E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Should we be making better use of public data in health research?

Paul Boyle

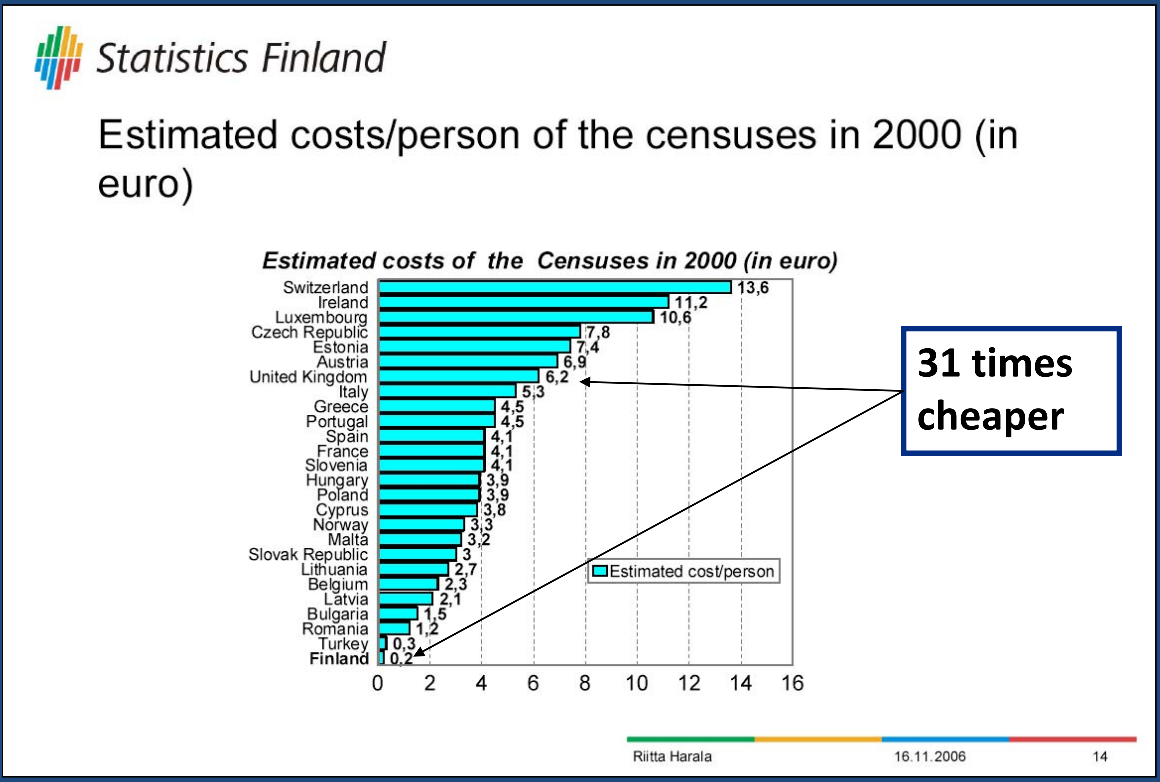
The value of routine administrative data

Routine administrative data are no panacea, but they do provide large (national?) samples

Representative and longitudinal

Little response bias or attrition

Cheap to create and update



Statistics Finland

Estimated costs/person of the censuses in 2000 (in euro)

31 times cheaper

Riitta Harala 16.11.2006 14

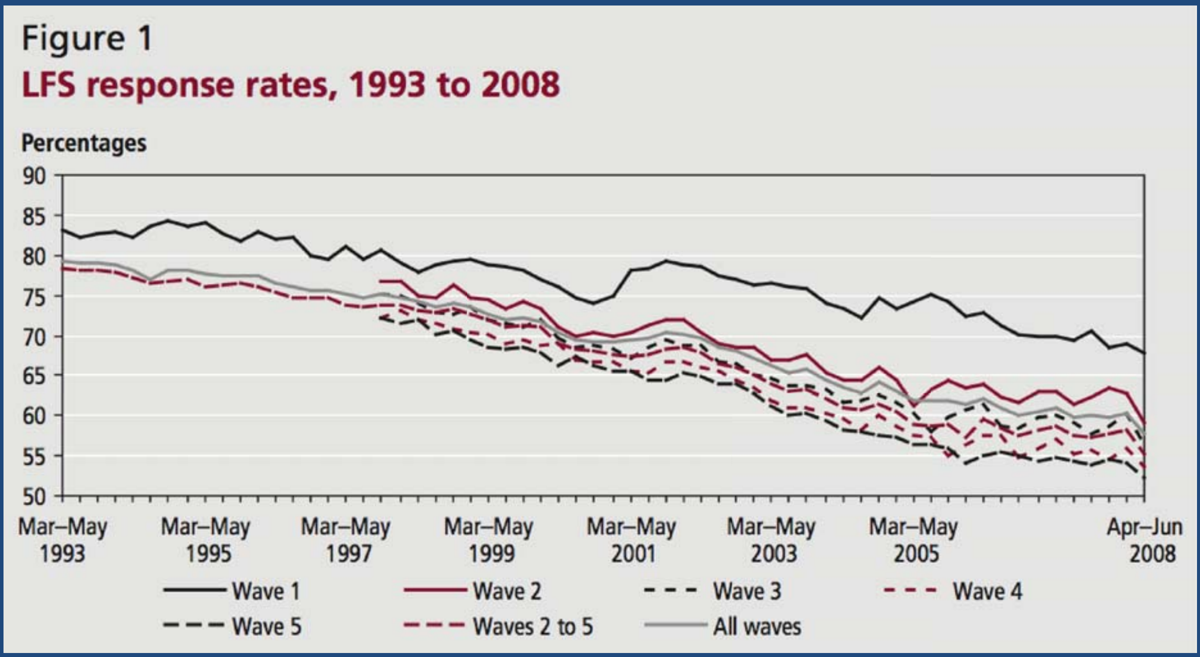
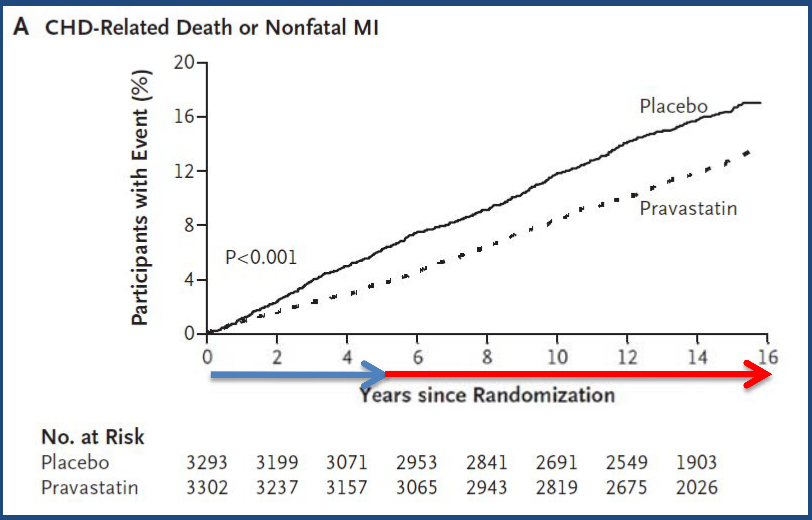


