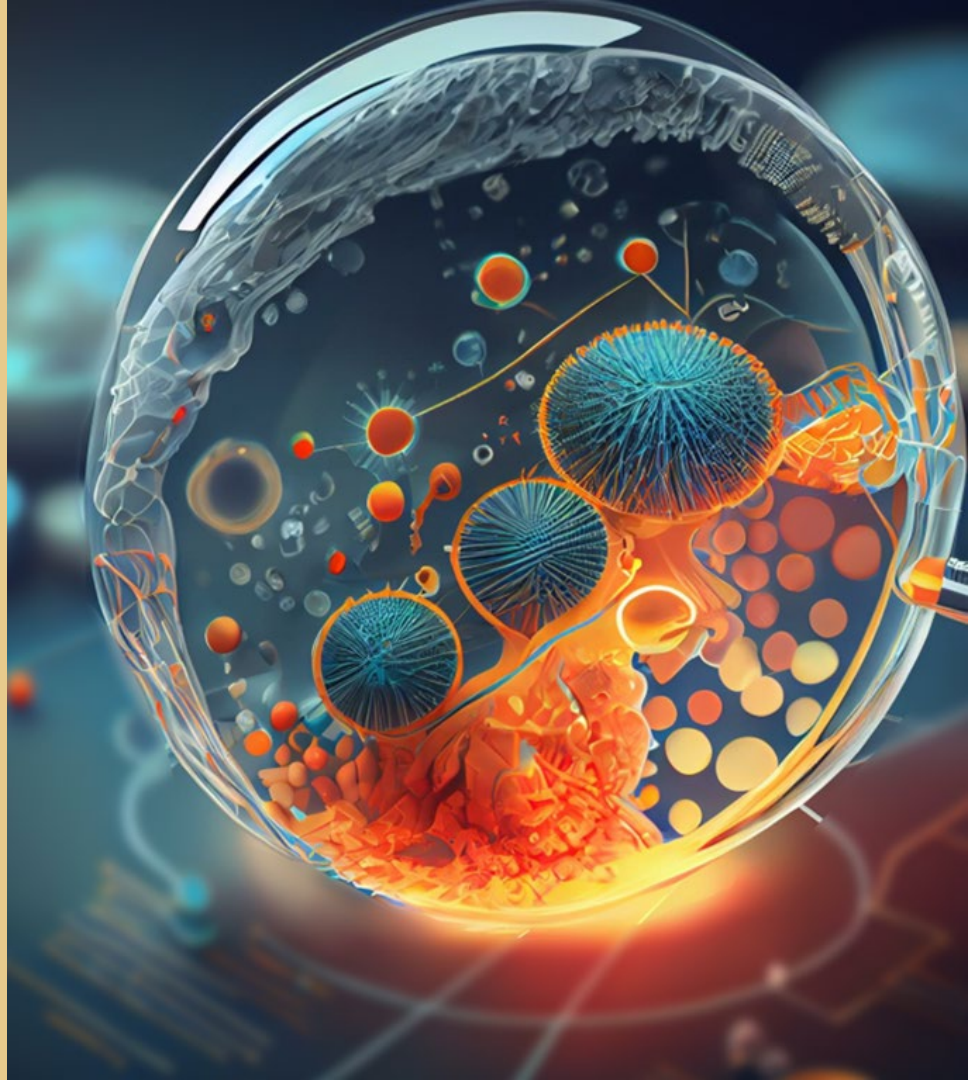




Data for the Common Good

Transforming health
through data

Samuel Volchenbom, MD, PhD
25-September-2024



Objectives

- Appreciate that pediatric cancer is a rare disease
- Understand Data for the Common Good
- Clinical trials matching for children with cancer
- Learn about obtaining richer EHR data
- Understand why long-term follow up for survivors is challenging



DATA FOR THE COMMON GOOD





DATA FOR THE COMMON GOOD

We build **communities, platforms, and ecosystems**
that **maximize the potential of data**
to **drive discovery and improve human health.**



Accessibility

lifting barriers to research



Usability

high-quality, interoperable data



Diversity

data that better represent the real world



Responsibility

making it easy to do the right thing



Community

we build consensus, not just tools



Impact

using data to its **maximum potential**

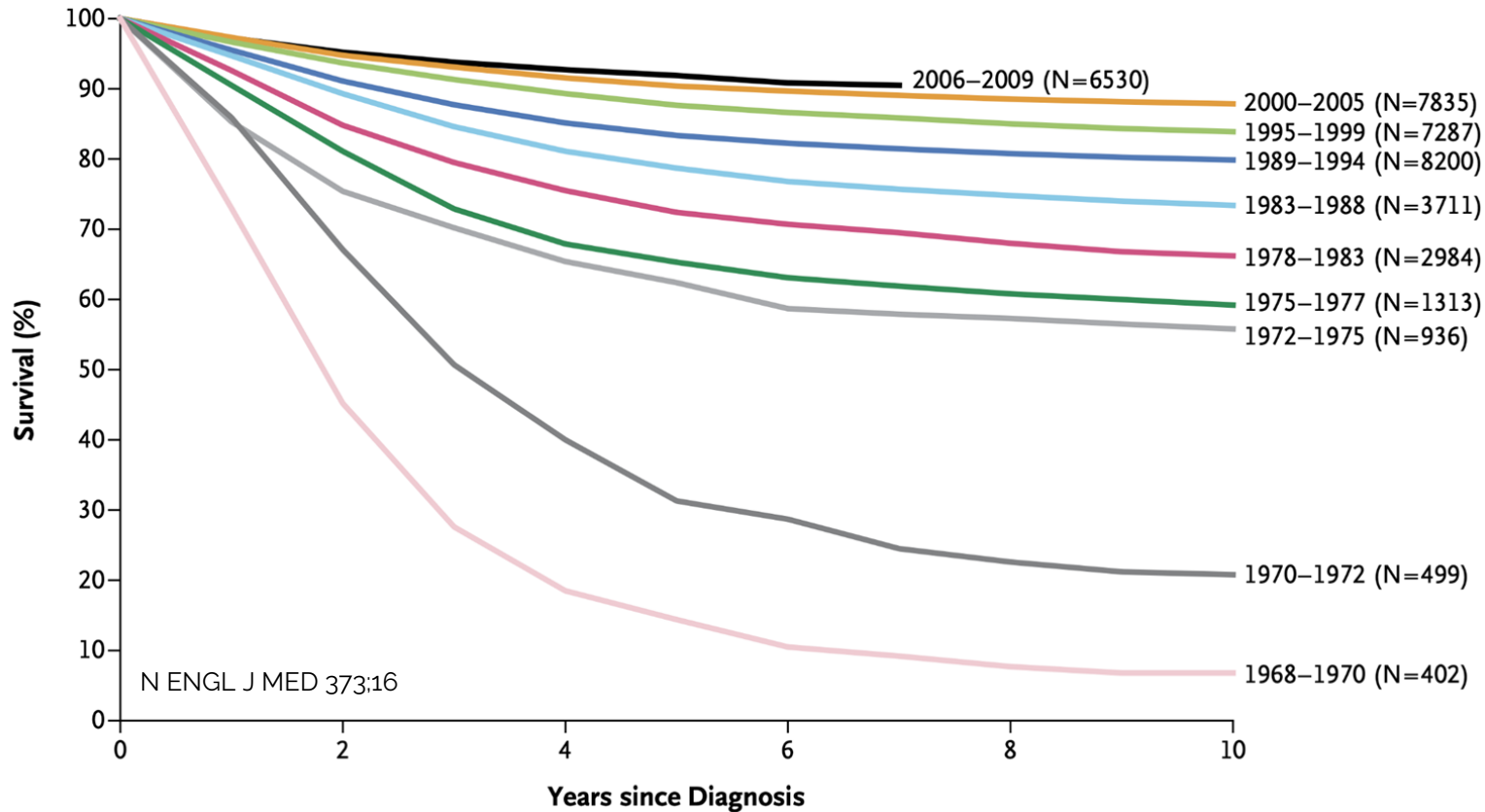


Cancer is the leading disease cause of death in US children

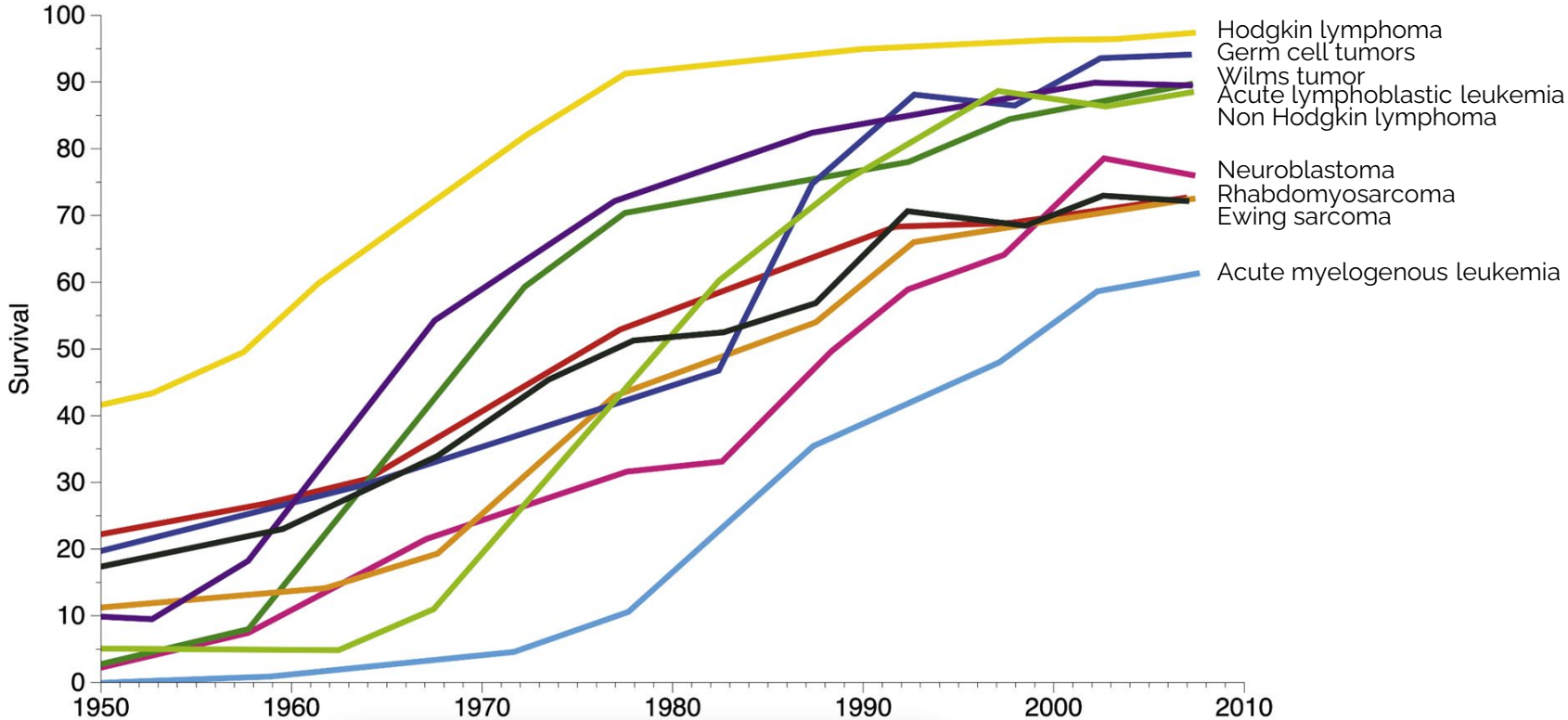


- **This year**, 16,000 children and adolescents in the U.S. will be diagnosed with cancer, and almost 2,000 will die.
- Although cure rates have improved, kids with relapsed or refractory disease have **dismal outcomes**.

U.S. Pediatric Acute Leukemia Survival



U.S. Pediatric Cancer Survival Rates





Despite these improvements, globally 43% of children with cancer **die without a diagnosis of cancer.**

In Sub-saharan Africa, this number is 57%

In the US and Western Europe - 3%

Ward ZJ, et al., Lancet Oncology, 2019 Apr;20(4):483-493

Why have survival rates improved so dramatically?

- New drug development

Why have survival rates improved so dramatically?

- ~~New drug development~~

Why have survival rates improved so dramatically?

- ~~New drug development~~
- Early detection / surveillance

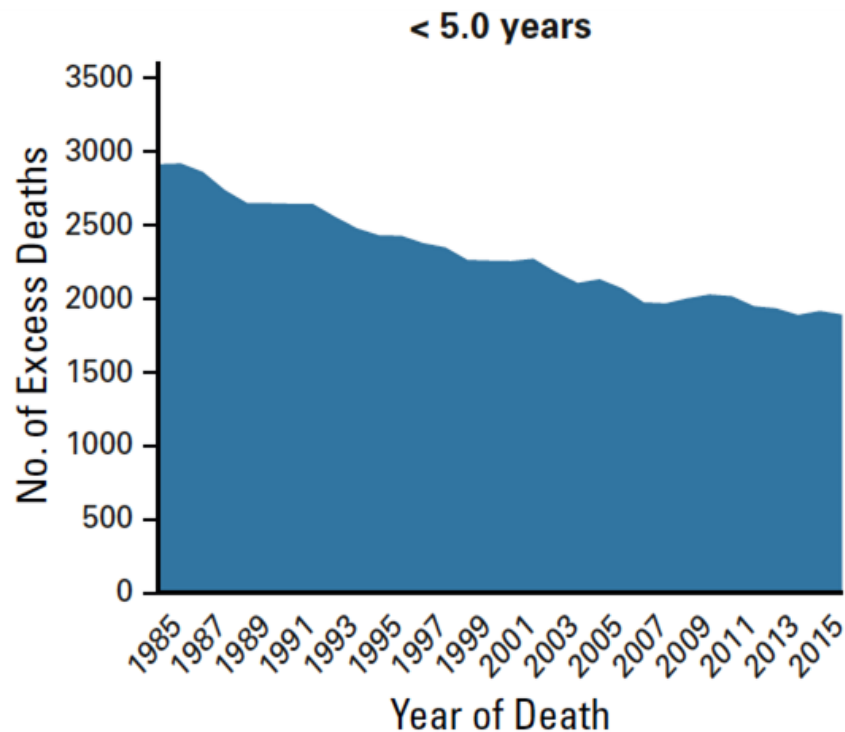
Why have survival rates improved so dramatically?

- ~~New drug development~~
- ~~Early detection / surveillance~~

Why have survival rates improved so dramatically?

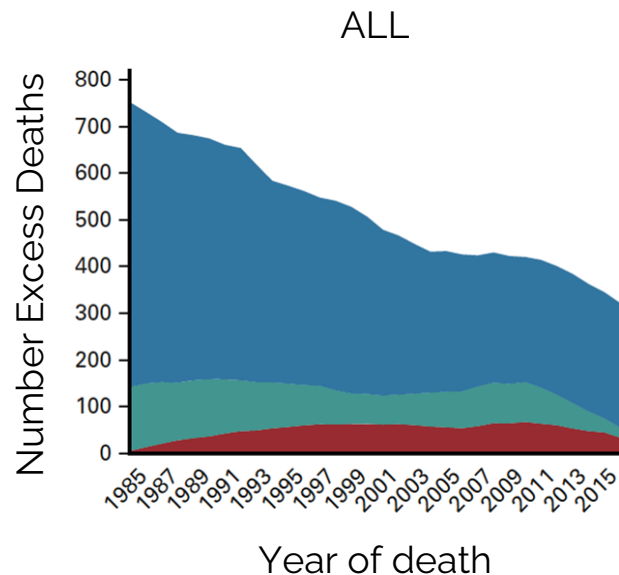
- ~~New drug development~~
- ~~Early detection / surveillance~~
- Cooperative group clinical trials
- Better understanding of biology
 - Molecularly-targeted agents
- Improved risk stratification
- More sensitive detection of treatment efficacy / failure
- More effective supportive services
- Rise of patient advocacy

Pediatric cancer - shifting the survival curves



AnnaLynn M. Williams, et al., JCO, Volume 39, Issue 20 2227

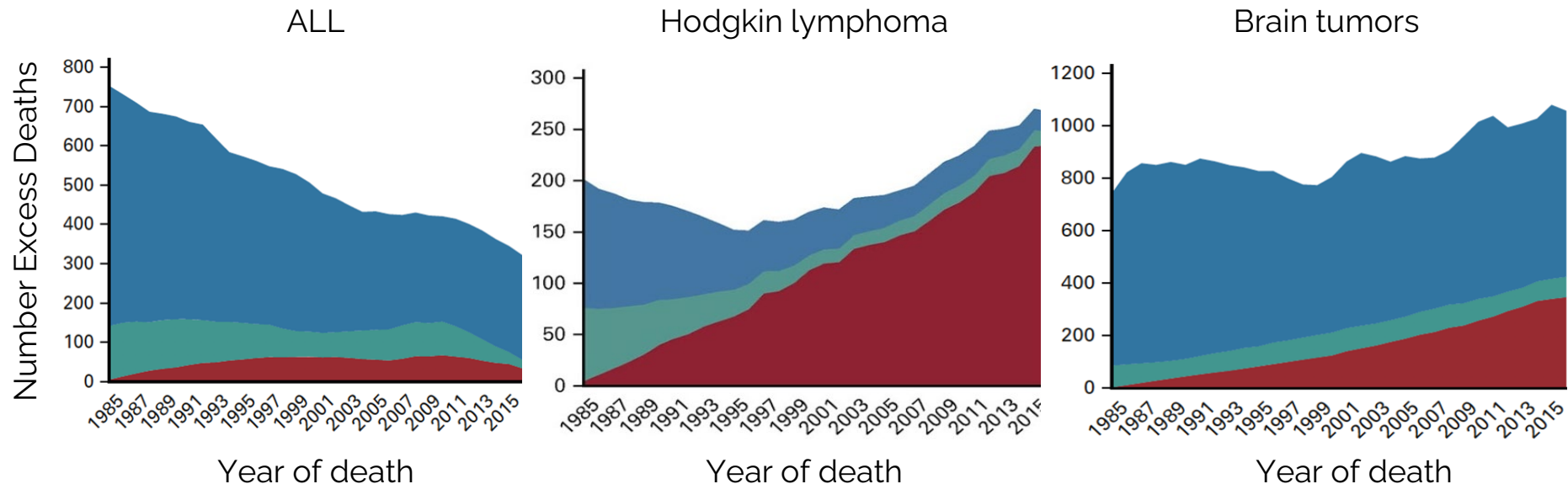
Pediatric cancer - shifting the survival curves






AnnaLynn M. Williams, et al., JCO, Volume 39, Issue 20 2227

 <5.0 years  5.0 to 10 years  ≥ 10.0 years

Pediatric cancer - shifting the survival curves



 **<5.0 years**  **5.0 to 10 years**  **≥ 10.0 years**

AnnaLynn M. Williams, et al., JCO, Volume 39, Issue 20 2227

The keys to improving outcomes for rare diseases are collecting, aggregating, and democratizing access to high-quality data.



