THE SCOTTISH LONGITUDINAL STUDY
A new opportunity for research in Scotland

Paul Boyle

- The (Scottish) Longitudinal Study
- Why does Scotland need a SLS?
- So, what is the SLS?
- Strengths and weaknesses
- How does the SLS differ from the LS?
- How far have we got?
- Governance issues
- Security issues
- Accessing the SLS
- The future...?
- Conferences

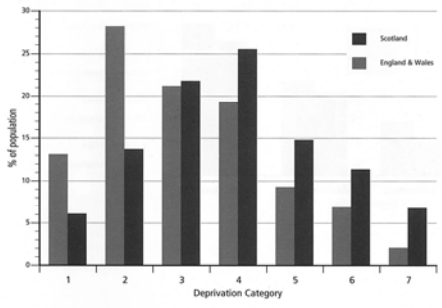
The (Scottish) Longitudinal Study

- The England and Wales Longitudinal Study (LS) established following 1971 Census
  - To study occupational mortality and fertility
  - Scotland included originally
  - Withdrew for funding / sample size reasons
  - Original files destroyed
- Re-establishing the SLS
  - Funded by SHEFC, CSO, Scottish Executive, GROS and ESRC
  - Working in close collaboration with GROS, ISD and ONS

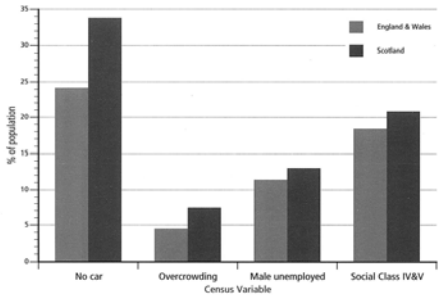
- People
  - Director: Paul Boyle
  - Project Manager: Lin Hattersley
  - Research Fellow: Zengyi Huang
  - Database Manager: Joan Nolan
  - Visiting Senior Lecturer: Vernon Gayle
  - Statistician: Gillian Raab
  - Research Fellow: Peteke Feijten
  - Research Fellow: Gerel Altankhuyag
  - 20 form pickers / clerical assistants

- Management committee
  - Paul Boyle (University of St Andrews)
  - Allan Findlay (University of Dundee)
  - Robin Flowerdew (University of St Andrews)
  - Vernon Gayle (University of Stirling / St Andrews)
  - Sally Macintyre (University of Glasgow)
  - Steve Platt (University of Edinburgh)

Why does Scotland need the SLS?

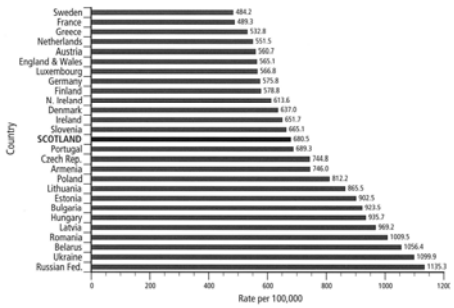


1991 deprivation in Scotland compared to England & Wales

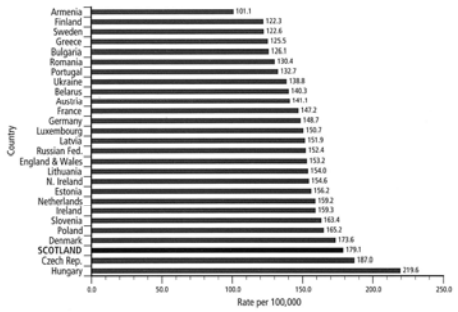


1991 deprivation census variables in Scotland and England & Wales

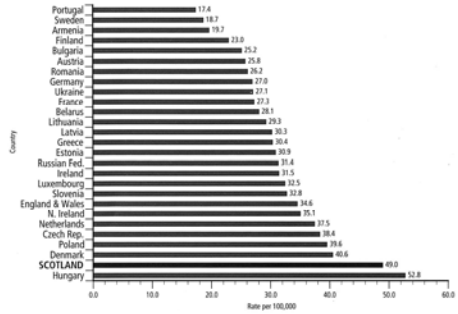
- Scotland is the only country in Europe that is losing population
- The Scottish population is ageing faster than in the rest of the UK
- Fertility rates are lower in Scotland that the rest of the UK