Figure 1

LFS response rates, 1993 to 2008



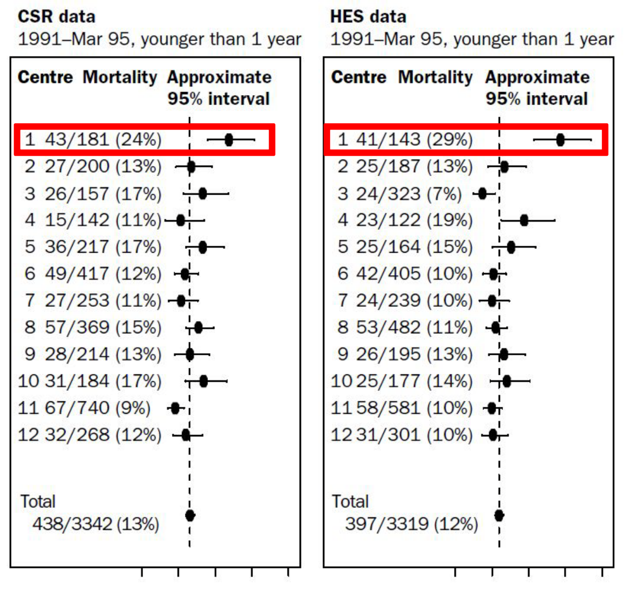
A CHD-Related Death or Nonfatal MI

West of Scotland Coronary Prevention Study was a RCT comparing pravastatin with placebo in men with hypercholesterolemia

Ford I *et al.* Long-term follow-up of the West of Scotland Coronary Prevention Study *N Engl J Med* 357 1477-86

Paediatric cardiac surgery at the Bristol Royal Infirmary

UK Cardiac Surgical Register and Hospital Episode Statistics



CSR data
1991–Mar 95, younger than 1 year

Centre	Mortality
1	43/181 (24%)
2	27/200 (13%)
3	26/157 (17%)
4	15/142 (11%)
5	36/217 (17%)
6	49/417 (12%)
7	27/253 (11%)
8	57/369 (15%)
9	28/214 (13%)
10	31/184 (17%)
11	67/740 (9%)
12	32/268 (12%)
Total	438/3342 (13%)

HES data
1991–Mar 95, younger than 1 year

Centre	Mortality
1	41/143 (29%)
2	25/187 (13%)
3	24/323 (7%)
4	23/122 (19%)
5	25/164 (15%)
6	42/405 (10%)
7	24/239 (10%)
8	53/482 (11%)
9	26/195 (13%)
10	25/177 (14%)
11	58/581 (10%)
12	31/301 (10%)
Total	397/3319 (12%)

Paul Aylin *et al.* 2001 Comparison of UK paediatric cardiac surgical performance by analysis of routinely collected data 1984–96: was Bristol an outlier? *Lancet 358 181–87*

Amyotrophic Lateral Sclerosis (ALS)

Most common form of motor neuron disease

A rapidly progressive, fatal neuromuscular disease, with no known cause or cure

Around 5-10% of ALS cases are likely inherited, but the relative influence of genes vs environment yet to be determined

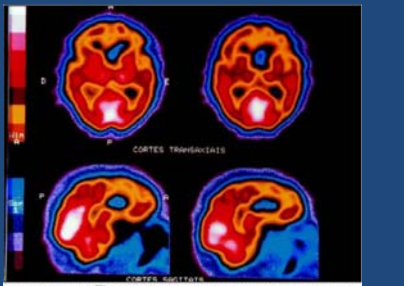
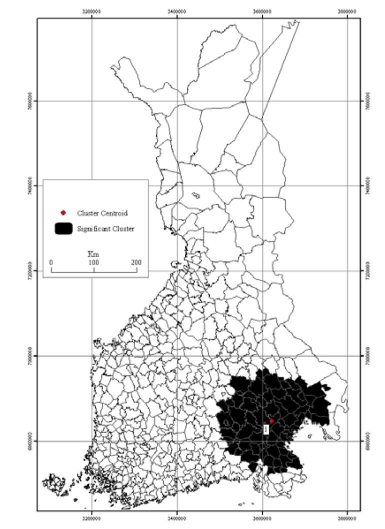


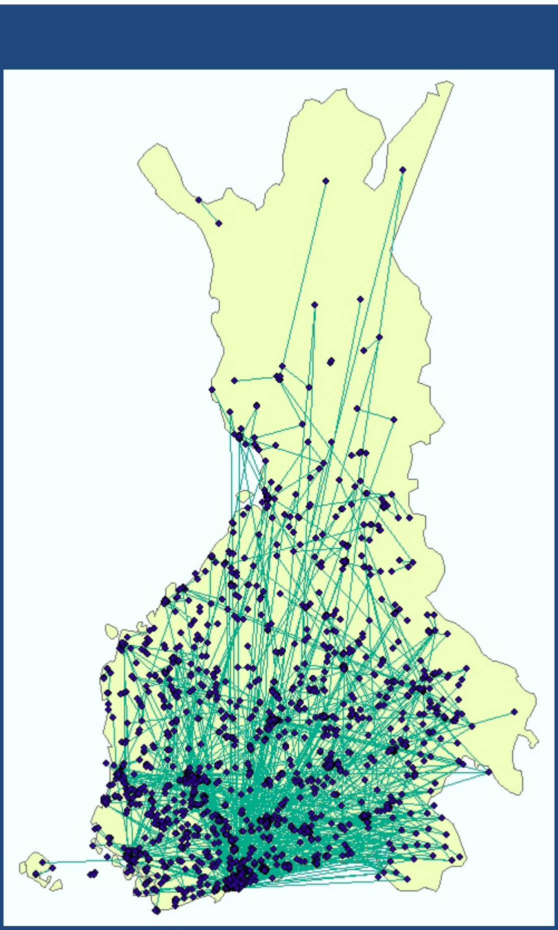
Fig 3. Brain SPECT using ^{99m}Tc HMPAO: transaxial and coronal images showing bilateral frontal and temporal hypoperfusion, mainly on the left side. Yellow regions represent decreased relative perfusion (see color scale on the left side).