A perfect storm

Rare diseases, siloed data,
manual processes, lack of standards,
difficult research questions



D4CG structure



**DATA FOR THE
COMMON GOOD**



Cancer experts reach out to be part of PCDC

PTLD

acute
lymphocytic
leukemia

acute myeloid
leukemia

INTERACT

bone tumors
(OS and EWS)

HIBiSCus

Fanconi anemia

FRIENDS

central nervous
system tumors

INSPIRE

germ cell
tumors

MaGIC

Hodgkin
lymphoma

NODAL

NBL

INRG

nasopharyngeal
carcinoma

NOBLE

oncofertility

**Reproductive
HOPE**

Langerhans cell
histiocytosis

cancer
predisposition

C3P

retinoblastoma

Global REACH

soft-tissue
sarcoma

INSTRuCT

myelodysplastic
syndrome

D4CG structure



**DATA FOR THE
COMMON GOOD**

PEDIATRIC CANCER
DATA COMMONS

GEARBOX



D4CG sought as a data commons provider



—————> Fanconi anemia data commons



—————> Crohn's disease data commons



—————> Monogenic epilepsy commons



—————> Monogenic diabetes commons

D4CG structure



DATA FOR THE COMMON GOOD

PEDIATRIC CANCER
DATA COMMONS

GEARBOX



Sociome

*penetrating
brain injury*

*monogenic
epilepsy*

PREDICT
*monogenic
diabetes*

other rare diseases

...and more to come



Step 1: Establish a data commons consortium

We employ a “big tent” philosophy.



Step 1: Establish a data commons consortium

We employ a “big tent” philosophy.



Pharma

Registries

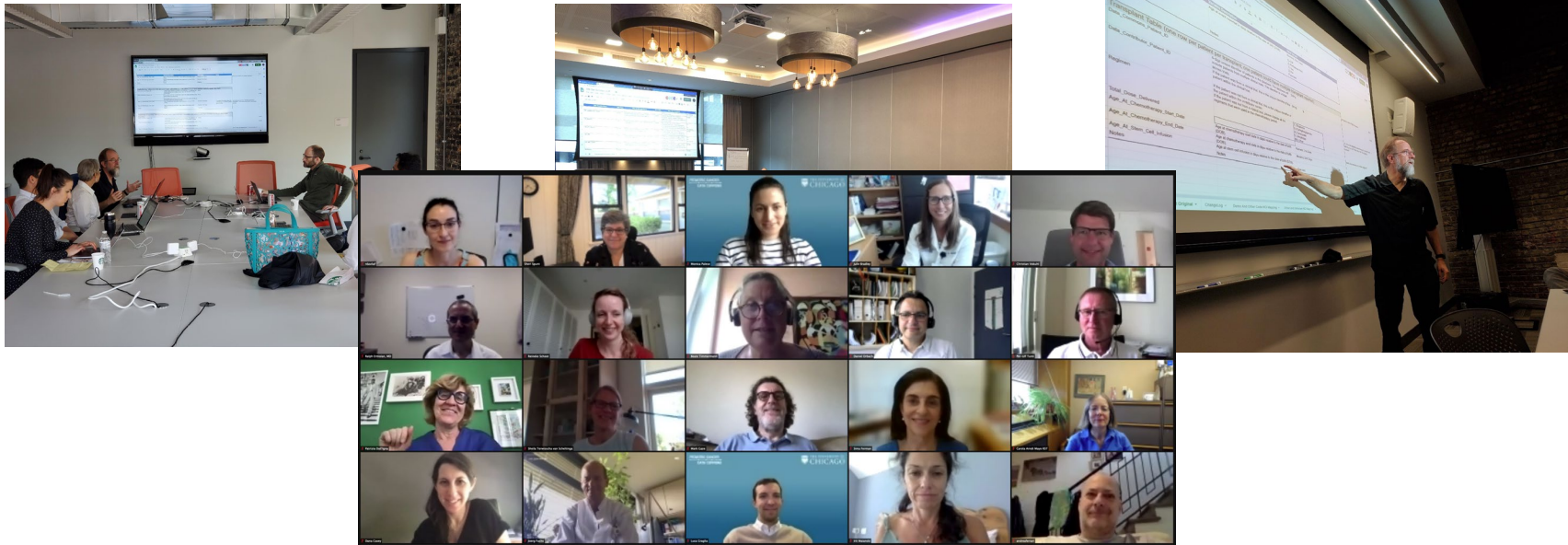
Cooperative groups

Academic medical centers

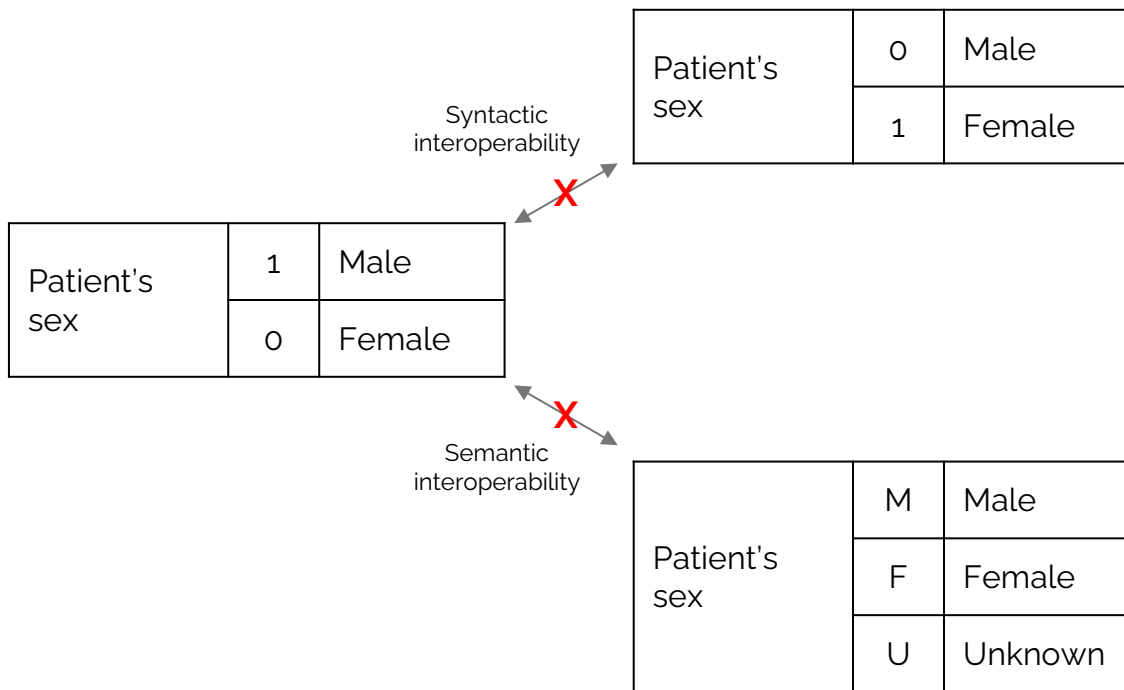
Step 2 - Build a data dictionary



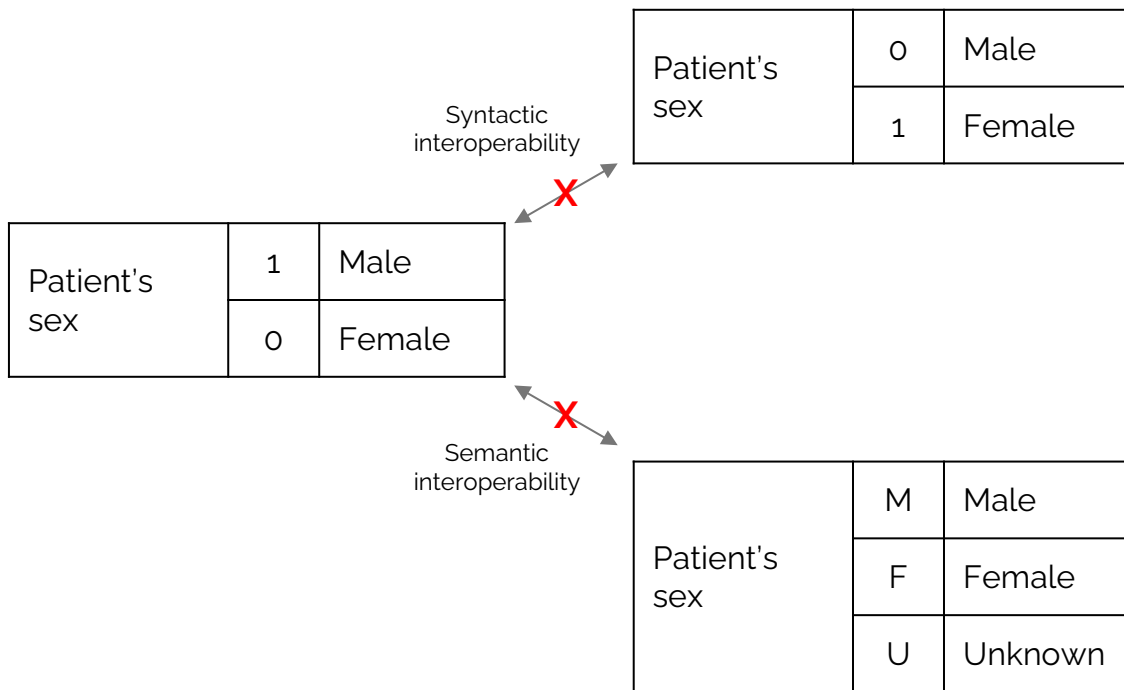
Data dictionary development



The importance of data standards

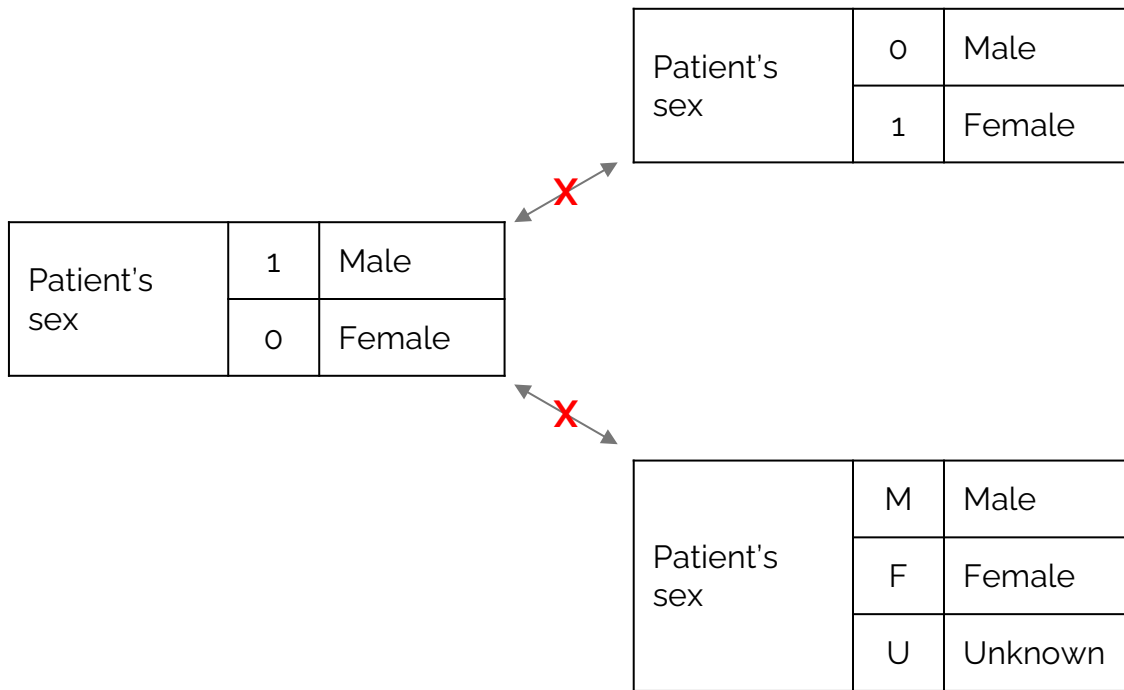


The importance of data standards



Institution	Local value
Hospital A	0
Hospital A	1
Hospital A	1
Hospital B	1
Hospital B	
Hospital B	0
Hospital C	F
Hospital C	F
Hospital C	U

The importance of data standards



Sex	
Description	NCIt Code
Male	C20197
Female	C16576
Unknown	C17998

The importance of data standards

NIH NATIONAL CANCER INSTITUTE www.cancer.gov
 NCI Term Browser EVS Enterprise Vocabulary Services
 Terminologies Value Sets Mappings
 NCIthesaurus
 Version: 21.05d (Release date: 2021-09-27)
 Search: male
 Contains Exact Match Begins With
 Name Code Property Relationship
 Back to search results Advanced Search
 Hierarchy Value Sets Maps Help
 Quick Links
 Male (Code C20197)
 View in Hierarchy | View History | View Graph | Add to Cart | Suggest Changes
 Terms & Properties Synonym Details Relationships Mappings View All
 Terms & Properties
 Preferred Name: Male
 Definition: A person who belongs to the sex that normally produces sperm. The term is used to indicate biological sex distinctions, cultural gender role distinctions, or both.
 CDISC Definition: A person who belongs to the sex that normally produces sperm. The term is used to indicate biological sex distinctions, cultural gender role distinctions, or both. (NCI)
 Label: Male
 NCI Thesaurus Code: C20197 (Search for linked ccdsr metadata) (search value sets)
 NCI Metathesaurus Link: C0086582 (see NCI Metathesaurus info)
 Synonyms & Abbreviations: (see Synonym Details)
 Human, Male
 M
 Male
 MALE
 male
 External Source Codes:
 UMLS CUI C0086582
 Other Properties:

Patient's sex	1	Male
	0	Female

Patient's sex	F	Female
	U	Unknown

Sex	
Description	NCIt Code
Male	C20197
Female	C16576
Unknown	C17998

Building a consensus data dictionary

Off Protocol Therapy/Study: one row per subject per off protocol therapy/study per reason off						
AGE_OFF	Number	Age in Days When Off Protocol Therapy or Study	C172678	Age of subject (in days) when		
DISEASE_PHASE	Code	Disease Phase	C168878	The stage or period of an in	Initial Diagnosis	C156813
					Relapse	C38155
DISEASE_PHASE_NUMBER	Number	Disease Phase Number	C173258	The number of the disease		
COURSE	Code	Protocol Treatment Course	C168807	The type of protocol treatm	Prephase	C168826
					Induction	C158876
					Intensification	C173105
					Consolidation	C15679
					Stem Cell Transplant Conditioning	C168794
					Maintenance	C15688
					Palliative Treatment	C15292
Other	C17649					
COURSE_NUMBER	Number	Course Number	C166235	The number assigned to a c		
OFF_TYPE	Code	Off Protocol Therapy or Study	C173256	The code used to designat	Protocol Therapy	C173257
					Study	C29851
REASON_OFF	Code	Off Protocol Therapy or Study Reason	C173519	The reason a subject went	Death	C93546
					Lost to Follow-Up	C70740
					Completion of Planned Therapy	C168935
					Physician Decision	C48250
					Withdrawal of Consent	C48271
					Subject/Guardian Refused Further Treatment	C168934
					Disease Progression	C35571
					Relapse	C38155
					Adverse Event	C41331
					Secondary Malignancy	C4968
					Other	C17649
					Unknown	C17998
					Not Reported	C43234

<http://sam.am/datadictionaries>

NCI thesaurus
Version: 22.06d (Release date: 2022-06-29)

Search: C168935

Contains
 Exact Match
 Begins With
 Name
 Code
 Property
 Relationship

Advanced Search

Hierarchy | Value Sets | Maps | Visited Concepts | Help

Quick Links

End of Planned Treatment (Code C168935)
View in Hierarchy | View History | View Graph | Add to Cart | Suggest Changes

Terms & Properties | Synonym Details | Relationships | Mappings | View All

Terms & Properties

Preferred Name: End of Planned Treatment

Definition: The end of the planned treatment.

