Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Australian Centre for Agricultural Health and Safety, Moree
School of Public Health, University of Sydney

Julie Depczynski
Tony Lower

THE UNIVERSITY OF SYDNEY

Australian Centre for Agricultural Health & Safety

THE UNIVERSITY OF SYDNEY

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Our study - objectives

To determine if any significant differences exist between farm residents, rural non-farm and urban residents in relation to:

- incidence and mortality of selected major cancers, including prostate, colo-rectal, breast cancer, melanoma and nHL
- screening rates, stages of presentation, diagnosis, treatment and survival rates of major cancers
- possible behavioural, environmental or socio-economic risk factors for cancer

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

How do we find this out?

1. Internal comparisons of farmers, rural non-farmers and urban people within the '45 and Up' Study cohort on self-reported cancer, screening practices and behavioural risk factors.

   - 267,120 participants from 2005-2009.
   - Cohort size and over-sampling of rural areas enables rural-focused studies.
   - "Current housing" question with option "house on a farm," enables identification of farm men (n=9,583) and women (n=10,861) by 'residential status'.

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

How do we find this out?

2. Data linkage. '45 and Up' Study primary data source /collection matched with records of 7 other health datasets.
   - NSW Pap Test Register
   - NSW Central Cancer Registry
   - NSW Admitted Patients Data Collection
   - MBS
   - PBS
   - NSW RBDM
   - ABS Deaths (NSW)

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Why use multiple datasets?

- Health data fragmented - collected by different agencies for different purposes
- Linkage adds value – quality and detail, reduced bias, cost-effectiveness, follow-up on health outcomes, data validation

How are records from datasets linked?

- Master Linkage Key (MLK) developed by the 'Centre for Health Record Linkage' (CHeReL)

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Centre for Health Record Linkage (CHeReL)

- Established in 2006. Managed by the NSW Ministry of Health
- Routinely matches personal information from core health and administrative datasets, which can then be linked using the Master Linkage Key (MLK). 15 NSW datasets now include an MLK (70 mill records).
- CHeReL provides:
  - advise on the design, process, feasibility and cost of linkage study proposals
  - a general linkage service for specific projects
- A cost involved (quotes)
- Privacy is strictly maintained by ethics approval processes and strict separation of personal information from research analysis ...

THE UNIVERSITY OF SYDNEY

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Linkage process and privacy *

1. Dataset custodians (eg. NSW CCR, NSW RBDM) provide encrypted source record number and demographics for each record to CHeReL (not clinical)

2. CHeReL links records using probabilistic matching of the demographic details. CHeReL person number assigned. The source record number and CHeReL person number form the Master Linkage Key (MLK).

3. When project approved , CHeReL assigns a project-specific person number (PPN) for each person in the linked datasets (unique to each project)

4. The data custodian merges the project person number with the clinical variables that have been approved for use in the project. The source record number is removed, leaving researcher with PPN and clinical data only.

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Application process* (http://www.cherel.org.au)

STEP 1 Plan your study using dataset, variable and linkage information on the CHeRel website

STEP 2 Contact the CHeReL to discuss your project and to obtain a “ball park” quote

STEP 3 Complete an Application for Data (AFD) form and email to the CHeReL for their feedback

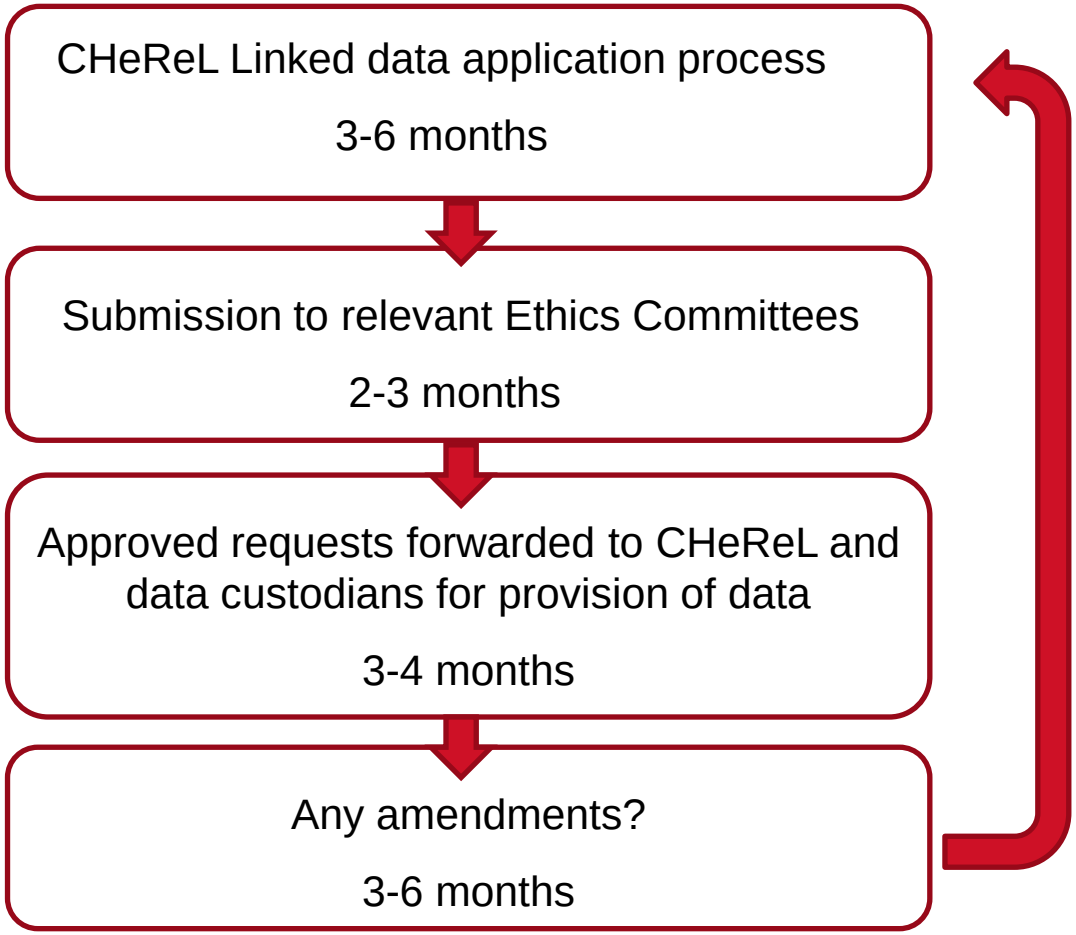
STEP 4 Amend the AFD incorporating feedback and resubmit to the CHeReL, along with the National Ethics Application Form and Research Protocol

