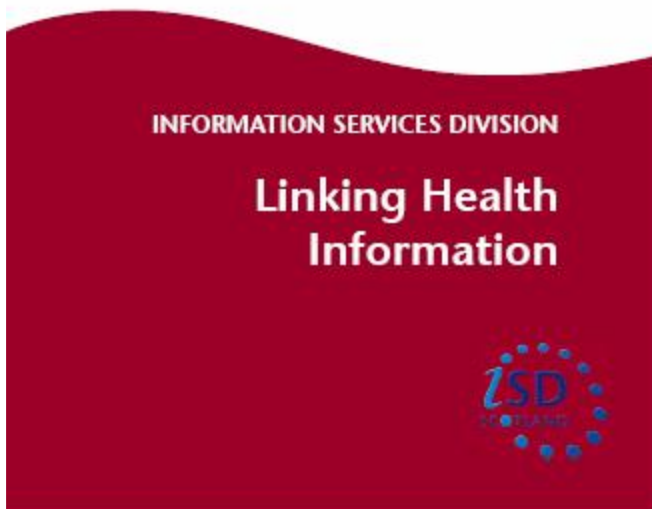




Linking Health Information in Scotland

Carole Morris
Information Services (ISD)



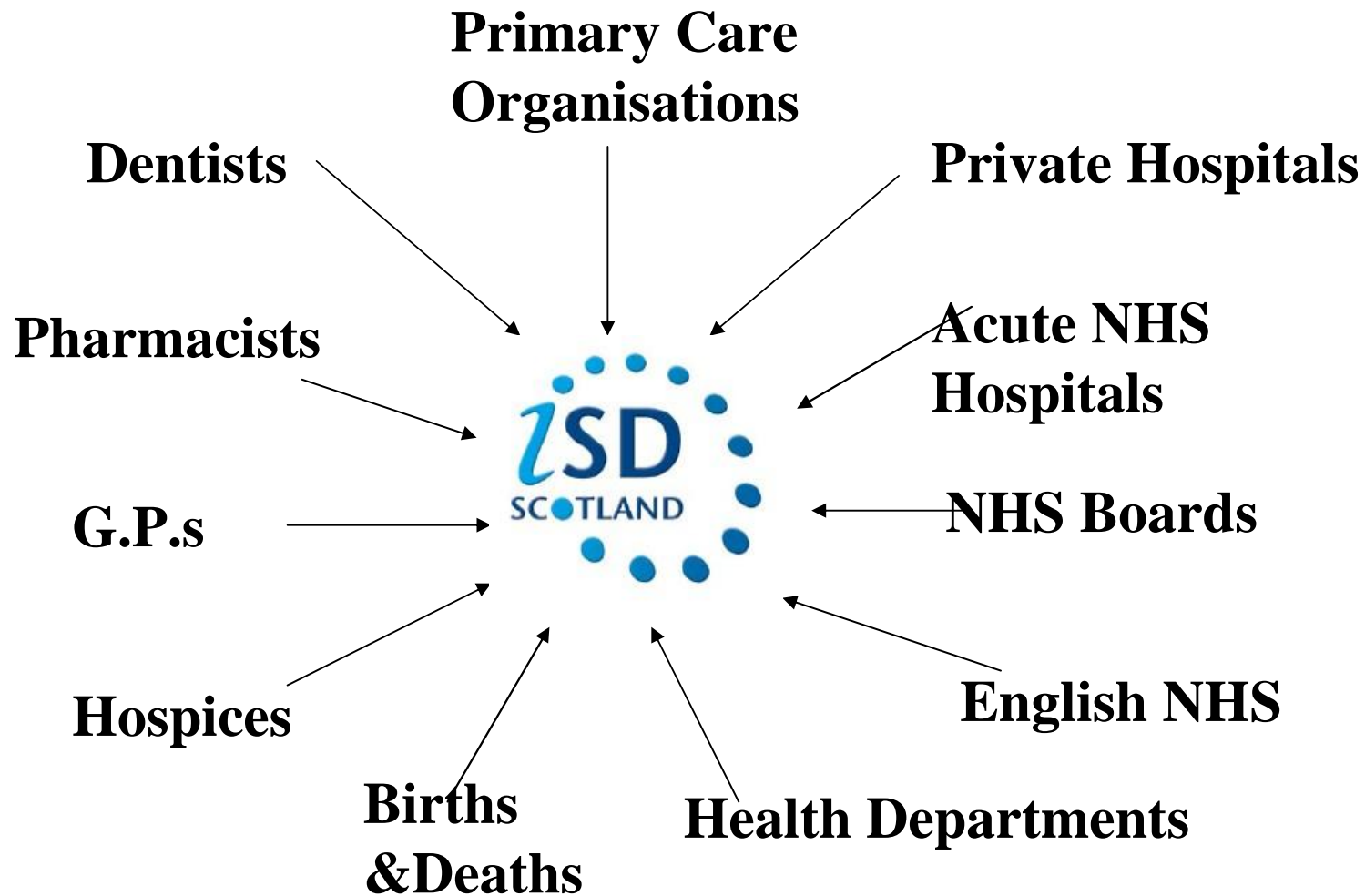
Linking Health Information in Scotland



- **Information Services Division (ISD)**
 - National data flows
- **Record linkage in Scotland**
 - What makes record linkage possible?
 - ISD's linked datasets
 - Linked data in the NHS
 - Examples of linked data
- **Chief Scientist Office (CSO)**
- **NHS data warehouse**

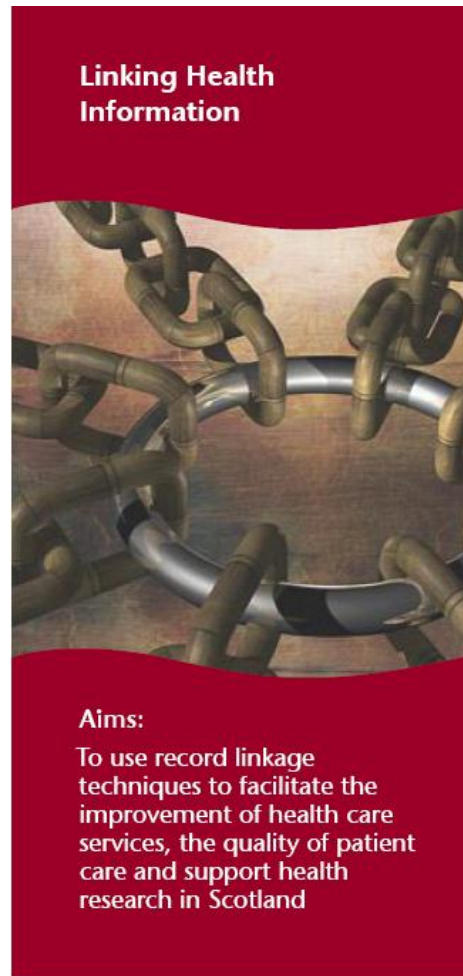


National data flows





Linking Health Information in Scotland



‘To use record linkage techniques to facilitate the improvement of healthcare services, the quality of patient care and support health research in Scotland’