Label: End of Planned Treatment

NCI Thesaurus Code: C168935 ([Search for linked caDSR metadata](#)) ([search value sets](#))

NCI Metathesaurus Link: CL1379002 ([see NCI Metathesaurus info](#))

Synonyms & Abbreviations: ([see Synonym Details](#))

Completion of Planned Therapy

End of Planned Treatment

External Source Codes:

NCI META CUI: CL1379002

Other Properties:

Name	Value (qualifiers indented underneath)
code	C168935
Contributing_Source	PCDC
Semantic_Type	Idea or Concept

Additional Concept Data:

Defined Fully by Roles: No

URL: https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C168935

Age of subject (in days) when		
The stage or period of an in	Initial Diagnosis	C156813
	Relapse	C38155
The number of the disease		
The type of protocol treatm	Prephase	C168826
	Induction	C158876
	Intensification	C173105
	Consolidation	C15679
	Stem Cell Transplant Conditioning	C168794
	Maintenance	C15688
	Palliative Treatment	C15292
	Other	C17649
The number assigned to a c		
The code used to designat	Protocol Therapy	C173257
	Study	C29851
The reason a subject went	Death	C93546
	Lost to Follow-Up	C70740
	Completion of Planned Therapy	C168935
	Physician Decision	C48250
	Withdrawal of Consent	C48271
	Subject/Guardian Refused Further Treatment	C168934
	Disease Progression	C35571
	Relapse	C38155
	Adverse Event	C41331
	Secondary Malignancy	C4968
	Other	C17649
	Unknown	C17998
	Not Reported	C43234

<http://sam.am/datadictionaries>

Data collection is highly localized

<input type="checkbox"/> Male <input type="checkbox"/> Female <input type="checkbox"/> Unknown <i>English</i>	<input type="checkbox"/> 男 <input type="checkbox"/> 女性 <input type="checkbox"/> わからない <i>Japanese</i>
<input type="checkbox"/> ♂ <input type="checkbox"/> ♀ <input type="checkbox"/> ? <i>Arabic</i>	<input type="checkbox"/> Homme <input type="checkbox"/> Femelle <input type="checkbox"/> Inconnue <i>French</i>
<input type="checkbox"/> Männlich <input type="checkbox"/> Weiblich <input type="checkbox"/> Unbekannt <i>German</i>	<input type="checkbox"/> 男性 <input type="checkbox"/> 女性 <input type="checkbox"/> 未知 <i>Chinese</i>

CDE mapping solves the localization problem

Male
 Female
 Unknown
English

男
 女性
 わからない
Japanese

♂ ≠
 ♀

Arabic

Homme
 Femelle
 Inconnue
French

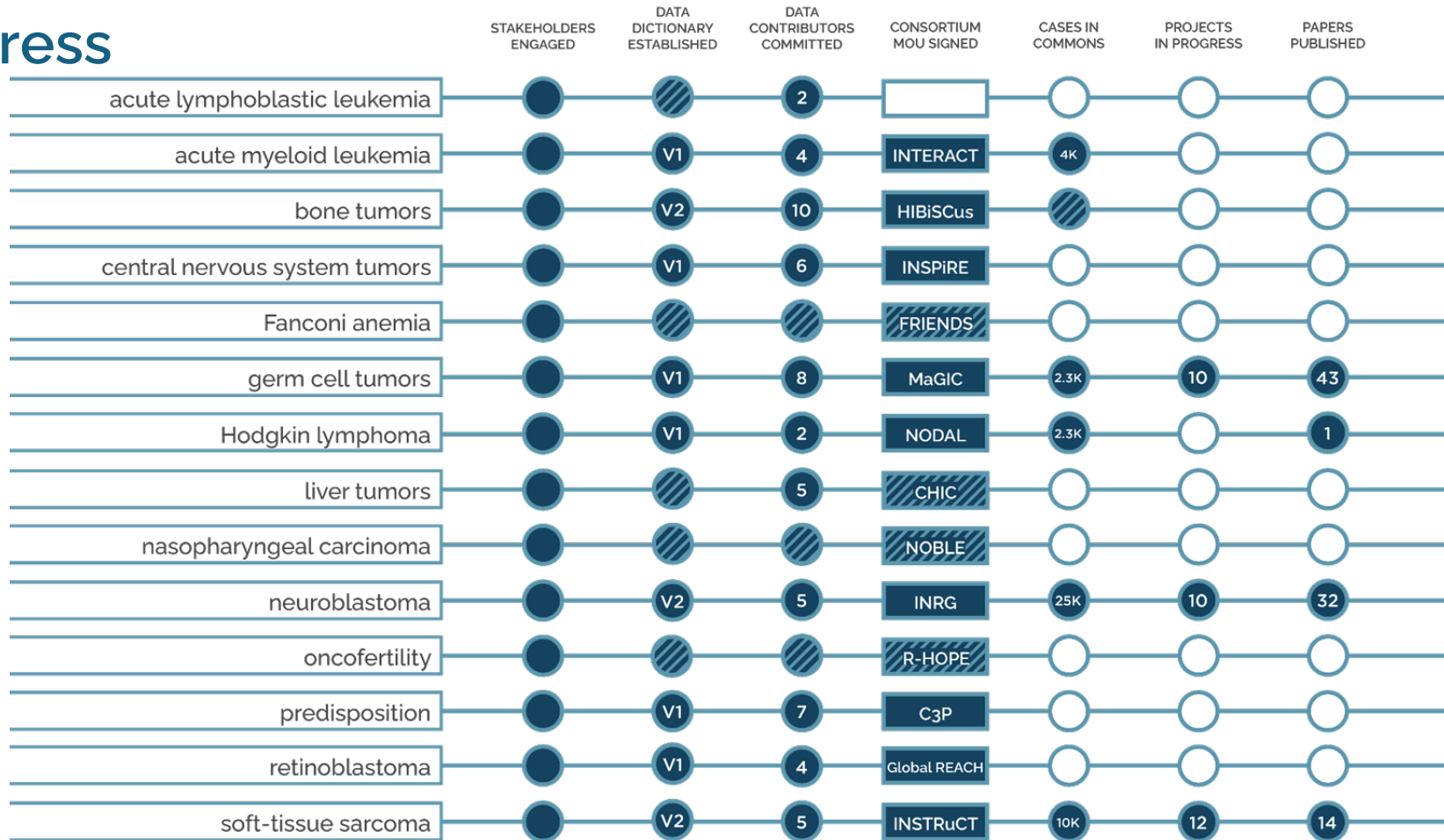
Männlich
 Weiblich
 Unbekannt
German

男性
 女性
 未知
Chinese

C20197
 C16576
 C17998

NIH NATIONAL CANCER INSTITUTE
 www.cancer.gov
 NCI Term Browser
 EVS Enterprise Vocabulary Services
 Terminologies Value Sets Mappings
 Search
 Contains Exact Match Begins With
 Name Code Property Relationship
 Back to search results Advanced Search
 Hierarchy Value Sets Maps Help
 Quick Links
 View in Hierarchy | View History | View Graph | Add to Cart | Suggest Changes
Male (Code C20197)
 Terms & Properties Synonym Details Relationships Mappings View All
Terms & Properties
 Preferred Name: Male
 Definition: A person who belongs to the sex that normally produces sperm. The term is used to indicate biological sex distinctions, cultural gender role distinctions, or both.
 CDISC Definition: A person who belongs to the sex that normally produces sperm. The term is used to indicate biological sex distinctions, cultural gender role distinctions, or both. (NCI)
 Label: Male
 NCI Thesaurus Code: C20197 (Search for linked caDSR metadata) (search value sets)
 NCI Metathesaurus Link: C0086582 (see NCI Metathesaurus info)
 Synonyms & Abbreviations: (see Synonym Details)
 Human, Male
 M
 male
 Male
 MALE

Progress



PCDC data dictionaries

- PCDC Master Data Dictionary
- Acute Lymphoblastic Leukemia (ALL)
- Acute Myeloid Leukemia (AML)
- Central Nervous System Tumors (CNS)
- Ewing Sarcoma (EWS)
- Germ Cell Tumors (GCT)
- Hodgkin Lymphoma (HL)
- Neuroblastoma (NBL)
- Non-Rhabdomyosarcoma Soft Tissue Sarcoma (NRSTS)
- Osteosarcoma (OS)
- Retinoblastoma (RB)
- Rhabdomyosarcoma (RMS)

<http://sam.am/datadictionaries>

Step 3 - Governance



PCDC worldwide participation

N. America

- CBTN
- COG
- DFCI
- IDIPGR
- NRG
- PNOC
- RBTC
- St. Jude



Europe

- AIEOP
- CCLG
- COSS
- CRCTU
- EEC
- EpSSG
- EuPAL
- GPOH
- GEIS
- GSF-GETO
- ISG
- MRC
- NCRI
- SIOPe
- SIOPEN
- SSG
- SFCE
- UNICANCER



Asia

- JCCG
- SIOPEN
- EpSSG
- COG

S. America

- SOBOPE
- EpSSG
- GALOP
- GLATO

Oceania

- COG
- EpSSG

DCAs

US:

- 11 master agreements
- 22 addenda

Non-US:

- 14 master agreements
- 13 addenda

DUAs

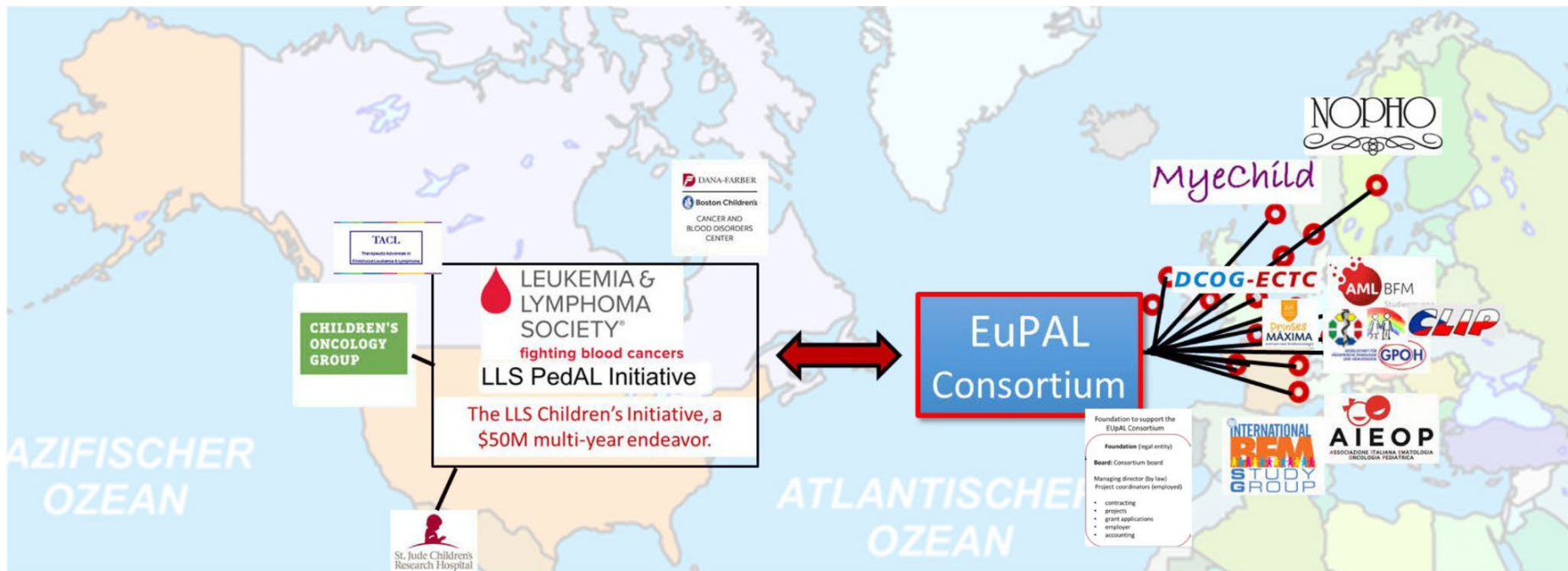
US:

- 9 master agreements
- 24 addenda/projects

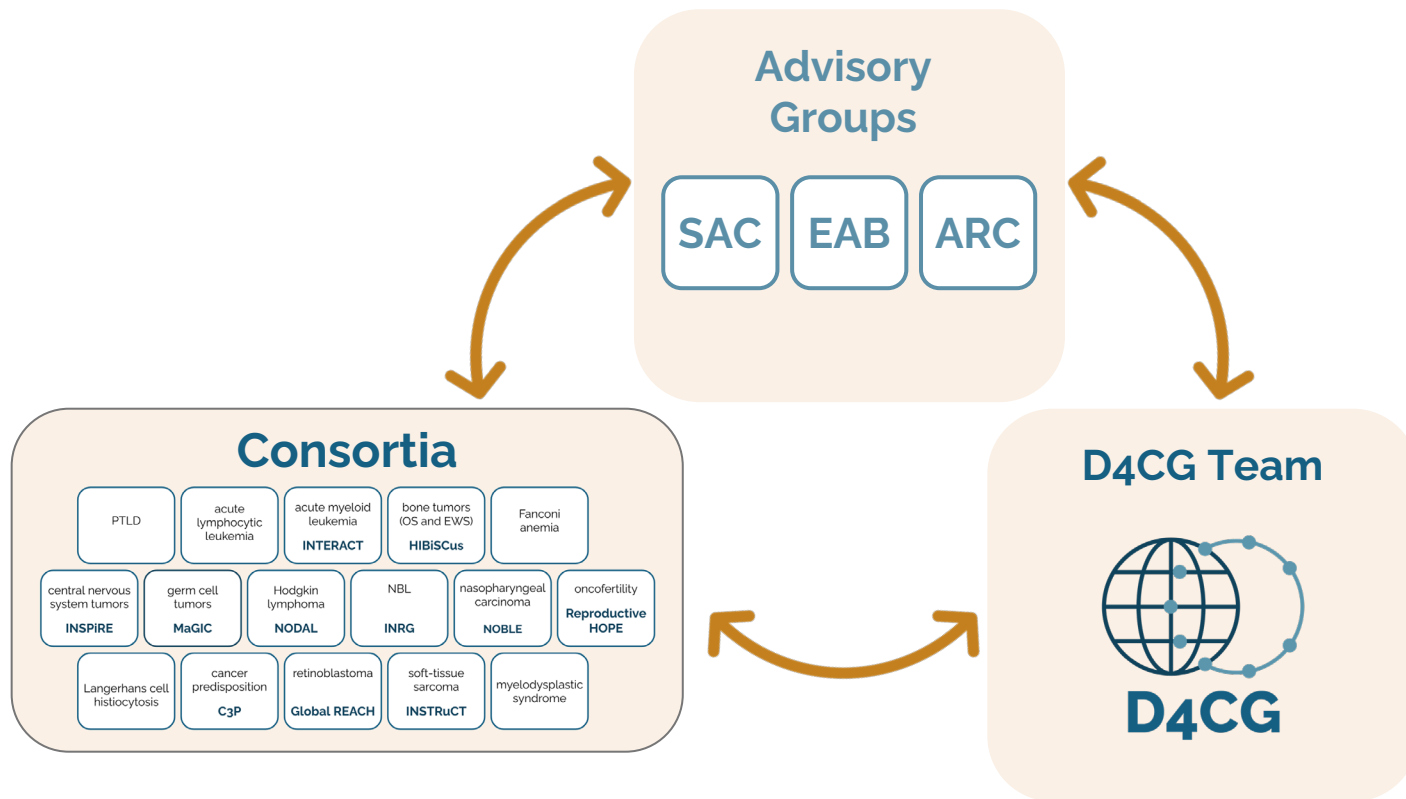
Non-US:

- 7 master agreements
- 8 addenda/projects

INTERACT AML Consortium



PCDC structure



Step 4 - The data commons portal



Step 4: Develop and deploy the technical infrastructure: the data portal

The screenshot displays the PCDC Data Commons portal interface. The main view shows a 'Filter Set Workspace' with a filter set named 'Active #1' containing the condition 'Consortium is "INRG"'. The total number of subjects is 24,682. A modal window is open, showing a list of stages with their respective subject counts:

Stage	Count
<input checked="" type="checkbox"/> Stage 1	4,122
<input checked="" type="checkbox"/> Stage 2a	1,279
<input checked="" type="checkbox"/> Stage 2b	1,777
<input checked="" type="checkbox"/> Stage 3	3,700
<input type="checkbox"/> Stage 4	11,413

The modal also includes a 'Save changes to the current Filter Set' form with fields for Name (INSS Stage 1, 2a, 2b, 3) and Description (Describe the Filter Set (optional)). The filters section shows the current filter set: 'Consortium is "INRG" AND Stage System is "INSS" AND Stage is any of "Stage 3", ...'. Buttons for 'Back to page' and 'Save changes' are visible at the bottom of the modal.

