031: Friedman DL, Chen L, Wolden S, Bulton A, McCarter K, Fitzgerald T.J, Kessel S, De Alarcon PA, Chen AR, Kohnirsky N, Ehrlich P, Hutchison RE, Corline LS, Schwartz CL, Children's Oncology Group. Dose-intensive response-based chemotherapy and radiation therapy for children and adolescents with newly diagnosed intermediate-risk Hodgkin lymphoma: a report from the Children's Oncology Group Study AHOD0031. J Clin Oncol. 2014 Nov 10;32(32):3651-6. Epub 2014 Oct 13. [link to original article#](#) **does not contain protocol** [link to PMC article#](#) [PubMed#](#) NCT00025255

Evaluate Response

ABVE-PC x 2 Cycles

ABVE-PC: [A](#)driamycin (Doxorubicin), [B](#)leomycin, [V](#)incristine, [E](#)toposide, [P](#)rednisone, [C](#)yclophosphamide

[back to top](#)

Study	Evidence
Friedman et al. 2014 (COG AHOD0031) #P	Non-randomized portion of phase 3 RCT

HemOnc.Org. Classical Hodgkin lymphoma, pediatric. Edited June 27, 2024.
https://hemonc.org/wiki/Classical_Hodgkin_lymphoma,_pediatric#COG_AHOD0031.

Uses of HemOnc.org Regimen Vocabulary



Disease → Common
Cancer Regimens

Identify **regimens** that a patient with a given cancer may have received



Cancer
Regimen → Components

Resolve missing or **conflicting data** when the regimen is known



Regimen Components →
Likely Regimens

Identify **possible regimens** that a patient received based on chemotherapy components



Real World
Data → Real World
Evidence

Enables **comparative analysis** of cancer treatments real patients received using registry data

Mapping of NCCR treatment data to HemOnc.org regimen vocabulary may augment the use of NCCR for comparative analysis of cancer treatments.

Minimal Common Oncology Data Elements (mCODE)

David H. Noyd, M.D., M.P.H.

Minimum Common Oncology Data Elements (mCODE)

A FHIR-based core set of common data elements for cancer that is standardized, computable, clinically applicable and available in every electronic health record for patients with a cancer diagnosis

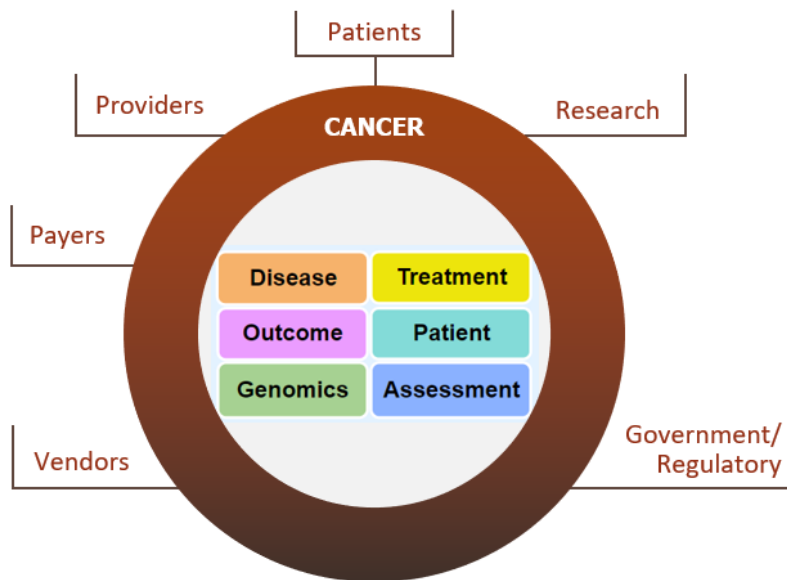
A **standard health record** for oncology

The **minimal set of data elements** applicable to all cancers, and collected for:

