Workshop #1

Harmonization of Clinical and Biological Data

Session 2

Eva Steliarova-Foucher, PhD

Serban Negoita, MD, DrPH

Workshop General Goals and Structure

Goal: Discuss strategies and plan an international partnership to advance the harmonization of biological and clinical data in support of childhood cancer research

Session 1: Core data elements: patient, tumor, prognostic factors

Session 2: Treatment and outcome + standards mapping

Session 3: Genetic/molecular data + challenges to harmonization

Session 4: Summary of discussions + pilot project

Intro to session 2 Topics

- 24% of young patients who had their tumors tested for genetic changes were eligible to receive one of the targeted therapies
- Paucity of standards for collecting treatment data
- Surveillance data limited capacity to collect/analyze long-term treatment
- Treatment data perceived as increasing the risk of patient reidentification
- Number of patients receiving testing and targeted therapy is small, increased need for data collection standardization

Intro to session 2 Topics (continued)

- Studying survival as an outcome, by stage and geographic area/healthcare delivery system is important (see BENCHISTA, CRICCS)
- Clinical outcomes variables particularly relevant for childhood cancers:
  - Recurrence/relapse
  - Comorbidities/quality of life
  - Large variability in limited sample sizes

Session-specific questions for consideration

- What data standards support code data elements discussed in session 1 and 2?
- Discuss the strength and limitations of existing standards and their creation
- Identify standard setters and adhering organizations/agencies; feasibility of using the standards for international projects
- Identify conversions/ crosswalks between items/standards used by various organizations/agencies

Session 2 Topics and Discussants

- **Dr. Paul Gibson**

Treatment and Outcomes Data

- **Dr. Bastien Rance**

Crosswalks and Data Standards in Oncology

Discussants

Paul GIBSON
Canada
McMaster Children's Hospital
Pediatric Oncologist

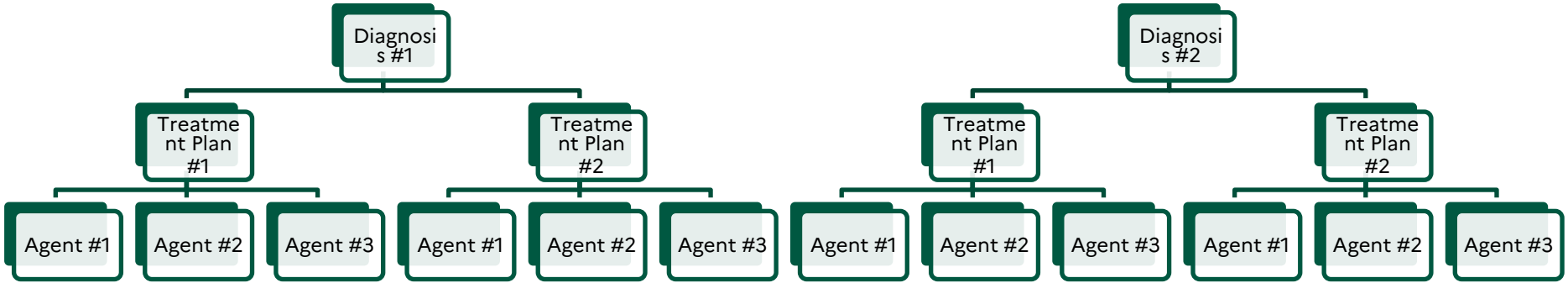
Bastien RANCE
France
Université Paris Cité
AP-HP Paris Hospital
Associate professor of medical informatics

Treatment and Outcomes: Canadian Approach to challenging measurements

Treatment and Outcomes

- Why collect?
  - Study of late effects (including dose-response)
  - Real World Evidence/Post Trial/Market review of therapies
  - Pool experience in Rare/Ultra-rare diagnoses
    - What has been tried? What worked?