<http://portal.pedscommons.org>

<http://sam.am/EHA2024>

commons.uchicago.edu



Data in the portal

▼ Consortium 1 selected X

Filter Mode **Include** Exclude

<input type="checkbox"/>	INRG	25,404
<input type="checkbox"/>	INSTRuCT	10,311
<input checked="" type="checkbox"/>	INTERACT	4,011
<input type="checkbox"/>	MaGIC	2,293
<input type="checkbox"/>	NODAL	2,437

▼ Data Contributor

Filter Mode **Include** Exclude

<input type="checkbox"/>	BFM-SG	1,405
<input type="checkbox"/>	COG	1,367
<input type="checkbox"/>	DCOG	117
<input type="checkbox"/>	JPLSG	443
<input type="checkbox"/>	PPLLSG	142
<input type="checkbox"/>	SJCRH	537

▼ Study Id

Filter Mode **Include** Exclude

<input type="checkbox"/>	AAML03P1	339
<input type="checkbox"/>	AAML0531	1,028
<input type="checkbox"/>	AML02	238
<input type="checkbox"/>	AML08	299
<input type="checkbox"/>	AMLBFM2004	830
<input type="checkbox"/>	AMLBFM2012	149
<input type="checkbox"/>	AMLBFMRegistry2012	426
<input type="checkbox"/>	JPLSGAML05	443

INTERACT (AML) Data in the PCDC

	COG	BFM-SG	JPLSG	SJCRH	DCOG	PPLLSG
Subject Characteristics	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓
Disease Phase Timing	✓	✓	✓	✓	✓	✓
Survival Characteristics	✓	✓	✓	✓	✓	✓
Disease Characteristics	✓	✓	✓	✓	✓	✓
Molecular Analysis	✓	✓		✓	✓	✓
Stem Cell Transplant	✓		✓	✓		
Minimal Residual Disease	✓	✓		✓		✓

Step 4: Develop and deploy the technical infrastructure: the data portal

Survival type: Overall Survival

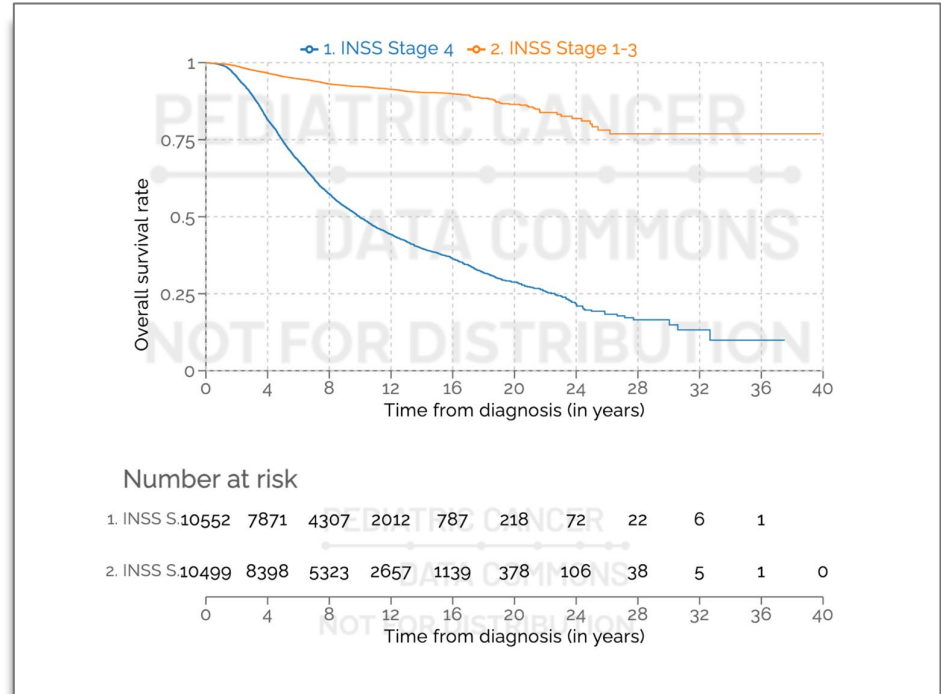
Start time (year): 0

End time (year): Optional, max value if left blank

Time interval (year): 4

Select Filter Set to analyze: Add

- > 1. INSS Stage 4 10552/11413 X
- > 2. INSS Stage 1-3 10499/10878 X



<http://portal.pedscommons.org>

PCDC Publications: By the Numbers

	Approved project requests	Projects in progress	Publications
INRG (NBL)	1	8	32
INSTRuCT (rhabdo and non-rhabdo)	13	12	13
MaGIC (germ cell tumors)		10	43
NODAL (Hodgkin lymphoma)	2	0	0

Totals: 16 approved project requests
30 projects in progress
88 publications

INRG - Defining more personalized and precise treatment

Complete list of projects available at <https://sam.am/pcdcresearch> and publications at <https://commons.cri.uchicago.edu/publications/>

- **Prognostic factors**

- Clinical and biological features
- Age, category , grade, and MKI
- Serum LDH and serum ferritin
- Segmental chromosomal alterations
- Age-at-diagnosis
- Racial and ethnic disparities
- MYCN amplification
- Participation in clinical trials

- **INRG consensus papers**

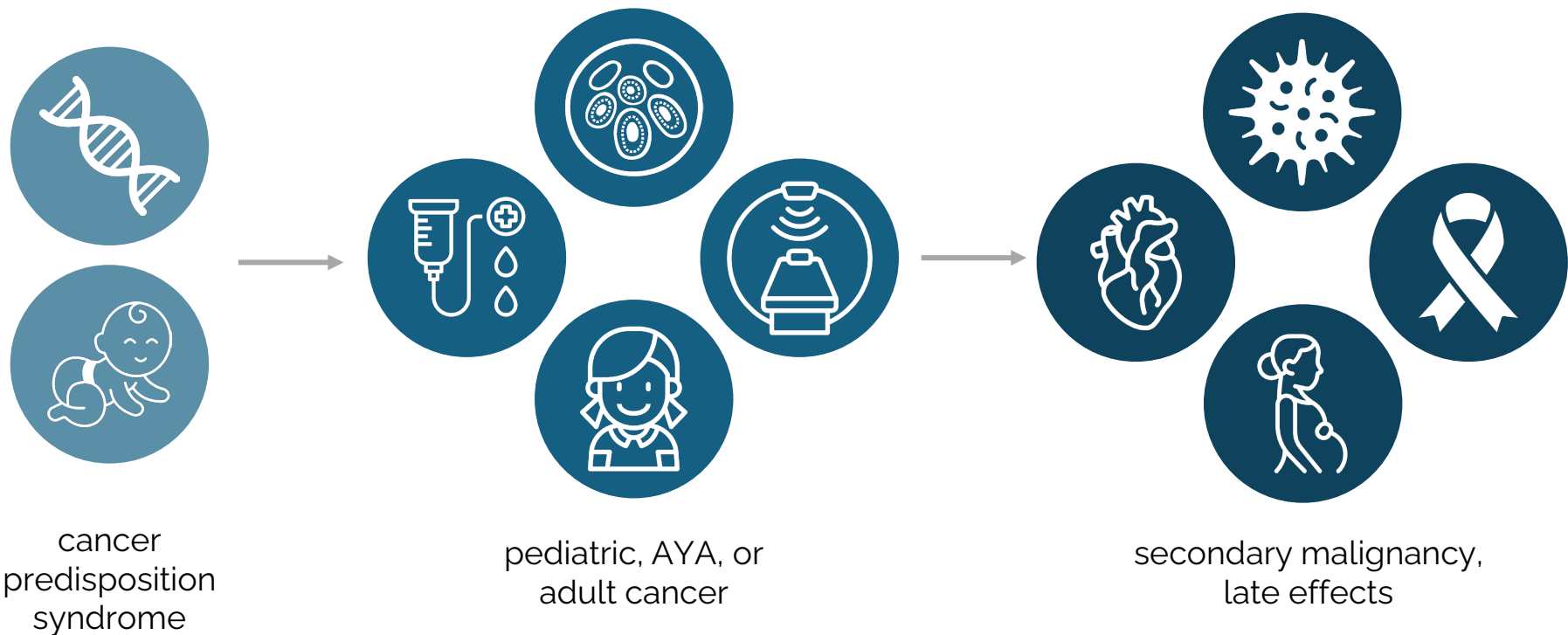
- Guidelines for imaging and staging
- NBL molecular diagnostics
- Criteria for sensitive detection of minimal NBL cells in bone marrow, blood, and stem cells

- **Immunogenomic determinants of tumor microenvironment**

Neuroblastoma

- Incidence: 500/yr in U.S.
- 25,000 patients in commons

PCDC connects the dots



AYA Cancer Incidence (15-39)

- 90,000 AYAs diagnosed with cancer each year in the US
- * >2.3 million AYA cancer survivors in the US
- Cancer subtypes & biology differ between AYAs, older adults, and younger children
- More patients diagnosed with AML between 15 and 39 are treated in the adult setting than the pediatric setting

Courtesy M. Roth



D4CG differentiators

- We take a **data first** approach
- No data are harmonized or ingested until a **stable, consensus data dictionary** is in place
- We build sustainable dictionaries that meet the **highly-specific needs** of researchers and clinicians
- We balance **autonomy** and **centralization**
- **Change management and quality control** are an important part of our value proposition
- **Governance and provenance** of the data dictionaries and the data are critical to success

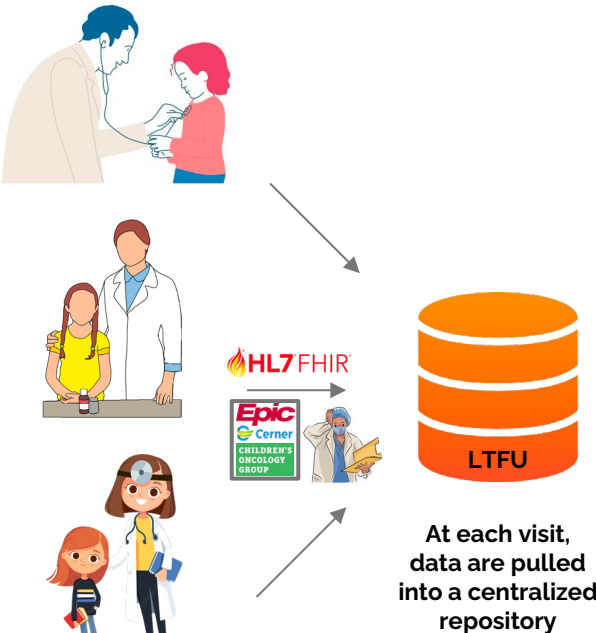
A vision for longitudinal care



A vision for the future for long-term follow up

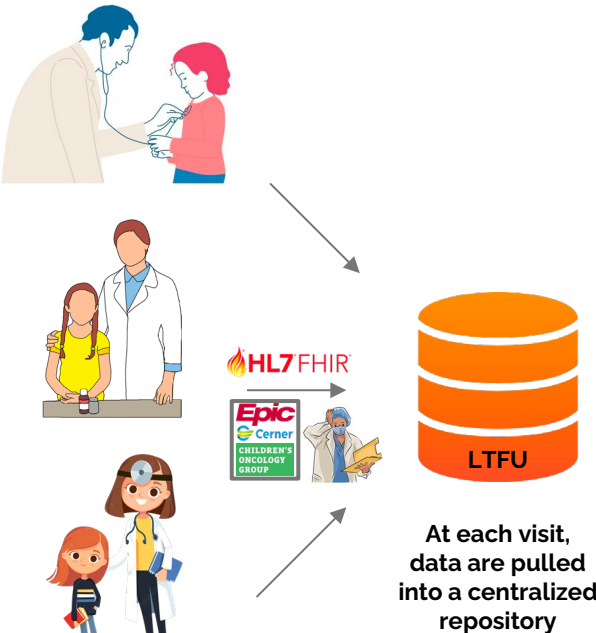


A vision for the future for long-term follow up



Patient seen at multiple institutions

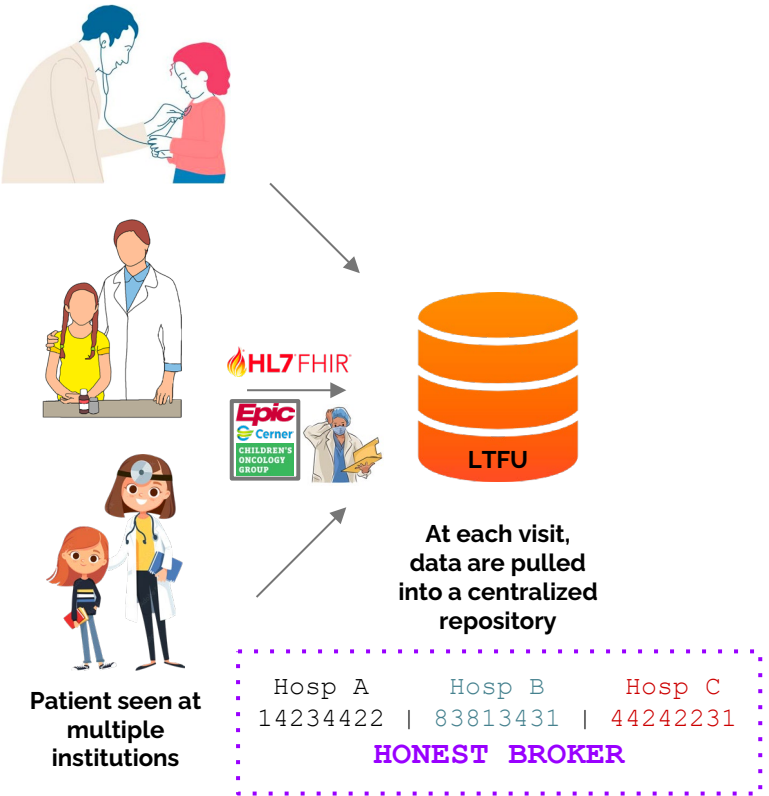
A vision for the future for long-term follow up