STEP 5 The CHeReL forwards these documents to the data custodians for approval, and provides a technical feasibility letter and written quote

STEP 6 When approvals and the feasibility letter have been received, submit all required documents to the ethics committee

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Timeframe?



Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Rural access to linked data

› The Secured Unified Research Environment (SURE)* was developed by the Sax Institute and launched in July 2012.

SURE Secure Unified Research Environment

› SURE is a secure computing environment that researchers log in to remotely access and analyse approved linked data files.

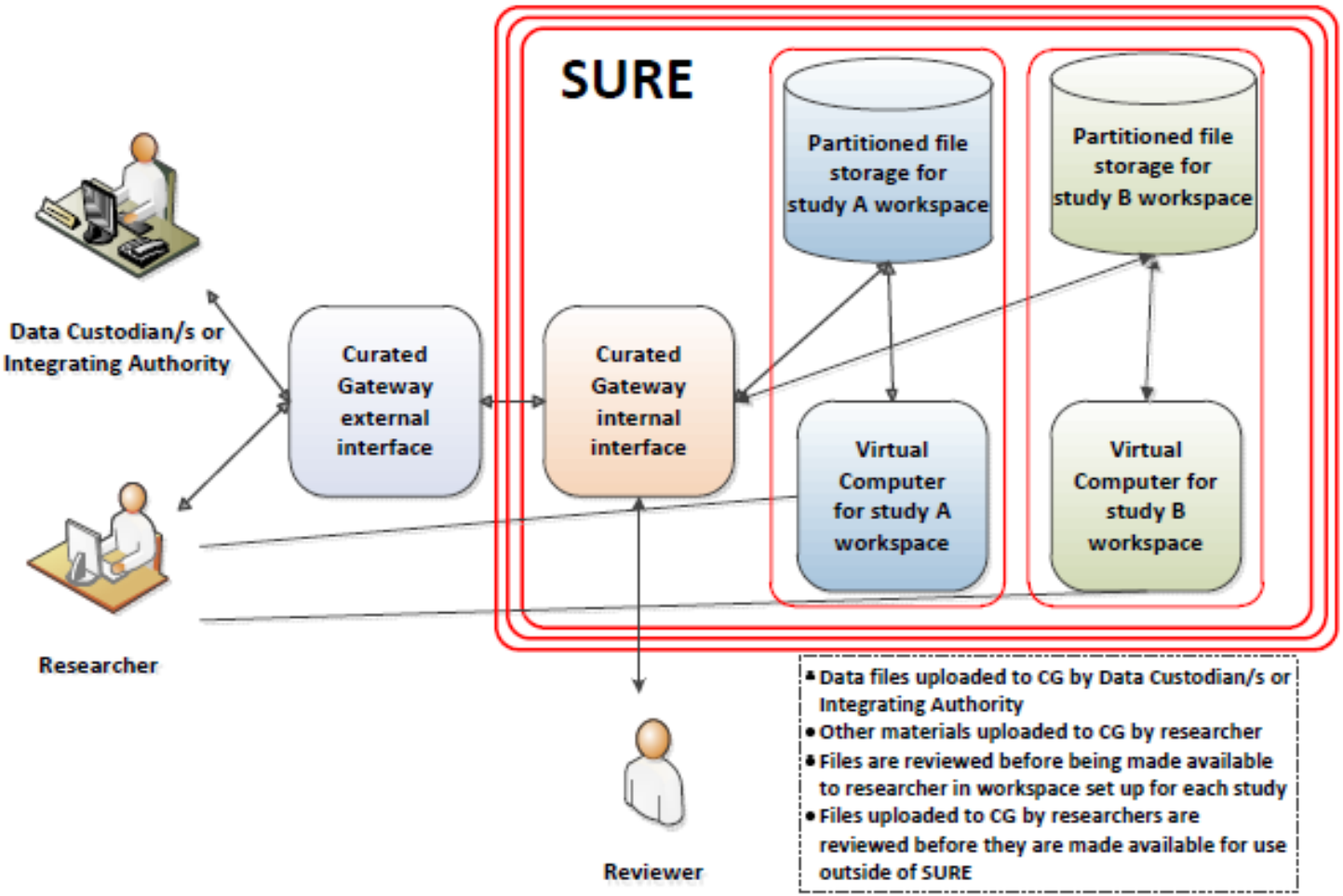
› SURE is accessed via AARNet or via an encrypted internet connection from researchers' local computer.