What makes record linkage possible?

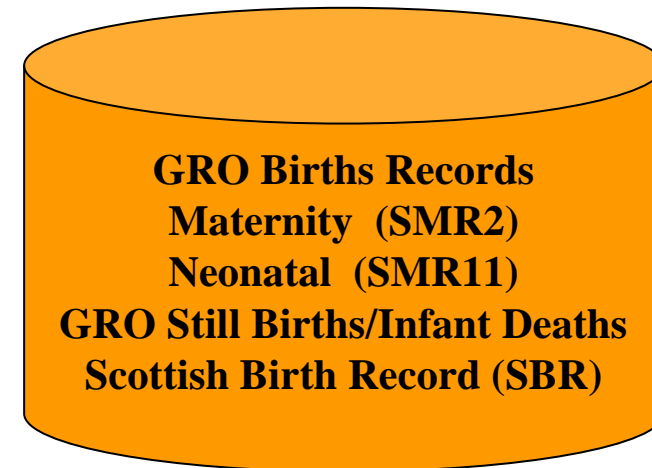
- **When was record linkage first done ?**
 - Initially developed in 1960's
 - National datasets held centrally in machine readable form
 - Routine Recording of Patient Identifiers
- **How has record linkage evolved?**
 - Computing capacity & data storage
 - Increasing demand for epidemiological and health service research
 - Creation of permanent, dynamic “linked” datasets
- **Why use “Probability Matching” ?**
 - Exact matching leads to inexact results
 - Quantifies the implications of levels of agreement and disagreement