Treatment Data Collection: Canadian Approach



Treatment Data: Systemic Therapy

- Treatment Plan: Name (a problem!), Start and Stop Dates
  - Clinical Trial? (Study data concerns)
- Agents (Chemotherapy)
  - Name
  - ?Date of First dose?
  - Cumulative doses (All or select agents?)
    - Anthracyclines
    - Alkylators
    - Platinum Agents
    - HD methotrexate

Treatment Data: Local Control

- Radiation Therapy
  - Modality (Photon, Proton, Brachy, etc.)
    - Nuclear Medicine Therapies?
  - Site, Dose, ?Fractions?, ?Boost?
- Surgery
  - What Types?
    - Biopsy?
    - Tumour Resection
    - Supportive Care/Toxicity induced?
  - Site?
  - Outcome?

9.6 Radiation site: Select all that apply
Specify general site: e.g. kidney = abdomen (left or right)

- ☐ Abdomen – hemi
- ☐ Abdomen/flank – whole
- ☐ Abdomen/flank – left
- ☐ Abdomen/flank – right
- ☐ Brain: infratentorial
- ☐ Brain: partial
- ☐ Brain: supratentorial
- ☐ Brain: whole
- ☐ Chest wall – left
- ☐ Chest wall – right
- ☐ Craniospinal
- ☐ Face
- ☐ Inverted Y nodes
- ☐ Limb – lower – left
- ☐ Limb – lower – right
- ☐ Limb – upper – left
- ☐ Limb – upper – right
- ☐ Liver
- ☐ Lung-bilateral
- ☐ Lung-left
- ☐ Lung-right
- ☐ Other, specify: _____
- ☐ Lymph nodes - abdominal
- ☐ Lymph nodes - axilla
- ☐ Lymph nodes - head and neck
- ☐ Lymph nodes - inguinal/femoral
- ☐ Lymph nodes - Mediastinum/hilar
- ☐ Lymph nodes - pelvic
- ☐ Lymph nodes - other
- ☐ Mantle nodes
- ☐ Mediastinum
- ☐ Nasopharynx
- ☐ Neck
- ☐ Orbit – Left
- ☐ Orbit – Right
- ☐ Parotid
- ☐ Pelvis
- ☐ Scalp
- ☐ Skull
- ☐ Spine- cervical
- ☐ Spine- lumbar
- ☐ Spine- sacrum
- ☐ Spine- thoracic
- ☐ Spine- whole
- ☐ Spleen
- ☐ Testis
- ☐ Not available

Outcomes

- Easy: Death, Relapse
- Harder: “Refractory”: How do we define? Disease Specific?
- Even HARDER: Progression (i.e. Low Grade Gliomas)
  - Clinical vs. radiologic?
- Complications:
  - Which ones? How specific?

Acknowledgements

- Cancer in Young People in Canada (CYP-C) Management and Partners Committees
- Public Health Agency of Canada
- Pediatric Oncology Group of Ontario POGONIS Team

Discussants

Paul GIBSON
Canada
McMaster Children's Hospital
Pediatric Oncologist

Bastien RANCE
France
Université Paris Cité
AP-HP Paris Hospital
Associate professor of medical informatics

Crosswalks and data standards in oncology

Bastien Rance, PhD
AP-HP, Université Paris Cité

National registries and their adoption of standard data elements or standard

Numerous national registries **share data** at international level (e.g. with the International Agency for Research on Cancer, in CI5 - cancer incidence in five continents)

Minimal list of items with value sets (e.g. ICD-O-, ICD-10, TNM)

Wide *heterogeneity* in the data beyond the minimal dataset

Difficulties to produce an exhaustive picture of the adoption or use of controlled vocabularies and data standards

Initiatives in place or on-going to create exhaustive national registries, but are unclear regarding the use of data models, standards terminologies and value sets

Historical initiatives toward data sharing and data standardization, e.g.:

- CaBIG (National Cancer Institute's Cancer Bioinformatics Grid)
  - Common Data Elements and controlled terminologies

