NIH NATIONAL CANCER INSTITUTE

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INSTITUT NATIONAL DU CANCER

Paris conference on GPCD strategy

Workshop sessions key takeways

November 7/8 2023

Facilitators

Serban NEGOITA
US National Cancer Institute

Eva STELIAROVA-FOUCHER
International Agency for Research on Cancer

Discussants

Discussants session 1

Gudrun SCHLEIERMACHER
France
Curie Institute
Practitioner and assistant director at SIREDO center

Sumit GUPTA
Canada
Hospital for Sick Children in Toronto
Staff Oncologist and Clinician Investigator, Division of Haematology/Oncology

Discussants session 2

Paul GIBSON
Canada
McMaster Children's Hospital
Pediatric Oncologist

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

Discussants session 3

Arnaud PETIT
France
Armand Trousseau Hospital
French Society for Childhood and Adolescent Cancer and Leukemia
Pediatric Oncologist

Frank WESTERMANN
Canada
Germany
German Cancer Research Cancer (DFKZ)
Head of the Department of Neuroblastoma Genomics

Main topics discussed

- Integration of data from registries, clinical trials & molecular studies
- Data models capturing temporal relationships
- Balancing depth of data vs sample size; focus shift from prognostic (diagnosis) to predictive (relapse)
- Changes in classification of (morphology) need to be validated, critical for risk stratification
- Structured / hierarchical systems for data collection
- Stage data are critical, other data elements available and relevant for prognostic
- Need to capture longitudinal treatment: treatment plan, dates, agents dose
- Treatment of complications (surgical intent), details of radiation therapy plan
- Outcomes to include progression and complications
- Development of systems for storing and sharing genomic data

Key barriers to be adressed

- Wide heterogeneity in the data
- Definition of the minimal dataset
- Need to test/validate collection of variables critical for classification/stratification
- Inconsistent standards used at national or regional level, few global standards
- No central repository of standards and vocabulary used by surveillance systems / clinical trial groups (who uses what standards?)
- Missing mapped variables across countries and databases
- Data recode is labor-intensive

Potential solutions

- Development of interoperable national standards (e.g. French OSIRIS interoperable with HL7 FIHR)
- Development of controlled vocabulary – both definitions and formats
- Expand existing data warehouses (e.g. OMAP) with cancer episodes
- Promotion of exchanges of data models and elements and the development of linkages and mapping as a step toward data sharing
- Development guidelines for disease-specific treatment data collection, including relapse and complications (ENCR)
- Use of AI in automatic data extraction

Next steps

Establish a Childhood Cancer Data Harmonization Task Force to oversee:

- **Mapping of data resources (data requirements, data standards, vocabulary, linkage, classification/stratification, etc) for childhood cancers**
- **Mapping of stakeholders**
- **A neuroblastoma project to pilot interdisciplinary international data collection and standards definition**
- **A medulloblastoma project**
- **Updated round of discussions: SES, long-term follow-up**

Facilitators

Eric DURBIN
Kentucky Cancer Registry

Johanna GODERRE
US National Cancer Institute

Discussants session 1

Peter GOODHAND
Canada
Global Alliance for Genomics and Health (GA4GH)
CEO

Jan NYGÅRD
Norway
Cancer Registry of Norway
Head of the Registry
Informatics Department

Discussants session 2

Marie CASTETS
France
INSERM – Share4kids
Co-director

Paul RINAUDO
France
ADLIN Science
CEO

Jacqueline CLAVEL
France
INSERM
Epidemiologist and Research professor

Discussants session 3

Stephanie HILL
USA
North American Association of Central Cancer Registries (NAACCR)
Associate Director

Paul SAULTIER
France
AP-HM, Inserm, LEA platform
Associate professor

Main topics discussed from Presentations

- Successful international collaboration towards genomic data standards
- Free and open comprehensive cancer data sharing, Norway as a model for the world
- An innovative approach to childhood cancer data sharing involving public and private partnerships
- North American cancer registry data standards
- Promoting and studying long-term health in survivors of childhood leukemia, the LEA Project
- Additional data needed for research:
  - Emerging diagnostic data such as single cell sequencing results
  - Greater granularity underlying some coded values
- Data visiting versus federated data access

Key barriers to be addressed

- Lack of trust among potential data sharing participants
- Restrictions due to intellectual property concerns
  - Publishing, career advancement
- Legal challenges: Lack of understanding about regulations, such as GDPR
- Large volumes of molecular/diagnostic data that lack clinical annotation
  - Standardized interoperable metadata needed to define provenance, content of datasets, accessibility and proper context for secondary use
- Characterizing true patient population underlying clinical trial/clinical datasets
- Burden of electronic health record entry for clinicians
- Quantifying and improving of quality of data
  - Clinical trial data not widely/universally accessible for research (secondary use)
- Balancing real time data needed for patient clinical care and secondary use as part of a learning health system
- Profit overriding adoption of universal data standards
- Difficulty of communication among participants from different disciplines
  - Speaking different languages

Potential solutions

- Face-to-face interactions are needed to establish and build trust among partners
  - Multidisciplinary communication among clinicians, scientists, informaticists, biostatisticians and others
  - Teaching/educating partners to understand each other's language
- Use "high quality" cancer registry data to link and annotate molecular/omics datasets
- Clearly define meaning of data quality, depending on context of use
- Promote and incentivize use of existing international standards, dictionaries, guidelines, etc. for short- and long-term follow-up of pediatric cancer patients
- Universal patient identifiers needed to facilitate continuity of care and patient follow-up
- International data standards to increase fair competition and technical advances
- Improved training and education for collection and use of standardized data
- AI to improve structured clinical data capture (in real time)
- Increase funding for professional cancer registrars in some European countries