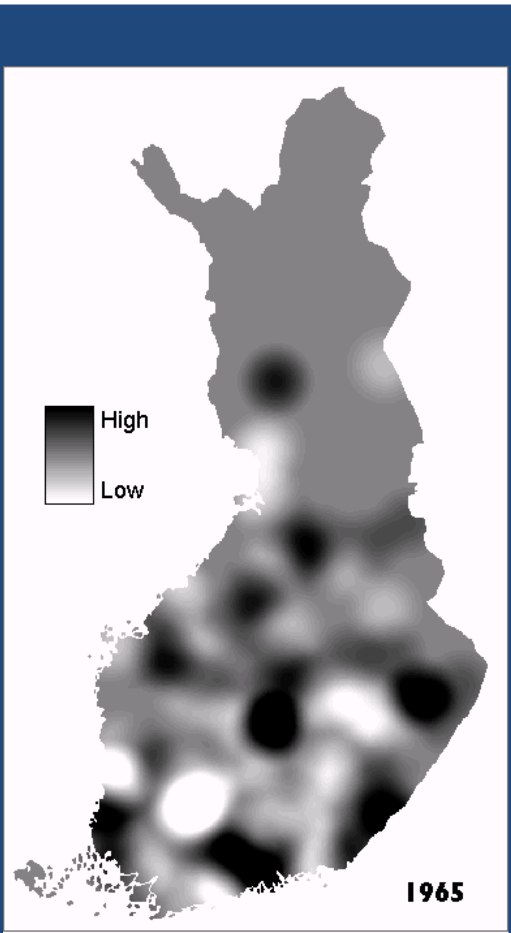
Previous study found significant clustering of ALS at the time of birth in south-east Finland

This could support either a genetic or an environmental hypothesis…



Sabel CE, Boyle PJ, Löytönen M, Gatrell AC, Jokelainen M, Flowerdew R and Maasilta P 2003 The spatial clustering of Amyotrophic Lateral Sclerosis in Finland at place of birth and place of death *American Journal of Epidemiology* 157: 898-905





The cases were significantly more likely to remain in SE Finland after birth, compared to the geographically matched controls

Moving away may be protective, pointing towards an environmental risk factor after birth

Sabel CE, Boyle PJ, Raab G, Löytönen M and Maasilta P 2009 Modelling individual space-time exposure opportunities: A novel approach to unravelling the genetic or environment disease causation debate *Spatial and Spatio-Temporal Epidemiology* 1 85-94

So,why don’t we make more use of routine administrative data in the UK?

Legal and cultural concerns

Various relevant laws including: the Data Protection Act, common law, European legislation and UK statutes

Hence the legal basis for such sharing is not clear cut (Thomas and Walport, 2008)

Uncertainty over definitions and interpretations: ‘what is personal data?’, ‘is consent required?’

DPA does not prevent legitimate research from using personal data

Section 33 provides limited exemptions to some of the data protection principles where personal data are to be processed for ‘research purposes’

Researchers do not need the consent of data subjects to process personal data where the section 33 ‘research exemption’ applies

To qualify for the ‘research exemption’ under section 33, the researcher needs to confirm that the personal data will not be processed:

1. In order to support measures or decisions with respect to particular individuals

2. In a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject

If the relevant conditions of use of the personal data are met, then personal data:

May be used for purposes other than they were originally collected for

May be kept indefinitely

Are exempt from a data subject’s rights of access, where the research/statistical results do not identify the data subject

Public concerns

protection of people’s privacy …

versus

… creation of bona fide and valuable knowledge about population and society

“Despite my background as a civil libertarian... I question the primacy of individual freedom (and its associated concepts – autonomy, privacy, and liberty) as the prevailing social norm. Freedom is a powerful and important idea, but I think scholars have given insufficient attention to equally strong values that are captured by the notions of partnership, citizenship, and community....”

Lawrence Gostin *Public Health Law* (University of California Press)

We need more active engagement with the public – a 'social contract' based on an informed understanding of research benefits

We have to explain how data are reliable, valuable, and can be properly managed

The failure to make better use of routinely-collected public data can be argued to be a criminal waste of public resources

The Scottish Longitudinal Study

SLS sample

Academic and government collaboration

Provides linked data from the Scottish Census and administrative records

Based on 20 ‘semi-random’ birthdays (5.3%)

Initial sample drawn from the 1991 Census (265,321) and 2001 Census (193,717)

SLS data sources

Census

- 1991 Census, 2001 Census:
  Data on age, sex, occupation, economic activity, social class, housing, ethnicity, marital status, household composition, health, education, country of birth, migration, workplace, religion etc.

- Information on SLS member and other household members

Population data

- Immigration
- Emigration

Vital statistics

- Births (SLS birthdate)
- Births (to sample members)
- Stillbirths
- Infant mortality
- Deaths
- Widow(er)hoods
- Divorces
- Marriages

Health data

- Cancer registrations
- Hospital episodes

Education data

- School census
- Exam results
- Absences, exclusions

Governance issues

Steering committee to oversee data protection, confidentiality and security issues

SLS Research Board to oversee applications to use the data

Three successful applications to the Privacy Advisory Committee (PAC), two to the Multi-centre Research Ethics Committee (MREC), one to the Information Commissioner

Security issues

The SLS is an psuedo-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 (NHSCR) is used to undertake the linkage between different datasets (Chinese wall)