- **Inpatients and day cases**
 - discharges from non-obstetric specialties
 - Cancer registrations
 - GRO death records
- **Database content**
 - Patient identifiable
 - Demographic & socioeconomic
 - Episode management
 - Clinical
- **Clinical support information**
 - Diagnoses (ICD-9 / ICD-10) – 6 diagnostic positions
 - Operations (OPCS-3 / OPCS-4) - 2/4 pair codes



Linked Maternity and Neonatal Dataset

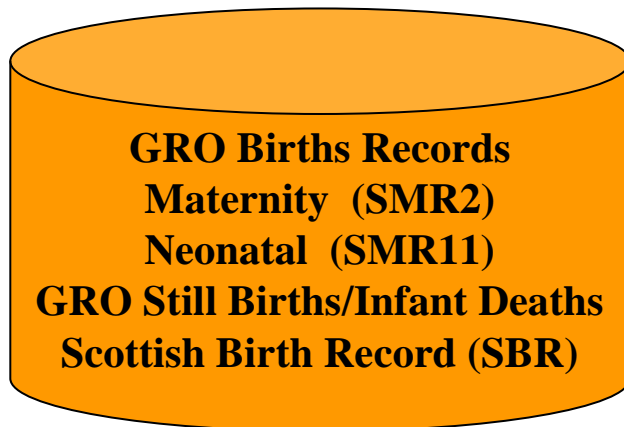
- **SMR2 Maternity Records**
 - Antenatal, delivery and postnatal discharges
- **SMR11 Neonatal Records**
 - Completed for all live births up to April 1996, sick babies thereafter
- **GRO Birth Records**
- **Stillbirth & Infant Deaths/ SSBID**
 - Cover deaths in the first year of life
- **Scottish Birth Record (SBR)**
 - Phased introduction in 2002/2003



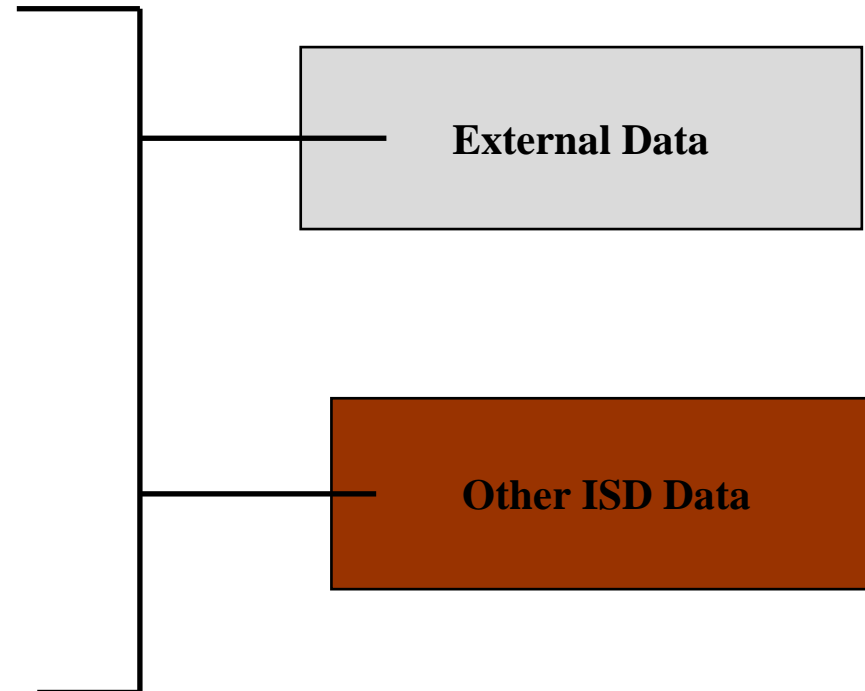
- **Linked “Acute” Database**



- **Linked “Maternity” Database**



Linkable data



- ***“The NHS Scotland is acknowledged to have amongst the highest quality data on health and healthcare in the world, over a considerable period of time”***

ISD, 2006

- Recent DQA Audit based on a 2% sample of SMR01 records
 - Principal Diagnosis recording ~ 90% accuracy
 - Main Operation recording ~ 91% accuracy
 - Postcode recording ~ 97% accuracy