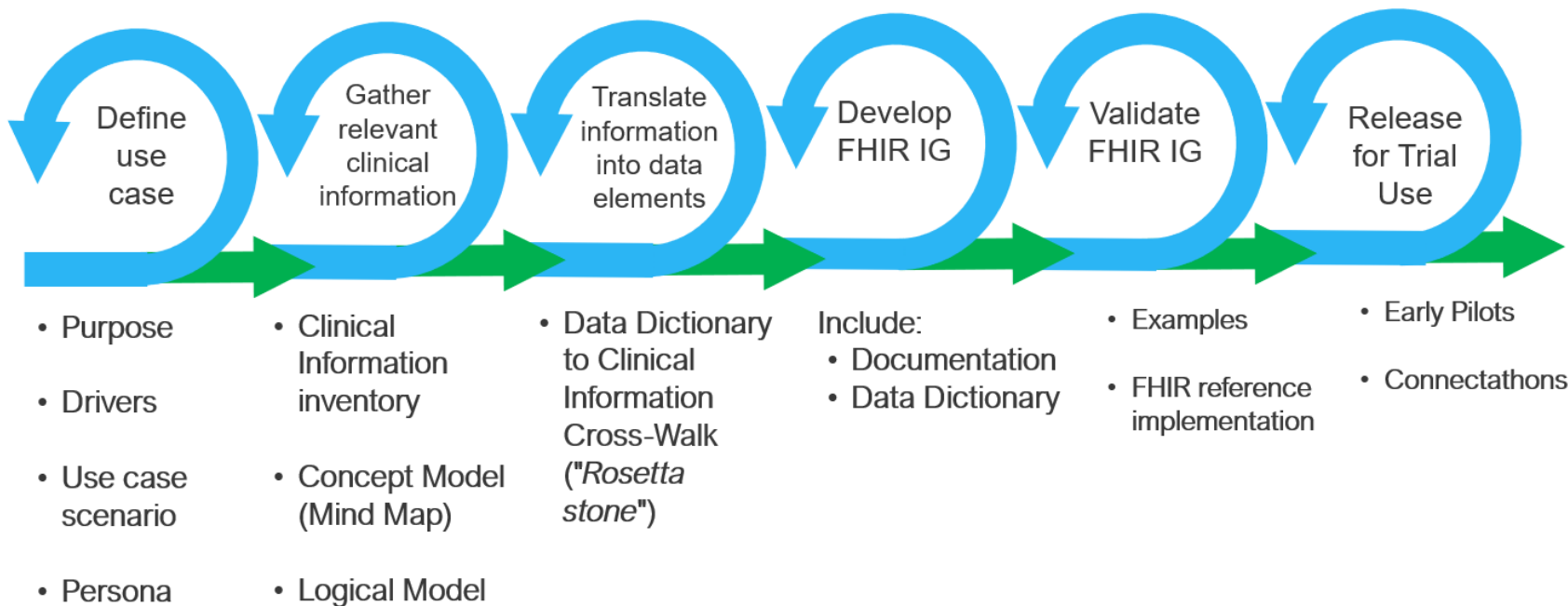
Standardized information exchange

Use-case driven and targeted use

Oncology data element domains:
patient, disease, treatment, outcomes, genomics, lab/vital



FHIR Implementation Guide Process (Initial Methodology)



CodeX | Approved for public release. Distribution unlimited 22-03274-12



Childhood Cancer-Data Integration for Research, Education, Care, and Clinical Trials (CC-DIRECT)

