Difficulties and rewards of linking to administrative data: experience from ELSA

Gemma Tetlow
Outline

• Overview of linking surveys to administrative data in the UK
  – What data is available? (non-exhaustive, older-age focussed)
  – What are the barriers to administrative data–survey linkages?
  – How can we overcome the barriers?

• Experience and progress with ELSA
  – Permissions
  – Linkages achieved to date
  – Potential future linkages
Background

- Limited experience in the UK of linking survey data to administrative data
  - Has mostly been done for one-off projects, often funded/supported by the government department who hold the administrative data
- Increasingly household surveys are asking respondents’ permission to link to administrative records
- Growing desire within research community to make administrative data more accessible
- Willingness/enthusiasm also in some parts of government
  - But counter-balanced by legal / privacy concerns
Administrative data sources in the UK (1)

1. National Insurance records (HM Revenue and Customs)
   - Information required to judge contributions to social security system
   - Introduced in 1948
   - Contributions were flat-rate until 1975, then earnings related
   - Provides information on:
     - Weeks of paid employment from 1948 onwards
     - Capped earnings from 1975 to 1997; uncapped from 1997 onwards
     - Accrued state pension entitlements
     - Periods of membership of private pension schemes
     - Periods in receipt of certain contributory out-of-work benefits (e.g. Statutory Sick Pay, disability and unemployment benefits)
     - (Marital history, home address history, country of origin)
     - (Employer information)
     - (Private/occupational pension scheme information)
Administrative data sources in the UK (2)

2. Other tax-related data (HM Revenue and Customs)
   - Income tax
   - Capital gains tax
   - Self-employment profits and expenses
   - Tax credits
   - Third-party data held by HMRC: wealth held in and contributions to private pensions, other savings vehicles
Administrative data sources in the UK (3)

3. Benefit records (Department for Work and Pensions)
   – Record of (contributory and non-contributory) welfare benefits received since 2003
   – Provides information on:
     • Type of benefit received
     • Start (and end) dates
     • Amount received
     • Reason for award (for disability benefits)
   – Amount received from some benefits not available
     • Housing Benefit
     • Council Tax Benefit

© Institute for Fiscal Studies
4. Hospital Episodes Statistics (Department of Health)
   - Every episode of publicly-funded inpatient treatment since 1989
     • In both public and privately-owned hospitals
   - Outpatient appointments / treatments also covered since 2006
     (incomplete data available back to 2003)
   - Accident and Emergency admissions since 2007 (experimental)
   - Provides information on
     • Date of birth, sex, geographic location of patient
     • Nature of treatment received, co-morbidities
     • Where treatment received
     • Primary care physician
     • Waiting time

5. Other health-related data
   - Mortality records (Office of National Statistics)
   - Cancer registry (Department of Health)
Difficulties in using administrative data

- **Practicalities of the linkage**
  - Data is held by different government agencies: must be approached individually
  - Datasets contain different unique identifiers (e.g. National Insurance number, NHS number)

- **Gaining access to the data**
  - Legal constraints / privacy issues: different rules for different government departments
  - Some government departments charge for data linkage/extraction (particularly where data storage has been outsourced)

- **Using the data for research purposes**
  - Data structure is complex
  - Data often does not contain all the information researchers require (e.g. truncated earnings information)
Approaches to overcoming the difficulties (1)

• Over-coming legal barriers
  – Getting individuals’ consent: agree acceptable wording for individual consent to data linkage
    • Government lawyers have previously changed their minds after a survey and permissions have been collected!
  – Government departments should be as explicit as possible about what purposes their data can be used for
    • Some data can only be used for research that is of benefit to the relevant ‘client group’
    • Currently some data only available to researchers contracted by a government department
  – Consider new legislation?
Approaches to overcoming the difficulties (2)

• Addressing concerns about privacy
  – Require approval from data release panel before access granted
  – Data typically must be accessed in a safe setting, for example
    • Secure, stand-alone computers
    • Remote access to a secure server
  – Outputs cannot be removed without permission
  – Penalties imposed for breaking the rules
  – Currently different restrictions about how and where data can be
    accessed apply for different datasets: ongoing efforts to harmonise

• ESRC-convened Administrative Data Taskforce
  – Bringing together researchers and data holders to try to develop
    general principles for using administrative data for research
  – Payoffs could be huge but progress is slow
Approaches to overcoming the difficulties (3)

• Costs of linking data
  – Carry out linkages for a number of surveys simultaneously
  – Trade-off between matching all available data and minimising cost by focusing on the most essential