Delivering a Record Linkage Request

- **Commissioned by Researchers**
 - Health Service
 - External Agencies (Universities, Private sector)
- **Assess feasibility**
 - agree detailed specification and funding
- **Data Protection & Confidentiality**
 - Privacy Advisory Committee (PAC)
 - Impose conditions (anonymisation)
- **Organisational protocols**
 - Access & storage of confidential data
 - Training

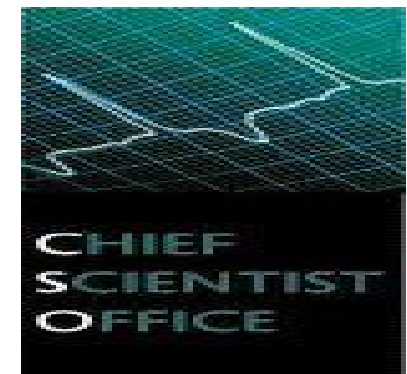




Chief Scientist Office (CSO)



- Part of the SG Health & Wellbeing Directorate
- “supports and promotes high quality research aimed at improving the services offered by NHSScotland, and the health of the people of Scotland.”
- In 2008-9 will invest over £66 million in NHS related research
- Fund around 190 projects at any one time



Examples of Linked Analyses

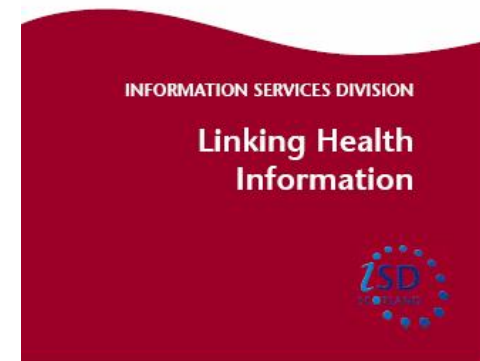
- **Patient Activity**
 - Patients, stays and episodes
 - Incidence and prevalence
 - Multiple admissions
- **Pathways of care**
 - State diagram
 - Survival analysis
- **Case control & cohort studies**
- **Clinical outcome indicators**





Examples (cont)

- **Epidemiology**
- **Quality Improvement**
- **Setting/ Monitoring government policy**
- **Specialised linkages**
 - National Audits
 - Scottish Health Survey