1996 age-standardised all cause mortality per 100,000 in Europe



1996 age-standardised mortality for all malignant neoplasms per 100,000 in Europe



1996 age-standardised mortality for malignant neoplasm of the trachea, bronchus and lung per 100,000 in Europe

- Scottish women have the lowest life expectancy in the EU
- Scottish men have the second lowest life expectancy in the EU
- Women in Scotland can currently expect to live five years less than women in Sweden
- Over the 20th century, Scotland's disadvantaged position is a relatively new phenomenon

So, what is the SLS?

So, what is the SLS?

www.lscs.ac.uk

- Provides linked data from the Scottish Census and administrative records
- Sampling based on 20 'semi-random' birthdays
- Initial sample drawn from the 1991 Census
- Similar sample drawn from 2001 Census

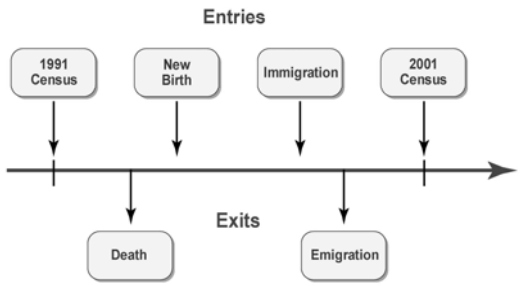
Data sources

- Census
  - 1991 Census, 2001 Census
  - Including data on occupation, economic activity, social class, housing, ethnicity, age, sex, marital status, household composition, health, education, country of birth, migration, workplace, religion etc.
- Population data
  - Immigration
  - Emigration
- Vital statistics
  - Births (SLS birthdate)
  - Births (to sample members)
  - Stillbirths
  - Infant mortality
  - Deaths
  - Widow(er)hoods
- Health data
  - Cancer registrations

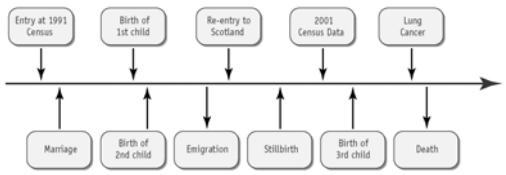
Data sources

- Census
  - 1991 Census, 2001 Census
  - Including data on occupation, economic activity, social class, housing, ethnicity, age, sex, marital status, household composition, health, education, country of birth, migration, workplace, religion etc.
- Population data
  - Immigration
  - Emigration
- Vital statistics
  - Births (SLS birthdate)
  - Births (to sample members)
  - Stillbirths
  - Infant mortality
  - Deaths
  - Widow(er)hoods
  - Marriages
- Health data
  - Cancer registrations
  - Hospital episodes

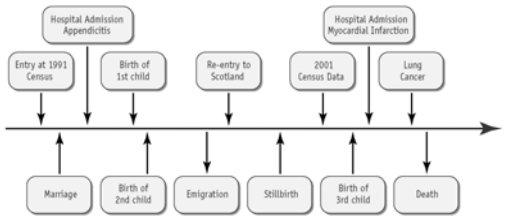
Entries & Exits to the Study



Example Event History of a Female SLS Member aged 21 at 1991 Census



Example Event History of a Female SLS Member aged 21 at 1991 Census



Strengths

- Sample size much larger than most surveys
  - BHPS has ~10,000 people in GB
  - SLS has ~265,000 traced members + ~505,000 household members in Scotland (1991)
- The census is compulsory
- Linkage and trace rates are high
- Includes those in communal establishments
- Ability to link hospital episodes data to socio-economic characteristics

Weaknesses

- Restricted range of variables
  - Income
  - Smoking
- Census information only collected every decade
- Not possible to return to the sample to ask extra questions
- The data are highly confidential

How does the SLS differ from the LS?