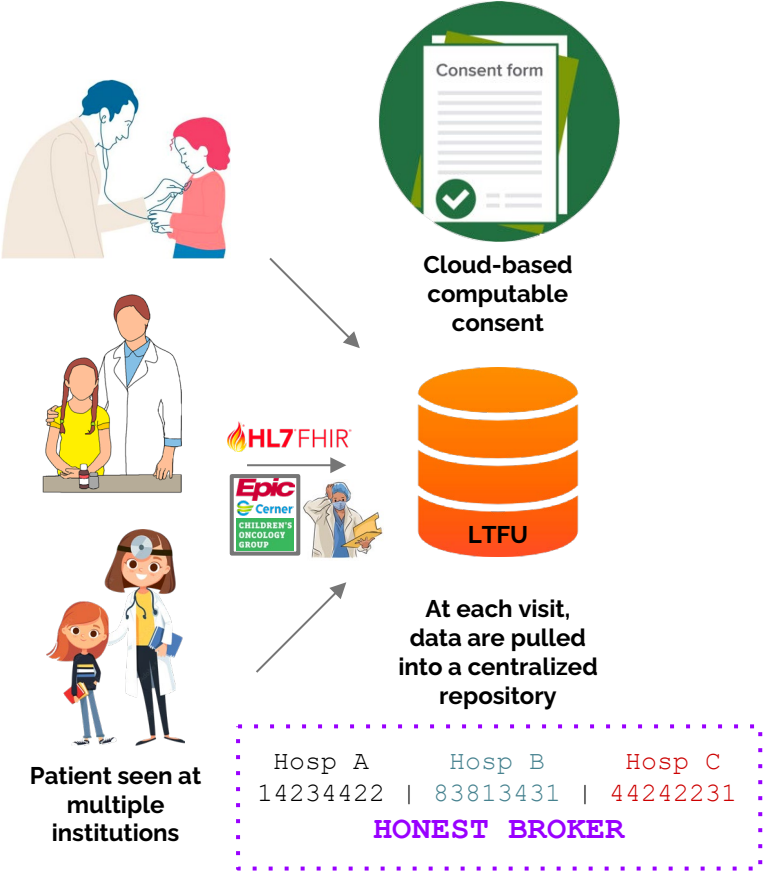
Patient seen at multiple institutions



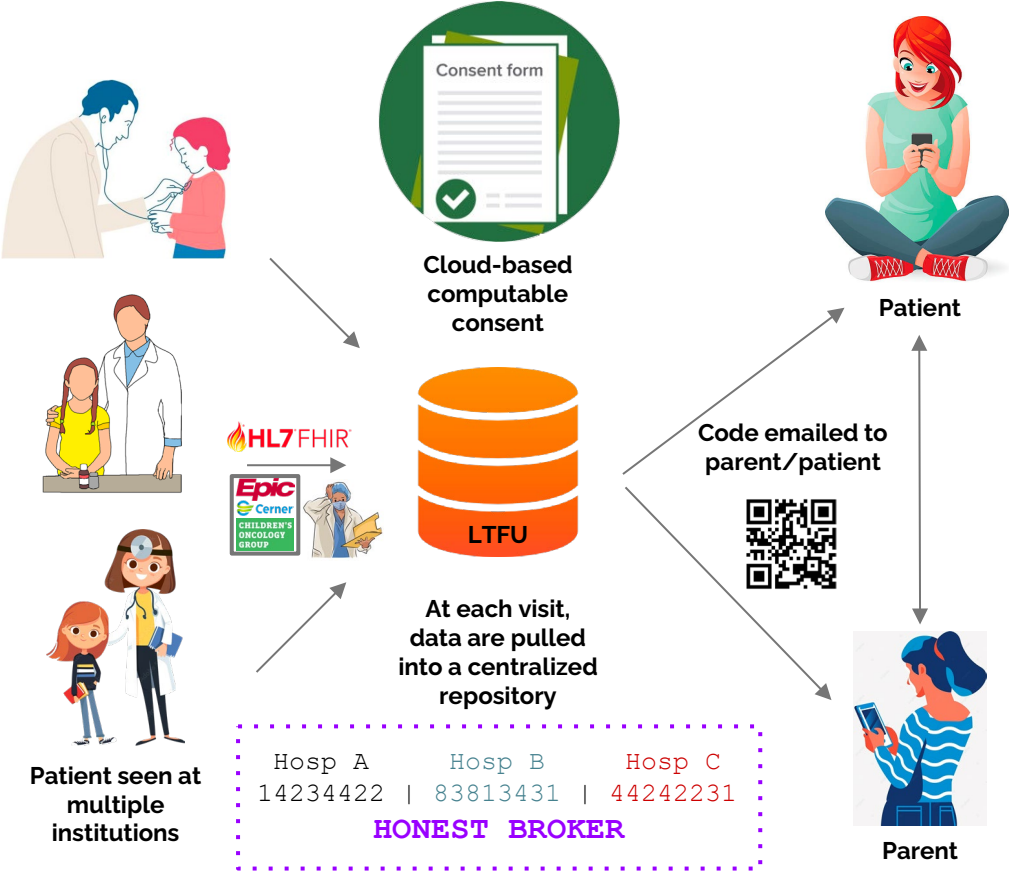
A vision for the future for long-term follow up



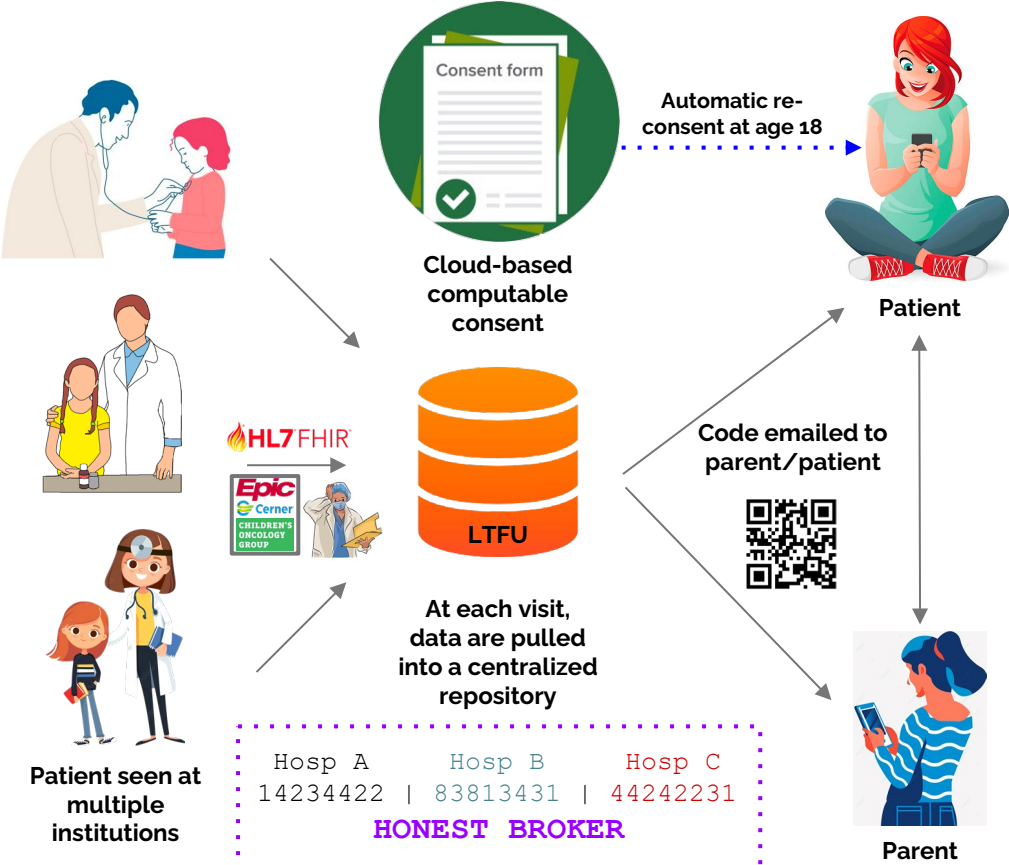
A vision for the future for long-term follow up



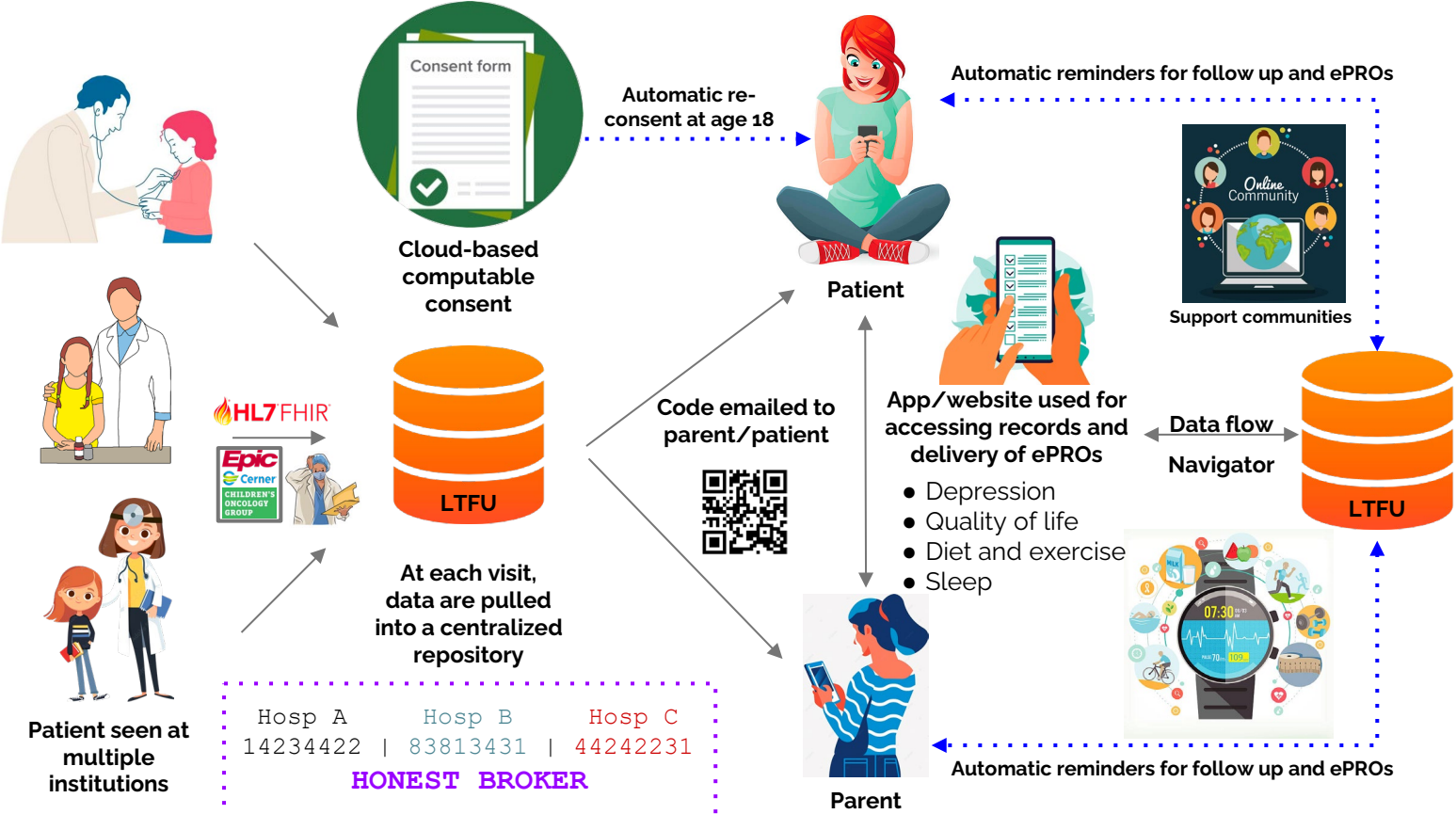
A vision for the future for long-term follow up



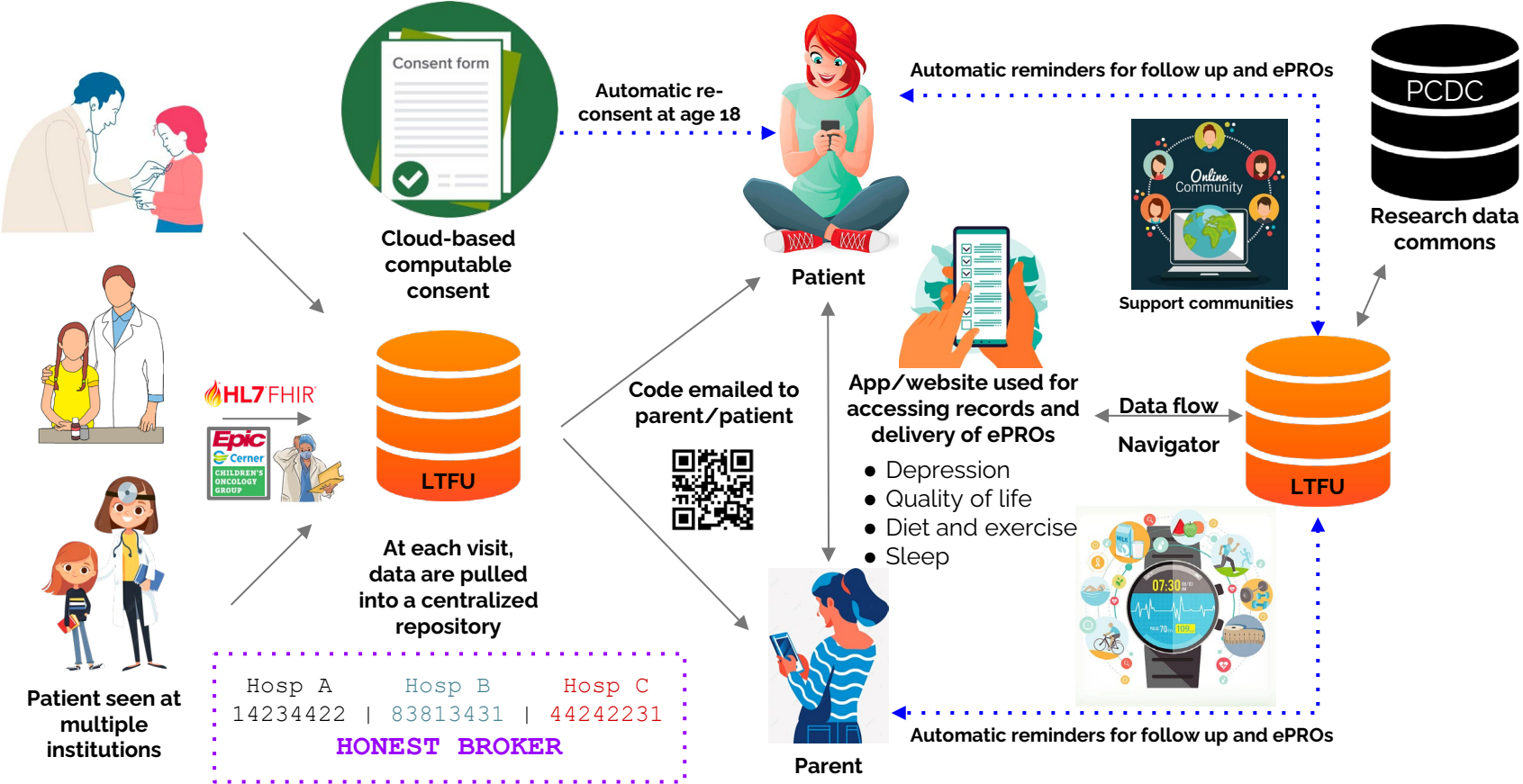
A vision for the future for long-term follow up



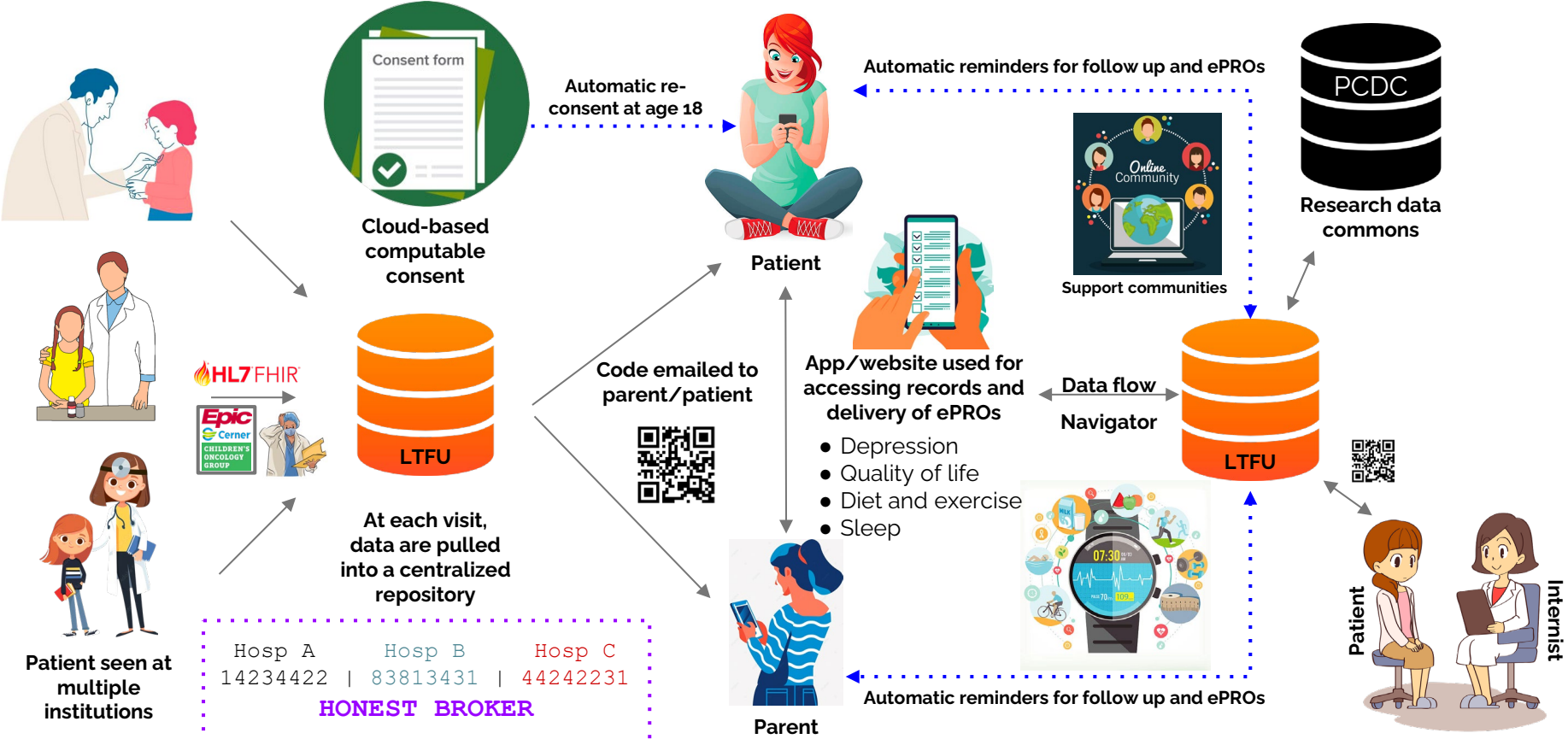
A vision for the future for long-term follow up