Data are held in a keypad-secure environment

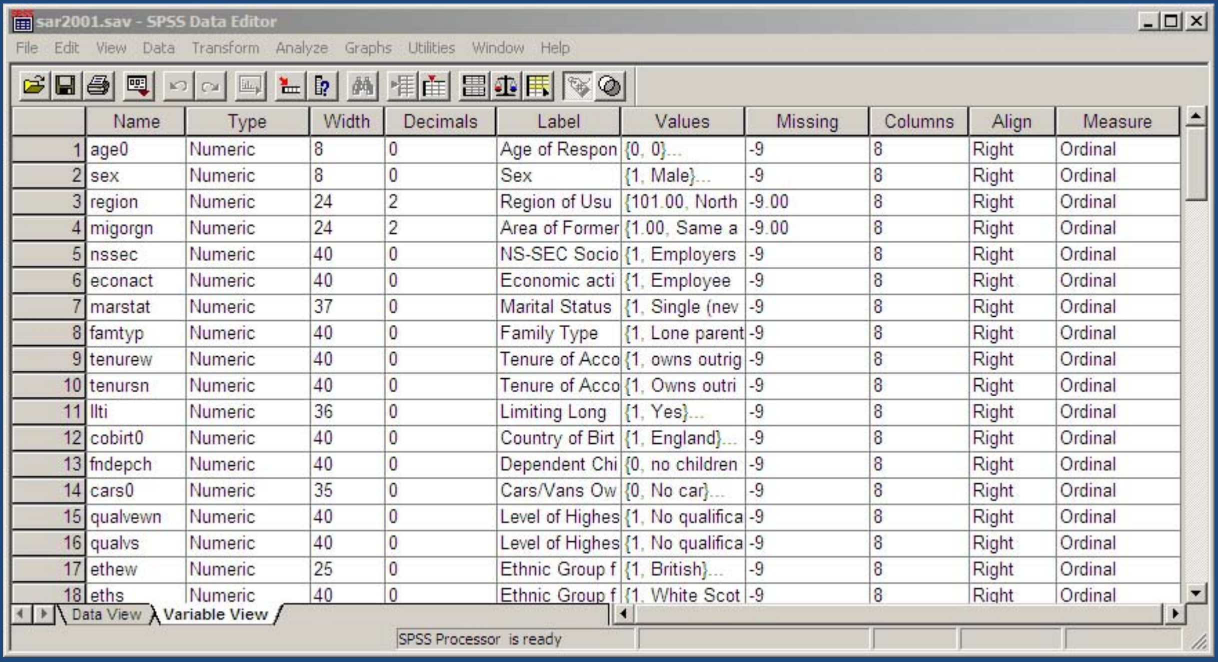
Accessing the SLS

A culture of data sharing

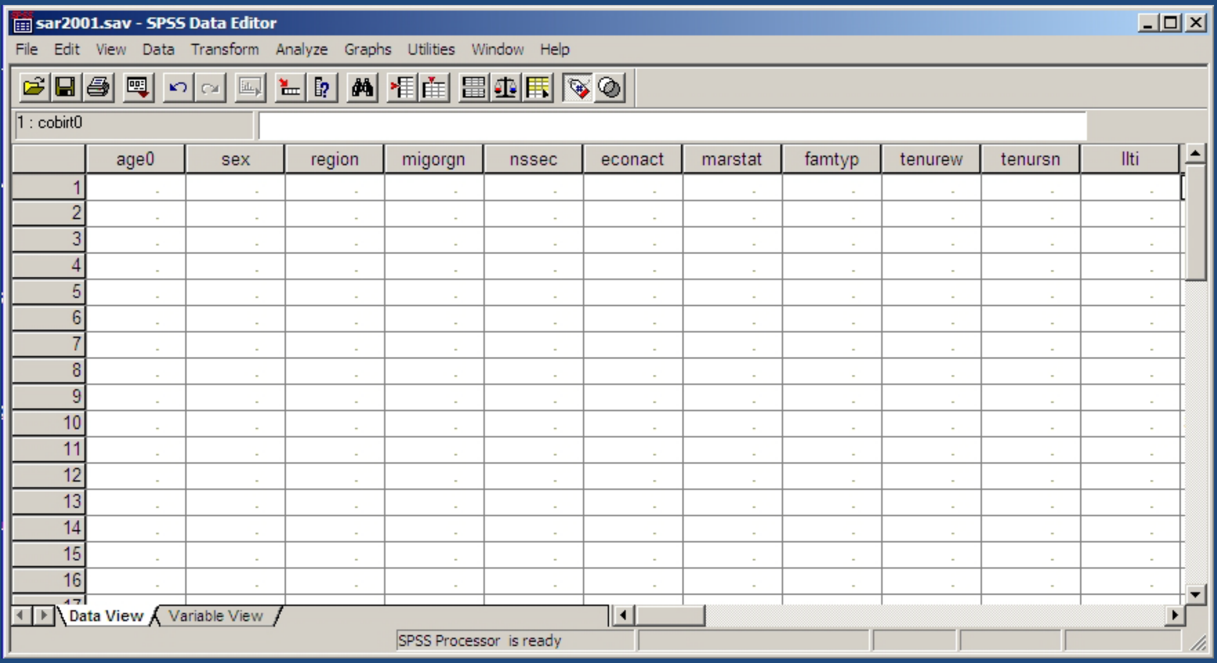
Funded support team provides access

Two methods of data analysis

- ‘Remote access’: Actual data not released – only an image of the dataset (in SPSS, SAS or STATA)
- ‘Safe-setting’: In-house modelling of individual-level data


sar2001.sav - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	age0	Numeric	8	0	Age of Respon	{0, 0}...	-9	8	Right	Ordinal
2	sex	Numeric	8	0	Sex	{1, Male}...	-9	8	Right	Ordinal
3	region	Numeric	24	2	Region of Usu	{101.00, North	-9.00	8	Right	Ordinal
4	migorgn	Numeric	24	2	Area of Former	{1.00, Same a	-9.00	8	Right	Ordinal
5	nssec	Numeric	40	0	NS-SEC Socio	{1, Employers	-9	8	Right	Ordinal
6	econact	Numeric	40	0	Economic acti	{1, Employee	-9	8	Right	Ordinal
7	marstat	Numeric	37	0	Marital Status	{1, Single (nev	-9	8	Right	Ordinal
8	famtyp	Numeric	40	0	Family Type	{1, Lone parent	-9	8	Right	Ordinal
9	tenurew	Numeric	40	0	Tenure of Acco	{1, owns outrig	-9	8	Right	Ordinal
10	tenursn	Numeric	40	0	Tenure of Acco	{1, Owns outri	-9	8	Right	Ordinal
11	llti	Numeric	36	0	Limiting Long	{1, Yes}...	-9	8	Right	Ordinal
12	cobirt0	Numeric	40	0	Country of Birt	{1, England}...	-9	8	Right	Ordinal
13	fndepch	Numeric	40	0	Dependent Chi	{0, no children	-9	8	Right	Ordinal
14	cars0	Numeric	35	0	Cars/Vans Ow	{0, No car}...	-9	8	Right	Ordinal
15	qualvewn	Numeric	40	0	Level of Highes	{1, No qualifica	-9	8	Right	Ordinal
16	qualvs	Numeric	40	0	Level of Highes	{1, No qualifica	-9	8	Right	Ordinal
17	ethew	Numeric	25	0	Ethnic Group f	{1, British}...	-9	8	Right	Ordinal
18	eths	Numeric	40	0	Ethnic Group f	{1, White Scot	-9	8	Right	Ordinal