- Sample percentage larger (5.3% vs 1%)
- 20 SLS birthdays, but includes the four LS birthdays
- Fewer censuses captured
  - SLS 1991 & 2001
  - LS 1971, 1981, 1991, 2001
- Some census variables in the LS not coded in the SLS
  - e.g. 1991 place of work
- Some variables in the SLS not coded in the LS
  - e.g. hospital admissions and marriages
- The SLS is cheaper!

How far have we got with the 1991 census?

- Identification of the sample
  - Electronic records extracted from 1991 Census
  - Forms were 'picked'
  - Flagging data passed to NHSCR
  - 1991 sample traced and flagged
- Coding 1991 'difficult to code' information
  - Only originally coded for 10% Census
  - Designed interface for data input
  - Implemented occupation and industry coding software
  - Basic coding completed
  - Occupation and industry coding completed
- Programming derived variables
  - Completed

How far have we got with the 2001 census?

- Identification of the sample
  - Electronic records extracted from 2001 Census
  - Forms not accessed as 100% coded and names captured
  - Flagging and tracing data passed to NHSCR
  - 2001 sample traced
- Post-edit, pre-imputation data requested and received
- Post-imputation data also requested but not received
  - Confidentiality (record matching would reveal disclosure controls)
- Hence, in-house creation of all census-derived variables
  - ~80 completed
- SLS derived variables
  - Completed

- 1991 tracing and flagging through NHSCR
  - 274,055 initial sample
  - 270,385 excluding 377 duplicates and 3,293 dummies
  - 265,321 flagged at NHSCR (98.1% tracing rate)
  - 5064 no trace
  - 5.41% gross sampling fraction
  - 5.31% net sampling fraction

Age	Total			
	Traced cases	Un-traced cases	Selected sample	'Not traced' rate
0 – 4	17246	109	17355	0.63
5 – 9	17124	120	17244	0.70
10 – 14	16828	111	16939	0.66
15 – 19	17709	223	17932	1.24
20 – 24	19390	461	19851	2.32
25 – 29	20784	417	21201	1.97
30 – 34	20113	395	20508	1.93
35 – 39	18179	330	18509	1.78
40 – 44	18805	389	19194	2.03
45 – 49	15976	383	16359	2.34
50 – 54	14898	409	15307	2.67
55 – 59	14276	385	14661	2.63
60 – 64	13822	374	14196	2.63
65 – 69	12881	381	13262	2.87
70 – 74	10170	232	10402	2.23
75+	17120	345	17465	1.98
Total	265321	5064	270385	1.87

COUNTRY OF BIRTH	Total			
	Traced cases	Un-traced cases	Selected sample	'Not traced' rate
England & Wales	19044	746	19790	3.7695806
Scotland	237765	3138	240903	1.302599
Northern Ireland	1388	122	1510	8.0794702
Other UK	6	0	6	0
United Kingdom	258203	4006	262209	1.5277889
Outside UK	7118	1058	8176	12.940313
Irish Republic	1102	143	1245	11.485944
Old Commonwealth	801	56	857	6.5344224
New Commonwealth & Pakistan	2089	337	2426	13.891179
India	455	77	532	14.473684
Pakistan	419	113	532	21.240602
African Commonwealth	376	42	418	10.047847
Caribbean Commonwealth	93	9	102	8.8235294
Remainder New Commonwealth	746	96	842	11.401425
European Community	1265	121	1386	8.7301587
Other Europe	361	56	417	13.429257
USA	473	197	670	29.402985
Elsewhere & not stated	1027	148	1175	12.595745
Total	265321	5064	270385	1.872885

- 2001 tracing and matching through NHSCR
  - 268,428 initial sample (including duplicates)
  - 184,537 total traced (69%)
  - 165,227 automatic matched and traced at NHSCR (62%)
  - 19,310 traced at NHSCR (new entries) (7%)
  - 84,615 probability matched against CHI (31%)
  - 57,256 matched and sent to NHSCR (21%)
  - 27,359 unmatched, sent to Southport (10%)
  - Still dealing with around 8,000 problem cases (returned to original census forms)

- Vital statistics
  - 1991-2004 vital events provided so far
  - Linkage at NHSCR in progress for 2005

Hope to ‘complete’ the job by the end of 2006!