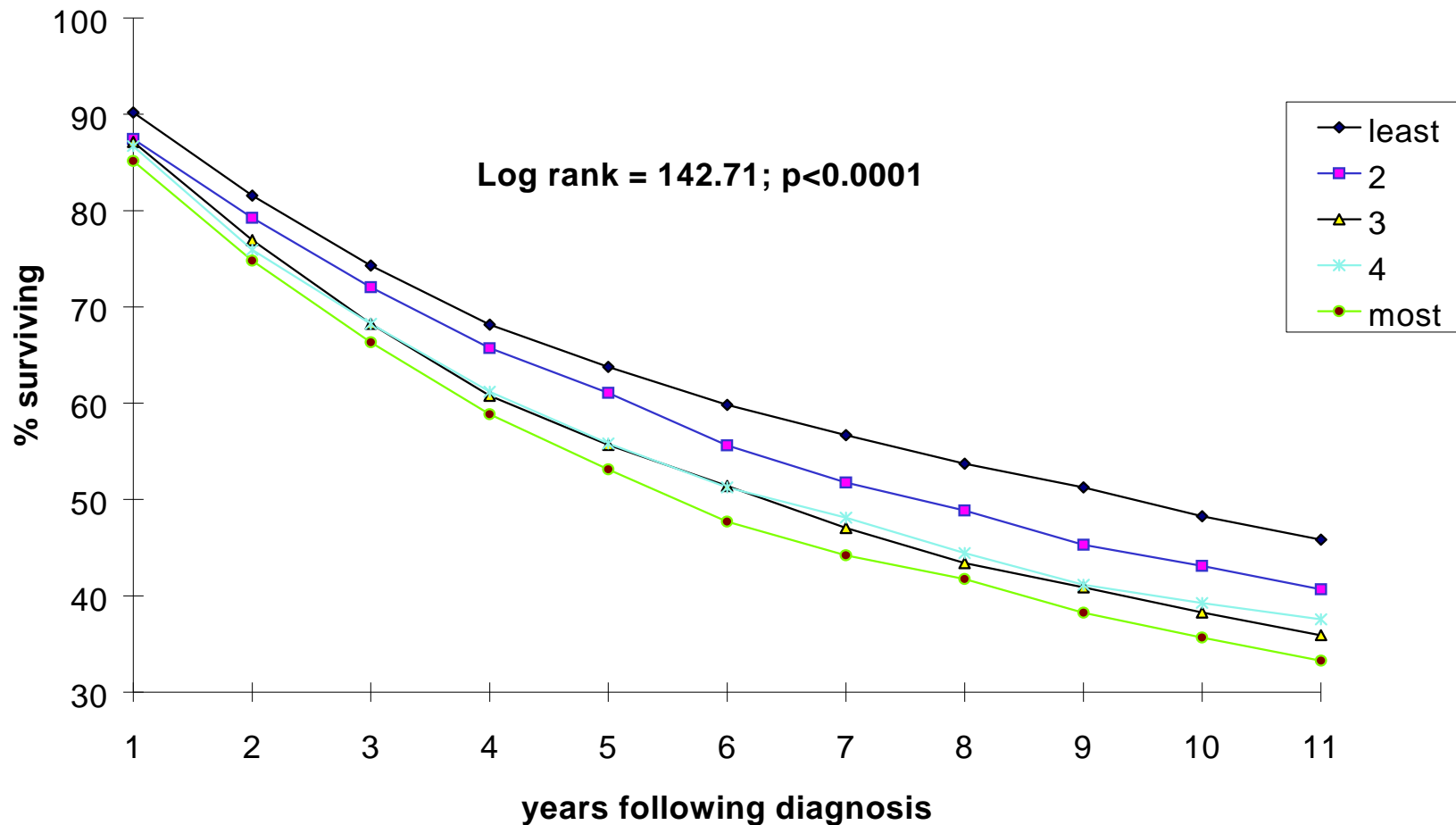
Acute hospital (non-obstetric, non-psychiatric) inpatient/day case discharges for COPD

NHS board of residence:

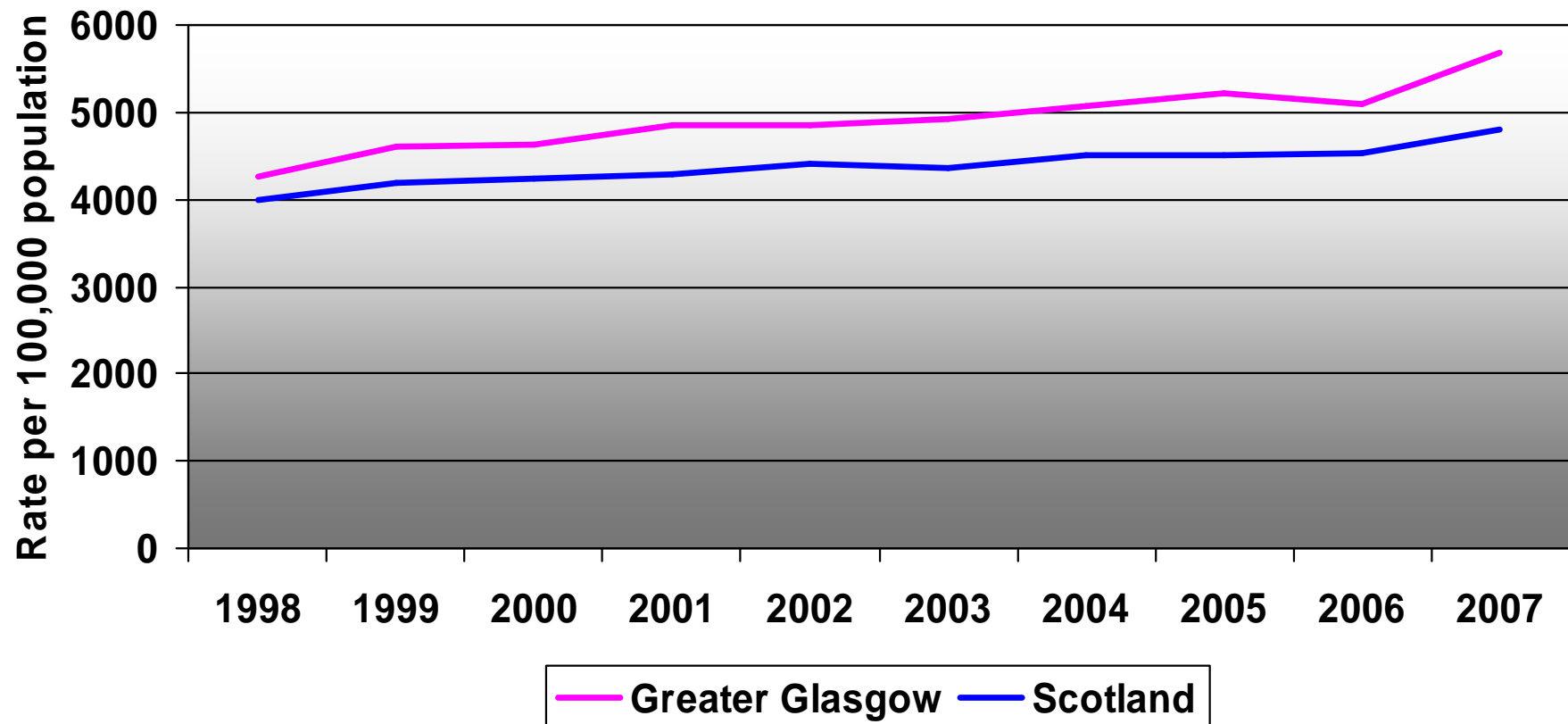
year ending 31 March 2007

NHS board of residence	Episodes	Continuous Stays	Patients
Ayrshire and Arran	2468	2094	1399
Borders	490	443	349
Dumfries & Galloway	700	554	428
Fife	1941	1365	1012
Forth Valley	1533	1256	926
Grampian	2605	1960	1466
Greater Glasgow & Clyde	8925	6515	4527
Highland	1840	1253	950
Lanarkshire	4064	2670	1915
Lothian	4859	3509	2471
Orkney Islands	84	76	61
Shetland Islands	56	54	44
Tayside	2434	1790	1267
Western Isles	158	148	113

Deprivation and survival for breast cancer (1981-1990)



Number of patients aged 65+ with 2+ emergency admissions within 1 year



Faster treatment for chest pain

Are “low-risk” patients safely reassured ?

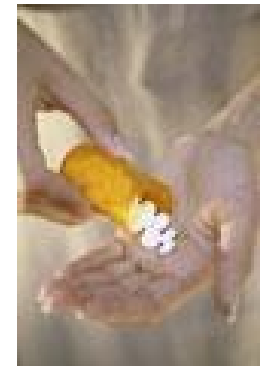
Study Population and Data:

- \approx 4,000 patients attending the RACPC between 1999 and 2000.
- Linked hospital morbidity, death, revascularisation records.
- Retrospective cohort study



Scientific Aims:

- To examine the long-term outcome of patients evaluated in a RACPC
- Are low-risk patients safely reassured?
- To allow rapid diagnosis, treatment and risk stratification for patients





Faster treatment for chest pain

Are “low-risk” patients safely reassured ?



Benefits/ Conclusions:

- RACPC was effective at triaging patients with chest pain
- Low risk patients unlikely to have adverse coronary outcomes
- Reduce waiting times and prevent unnecessary admissions





Studies Linking ISD Data Epidemiologically (SLIDE)



- **Funded by British Heart Foundation**
 - Alan Finlayson, Jim Chalmers, James Boyd, Adam Redpath. *ISD Scotland*
 - Simon Capewell. *University of Liverpool*
 - John McMurray, Kate MacIntyre, Simon Stewart, Jill Pell. *University of Glasgow*
- **SLIDE Aims**
 - Examine ischaemic heart disease (coronary artery disease) effects in Scotland
 - present picture
 - recent trends
 - Using routine linked data

“I don’t like Mondays!”

Background

- High risk of death at end of weekends in Russia
- Other studies show heart disease problems on Mondays
- Possibly related to heavy “binge” drinking
- Was it the same in Scotland?