Variable	Format	Unknown/Missing	Definition/Notes
Ethnic group		99	If available, Read more
*Patient ID		Not allowed	Unique ID of the patient in the registry Read more
Tumour sequence #	00 = single tumour 01 = 1st of several tumours 02 = 2nd of several tumours...	99	If only one tumour is identified for a patient, this variable should contain "00". If several tumours are identified, numbering should start at "01". Read more
*Date of Birth	YYYYMMDD	99999999	Read more
*Sex	1 = Male 2 = Female	9	
*Date of Incidence	YYYYMMDD	Not allowed	Read more
*Age in Years		999	Last completed year of age: <1 = 0, >99 = 100. Read more
*ICDO-3 Topography	ICDO-3 Definition (with letter C)	Not allowed	E.g. C531
*ICDO-3 Morphology	ICDO-3 Definition	Not allowed	E.g. 8170
*ICDO-3 Behaviour	ICDO-3 Definition	Not allowed	E.g. 3
*Basis of Diagnosis	0 Death certificate only 1 Clinical 2 Clinical investigation 4 Specific tumor markers 5 Cytology 6 Histology of a metastasis 7 Histology of a primary tumor 9 Unknown	9	ICDO-3 Definition Read more
Vital Status	1 = Alive 2 = Dead 3 = Lost to follow-up	9	Read more
Date of Last Contact	YYYYMMDD	99999999	Read more

http://www.iacr.com.fr/images/doc/CI5/CI5-XII%20Call%20for%20Data%20ENGLISH_FINAL20211023.pdf

The French OSIRIS initiative

Started in 2013/2014 by the National SIRICs (Integrated Cancer Research Sites, French comprehensive cancer centers). Funded by the *French National Cancer Institute*.
Results of a large collaboration, and numerous discussions with all the partners
Aim at simplifying **data sharing**

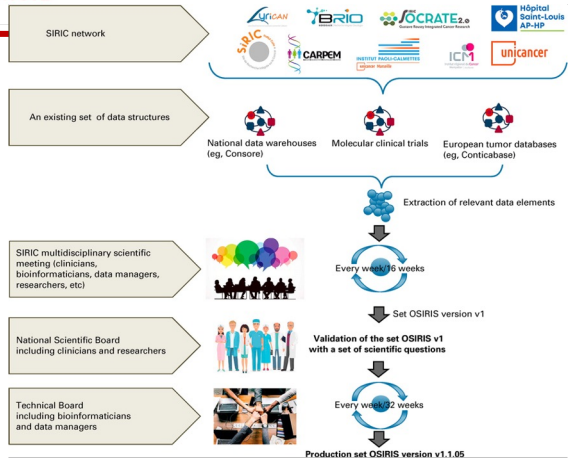
A modular minimal dataset in oncology including:

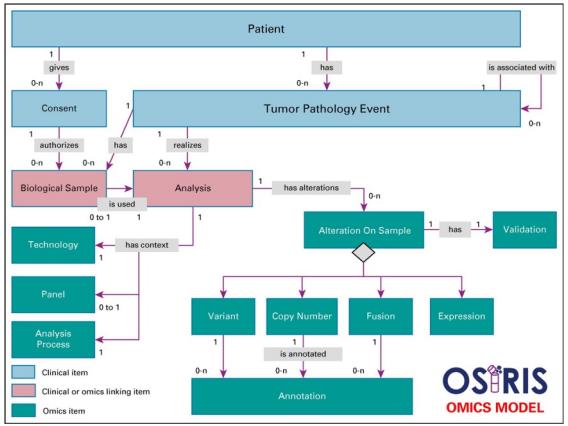
- Clinical
- Molecular

And a **Data Model**

Together with the minimal dataset, includes mapping to **standard terminologies**, definitions, implementation guides

Compatible with HL7 FHIR





OSIRIS: A Minimum Data Set for Data Sharing and Interoperability in Oncology
Julien Guérin, Yec'han Laizet, Vincent Le Texier, et al. JCO Clinical Cancer Informatics 2021 :5, 256-265

mCode (Minimal Common Oncology Data Elements)

Created in 2018 and produced by a working group of the American Society of Clinical Oncology (ASCO).
Aims at standardization of ensure consistency of data collection and enable data sharing