Governance issues

- Steering committee to oversee data protection, confidentiality and security issues
  - Peter Scrimgeour (GROS)
  - Ganka Mueller (GROS)
  - Robert Brown (GROS)
  - Ed Turnbull (GROS)
  - Muriel Douglas (NHSCR)
  - Paul Boyle (SLS)
  - Lin Hattersley (SLS / GROS)
  - Rod Muir (ISD, Cauldicot Guardian)
  - Louisa Blackwell (ONS)
  - Barbara Kelly (Lay member)

- Three successful applications to PAC
- Two successful applications to national MREC
- One successful visit to the Information Commissioner
- Established a SLS Research Board to assess all applications to use SLS data
- Data access protocols drafted

Security issues

- The SLS is an anonymous dataset; we do not hold name or address information
- Only those with a 'need to know' are aware of the 20 birthdays
- A third party is used to undertake the linkage between different datasets (NHSCR)
- Data are held in a keypad-secure environment
- Three databases (maintenance, development, production)

- Computers are on a password-protected, stand-alone network
- Team is based within the offices of GROS
- Data transfer is logged and monitored
- Abide by ONS protocols on data sharing, access and security
- Release of the results of data analysis overseen (cross-tabs)

Accessing the SLS

- A culture of data sharing
- The data will be kept in a 'secure environment'
- A support team already established to provide access to the data
- A web-based 'data dictionary' currently being completed
- Two methods of data analysis (current plan)
  - 'Remote access': Actual data will not be released – only an image of the dataset will be released (in SPSS, SAS or STATA)
  - 'Safe-setting': In-house modelling of individual-level data (by support team)

Administrator receives initial request for use of SLS (usually via email). This is logged on web-based tracking system; all key stages indicated

Support person contacts user via email

User encouraged to attend user-training session (held every few months) if such a session has not been attended previously

Project details discussed, project request form completed

Project request form submitted to SLS 'Research Board'

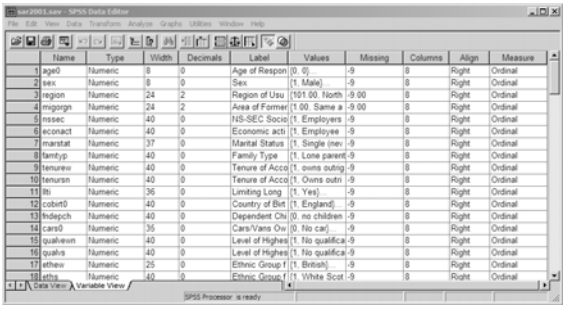
Users informed of rejected proposals and revisions discussed

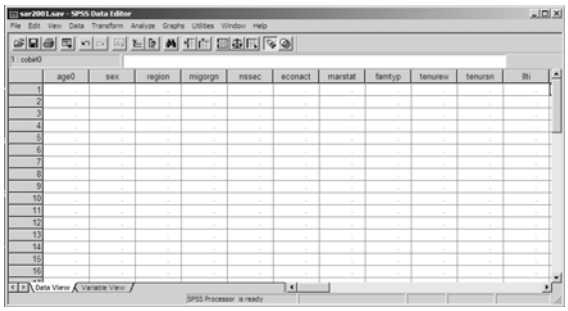
Dataset created by user support person, including bespoke variables. 'Dataless' file and basic variable information sent to user

User defines analysis using SAS, SPSS, STATA 'dataless' file; support person carries out user-defined analysis

Support person checks results are not disclosive and returns results to user

Project results written up into presentations / papers and discussed with support person





The future...?

