IMPROVING BIOBANK STANDARDS CONFERENCE
12th June 2018

The Final Frontier: Near real-time clinical data linkage from the eMR for biobanks

Kevin Spring, Nicki Meagher, Wayne Ng, Deborah Marsh, Susan Goode, Anna deFazio, Sarah Nielsen, Jitendra Jonnagaddala, John Pimanda

Background

- Translational research programs increasingly demand high quality-assured, fit-for-purpose clinically annotated biospecimens
- Clinical annotation comprises histopathologic, treatment and detailed follow-up data
- As a result biobank capabilities need to move beyond simply knowing sample location and requires sophisticated informatics/LIMS solutions to meet this demand
- Achieving this new standard is challenging and requires an understanding of the access issues, data variation and source systems to extract and harmonize required clinical data

Models of data collection

DATA LINKAGE

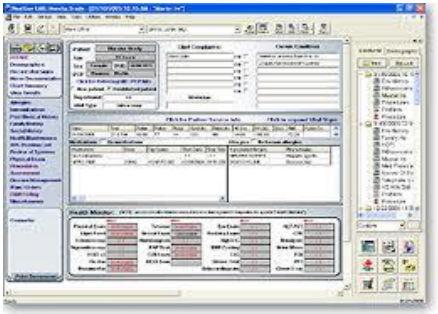
MANUAL DATA MANAGEMENT

BIOBANK participants

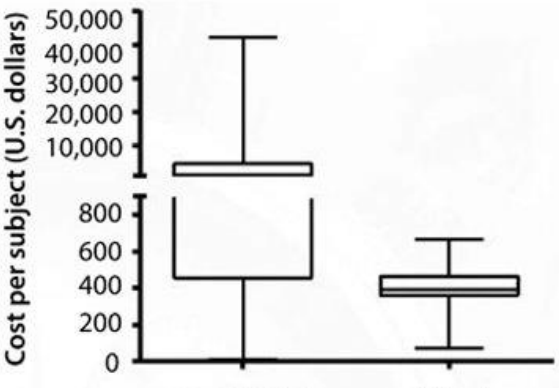
State-based datasets (Registries)

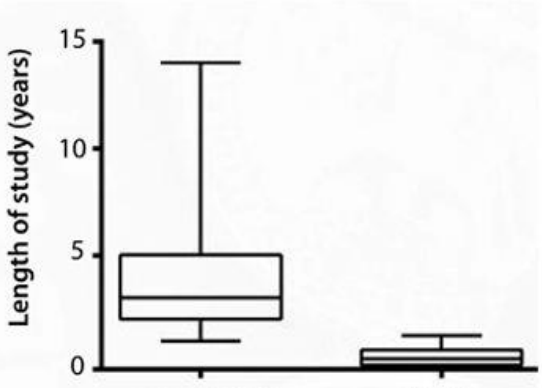
Commonwealth datasets (Medicare/PBS)

EMR EXTRACTION?



EMR linkage cost and time efficiency





Bowton et al *Sci Transl Medicine* 2014

Challenges/Considerations

Multiple systems within and across LHDs

- Integration challenges
- Patient Identfiers (MRN/UPI), Some surgery only patients not in OIS
- Varied accessibility

Governance of data sources

- Approval from ethics and data custodian
- Policy on usage for research
- Security of data

Completeness and quality of data

- Gaps between desired data items and the actual data availability
- Gaps between desired data item format and the actual data format (such as text in notes)

Project Aims - phase 1

2017: Roadmap of essential data items across eMRs in 4 LHDs

1. **Define the minimum data set required by all cancer biobanks**
2. **Map the location of data items in four LHD based information systems**
3. **Investigate data system access opportunities and barriers at four LHDs**

Identified Minimum Data Set

eMR Project Data Items (common data variables requested by researchers)	
DEMOGRAPHICS/PATIENT INFOMRATION	**TREATMENT**
Patient unique identifier	Surgery details
Patient date of birth	Radiation therapy details
Gender	Systemic therapy details
Height at time of diagnosis	Other treatment details
Weight at time of diagnosis	
Race/Ethnicity	
Family history	
Smoking history	
Previous or concurrent malignancy (all cancer types)	
Site of previous or concurrent malignancy (all cancer types)	
Details of treatment for previous or concurrent malignancy (year, treatment type)	
Germline test details	
TUMOUR CHARACTERISTICS (primary)	**OUTCOME - recurrence, progression, relapse & survival data**
Date of diagnosis	Progression
Primary site	Date of progression
Tumour grade / morphology	Basis for determining disease progression
Clinical trial details	Basis for disease progression - Other, Specify
TNM Stage/Group	Patient status (alive or dead)
Other staging system	Patient date of death
Molecular test details	Cause of death
Immunohistochemistry test details	