• Data complexity
  – Exploit existing knowledge among civil servants: these data are used internally by government departments already
  – Learning by doing: the more data is used by researchers, the more will be known about it
  – Pool knowledge: researchers who have started using these data should produce documentation, share information, code etc.
    • Avoids duplicating effort
    • Allows common derived variables to be produced for different household surveys
    • Somewhat *ad hoc* at the moment but increasingly efforts to document and pool knowledge: e.g. new project at the Institute of Education to produce combined resource for the birth cohort studies
ELSA: Getting permission

• Respondents were asked for consent to link to
  – Health records
  – National Insurance contribution records
  – Benefits and tax records

• Respondents asked for consent when first interviewed: signed form collected by interviewers
  – If did consent, verbally reminded at subsequent interviews: must write to formally withdraw consent
  – If did not consent, asked again in subsequent interviews

• Permission valid until explicitly revoked by the respondent

• About 80% of ELSA respondents consented to linking their data to administrative records
  – Consent rates slightly higher for health data than economic data

© Institute for Fiscal Studies
ELSA: Progress on linking datasets (1)

1. National Insurance records (HMRC)
   - Data up to March 2005 has been matched for wave 1 respondents
     • Work in progress to link more recent data and for respondents to later waves
   - Researchers can apply to the ELSA Linked Data Access Committee for permission to use this data
   - Data can only be accessed in an approved enclave
     • NatCen has one that can be used by external researchers (London)
     • IFS has one for internal researchers only (London)
     • Other institutions can apply to have their own enclaves approved

2. Benefits records (DWP)
   - Data will soon be matched
   - Raw data will then be available to researchers: apply to ELDAC for permission
   - Data will have to be accessed in an approved enclave
ELSA: Progress on linking datasets (2)

3. Hospital Episodes Statistics (Department of Health)
   - Data now linked
   - Raw data will soon be available to researchers: application process still being finalised
   - Data will have to be accessed in an approved enclave

4. Mortality records (Office of National Statistics)
   - Data linked and regularly updated
   - Quarter and year of death, and broad cause of death available

   • Derived data
     - Gradually creating some generally useful derived variables
     - Work in progress but they will gradually become available to other researchers
     - Encourage researchers who use the data to make their own derived variables or code available to others
ELSA: Applying for administrative data

• ELSA Linked Data Access Committee
  – Comprises representatives of the survey team and relevant government departments who own the administrative data
  – Meets every 8 weeks
  – Typically approves data applications at the next available meeting

• Practicalities
  – Still finalising application process and hope to speed up process, but for the moment...
  – Application forms can be obtained from Andrew Phelps (Andrew.Phelps@natcen.ac.uk)
  – Researchers must outline: research question, what data is required, why this specific data is needed to answer the question, where intend to access the data
  – Data will then be transferred to an approved enclave (in the UK)
ELSA: How has the data been used so far? (1)

• Lifetime earnings and wealth: comparing England and the US
  – Using ELSA linked to National Insurance records (and HRS linked to Social Security records)
  – Antoine Bozio (Paris School of Economics) & Gemma Tetlow (IFS)

• Optimality of retirement saving
  – Using ELSA linked to National Insurance records
  – Rowena Crawford & Cormac O’Dea (IFS)

• Use of individual-level accuracy measures in the analysis of ELSA earnings data
  – Using ELSA linked to National Insurance data
  – Chris Skinner (LSE)
ELSA: How has the data been used so far? (2)

- Examining longitudinal patterns of misreporting in ELSA
  - Using ELSA linked to benefit records
  - Mike Brewer (Institute of Social and Economic Research)
  - (Starting 2013)
- Examining determinants of benefit non-take-up
  - Using ELSA linked to benefit records
  - Mike Brewer (Institute of Social and Economic Research)
  - (Starting 2013)
How could the data be used in future?

• What administrative data can add to survey data
  – Information on life experiences before and between survey data collection
  – Increased periodicity of data available
  – Information that is too difficult or costly to collect in the survey
  – Precise measurement of incentives provided by tax/benefit policies
  – Consistent measures across different surveys

• What survey data can add to administrative data
  – Many more covariates
  – Data on other family / household members

• Data validation and methodological improvements
  – Compensate for item non-response
  – Validate survey measurements
  – Test survey design and collection techniques
Conclusions

• Linking administrative data to household surveys is still relatively new in the UK

• Some progress being made
  – Co-ordinated efforts by research community to engage with government data holders
  – Individual surveys pursuing links on their own in the meantime

• ELSA relatively far advanced
  – Mortality data, National Insurance records and Hospital Episodes Statistics already linked
  – Benefits records should also be linked soon

• Many difficulties in both getting access to the data and using it for research purposes
  – Ongoing learning process!