SAHMRI
South Australian Health & Medical Research Institute

Connecting Colorectal Cancer data and making it FAIR

Steve Wesselingh

Context

- Colorectal Cancer (CRC) is one priority area of the SA Academic Health Science and Translation Centre
- The *Beat Bowel Cancer Project* has the ambitious target of eliminating premature and preventable mortality from CRC
- There is a high burden of disease in Australia, and bowel cancer is Australia's second leading cause of cancer death
- CRC is highly amenable to intervention:
    - When detected earlier and at a more treatable stage, survival is high (stage I survival: 95%; cf stage IV: 18%)
    - Amenable to primary prevention – 49.8% attributable to lifestyle risk factors
    - Amenable to secondary prevention – only 39% of eligible Australians are participating in the National Bowel Cancer Screening program
    - Opportunity to improve clinical care – referral time to coloscopy (where indicated) and participation; adherence to treatment guidelines
    - New approaches: e.g. expanding high-risk group surveillance

Context

- To succeed, the *Beat Bowel Cancer Project* needs data to drive rational decision making about where to focus interventions

- Economic modelling is being undertaken to predict potential deaths prevented from:

  (i) increasing FIT participation rates; (ii) increasing colonoscopy participation; (iii) changing ages of screening eligibility; and (iv) cost-benefits of each scenario

- Data linkage is being undertaken to enable a 'whole-of-system' view of CRC, to identify unexplained variations in health care. Linkage includes:

  - Population cancer registry data (incidence, mortality and survival by cancer type, age, sex, SES, remoteness, Aboriginal status, and country of birth);

  - SA clinical cancer registry data (stage and grade); SA inpatient and radiotherapy data, clinician held registries, Commonwealth MBS/PBS claims data and National Bowel Cancer Screening Registry data.

Data linkage project

- State data linkage performed by SANT Datalink – complete.
- Better insights already gained into CRC survival across the SA population, adequately controlling for confounders, for example:
    - Lower survival observed among:
        - Under 50 years who normally have higher grade tumours
        - 80 years and over, whose treatment is often compromised by lower resilience and poorer health status
    - However, little evidence of variation in survival by socio-economic status and residential remoteness
- Link to Commonwealth data will be within the SURE environment at the SAX Institute.

Project achievements

- All state based data linked, analysed and reported on. New insights in terms of burdens of mortality.
- Commonwealth data approved for linkage, awaiting delivery
- (When finished) a national first – a population-based health-system-wide data set for research, health service planning and evaluation, monitoring effectiveness, cost-effectiveness and equity in service delivery
- The data linkage project will: inform our local clinicians; inform our health services; enable the Beat Bowel Cancer Project to target interventions to improve health outcomes.
- We will publish our findings in the scientific literature.
- But we were challenged by ANDS – can we go beyond this? Can we make our linked database FAIR?

Making data FAIR

Findable

Accessible

Interoperable

Re-usable

Introducing FAIR principles to the CRC Data Linkage project would mean not only bringing the CRC dataset together for the Beat Bowel Cancer Project, but also making the data set available for future research.

Anticipated to be achievable

Findable

- Minting a DOI (Digital Object identifier)
- Providing Rich Meta Data

Accessible

- Metadata will be accessible from Research Data Australia
- DOI will resolve to a landing page e.g. SA Translation Centre, which is maintained

Interoperable

- Data has already been assembled using firmly established methods and standard language
- Data collection will be described using FOR codes, MESH subject headings,
- Links established to investigators' ORCID IDs, relevant future Grant IDs and publications

Uncertain feasibility

Accessible

- The Data Collection would be accessible from within SURE, to other researchers with the appropriate ethics approvals

Re-usable

- The specific reuse conditions would be provided to each new research project that gains ethics and data custodian approvals to use the Data Collection via SURE

Lessons learned

- In exploring making this data FAIR, the challenges and questions that arose were:
  - Sensitive health data
  - Multiple data custodians
  - To move beyond the original, approved linkage project, custodians will all need to be consulted

  - We don't yet know the appetite of all the different custodians including clinician scientists and state and commonwealth government agencies for the linked data set to be made FAIR
  - We don't yet know the appetite of the data linkage platform partner (SURE) or the resourcing/capacity

  - The project continues to be a pathfinder in this field, addressing questions that have not been raised in linking human identifiable datasets

Acknowledgements

- BBCP-team, particulartly:
  - David Roder
  - Caroline Miller
- ANDS, paricularly
  - Keith Russell
- South Australian Academic Health Science and Translation Centre

SAHMRI
South Australian Health &
Medical Research Institute

www.sahmri.com