- Linkage of additional data into the SLS?
  - Addition of pre-1991, post-1974 fertility events
  - Educational data (school census and exam results)
  - Historical IQ tests (tested 1932, born in 1921)?
  - DWP 'claimant count cohort data'?
  - Small-area geographical estimates of income and health-related behaviours?
  - Linkage back to 1981?
  - Information on parents of SLS members from DIGROS?
  - Creation of a UK LS?

Launch / Training

- Launch and first training course, Easter 2007

Conferences

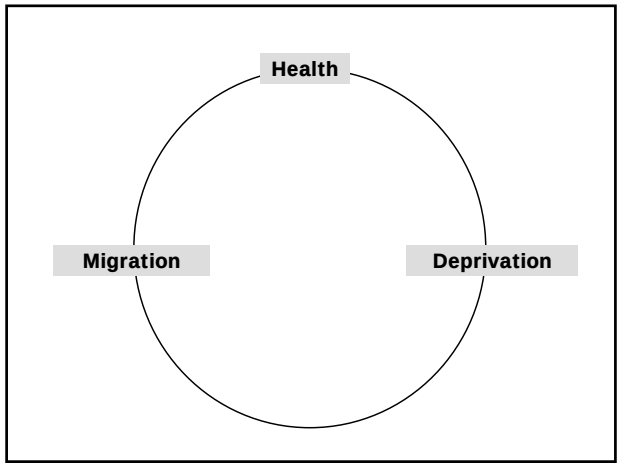
- The Long View: Longitudinal Studies in Scotland (February 2004)
- Health Research: Record Linkage Studies and Longitudinal Approaches (Summer 2007)

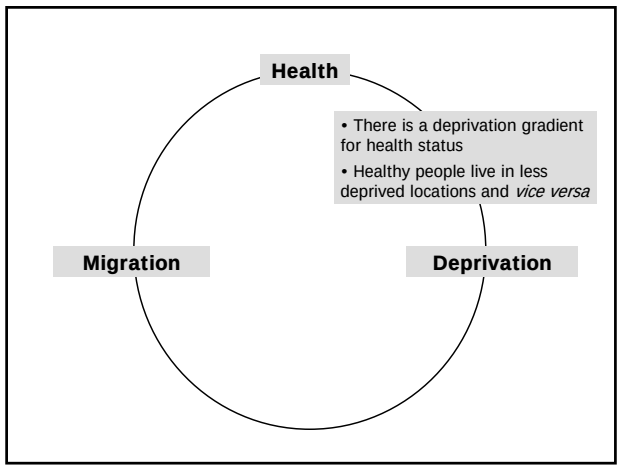
SCIENCE & TECHNOLOGY

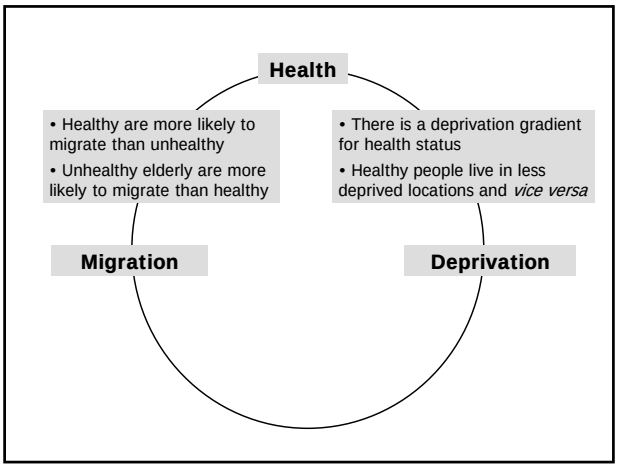
Can we beat the 'Scottish effect'?

