ESRC NCRM ADMIN node

John 'Mac' McDonald

The significance of data management for social survey research

ESDS, University of Essex, 27 January 2009

Administrative

Data -

Methods,

Inference &

Network

Director: Lorraine Dearden

ADMIN

☐ exploring how researchers should best use administrative datasets that have become available in the UK

☐ determining how researchers can enhance longitudinal survey data by exploiting available administrative data

☐ training and capacity building (TCB)

"Perhaps the greatest unexploited opportunity for survey projects lies in administrative data."
Randall J. Olsen, Conference on Longitudinal Social and Health Surveys in an International Perspective, Montreal, January 25-27, 2006

ADMIN ...

☐ NCRM BIAS node is primarily concerned with using Bayesian methods and combining individual survey data with aggregate data measuring the same variables

☐ **ADMIN will exploit administrative and survey data that have been linked at the individual level** (as well as more aggregated levels) and will predominantly rely on non-Bayesian methods

☐ **substantive research themes are concentrated in education, as this is the area where linked survey and administrative data are already available**

ADMIN - Methodology

☐ **Director: Sophia Rabe-Hesketh**

☐ administrative data and missing covariates

☐ imputation

☐ attrition bias

☐ sampling schemes

☐ measurement error

☐ data-linking methods

☐ small area mapping

ADMIN . . .

- 2 closely connected research strands, which are mirror images of one another, each involving 4 projects

- the first with overcoming the limitations of administrative data **by exploiting linked survey data**

- the second with overcoming the limitations of survey data **by exploiting linked administrative data**

- TCB - develop and train new researchers in quantitative social research using administrative data
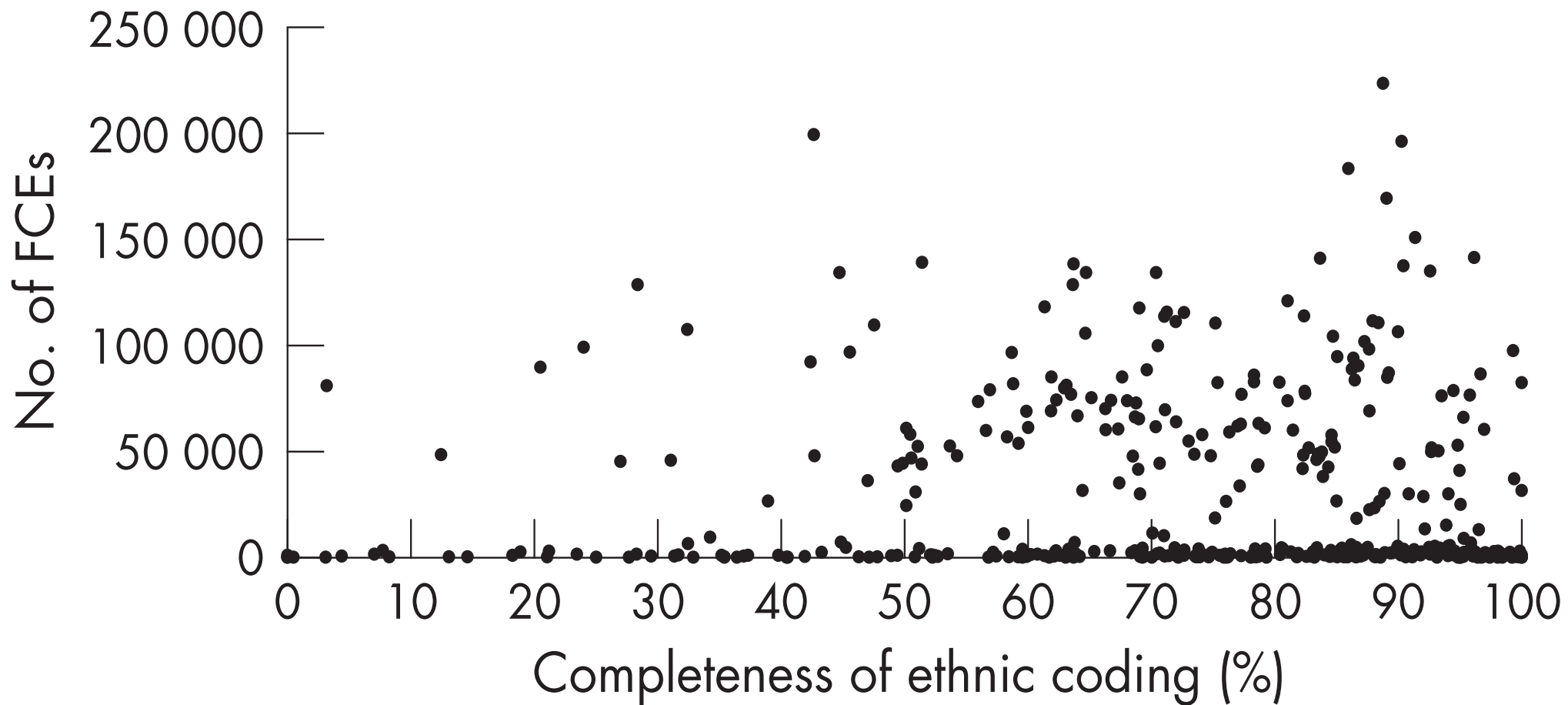
Figure 1 Completeness of ethnic coding, England, National Health Service (NHS) trusts, 2003/4. Source: Department of Health, Hospital Episode Statistics database. Based on 416 primary care trusts and hospital inpatient NHS trusts with >100 finished consultant episodes (FCEs; excluding 65 trusts with (100 FCEs). Overall completeness: 72.4%, based on 12 710 591 records; 59 trusts (14.2%) with >95% completeness; 108 trusts (26.0%) with >90% completeness.