Next steps

- Continue Paris Conferences
  - Form working groups to continue this collaboration
- Build consensus around “universal” data standards/data dictionaries
- Evaluate potential of UMLS metathesaurus as a repository to document childhood cancer data standards globally
- Use inventory of standards, guidelines, etc. emerging from this collaboration as a requirement for sponsor funding
  - Perform analysis demonstrating efficacy and value of this approach
  - Provide resources and tools to help facilitate adoption of these standards, guidelines, etc.
- Use case: Implement a US & European federated clinical & molecular data repository to address a specific challenge in childhood cancer (such as DIPG)

Facilitators

Michel COLEMAN
London School of Hygiene & Tropical Medicine

Chuck WIGGINS
New Mexico Tumor Registry

Discussants session 1

Suzi BIRZ
USA
University of Chicago
Pediatric Cancer Data Commons, Regulatory and
Data Governance Consultant

Joanne AITKEN
Australia
Cancer Council Queensland
School of Public Health
The University of Queensland
Director of Research, Honorary Professor

Discussants session 2

Philippe-Jean BOUSQUET
France
French National Cancer Institute
Director of Health survey, data-science, and assessment division

Chloé JEGO
France
INSERM
4.UNCAN.eu
Project manager

Rosa CASTRO
Belgium
Deutsche Stiftung Weltbevölkerung (DSW)
Senior Advocacy Officer

Discussants session 3

Hugo CROCHET
France
Léon Bérard Centre
Chief information system and data officer

Dr Tomohiro MATSUDA
Japan
National Cancer Center
Head, Office of International Affairs, Strategic Planning Bureau

Main topics

Australian Childhood Cancer Registry
BENCHISTA
Pediatric Cancer Data Commons
CI5C and CONCORD
Approaches to sharing databases

Topics

- Consent
- Public benefit vs. private risk
- Common good vs. individual risk,
- Reputational risk: institutions, communities
- Public approval for personal data research
- Standardised data dictionaries
- Publication policy should be required
- Legal steps required for data access

Key barriers

Permissions, protections and ownership

Legislation – fails to reflect needs for research

Lack awareness of law, science, technology, mechanisms of protection

Lack of uniform guidance

Real *vs.* perceived risks of data use

Identifiability – perceptions of legislators, administrators, controllers, researchers

Politicians – prioritise infections over cancer

Funding and sustainability

Public understanding of data sharing

Transparency statement involving patient advocates

Misunderstanding of risk

Potential solutions

Public attitude studies (what, why, how)

Cancer survivors – advocate for research, educate decision-makers…

Statutory cancer registration

Data Use Agreements – templates

Multi-project data warehouses

Trusted Research Environments – alternative infrastructures…

New GDPR guidance – EU Commission, European Data Protection Board?

Playbook on how GDPR can be deployed to maximise data use

Early wins

Next steps

- **Assemble working groups**
- **Obtain funding**
- **Deliverables…**

Facilitators

Jaime GUIDRY AUVIL
US National Cancer Institute

Gijs GELEIJNSE
Netherlands Comprehensive Cancer Organisation

Discussants session 1

Steven THOMAS
Canada
Statistics Canada
Section chief

Leucio Antonio CUTILLO
Italy
European Commission, Joint Research Centre (JRC)
Project officer and Scientific researcher / Disease Prevention Unit

Discussants session 2

Heidi HANSON
USA
Statistics Canada
Oak Ridge National Laboratory
Group Lead Biostatistics and Multilevel Systems Modeling, Senior scientist

Robert MILLER
USA
Minderoo Foundation
Principal, Unlocking Patient Data, Cancer Mission, Minderoo Foundation

Discussants session 3

Tamara MILLER
USA
Emory University School of Medicine
Pediatric Oncologist

Laura BOTTA
Italy
Istituto Nazionale dei Tumori Foundation
Senior biostatistician – Cancer epidemiology

Main topics discussed

Technologies and strategies to unlock, access and analyse data

- Pseudonymization
- Safe data
- Federated technologies
- Data Standardization and Harmonization
- Data extraction from the patient record
- Expérience: Centralized versus federated analyses

Key barriers to be adressed

- **Technology is not a barrier! We can break silos and address privacy and legal issues**
- **Innovation is also process and governance innovation**

But we need:

- Legal and regulatory framework!!
- Governance!!

And

- People
- Funding (few international calls)
- Technology needs to be globalized
- Focus
- Patient Advocate involvement

Potential solutions

- Show and Tell: define **driver projects** to demonstrate potential and pave the way for more complex, longer term ambitions
- Bridge data from cancer registries with OMOP-CDM and Federated Technologies
- Unlock data from the records using extractEHR
- A well-structured project (workstreams, work packages)
- Work multi-disciplinary!

Next steps

- **Coalition of the willing**
- **Formulate some diverse DRIVER projects that partners that can commit to**
- **Reuse technologies, legal frameworks and best practices from other international data sharing projects**
- **Building landscape of data, systems, formats and policies/regulatory frameworks including data access rules**
- **Prioritize sustainability and investigate public-private partnerships**