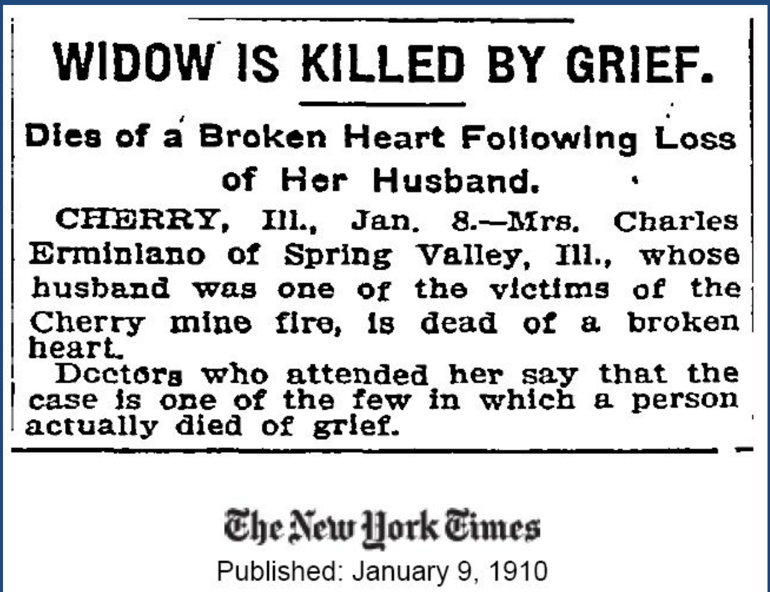
1 : cobirt0

	age0	sex	region	migorgn	nssec	econact	marstat	famtyp	tenurew	tenursn	llti
1	.	.	.	.	.	.	.	.	.	.	.
2	.	.	.	.	.	.	.	.	.	.	.
3	.	.	.	.	.	.	.	.	.	.	.
4	.	.	.	.	.	.	.	.	.	.	.
5	.	.	.	.	.	.	.	.	.	.	.
6	.	.	.	.	.	.	.	.	.	.	.
7	.	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	.	.	.	.	.	.	.
9	.	.	.	.	.	.	.	.	.	.	.
10	.	.	.	.	.	.	.	.	.	.	.
11	.	.	.	.	.	.	.	.	.	.	.
12	.	.	.	.	.	.	.	.	.	.	.
13	.	.	.	.	.	.	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	.
15	.	.	.	.	.	.	.	.	.	.	.
16	.	.	.	.	.	.	.	.	.	.	.

SPSS Processor is ready

Making use of the Scottish Longitudinal Study

Widowhood and life expectancy



WIDOW IS KILLED BY GRIEF.

Dies of a Broken Heart Following Loss of Her Husband.

CHERRY, Ill., Jan. 8.—Mrs. Charles Erminiano of Spring Valley, Ill., whose husband was one of the victims of the Cherry mine fire, is dead of a broken heart.

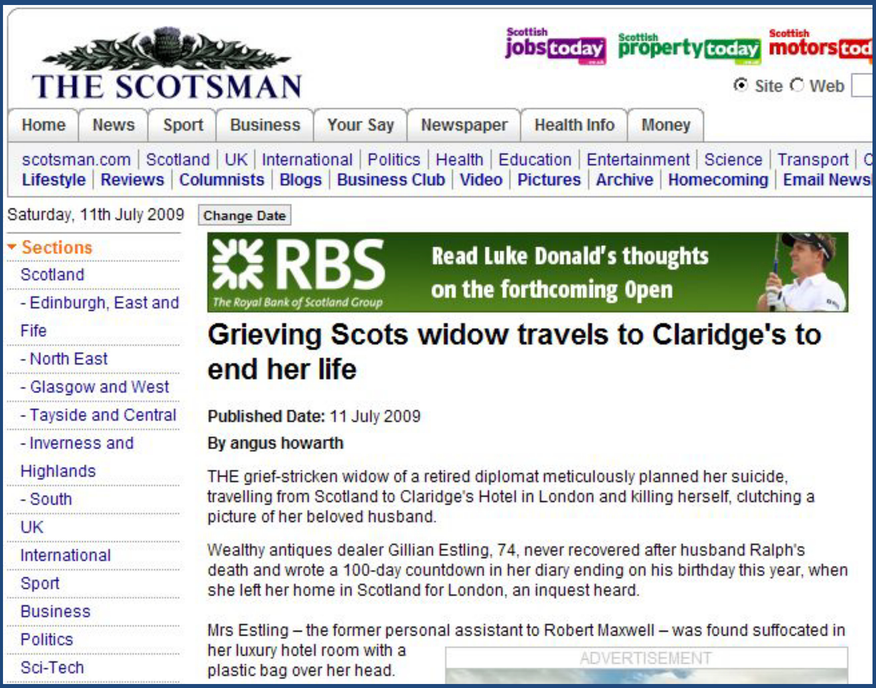
Doctors who attended her say that the case is one of the few in which a person actually died of grief.

The New York Times

Published: January 9, 1910

Boyle PJ, Feng Z and Raab G 2011 Does widowhood increase mortality risk? Comparing different causes of spousal death to test for selection effects *Epidemiology* 22 1-5

Widowhood and life expectancy


THE SCOTSMAN
Scottish jobs today
Scottish property today
Scottish motors tod
Site ○ Web
Home
News
Sport
Business
Your Say
Newspaper
Health Info
Money
scotsman.com | Scotland | UK | International | Politics | Health | Education | Entertainment | Science | Transport | C
Lifestyle | Reviews | Columnists | Blogs | Business Club | Video | Pictures | Archive | Homecoming | Email News
Change Date
Sections
Scotland
- Edinburgh, East and Fife
- North East
- Glasgow and West
- Tayside and Central
- Inverness and Highlands
- South
UK
International
Sport
Business
Politics
Sci-Tech
RBS The Royal Bank of Scotland Group
Read Luke Donald's thoughts on the forthcoming Open
ADVERTISEMENT

Saturday, 11th July 2009

Grieving Scots widow travels to Claridge's to end her life

Published Date: 11 July 2009

By angus howarth

THE grief-stricken widow of a retired diplomat meticulously planned her suicide, travelling from Scotland to Claridge's Hotel in London and killing herself, clutching a picture of her beloved husband.

Wealthy antiques dealer Gillian Estling, 74, never recovered after husband Ralph's death and wrote a 100-day countdown in her diary ending on his birthday this year, when she left her home in Scotland for London, an inquest heard.

Mrs Estling – the former personal assistant to Robert Maxwell – was found suffocated in her luxury hotel room with a plastic bag over her head.