LHD Source Systems

Local Health District	Source systems		
	Pathology	Inpatient	Oncology
South East	OmniLab	Powerchart	MOSAIQ/ARIA
South West Sydney	PathNet/Powerchart	Powerchart	MOSAIQ
Hunter	AusLab	iPM/CAP	ARIA
Northern	AusLab	Powerchart	ARIA (Rad Onc only)

Access to Source Systems

Biobank	Source Systems		
	Pathology	Inpatient	Oncology
TCRN	Yes	Yes	No
CONCERT	Limited	Yes	Yes
HCRA	Yes	Yes	Yes
SYD VITAL	Limited	Yes	No

Minimum Data Set - common data items

eMR Project Data Items (common data variables requested by researchers)	Site A (TCRN)	Site B (CONCERT)	Site C (HCRA)	Site D (SYD VITAL)
DEMOGRAPHICS/PATIENT INFOMRATION				
Patient unique identifier				
Patient date of birth				
Gender				
Height at time of diagnosis				
Weight at time of diagnosis				
Race/Ethnicity				
Family history				
Smoking history				
Previous or concurrent malignancy (all cancer types)				
Site of previous or concurrent malignancy (all cancer types)				
Details of treatment for previous or concurrent malignancy (year, treatment type)				
Germline test details				
TUMOUR CHARACTERISTICS (primary)				
Date of diagnosis				
Primary site				
Tumour grade / morphology				
Clinical trial details				
TNM Stage/Group				
Other staging system				
Molecular test details				
Immunohistochemistry test details				
TREATMENT				
Surgery details				
Radiation therapy details				
Systemic therapy details				
Other treatment details				
OUTCOME - recurrence, progression, relapse & survival data				
Progression				
Date of progression				
Basis for determining disease progression				
Basis for disease progression - Other, Specify				
Patient status (alive or dead)				
Patient date of death				
Cause of death				

Data item exists on eMR

Possible item/information embedded in text

Unable to access/not available

The Data Challenge

- Multiple systems within and across hospitals
- Patient identifier issues
- Data governance framework
- Data type – discrete data and free text
- Completeness/quality
- Timely longitudinal data updates

Importance of this project:

- Shed light on utility of eMR data for clinical annotation of biospecimens
- Framework to be scalable and system-neutral

Future Directions - phase 2

- **Collaboration between LHD data custodians, Biobanks and CINSW Clinical Cancer Registry**
  - **ensure biobank clinical cancer data linkage efforts are not duplicated**
- **Conduct multiple test cases using a defined cohort of biobank specimens and associated clinical data to test completeness of eMR data linking mechanism**
- **Source death data from national source**
- **Ensure compliance with *NSW Health Policy Directive PD2018_001* relating to disclosure of data for research**

Future Directions - phase 2

2018: Framework for near real-time semi-automated extracts

1. **Develop the framework and model of data extractions from OIS**

2. **Test extracts for 1-2 prospective biobank projects. Investigate the completeness and currency of dataset**

3. **Develop a standardised ethics application for biobanks to link National Death Index data**

Acknowledgements

CONCERT

Kevin Spring
Wayne Ng

HCRA

Rodney Scott
Susan Goode
Sarah Nielsen

TCRN

John Pimanda
Nicki Meagnher
Carmel Quinn
Jitendra Jonnagaddala

Sydney Vital

Deborah Marsh

Ussha Pillai

SW-TCRC

Anna deFazio

CINSW Biospecimen Stakeholder Network

All reference members of the BSN for their contribution

LHD Cancer Services & Pathology IT Teams

Funding

NSW GOVERNMENT

cancer institute NSW

Pilot Concordance Analysis

Records were linked for **n=186** HSA Biobank participants (mixed diagnoses) from a **manually curated** dataset with a data extract from the Cancer Services **eMR** MOSAIQ

Data item	Discrepancies (n, %)	Type of discrepancy	Example
Primary diagnosis (ICD-O-3 morphology)	12, 6%	11 granularity	8720/3 malignant melanoma NOS vs 8721/3 nodular melanoma
Cancer stage	20, 11%	9 granularity 3 missing (eMR) 8 discordant	IV vs. IVB
Grade	17, 9%	5 missing (eMR) 7 missing (curated)	
SUMMARY	49, 25%	Missing/discordant only = 15%	