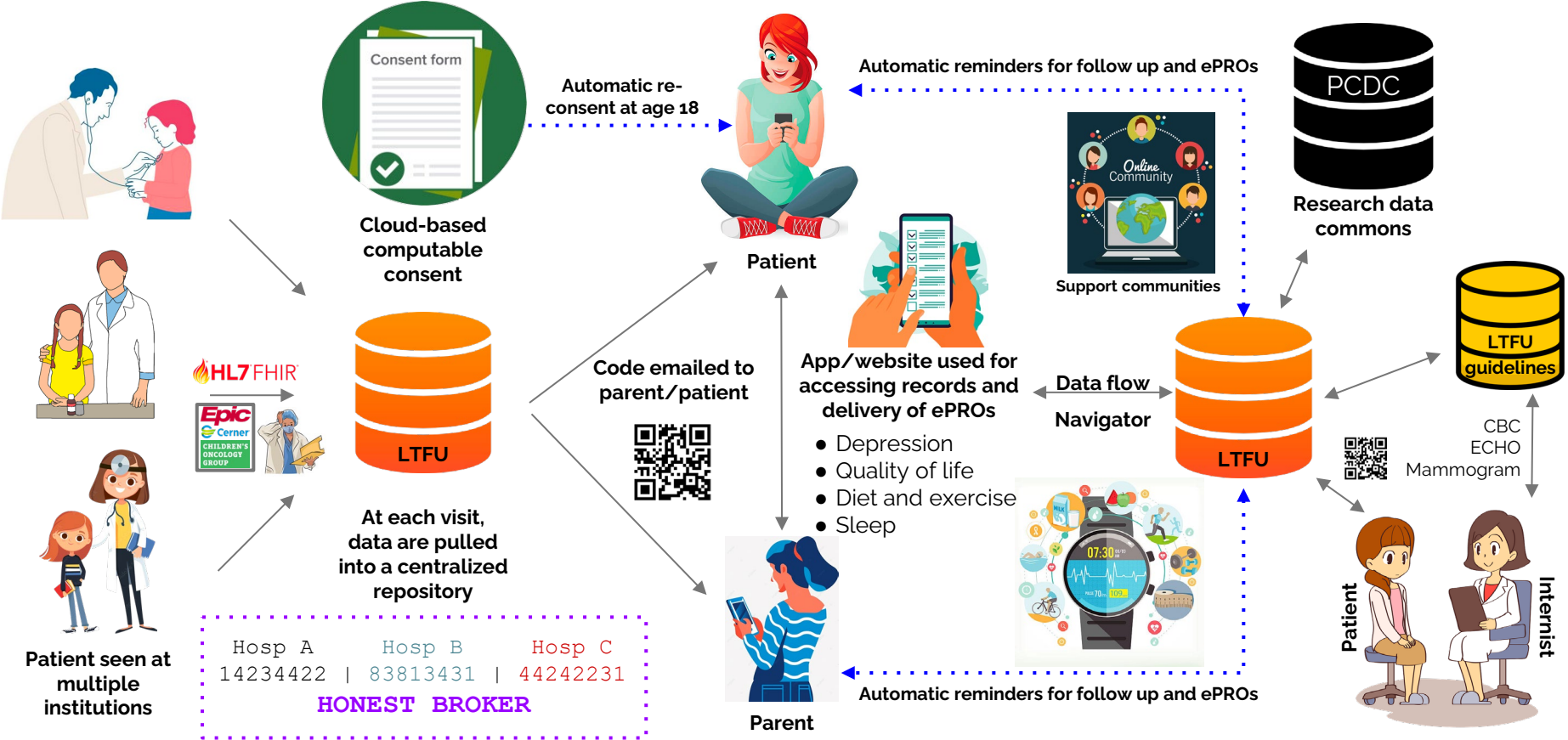
A vision for the future for long-term follow up



A vision for the future for long-term follow up



A vision for the future for long-term follow up



Hosp A Hosp B Hosp C
 14234422 | 83813431 | 44242231
HONEST BROKER

GEARBOX



THE UNIVERSITY OF
CHICAGO



DATA FOR THE
COMMON GOOD

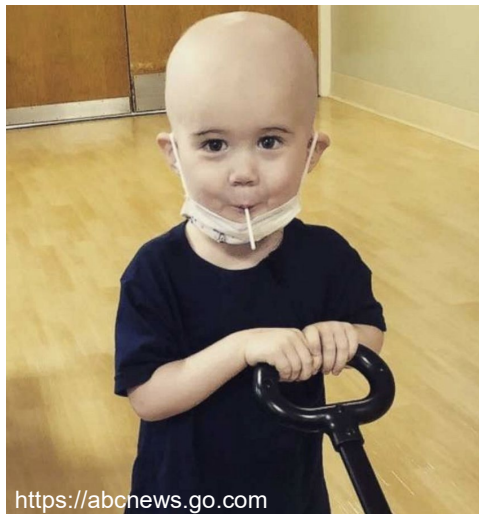
<http://sam.am/EHA2024>

commons.uchicago.edu

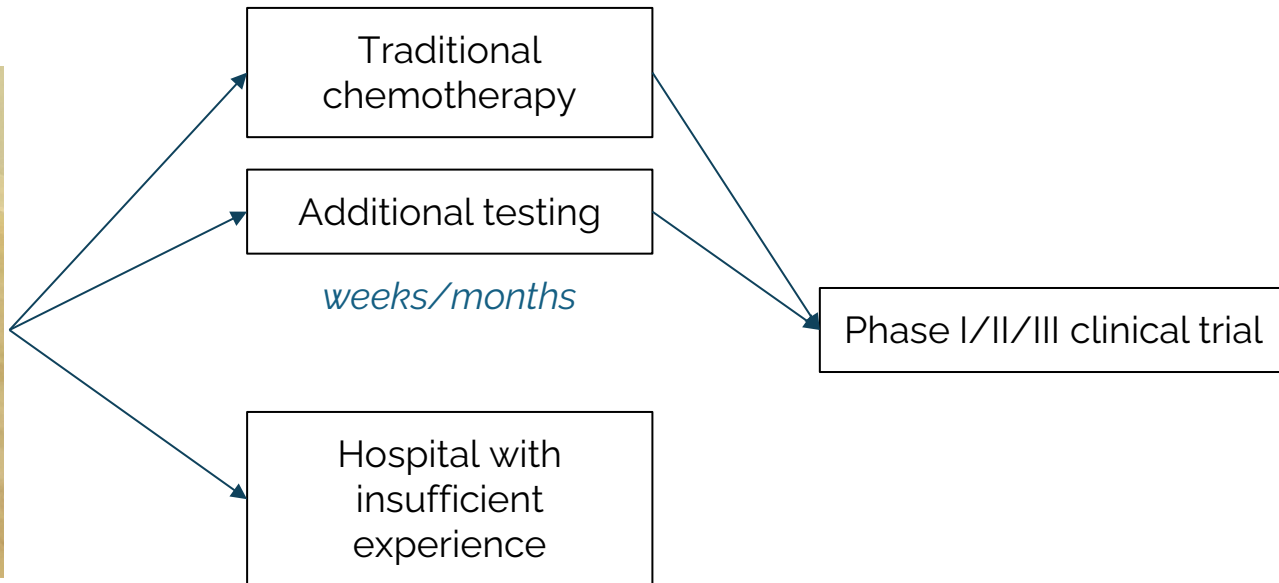


eha **Sf(PM)**

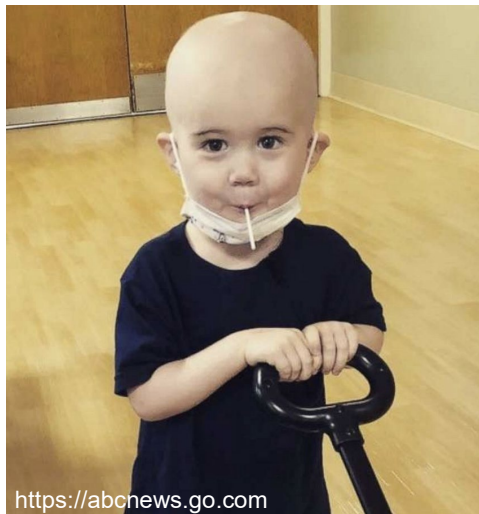
Relapsed patients struggle to find therapies



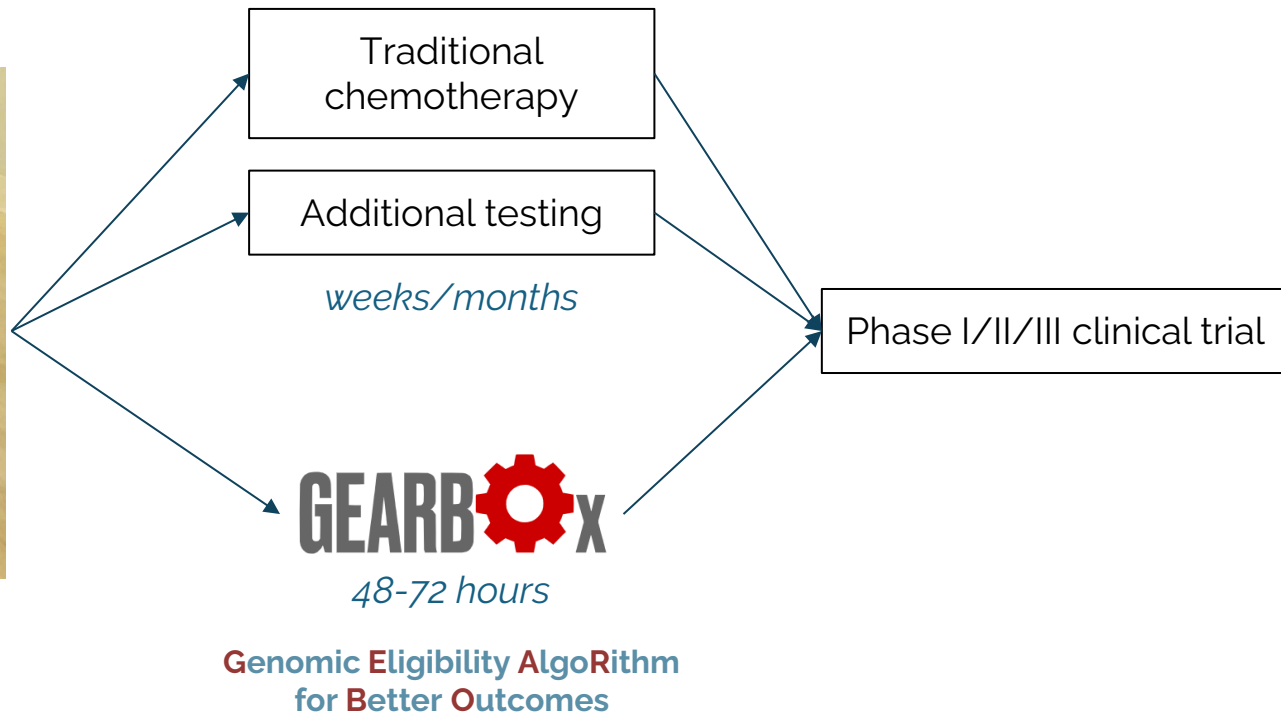
Child with relapsed AML



Relapsed patients struggle to find therapies



Child with relapsed AML



Eligibility criteria

3.2.1 Eligibility Screening

All B-ALL patients must be enrolled on APEC14B1 and consented to Eligibility Screening (Part A) prior to treatment and enrollment on AALL1731. See [Section 3.1.4](#) for timing details.

APEC 14B1 is not a requirement for B-LLy patients. B-LLy patients may directly enroll on AALL1731.

3.2.2 Age at diagnosis

Patients must be ≥ 365 days and < 10 years of age (B-ALL patients without DS)

Patients must be ≥ 365 days and ≤ 31 years of age (B-ALL patients with DS)

Patients must be ≥ 365 days and ≤ 31 years of age (B-LLy patients with or without DS)

3.2.3 White Blood Cell Count (WBC) Criteria

- B-ALL patients without DS must have an initial white blood cell count $< 50,000/\mu\text{L}$
- B-ALL patients with DS are eligible regardless of the presenting WBC

3.2.4 Diagnosis

- Patient has newly diagnosed B-cell ALL, with or without Down syndrome: $> 25\%$ blasts on a BM aspirate;

OR if a BM aspirate is not obtained or is not diagnostic of B-ALL, the diagnosis can be established by a pathologic diagnosis of B-ALL on a BM biopsy;

OR a complete blood count (CBC) documenting the presence of at least $1,000/\mu\text{L}$ circulating leukemic cells;

- OR Patient has newly diagnosed B-cell LLy Murphy Stages I or II (see [Appendix VII](#) for staging), with or without Down syndrome.

3.2.5 Exclusion Criteria

- 3.2.5.1 Patient must not have secondary ALL that developed after treatment of a prior malignancy with cytotoxic chemotherapy. Note: patients with Down syndrome with a prior history of transient myeloproliferative disease (TMD) are not considered to have had a prior malignancy. They would therefore be eligible whether or not the TMD was treated with cytarabine.

3.2.5.2 Prior Therapy

With the exception of steroid pretreatment (defined in [Section 3.3.3](#)) or the administration of intrathecal cytarabine, patients must not have received any prior cytotoxic chemotherapy for either the current diagnosis of B-ALL or B-LLy or for any cancer diagnosed prior to initiation of protocol therapy on AALL1731.

Please see [Section 4.1.4](#) for the concomitant therapy restrictions for patients during treatment.

- 3.2.5.3 For patients receiving steroid pretreatment (See [Section 3.3.3](#)), the following additional exclusion criteria apply:

- Non-DS B-ALL patients must not have received steroids for more than 24 hours in the 2 weeks prior to diagnosis without a CBC obtained within 3 days prior to initiation of the steroids.
- DS and non-DS B-LLy patients must not have received > 48 hours of oral or IV steroids within 4 weeks of diagnosis.

- 3.2.5.4 Patients who have received > 72 hours of hydroxyurea within 1 week (7 days) prior to the start of systemic protocol therapy.

- 3.2.5.5 B-ALL who do not have sufficient diagnostic bone marrow submitted for APEC14B1 diagnostic testing and who do not have a peripheral blood sample submitted containing $> 1,000/\mu\text{L}$ circulating leukemia cells.

- 3.2.5.6 Patient must not have acute undifferentiated leukemia (AUL).

- 3.2.5.7 Non-DS B-ALL patients with CNS3 leukemia (see definition in [Section 3.3.4](#), CNS status must be known prior to enrollment)

Note: DS patients with CNS3 disease are eligible but will be assigned to the DS-High B-ALL arm. CNS status must be determined based on a sample obtained prior to administration of any systemic or intrathecal chemotherapy, except for steroid pretreatment as discussed in [Section 3.3.3](#).

- 3.2.5.8 Non-DS B-ALL patients with testicular leukemia. (Note: DS patients with testicular disease are eligible but will be assigned to the DS-High B-ALL arm)

- 3.2.5.9 For LLy patients, the following additional exclusion criteria apply:

- T-Lymphoblastic Lymphoma.