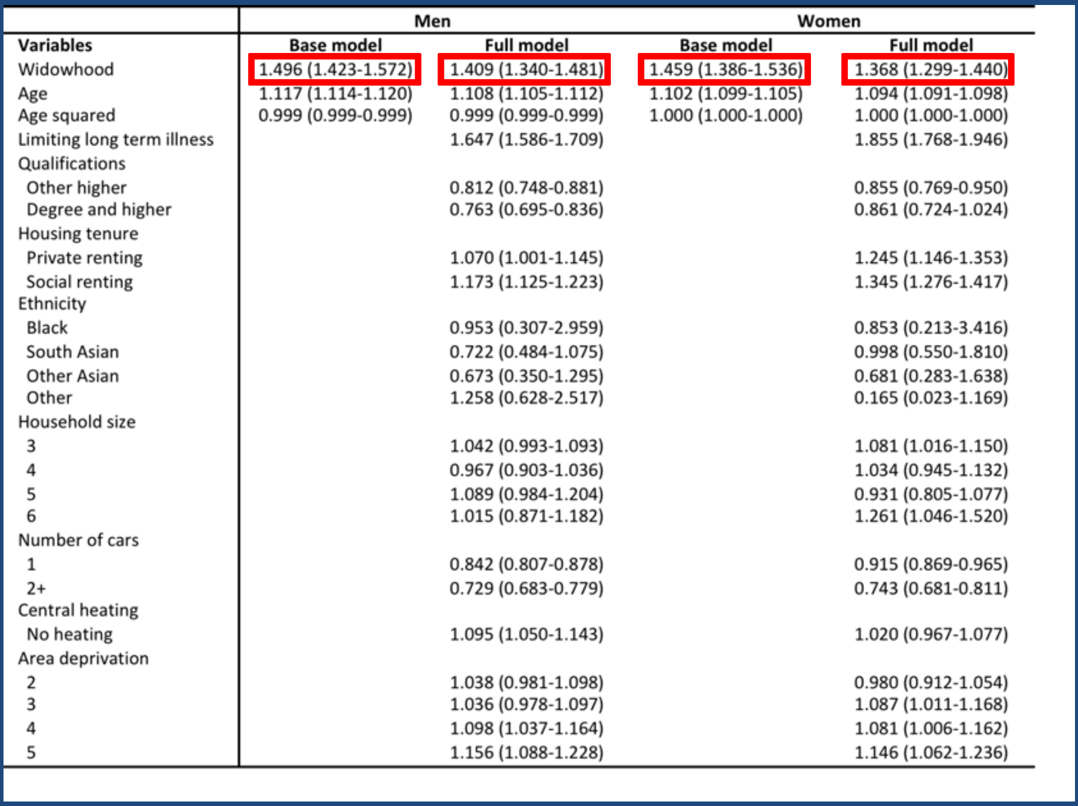
Boyle PJ, Feng Z and Raab G 2011 Does widowhood increase mortality risk? Comparing different causes of spousal death to test for selection effects *Epidemiology* 22 1-5

Many studies demonstrate raised risk (10-40%) of death following widowhood

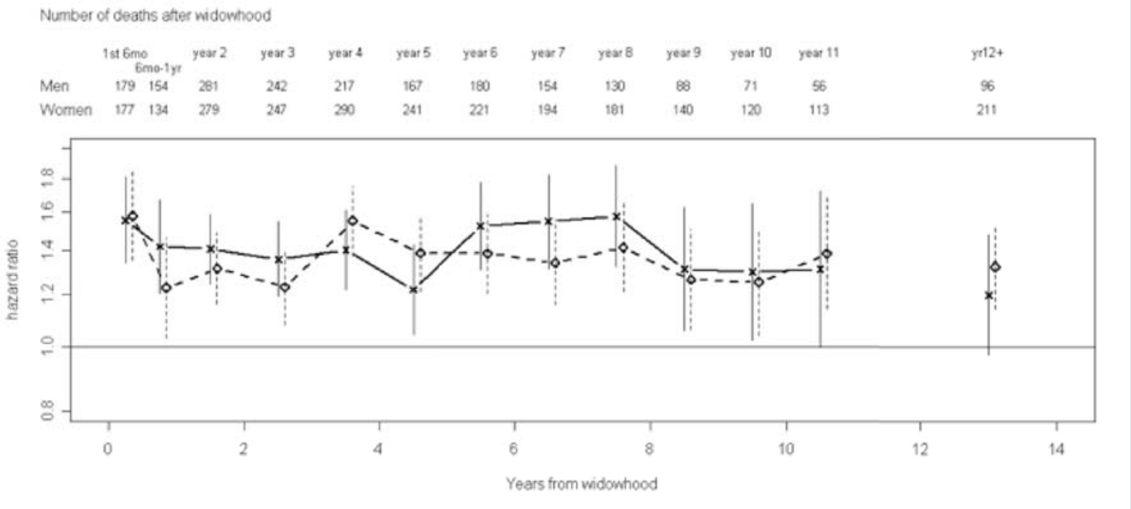
The emotional stress (‘broken heart’ effect)

Protective effects of marriage are eroded

Marriage selection



	Men		Women	
Variables	Base model	Full model	Base model	Full model
Widowhood	1.496 (1.423-1.572)	1.409 (1.340-1.481)	1.459 (1.386-1.536)	1.368 (1.299-1.440)
Age	1.117 (1.114-1.120)	1.108 (1.105-1.112)	1.102 (1.099-1.105)	1.094 (1.091-1.098)
Age squared	0.999 (0.999-0.999)	0.999 (0.999-0.999)	1.000 (1.000-1.000)	1.000 (1.000-1.000)
Limiting long term illness		1.647 (1.586-1.709)		1.855 (1.768-1.946)
Qualifications				
Other higher		0.812 (0.748-0.881)		0.855 (0.769-0.950)
Degree and higher		0.763 (0.695-0.836)		0.861 (0.724-1.024)
Housing tenure				
Private renting		1.070 (1.001-1.145)		1.245 (1.146-1.353)
Social renting		1.173 (1.125-1.223)		1.345 (1.276-1.417)
Ethnicity				
Black		0.953 (0.307-2.959)		0.853 (0.213-3.416)
South Asian		0.722 (0.484-1.075)		0.998 (0.550-1.810)
Other Asian		0.673 (0.350-1.295)		0.681 (0.283-1.638)
Other		1.258 (0.628-2.517)		0.165 (0.023-1.169)
Household size				
3		1.042 (0.993-1.093)		1.081 (1.016-1.150)
4		0.967 (0.903-1.036)		1.034 (0.945-1.132)
5		1.089 (0.984-1.204)		0.931 (0.805-1.077)
6		1.015 (0.871-1.182)		1.261 (1.046-1.520)
Number of cars				
1		0.842 (0.807-0.878)		0.915 (0.869-0.965)
2+		0.729 (0.683-0.779)		0.743 (0.681-0.811)
Central heating				
No heating		1.095 (1.050-1.143)		1.020 (0.967-1.077)
Area deprivation				
2		1.038 (0.981-1.098)		0.980 (0.912-1.054)
3		1.036 (0.978-1.097)		1.087 (1.011-1.168)
4		1.098 (1.037-1.164)		1.081 (1.006-1.162)
5		1.156 (1.088-1.228)		1.146 (1.062-1.236)



Number of deaths after widowhood

	1st 6mo	6mo-1yr	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10	year 11	yr12+
Men	179	154	281	242	217	167	180	154	130	88	71	56	96
Women	177	134	279	247	290	241	221	194	181	140	120	113	211

So where next?