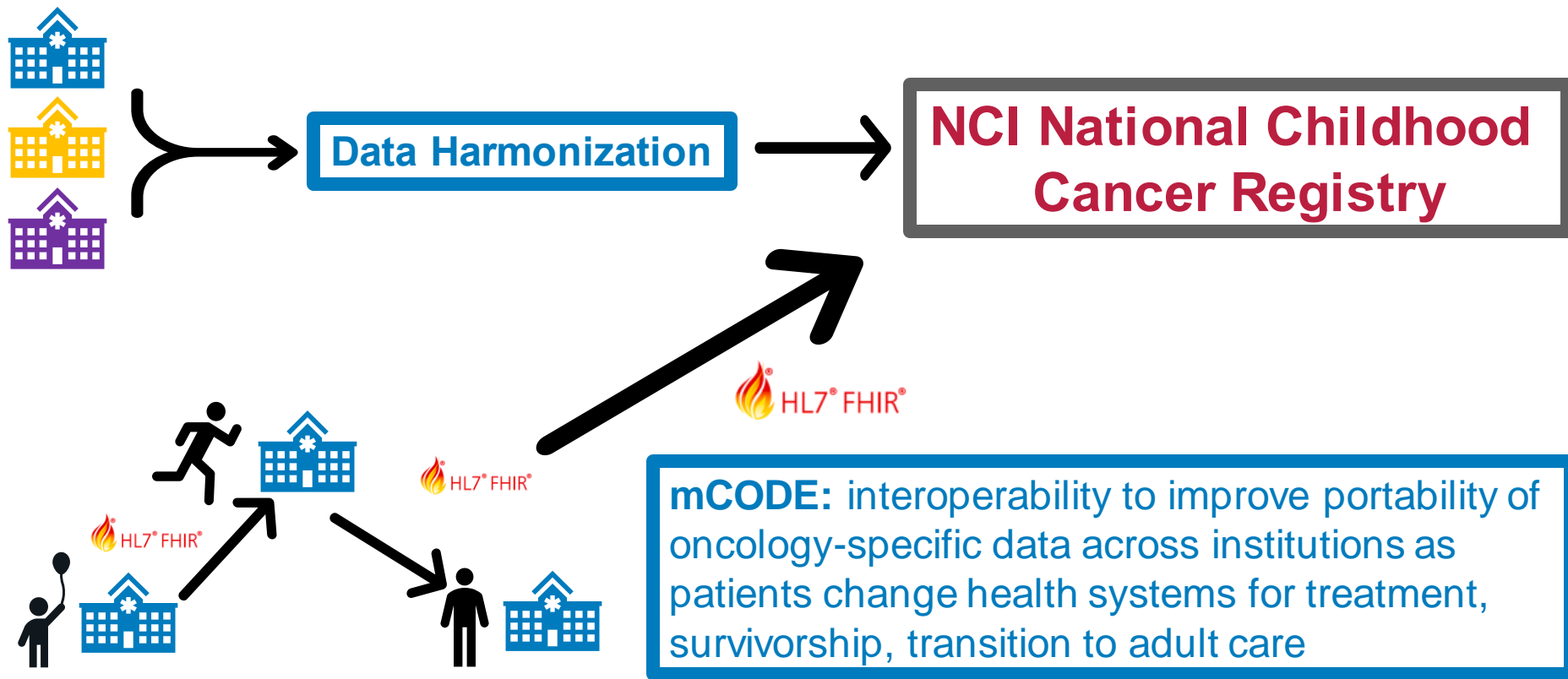
- Expand **mCODE** to include data elements specific to childhood cancer
 - Identify data required for pediatric cancer
 - Compare with existing FHIR resources
 - Prioritize data
 - Determine data flow requirements
 - Model and align with other work
 - Propose mCODE update



CC-DIRECT: Leveraging mCODE to create a standardized cancer health record

- Enable automatic extraction of discrete data elements from the electronic health record to align with data standards for NCCR
 - Generate real-world data evidence
 - Reduce time-intensive harmonization of data
 - Promote interoperability of data to support clinical and patient navigation from diagnosis through survivorship

Data Standards



Childhood Cancer Survivorship as a Use Case



Risk Prediction Calculators

Acute Ovarian Failure Risk
Prediction Calculator

[Learn More](#)

Breast Cancer Probability
Calculator

[Learn More](#)

Cardiovascular Risk Calculator

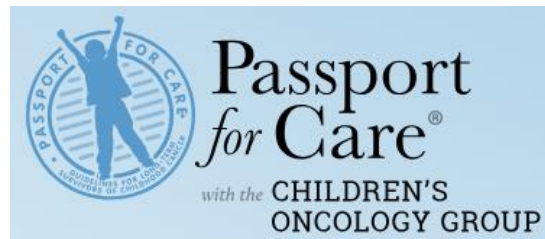
[Learn More](#)

Kidney Failure Calculator

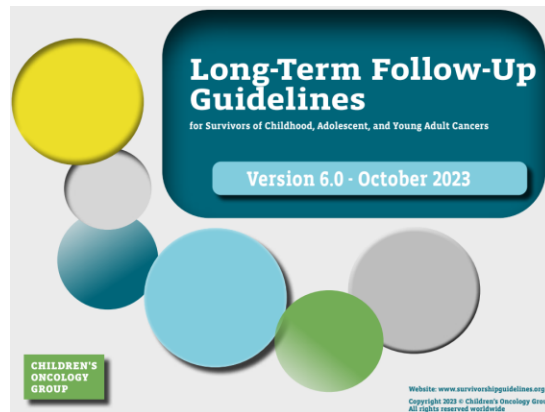
[Learn More](#)