GEARBOX - current state

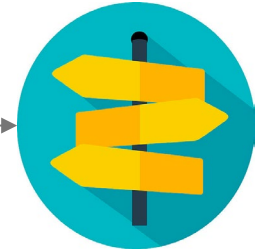
Row	Name	Status	Study Title	Conditions	Interventions	Locations
1	Active, not recruiting	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	AML, Childhood	Drug: 0007023 (BCL-2 Inhibitor)	Children's Hospital of Pittsburgh at UPMC, United States Birmingham, Alabama, United States Children's Hospital of Orange County, United States UCSF Benioff Children's Hospital San Francisco, California, United States (not 0 rows.)
2	Active, not recruiting	Randomized Controlled Trial: Comparison of Children's Supportive Care	Randomized Controlled Trial: Comparison of Children's Supportive Care	AML, Childhood	Drug: Interleukin-6, Drug: 5-azacytidine	Children's Hospital of Los Angeles, Los Angeles, California, United States Children's Hospital of Orange County, Orange, California, United States Johns Hopkins University, Baltimore, Maryland, United States (not 0 rows.)
3	Completed	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Pediatric Acute Myeloid Leukemia (AML)	Drug: Decitabine	Franklin D. Roosevelt University, Chicago, Illinois, United States Hennepin County, United States Aurora, Colorado, United States (not 0 rows.)
4	Completed	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Acute Myeloid Leukemia, AML, Childhood	Drug: Gemtuzumab, Drug: Daunorubicin, Drug: Interleukin-6	Egerton Hospital, Ayr, North Ayrshire, United Kingdom (not 0 rows.)
5	Completed	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Pediatric Acute Myeloid Leukemia	Drug: Cyclophosphamide, Drug: Doxorubicin, Drug: Mitoxantrone	Overman National University Hospital, Charlotte, North Carolina, United States Gangneung Hospital, Gangneung-si, Gangwon-do, Republic of Korea St. Mary's Hospital, Seoul, Korea, Republic of Korea (not 0 rows.)
6	Completed	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Phase 1/2a (Phase 1/2a) Study: Pediatric Patients 10-17	Childhood Acute Myeloid Leukemia	Drug: Doxorubicin, Drug: Cyclophosphamide, Drug: Vincristine	Children's Hospital of Orange County, Orange, California, United States Hennepin County, United States Lakeland Children's Hospital, Lakeland, Florida, United States (not 0 rows.)

List of clinical trials

Rules Extraction



Login



Patient navigator



Patient characteristics
Disease characteristics
Lab tests
Genomic testing

GEARBOX by LLS PedAL

ABOUT GEARBOX USER GUIDE Hello, demo LOG OUT

PATIENT INFORMATION

Demographics ^

What is the patient's current age (in years)?

What is the patient's current weight (in kg)?

Disease v

Prior treatment v

Organ function v

Biomarkers v

RESET

OPEN TRIALS

Matched (0) ^

Undetermined (9) ^

APAL2020SC ⓪ v

AAML2112 ⓪ v

APAL2020B ⓪ v

PEPN2113 ⓪ v

APAL2020D ⓪ v

AAML2020E ⓪ v

APAL2020F ⓪ v

APAL2020G ⓪ v

T2017-002 ⓪ v

Unmatched (0) ^

THE UNIVERSITY OF CHICAGO BIOLOGICAL SCIENCES PEDIATRIC CANCER DATA COMMONS LLS PedAL

Clinical trials
Information about enrollment
Study locations

GEARBox by LLS Pedal

This site is a prototype created for demo purposes only.

GEARBox ABOUT GEARBOX USER GUIDE Hello, demo LOG OUT

PATIENT INFORMATION

Demographics

What is the patient's current age (in years)?
10

What is the patient's current weight (in kg)?
40

Does most recent blast percentage measurement represent a 1 log increase from a measurement 7 days prior?
 Yes No Not sure

Disease

What is the patient's current disease?
Acute myeloid leukemia (AML)

How many occurrences of refractory disease, including the current, has the patient experienced?
0

How many confirmed or suspected relapses, including the current, has the patient experienced?
1

Is the patient currently in relapse (or suspected relapse)?
 Yes No Not sure

What is the most recent measurement of the patient's percentage of BM blasts?
25

Most recent blast percentage measured by how many methods (e.g. Flow, FISH, etc.)?
1

Has the patient experienced Grade 4 Sinusoidal Obstructive Syndrome (SOS)?
 Yes No Not sure

Does the patient have isolated EMD?
 Yes No Not sure

Does the patient have adequate BM function?
 Yes No Not sure

OPEN TRIALS

Matched (3)

- APAL2020SC
- APAL2020D
- APAL2020G

Undetermined (1)

- APAL2020B

Unmatched (5)

- AAML2112
- PEPN2113
- AAML2020E
- APAL2020F
- T2017-002

GEARBox by LLS Pedal

This site is a prototype created for demo purposes only.

GEARBox ABOUT GEARBOX USER GUIDE Hello, demo LOG OUT

PATIENT INFORMATION

Demographics

What is the patient's current age (in years)?
10

What is the patient's current weight (in kg)?
40

Does most recent blast percentage measurement represent a 1 log increase from a measurement 7 days prior?
 Yes No Not sure

Disease

What is the patient's current disease?
Acute myeloid leukemia (AML)

How many occurrences of refractory disease, including the current, has the patient experienced?
0

How many confirmed or suspected relapses, including the current, has the patient experienced?
1

Is the patient currently in relapse (or suspected relapse)?
 Yes No Not sure

What is the most recent measurement of the patient's percentage of BM blasts?
25

Most recent blast percentage measured by how many methods (e.g. Flow, FISH, etc.)?
1

Has the patient experienced Grade 4 Sinusoidal Obstructive Syndrome (SOS)?
 Yes No Not sure

Does the patient have isolated EMD?
 Yes No Not sure

Does the patient have adequate BM function?
 Yes No Not sure

OPEN TRIALS

Matched (3)

- APAL2020SC
 - Description
 - This study aims to use clinical and biological characteristics of acute leukemias to screen for patient eligibility for available pediatric leukemia sub-trials. Testing bone marrow and blood from patients with leukemia that has come back after treatment or is difficult to treat may provide information about the patient's leukemia that is important when deciding how to best treat it, and may help doctors find better ways to diagnose and treat leukemia in children, adolescents, and young adults.
 - Locations
 - Links
 - Oncology Patient Enrollment Network (OPEN)
 - LLS Clinical Trial Support Center
 - ClinicalTrials.gov
- APAL2020D
- APAL2020G

Undetermined (1)

- APAL2020B

Unmatched (5)

- AAML2112
- PEPN2113
- AAML2020E

Boolean logic

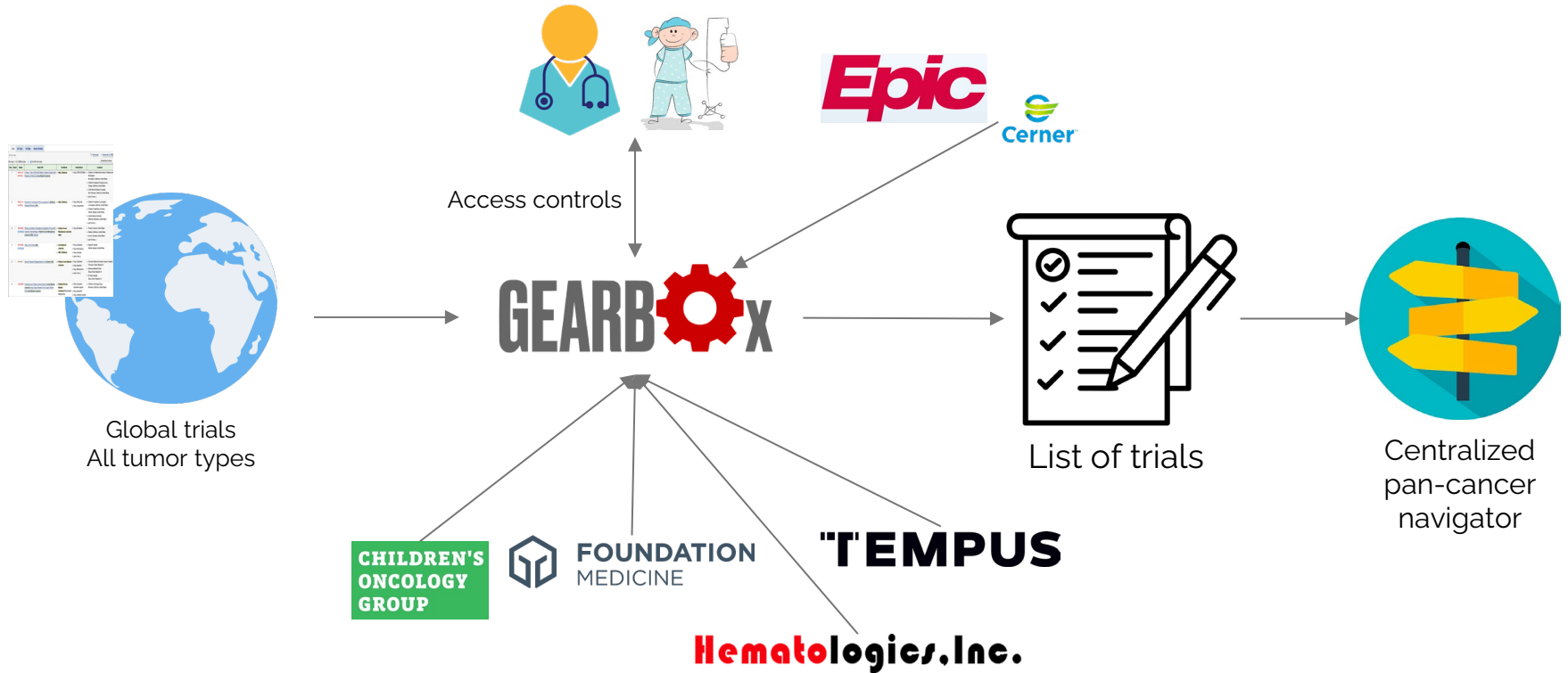
```
Eligibility Criteria for
VE
Li Eligibility Criteria for
VENAML: Study of Venetoclax in Combination With Chemotherapy in Pediatric Patients With Refractory or Relapsed Acute Myeloid Leukemia or Acute Leukemia of Ambiguous
Lineage
W
(
  (
    What is the patient's current age (in years)? is less than 24 ☺ AND
    (
      (
        What is the patient's current diagnosis? is equal to "Acute myeloid leukemia (AML)" ☺ AND
        (
          (
            Does the patient currently have, or have they in the past had, refractory disease? is equal to "Yes" ☹ AND
            Is the patient's disease currently refractory? is equal to "Yes" AND
            Current disease is refractory to how many cycles of therapy? is greater than/equal to 2
          ) OR
          (
            Does the patient currently have, or have they in the past had, confirmed or suspected relapse disease? is equal to "Yes" ☺ AND
            Is the patient currently in relapse (or suspected relapse)? is equal to "Yes" ☺
          )
        )
      )
    ) OR
    (
      What is the patient's current diagnosis? is equal to "Ambiguous lineage acute leukemia (ALAL)" ☹ AND
      (
        (
          Does the patient currently have, or have they in the past had, refractory disease? is equal to "Yes" ☹ AND
          Is the patient's disease currently refractory? is equal to "Yes" AND
          Current disease is refractory to how many cycles of therapy? is greater than/equal to 2
        ) OR
        (
          Does the patient currently have, or have they in the past had, confirmed or suspected relapse disease? is equal to "Yes" ☺ AND
          Is the patient currently in relapse (or suspected relapse)? is equal to "Yes" ☺
        )
      )
    )
  ) AND
  (
    Was the patient able to provide a recent bone marrow draw? is equal to "Yes" ☺ AND
    (
      (
        The most recent measure of bone marrow blast percentage was assessed by which method? is equal to "Morphology" ☺ AND
        What is the most recent measure of the patient's bone marrow blast percentage? is greater than/equal to 5 ☺
      )
    )
  )
)
```

Boolean logic








```
Eligibility Criteria for
V#
Lit Eligibility Criteria for
LINE VenAML: Study of Venetoclax in Combination With Chemotherapy in Pediatric Patients With Refractory or Relapsed Acute Myeloid Leukemia or Acute Leukemia of Ambiguous Lineage
W
(
  What is the patient's current age (in years)? is less than 24 ☺ AND
  (
    (
      What is the patient's current diagnosis? is equal to "Acute myeloid leukemia (AML)" ☺ AND
      (
        (
          (
            Does the patient currently have, or have they in the past had, refractory disease? is equal to "Yes" ☒ AND
            Is the patient's disease currently refractory? is equal to "Yes" AND
            Current disease is refractory to how many cycles of therapy? is greater than/equal to 2
          ) OR
          (
            Does the patient currently have, or have they in the past had, confirmed or suspected relapse disease? is equal to "Yes" ☺ AND
            Is the patient currently in relapse (or suspected relapse)? is equal to "Yes" ☺
          )
        ) AND
        (
          Was the patient able to provide a recent bone marrow draw? is equal to "Yes" ☺ AND
          (
            (
              The most recent measure of bone marrow blast percentage was assessed by which method? is equal to "Morphology" ☺ AND
              What is the most recent measure of the patient's bone marrow blast percentage? is greater than/equal to 5 ☺
            )
          )
        )
      )
    )
  )
)
```



GEARBOX - future state



② Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia

Samuel Kaskovich, MD¹; Kirk D. Wyatt, MD² ; Tomasz Oliwa, PhD³ ; Luca Graglia, MS⁴ ; Brian Furner, MS⁴ ; Jooho Lee, PhD⁴ ; Anoop Mayampurath, PhD⁵ ; and Samuel L. Volchenboum, MD, PhD⁴ 

DOI <https://doi.org/10.1200/CCI.23.00009>

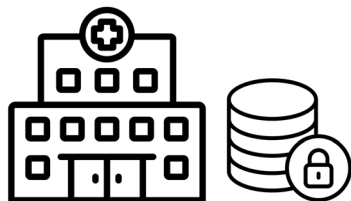
- Records from pediatric leukemia clinical trials were downloaded from ClinicalTrials.gov.
- Regular expressions were used to discretize and extract individual trial criteria.
- Multilabel SVM trained to classify sentence embeddings of criteria into clinical categories.
- Labeled criteria parsed using RegEx to extract numbers, comparators, and relationships.
- Multilabel SVM demonstrated a pooled accuracy of 75%.
- Text processing pipeline automatically extracted 68% of eligibility criteria rules (vs. 80% manual)
- Automated matching was accomplished in approximately 4 seconds (vs. hours for manual)

What's next for data collection?



Electronic Health Record (EHR) Data Transfer Models

Bulk Institutional Transfer












- Multiple connections
- Legal/regulatory barriers
- Competitive interests

Pilot study

JCO® Clinical Cancer Informatics May 30, 2024

⑧ Extracting Electronic Health Record Neuroblastoma Treatment Data With High Fidelity Using the REDCap Clinical Data Interoperability Services Module

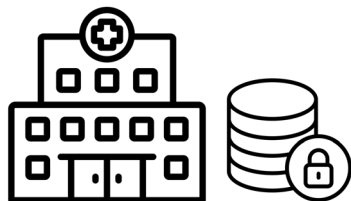
Brian Furner, MS¹ ; Alex Cheng, PhD² ; Ami V. Desai, MD, MSCE¹ ; Daniel J. Benedetti, MD, MA³ ; Debra L. Friedman, MD, MS³ ; Kirk D. Wyatt, MD⁴ ; Michael Watkins, PhD¹ ; Samuel L. Volchenboun, MD, PhD¹ ; and Susan L. Cohn, MD¹ 

Proof of concept study:

- Chemotherapy and immunotherapy data from patients at the University of Chicago (UChicago) and Vanderbilt University Medical Center (VUMC) enrolled on ANBL00B1 were extracted from the EHR into REDCap using CDIS
- The accuracy and completeness of the extracted medication data was assessed against gold standard data warehouses at UChicago and VUMC

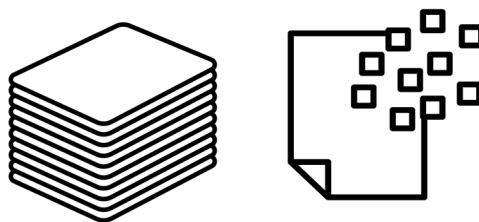
Electronic Health Record (EHR) Data Transfer Models

Bulk Institutional Transfer



- Multiple connections
- Legal/regulatory barriers
- Competitive interests

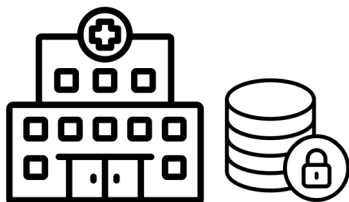
Individual Records Request



- Manual
- Slow
- Costly

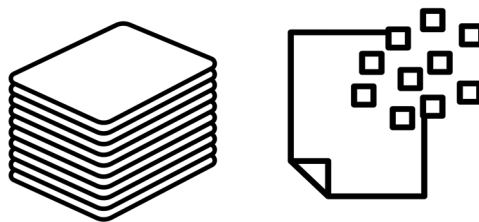
Electronic Health Record (EHR) Data Transfer Models

Bulk Institutional Transfer



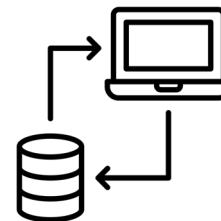
- Multiple connections
- Legal/regulatory barriers
- Competitive interests