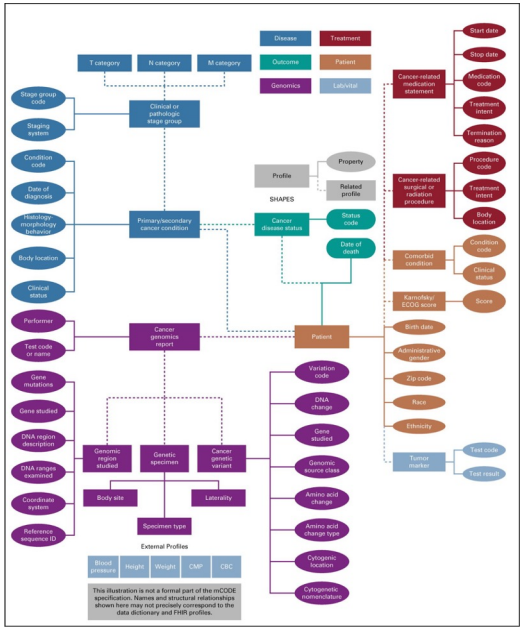
Both a **standard dataset** and **a model** (organized around the patient)

- Clinical
- Molecular

Consists in about 30 **FHIR profiles**, and data elements

Relies on standard terminologies

Provides both data elements and value sets



Improving Cancer Data Interoperability: The Promise of the Minimal Common Oncology Data Elements (mCODE) Initiative
Travis J. Osterman, May Terry, and Robert S. Miller. JCO Clinical Cancer Informatics 2020 :4, 993-1001

The OMOP Oncology CDM extension

Initiative within the Observational Health Data Sciences and Informatics (OHDSI) community.
Extension of the OMOP Common Data Model to better cover cancer data integration

Provides a shared model across OMOP data warehouses worldwide

Extends the model with a cancer episodes (an **event-based** approach)
Identifies key concepts and **standard terminologies** to better integrate cancer data

TABLE 1. Evaluation of Seven Vocabularies for Fitness for Purpose as Sources of Standard Concepts

Category	Evaluation	ICD-O-3	NCIt	NAACCR	CAP	AJCC	HemOnc	ATC
Topology	Presence	x	x	x	x			
	Completeness	x	x		x			
	Ontological principles	x						
Histology	Presence	x	x	x	x			
	Completeness	x	x		x			
	Ontological principles	x						
Staging categories	Presence		x	x	x	x		
	Completeness		x			x		
	Ontological principles					x		
Pathologic characteristics	Presence		x	x	x	x		
	Completeness		x	x	x			
	Ontological principles			x	x			
Genomic markers	Presence		x	x	x	x		
	Completeness							
	Ontological principles		x					
Drug classifications	Presence		x				x	x
	Completeness		x				x	x
	Ontological principles						x	x
Drug regimens	Presence		x				x	
	Completeness						x	
	Ontological principles						x	

Abbreviations: AJCC, American Joint Committee on Cancer; ATC, Anatomical Therapeutic Chemical; CAP, College of American Pathologists; HemOnc, Hematology/Oncology; ICD-O-3, International Classification of Diseases for Oncology, 3rd Edition; NAACCR, North American Association of Central Cancer Registries; NCIt, National Cancer Institute Thesaurus.

Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research
Rimma Belenkaya, Michael J. Gurley, Asieh Golozar, et al. JCO Clinical Cancer Informatics 2021 :5, 12-20

The need for crosswalks across countries

Several national and international initiatives of standard minimal datasets:

- Includes the basic common data elements (with ICD, ICD-O-3…)
- Aligned to controlled vocabularies
  - For data elements
  - And value sets

Propositions usually go further (e.g. include molecular data) than solutions already in use

Proposition of data models, probably needed to enable homogeneous data collection and terminology alignment process

Moderate adoption as of 2023: alignments are labor intensive, data complex to collect in uniform way

Two main objectives: data sharing (with exchange of the data) and data harmonization (with exchange of the data model and elements)