Programme I - Director: Anna Vignoles

Using survey data to enhance methods for the analysis of administrative data

☐ linked datasets provide the opportunity

– to assess the limitations of administrative data

– to develop quantitative methods of analysis that can overcome such limitations

☐ 4 projects in areas of research where administrative data are currently being used without full understanding of the potential biases being introduced by the weakness of administrative data, i.e. missing and limited covariates

Programme I ...

Using survey data to enhance methods for the analysis of administrative data

☐ Measuring the effects of family background on pupil attainment using administrative data (Dearden, Rabe-Hesketh, Vignoles)

☐ To what extent does school attendance reflect real school preferences? (Hansen, Machin)

☐ Is education contributing to migration flows in Northern Ireland? (Brown, McDonald, Dustmann)

☐ Can a survey really improve the quality of census/ administrative data at low levels of aggregation? (Brown)

Programme II - Director: James Brown

Using administrative data to enhance methods for the analysis of survey data

☐ linked datasets provide the opportunity

    – to assess the limitations of survey data, particularly panel survey data

    – to develop methods that can overcome problems commonly associated with survey data, such as

        ▷ attrition

        ▷ non-response

        ▷ measurement error

Programme II . . .

Using administrative data to enhance methods for the analysis of survey data

☐ Ethnic differences in child outcomes (Dearden, Hansen)

☐ Attrition, non-response and the determinants of school outcomes at 16 (Rabe-Hesketh, McDonald, Skrondal)

☐ Neighbourhood mapping using survey data (Brown, Tzavidis)

☐ Enhancing event history analysis of social surveys with administrative data (Brown, McDonald)

Training and capacity building

- **Director: John 'Mac' McDonald**

- short courses in methods for data linkage

- short courses in methods for the analysis of linked data

- emphasis on large-scale UK longitudinal linked surveys

- build national capacity in the analysis of linked longitudinal data via a visitor and fellowship programme

Training and capacity building . . .

5 core courses offered yearly

☐ Introduction to Data Linkage (1 day)

☐ The Value of Data Linkage for Research (1 day)

☐ Data Linkage - Methodological and Statistical Issues (2 days)

☐ Longitudinal Data Analysis (3 days)

☐ Event History Analysis (3 days)

Training and capacity building . . .

4+ non-core courses + 1 new course offered yearly

- ☐ Enhancing Longitudinal Surveys by Linking to Administrative Data (1 day)

- ☐ Longitudinal Data Linkage (LS, NILS, SLS) (2 days)

- ☐ Using Longitudinal Data Linkage to Evaluate Area-Based Interventions (1 day)

- ☐ Data Linkage with the NPD (1 day)

- ☐ new course(s)

Innovations in combining data on the MCS

www.ccsr.ac.uk/methods/archive/combiningdata/Dex.pdf

- ☐ MCS has been linked to birth records and NPD

- ☐ "While technology is offering these new possibilities to combine data, old problems remain. These are:
    - – 1. Confidentiality . . .
    - – 2. Geocoded data and boundaries . . .
    - – 3. Resources . . .

- ☐ **Add - Secure access to confidential data**

www.iser.essex.ac.uk/ulsc/mols2006/programme/data/papers/Calderwood.pdf

Figure 2: The extent of linkages in current major longitudinal studies in the UK

	ONS-LS	Scottish LS	ELSA	ISMIE	NSHD	NCDS	BCS70	MCS	FACS	ALSPAC	Whitehall II
Birth registration data	3	3			3	3	3	3			
Mortality data (NHSCR)	3	3	3		3	3	3	3		3	3
Cancer data			3		3	?	?			3	3
Hospital Episode Statistics		3	3					3			
National Pupil Database								3			
Benefits and tax credits			3	3					3		
Employment spells									3		
National Insurance records			3						3		
Earnings data held on employer records				3							
Private pension characteristics			3								

Millennium Cohort Study sweep 4

- those MCS2 non-responders who we converted into productive for MCS3 - we've maintained them as productive for MCS4

- we've brought about half of unproductive at MCS3 back into the sample

- **about 96% give consent for us to access child health & education records**

- **for adults, 90% of main respondents gave consent for health records, 85% for economic records (of which 75% gave a national insurance number)**

- these figures were about 4% lower for partners