Individual Records Request



- Manual
- Slow
- Costly

Individual API Transfer

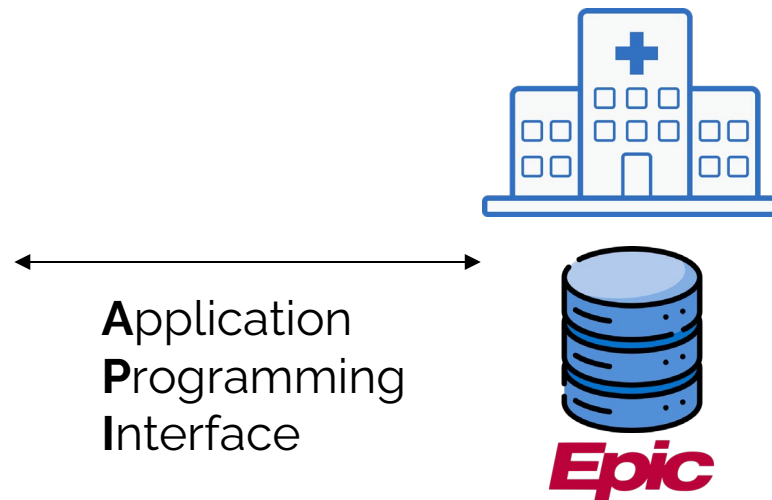
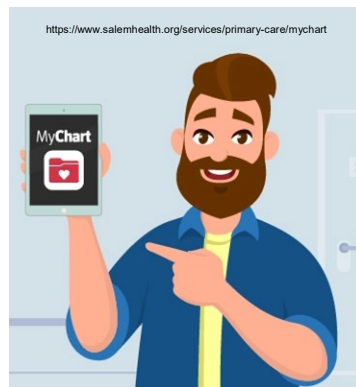


- Rapid
- Easily updated
- Limited dataset

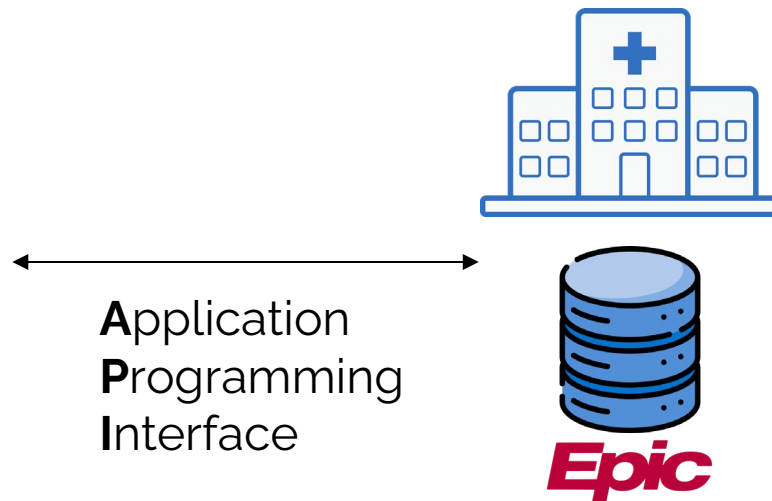
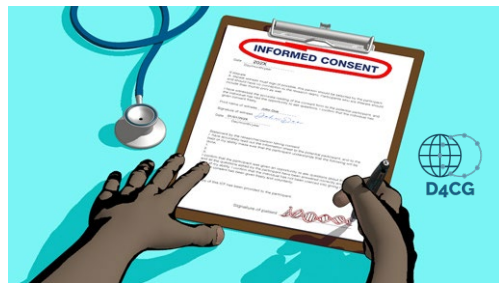
What is a patient access API?



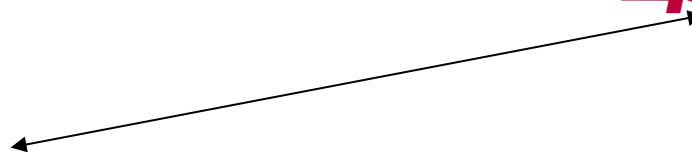
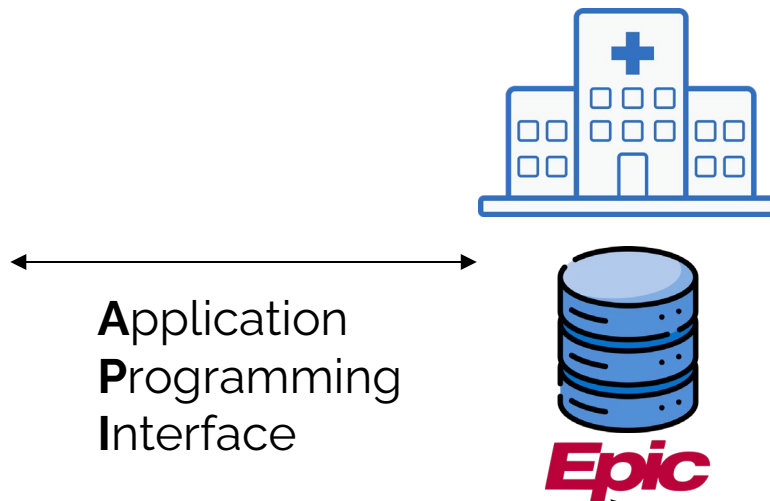
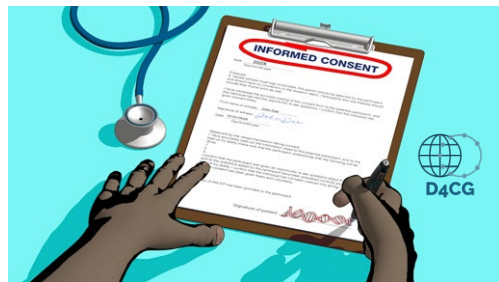
What is a patient access API?



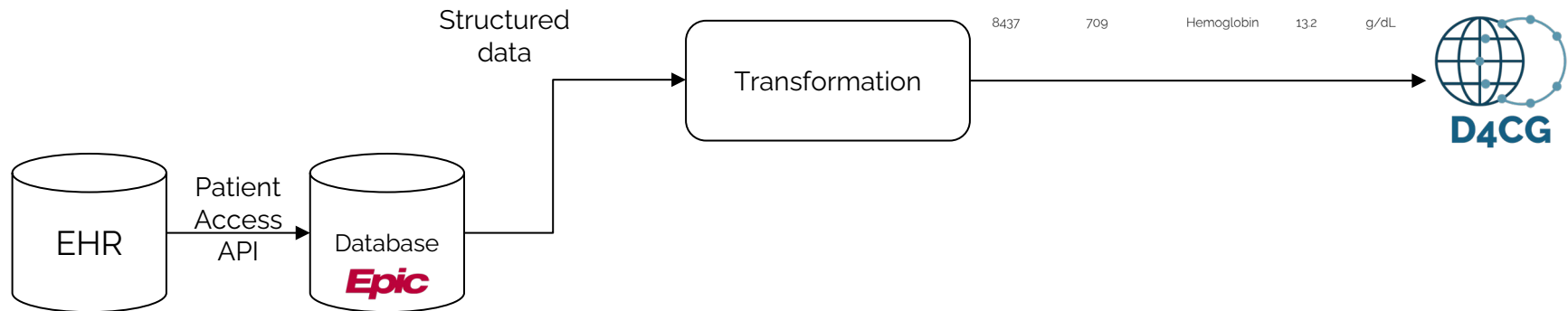
What is a patient access API?



Leverage patient access APIs



Data pipeline



Patient ID	Age (days)	Lab test	Value	Units
3452	320	Hemoglobin	12.5	g/dL
3452	548	Hemoglobin	11.0	g/dL
3452	1095	Hemoglobin	9.5	g/dL
8437	400	Hemoglobin	13.0	g/dL
8437	534	Hemoglobin	13.6	g/dL
8437	709	Hemoglobin	13.2	g/dL

Data pipeline



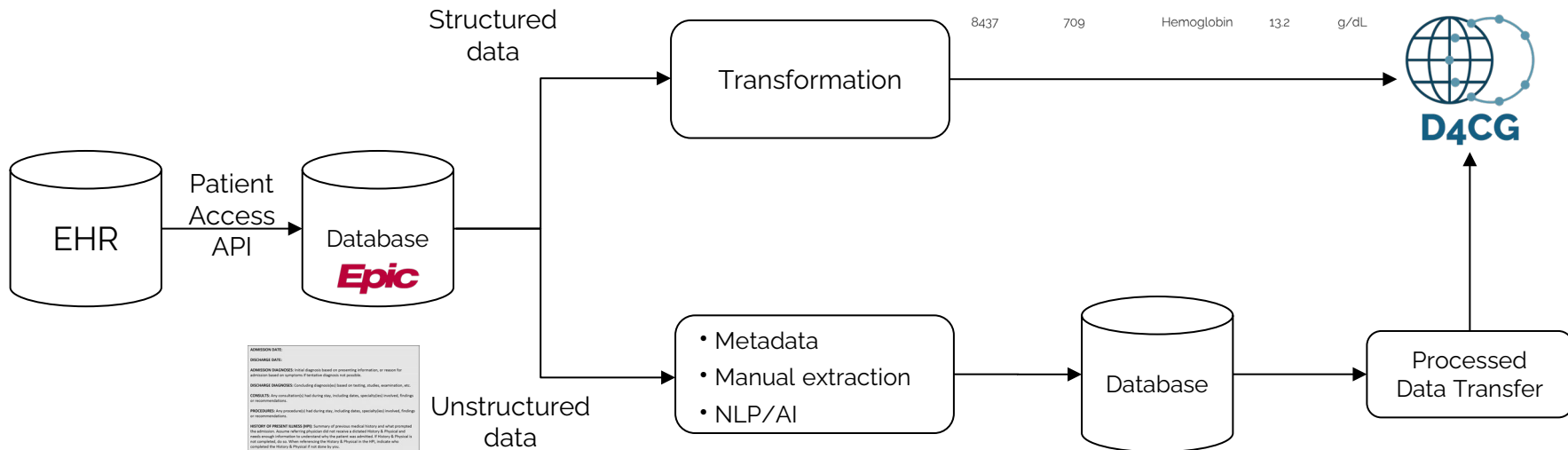
Patient ID	Age (days)	Lab test	Value	Units
3452	320	Hemoglobin	12.5	g/dL
3452	548	Hemoglobin	11.0	g/dL
3452	1095	Hemoglobin	9.5	g/dL
8437	400	Hemoglobin	13.0	g/dL
8437	534	Hemoglobin	13.6	g/dL
8437	709	Hemoglobin	13.2	g/dL

ADMISSION DATE:
 DISCHARGE DATE:
 ADMISSION DIAGNOSES: Initial diagnosis based on presenting information, or reason for admission based on symptoms if relational diagnosis not possible.
 DISCHARGE DIAGNOSES: Concluding diagnosis(es) based on testing, studies, examination, etc.
 CONSULTS: Any consultation(s) had during stay, including dates, specialty(ies) involved, findings or recommendations.
 PROCEDURES: Any procedure(s) had during stay, including dates, specialty(ies) involved, findings or recommendations.
 HISTORY OF PRESENT ILLNESS (HPI): Summary of previous medical history and what prompted the admission. Assume referring physician did not receive a dictated History & Physical and needs enough information to understand why the patient was admitted. If History & Physical is not completed, do so. When referencing the History & Physical in the HPI, include who completed the History & Physical if not done by you.
 HOSPITAL COURSE: Consider what information would be important for you as the primary, or receiving physician seeing the patient in follow-up. Be succinct and only include pertinent information.
 DISCHARGE TO: Home or facility, include homecare if applicable.
 DISCHARGE CONDITION: One line summary of patient's condition.
 DISCHARGE MEDICATIONS: Include doses, frequency, length of therapy, and any changes to pre-existing medications.
 DISCHARGE INSTRUCTIONS: List all instructions that were written on patient's discharge form.
 PENDING LABS: List all lab results that have not yet arrived at time of dictation, as well as any lab results that arrived between time of discharge and time of dictation.
 FOLLOW-UP: List all follow-up appointments with dates, times, names of physicians/services involved, and contact information.
 COPY TO: Require a copy sent to the primary care provider (PCP) which includes PCP's fax, address and phone number.

What about text notes?



Data pipeline



Patient ID	Age (days)	Lab test	Value	Units
3452	320	Hemoglobin	12.5	g/dL
3452	548	Hemoglobin	11.0	g/dL
3452	1095	Hemoglobin	9.5	g/dL
8437	400	Hemoglobin	13.0	g/dL
8437	534	Hemoglobin	13.6	g/dL
8437	709	Hemoglobin	13.2	g/dL

DISCHARGE DATE:
DISCHARGE DATE:
ADMISSION DIAGNOSIS: Initial diagnosis based on presenting information, or reason for admission based on symptoms. Primary diagnosis not possible.
DISCHARGE DIAGNOSIS: Concluding (diagnoses) based on testing, studies, examination, etc.
COMMENTS: Any comments not during stay, including dates, unspecified medical, findings or observations.
PROCEEDS: Any procedure(s) had during stay, including dates, unspecified (included, findings or observations).
HISTORY OF PRESENT ILLNESS (HPI): Summary of previous medical history and other pertinent background, including description of the nature of symptoms and physical findings, as well as the course of the illness, including the date of onset of symptoms, if known, and the date of admission to the hospital, if applicable, and the date of discharge, if applicable.
HOSPITAL COURSE: Consider what information would be important for you as the primary, or attending physician caring for the patient in future as he or she and you complete patient's discharge.
DISCHARGE TO: Home or facility, include homecare if applicable.
DISCHARGE CONDITION: One line summary of patient's condition.
DISCHARGE INSTRUCTIONS: Include dates, frequency, length of therapy, and any changes to previous instructions.
DISCHARGE INSTRUCTIONS: List of instructions that were written on patient's discharge form.
PHYSICIAN LABS: List of all tests that have not yet passed or time of discharge, as well as any tests done at or between time of discharge and time of admission.
PHYSICIAN OP: List of follow-up appointments with dates, times, names of physician/clinician, location and contact information.
CCFV TO: Request a copy sent to the primary care provider (PCP) which includes PCP file, address and phone number.



Learning points

- Pediatric cancer is rare
- While overall survival for pediatric cancer is good, many types still have dismal outcomes
- Collecting and harmonizing data are keys to overcoming the paucity of data
- Clinical trials are not standardized
- Automated EHR data extraction holds promise of richer and cleaner data
- Survivors require better resources for long-term follow up

 LEUKEMIA & LYMPHOMA SOCIETY™	 St. Baldrick's FOUNDATION Conquer Childhood Cancers	 FOR CHILDHOOD CANCER RESEARCH Rally! FOUNDATION	 Children's Cancer Research Fund™	 CANCER RESEARCH FOUNDATION	Chan Zuckerberg Initiative 	FOUNDATION gray.	THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST
 ICI INNOVATION IN CANCER INFORMATICS	 BRIGHTSIDE	 ANDREW McDONOUGH B+ FOUNDATION™	 INFINITE LOVE for KIDS FIGHTING CANCER	 THE TRUTH 365	 BRIDGE TO A CURE foundation	 FANCONI ™ CANCER FOUNDATION	 MIB AGENTS
 The Neuroblastoma Children's Cancer Society	 CHILDREN'S RESEARCH FOUNDATION <i>"So they may live"</i>	 Children's Neuroblastoma Cancer Foundation	 leidos	 SUMMER'S WAY	 MADDIE'S PROMISE	 FRIENDS OF T1 FOUNDATION	SEBASTIANSTRONG FOUNDATION
 CHILDREN'S ONCOLOGY GROUP Foundation	 THE MULLIN FUND Fighting To Beat Sarcoma	 Seattle Children's HOSPITAL • RESEARCH • FOUNDATION	 KAT'S CREW	 Alex's Lemonade Stand FOUNDATION FOR CHILDHOOD CANCER™	 LITTLE HEROES	Sarah Jane Adicoff Endowment for Research in Rhabdomyosarcoma	 Jeffrey Pride Foundation For Pediatric Cancer Research
The Brumfield Family	 SAMMY'S SUPERHEROES FOUNDATION	 AT THE FOREFRONT OF UChicago Medicine Comprehensive Cancer Center	 AT THE FOREFRONT OF UChicago Medicine Comer Children's Development Board	Aileen S. Andrew Foundation	 THE MATTHEW BITTKER FOUNDATION It's What Matters	 King Baudouin Foundation	Mr. Daniel Tierney
The William and Evelyn Fuchs Family Foundation	 kick cancer	A gift made in memory of Payton O'Brien	 WILLIAM GUY FORBECK RESEARCH FOUNDATION wgrf.org	United States Department of the Interior	 The Super Jake Foundation	 NIH NATIONAL CANCER INSTITUTE	

Thank you



Last but not least, we are deeply grateful to the **patients and families** who make our work possible.



THE UNIVERSITY OF
CHICAGO



DATA FOR THE
COMMON GOOD

<http://sam.am/EHA2024>

commons.uchicago.edu



eha **Sf(PM)**