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

SURE Secure Unified Research Environment

- Strong data confidentiality and security measures, good speed, storage and analytic software (eg. SAS).
- Users are provided with a remote virtual computer desktop and ‘Project Workspace’ for each research study.
- Users see a remote virtual desktop, on their local screen –
- Keystrokes /mouse on local computer transmitted to the remote virtual desktop computer in the SURE facility.

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers



Establishing a Project in SURE*

1. Plan and determine project requirements - data access, analysis and storage
2. Register and complete SURE Training (1 day). Obtain an initial quote.
   (Meanwhile: await approvals for project - ethics committees / data custodians)
3. Complete SURE Project Workspace Application Form. Sign SURE agreements. Receive invoice and complete payment.
4. Install required software and check IT system requirements. SURE provides access credentials
5. Ongoing processes – annual fees and project support (good)

› *Takes about 2-3 months, largely dependent on your ethics approval time and scheduling of training day. Process quite easy, uncomplicated, helpful people.*

CHeReL, SURE processes

- Lots of patience – but people were very helpful and want your project to succeed !!
- Contacts :
  - CHeReL:

    http://www.cherel.org.au

    cherel.mail@moh.health.nsw.gov.au
  - SURE:

    https://www.sure.org.au/

    sure-admin@saxinstitute.org.au

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Progress on our project ...

- Once data obtained, spend considerable time ‘checking’ it (eg. duplicates, incorrect dates of death etc.)
- Some data was insufficient for our needs and we had to re-apply for more!! (eg. RBDM vs. ABS mortality data, another year of CCR data)

(Prelim: Cancer incidence (DSR) 2006-2008. ‘45 and Up’ vs. Aust. Pop).

Men:

- Overall cancer, bowel and lung cancer were sig. lower in ‘45 Up’ cohort
- Melanoma and prostate cancer sig. higher in ‘45 Up’ cohort

Women:

- Overall cancer and lung cancer were sig. lower in ‘45 Up’ cohort
- Melanoma was significantly higher in ‘45 Up’ cohort

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Cancer incidence (ISR) 2006-2008

Farm men:

› 9281 farm men, 117 new cancer cases/yr (3.8% of farm men in cohort over 3 years)

Compared to 45 and Up cohort:

› Overall cancer ISR ratio in farm men 102.4% (100.7 - 104.1)

› **Prostate cancer incidence 9% higher in farm men and only cancer significantly higher in farm men than '45 Up' cohort**

› Melanoma 21% higher in farm men but not significant (lower cases = wider CI's)

› Lung cancer, nHL and MM <5 cases/yr

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

Cancer incidence (ISR) 2006-2008

Farm women:

- 10,550 farm women, 71 new cancer cases / yr (2.0% farm women in cohort over 3yrs)

Compared to 45 and Up cohort:

- Overall cancer ISR ratio in farm women 101.2% (99.0 – 103.4)
- **No cancers significantly higher / or lower in farm women (low case no.'s)**
- In farm women, breast cancer was12% lower and colo-rectal cancer 29%, but neither results were significant
- Lung, cervical, nHL & MM all less than 5 observed cases/ yr in farmers

Remote Access to Big Datasets:
Enabling rural-based research on cancer in NSW Farmers

› Main point:

Thanks to CHeReL (MoH) and the SURE facility (Sax Institute), you can now conduct / analyse large linkage studies in rural areas without having to re-locate to the city (YIPPEEE !!!)

› Next time .....

- Results from our Project - cancer incidence, mortality, screening, treatment and cancer risk factors in farmers and other rural people.

› Acknowledgements:

- Associate ProfessorTony Lower (ACAHS, University of Sydney)
- CHeReL : Reinier de Vos , Elizabeth Moore
- Sax Institute / SURE : Kate Humphries, Joanne Khoo