A national strategy for admin data?

Establish a taskforce including members from Government departments, funders and academic experts

Terms of reference likely include:

Development of common procedures

Development of agreed data linkage methods

Clarification of the legal situation

Clarification of when consent is required

Recommendations for legislation (if required)

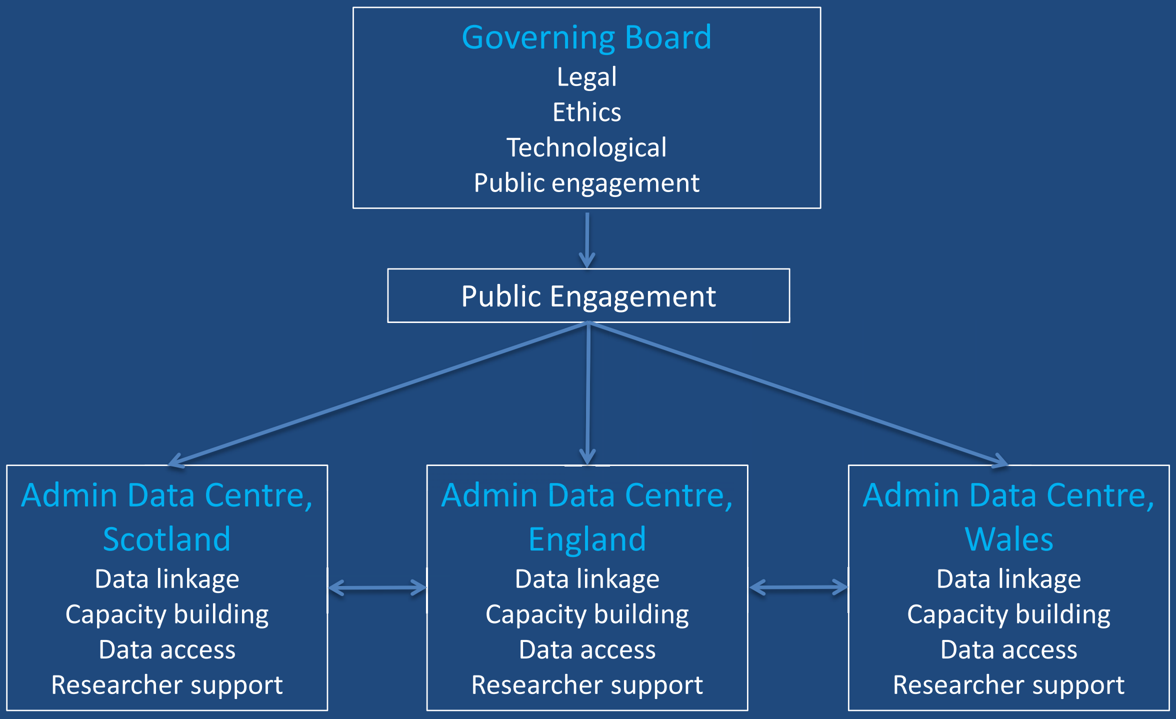
Development of a ‘metadata authority’