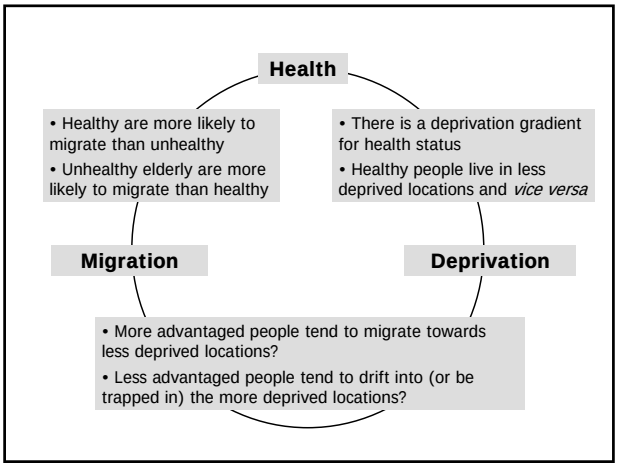
Over 250,000 residents could reveal why Scots are even more unhealthy than they should be, says IAN JOHNSTON

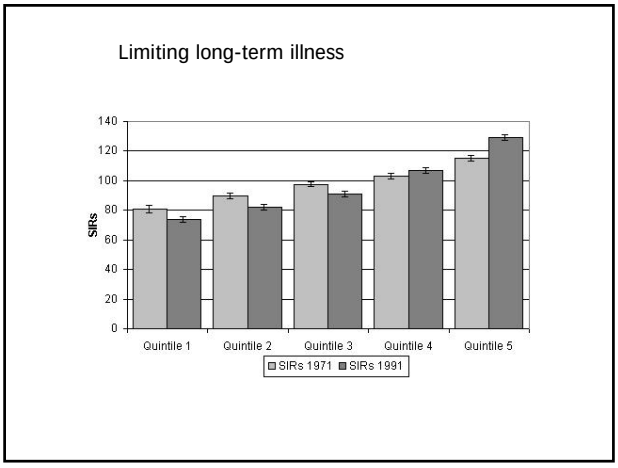
Elderly



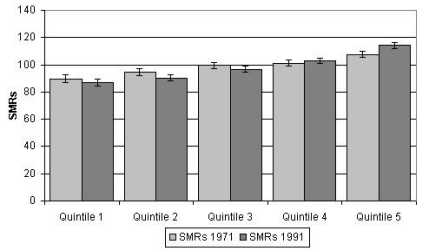



Health
• Healthy are more likely to migrate than unhealthy
• Unhealthy elderly are more likely to migrate than healthy
• There is a deprivation gradient for health status
• Healthy people live in less deprived locations and *vice versa*
Migration
Deprivation

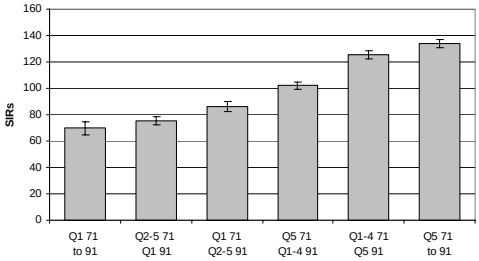

Health
• Healthy are more likely to migrate than unhealthy
• Unhealthy elderly are more likely to migrate than healthy
• There is a deprivation gradient for health status
• Healthy people live in less deprived locations and *vice versa*
Migration
Deprivation
• More advantaged people tend to migrate towards less deprived locations?
• Less advantaged people tend to drift into (or be trapped in) the more deprived locations?


Limiting long-term illness
140
120
100
80
60
40
20
0
SIRs
Quintile 1
Quintile 2
Quintile 3
Quintile 4
Quintile 5
SIRs 1971
SIRs 1991

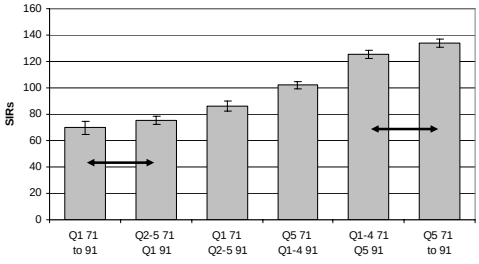
All-cause mortality



Limiting long-term illness



Limiting long-term illness



Limiting long-term illness