Primary Ovarian Insufficiency
Risk Calculator

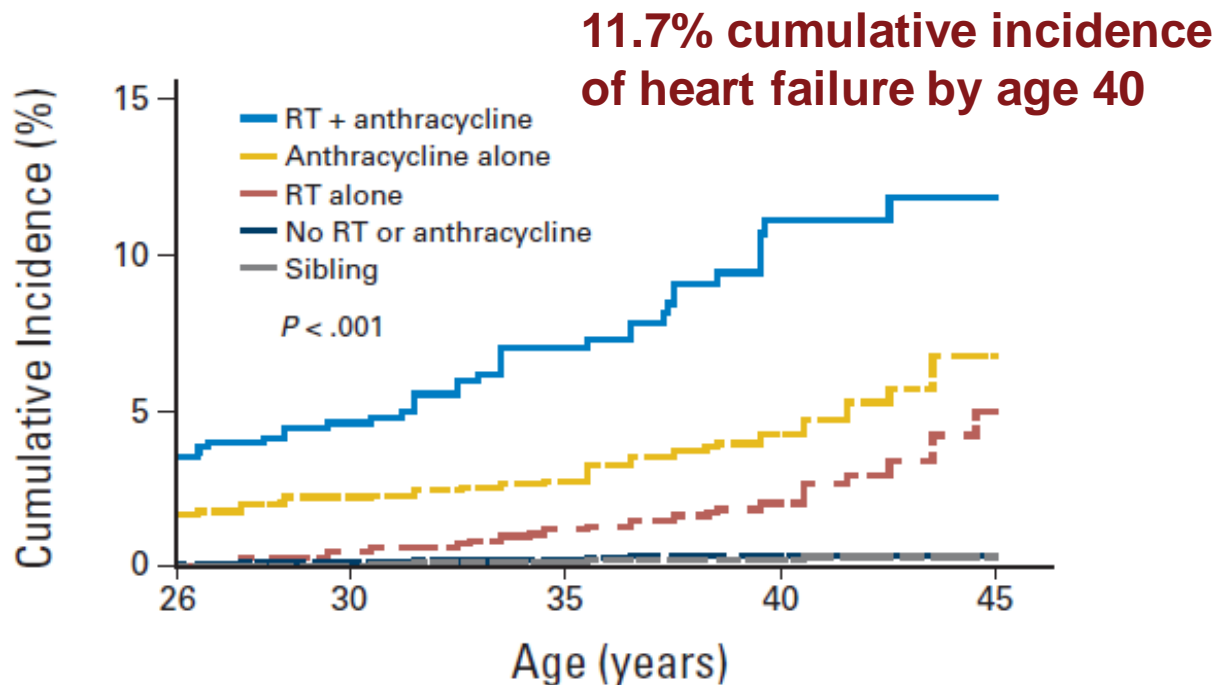
[Learn More](#)



Maps treatment
exposures to LTFU
Guidelines



Late Cardiovascular Disease Risk Stratification



Data Elements Included in NCCR, CC-DIRECT

Data Elements

Anthracycline dose and field-specific radiation doses known, age at diagnosis, gender

(Chow *et al* 2015; Armstrong *et al* 2020)

Take Home Points

- The **NCCR** enhances **central cancer registry data** with linked and harmonized data from **hospitals, NCI-Designated Cancer Centers, administrative data sources** and **multiple other sources**
- **Data standards** are essential for interoperability across the CCDI Data Ecosystem
- The **NCCR Data Platform** is a new data resource for cancer research
- Future direction includes integrating with **mCODE** and **HemOnc.org**

Q&A

Join Us at Our Next Event

CCDI Federated Data: Enhancing Data Discoverability

Tuesday, August 13, 2024, from 1:00–2:00 p.m. ET

A new advancement for querying genomic, clinical, imaging, and biospecimen data is through a standard application programming interface (API). We invite the broader community to hear from representative members of the demonstration project and learn about advancing research through CCDI data federation.

Learn more and register at events.cancer.gov/ccdi/webinar

How You Can Engage with CCDI



Learn about CCDI, register for upcoming events, and subscribe to our monthly newsletter:

cancer.gov/CCDI



Access CCDI data and resources:

ccdi.cancer.gov



Questions? Email us at:

NCIChildhoodCancerDataInitiative@mail.nih.gov

Thank you for attending!



**NATIONAL
CANCER
INSTITUTE**

cancer.gov

cancer.gov/espanol