Public awareness procedures

Guidance on data access

Recommendations for assessing data quality

Estimates of resource implications



Governing Board

Legal
Ethics
Technological
Public engagement

Public Engagement

Admin Data Centre, Scotland

Data linkage
Capacity building
Data access
Researcher support

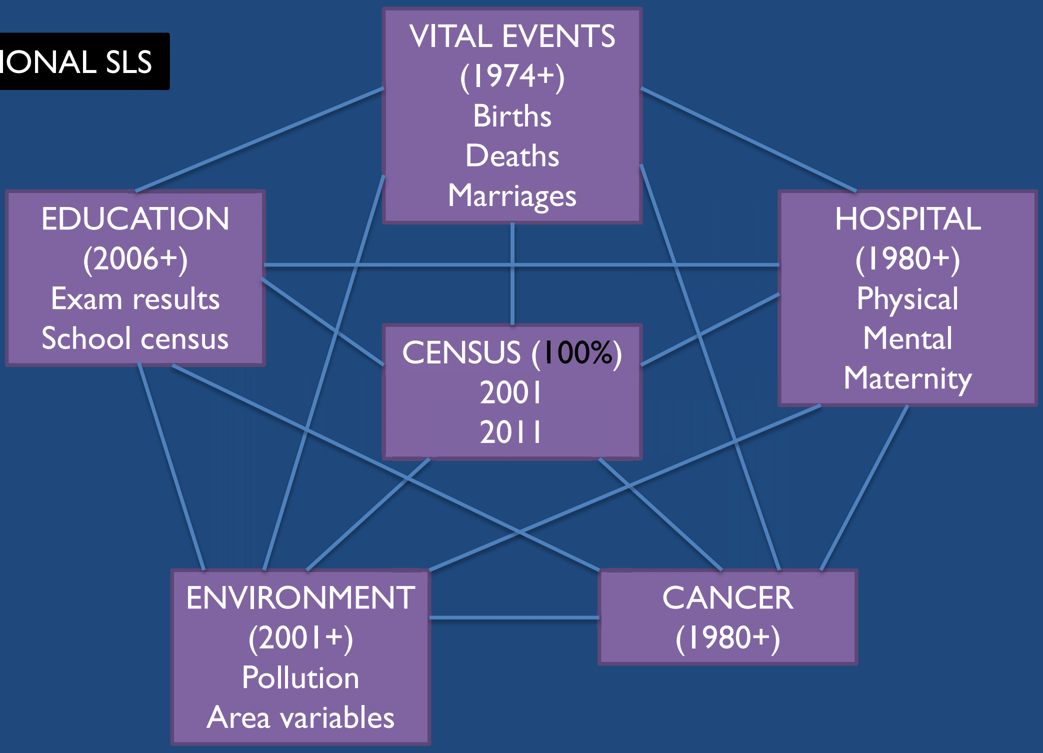
Admin Data Centre, England

Data linkage
Capacity building
Data access
Researcher support

Admin Data Centre, Wales

Data linkage
Capacity building
Data access
Researcher support

NATIONAL SLS



VITAL EVENTS
(1974+)
Births
Deaths
Marriages

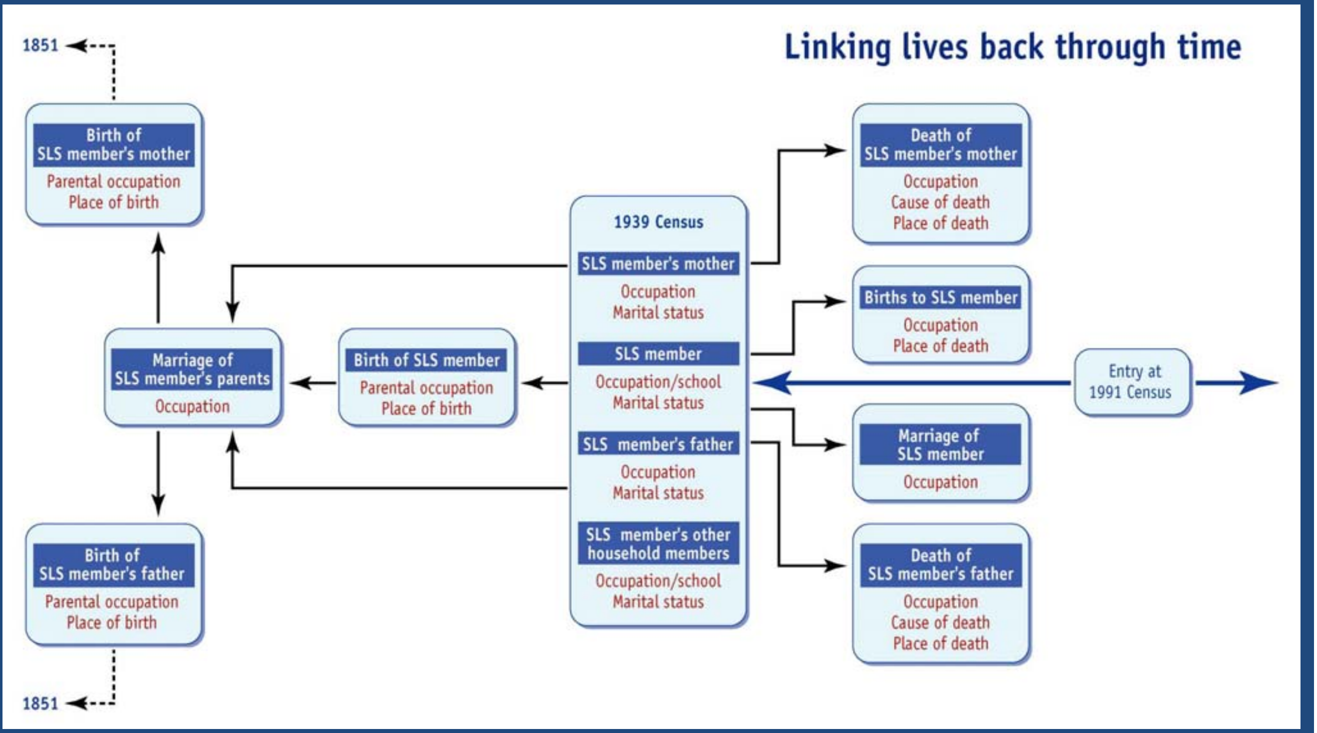
EDUCATION
(2006+)
Exam results
School census

HOSPITAL
(1980+)
Physical
Mental
Maternity

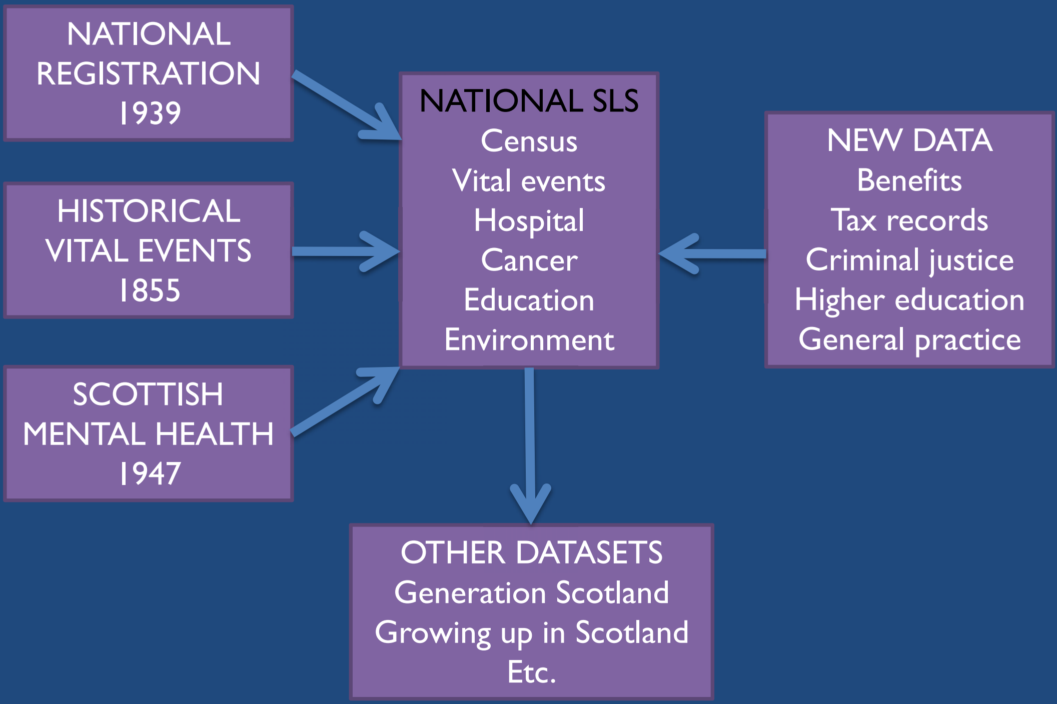
CENSUS (100%)
2001
2011

ENVIRONMENT
(2001+)
Pollution
Area variables

CANCER
(1980+)



SCOTLAND'S PEOPLE



NATIONAL REGISTRATION 1939

HISTORICAL VITAL EVENTS 1855

SCOTTISH MENTAL HEALTH 1947

NATIONAL SLS
Census
Vital events
Hospital
Cancer
Education
Environment

NEW DATA
Benefits
Tax records
Criminal justice
Higher education
General practice

OTHER DATASETS
Generation Scotland
Growing up in Scotland
Etc.

Conclusion

Routinely collected admin data are under-utilised

They are paid for from the public purse

Robust models exist for managing and providing access to anonymised data for valuable research

With appropriate academic / government partnerships we could make significant progress