Study Population

- Deaths outside hospital in people with no previous admissions for coronary heart disease
- 58,448 deaths over 10 year period



“I don’t like Mondays”—day of the week of coronary heart disease deaths in Scotland: study of routinely collected data

Christine Evans, James Chalmers, Simon Capewell, Adam Redpath, Alan Finlayson, James Boyd, Jill Pell, John McMurray, Kate Macintyre, Lesley Graham





The Health Impact of Safety Cameras



Purpose:

- To evaluate the health impact (in its widest sense) of the rollout of safety cameras in the 12 local authority areas in the Strathclyde Police region

External Source Data:

- STATS-19: Police records of personal injury collisions (PICs).
- Strathclyde Police Area
- 80,265 records – 1987-2005

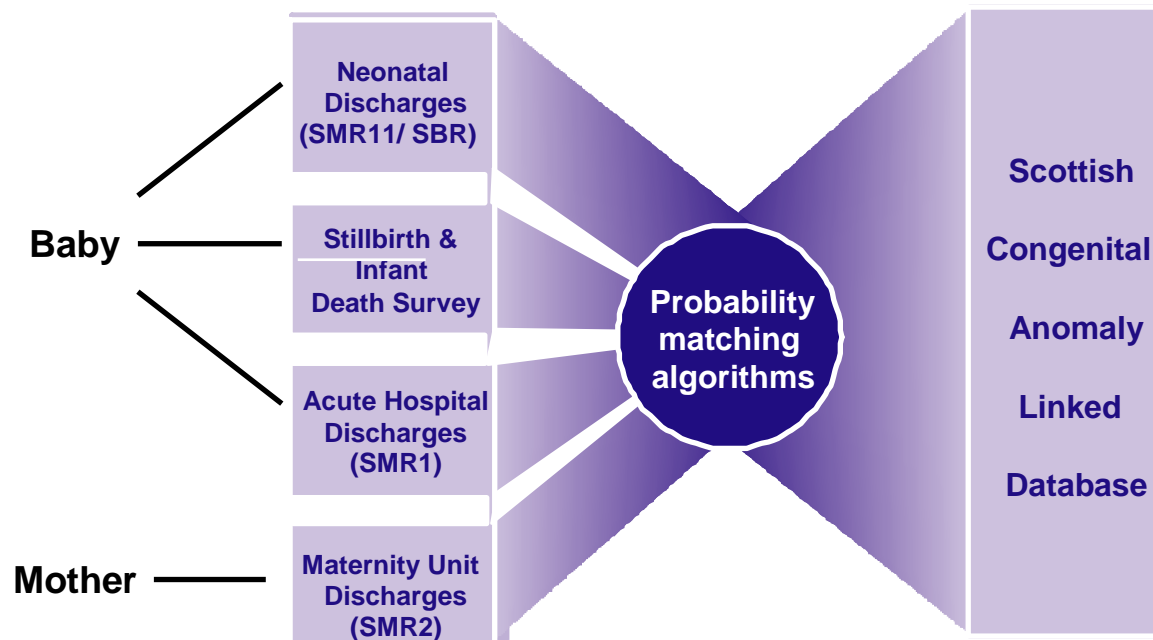
To be linked to :

- SMR01: NHS Scotland Hospital In-patient and Day-case Discharge Episodes
- GRO Scotland: Death Registrations



Congenital Anomalies in Scotland

The Scottish Congenital Anomaly Linked Database (SCALD) was created using the current system of medical record linkage in Scotland.



- Incidence rates published on a variety of anomaly categories for the first time since 1991.
- Regional and temporal variation tested.
- Catalyst for developing maternity and neonatal linked data.

**Babies born in Scotland & detected with congenital anomalies at birth or during infancy¹
rates per 1,000 births by anomaly and year: 1995-2004**

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Neural Tube Defects	0.69	0.78	0.52	0.63	0.66	0.43	0.44	0.47	0.50	0.37
Heart	5.19	5.27	5.33	4.35	4.42	5.28	5.38	4.75	5.83	6.37
Circulatory System	4.81	4.40	4.28	3.65	4.33	4.21	4.22	4.10	4.30	5.13
Trisomy 13	0.10	0.07	0.09	0.07	0.13	0.16	0.06	0.12	0.14	0.12
Trisomy 18	0.34	0.32	0.21	0.27	0.15	0.19	0.38	0.18	0.32	0.17
Down's Syndrome	1.25	1.10	1.54	1.08	1.19	1.27	1.06	0.73	0.99	1.26

Scottish Perinatal & Infant Mortality & Morbidity Report 2005

1. All infants are followed up from birth for a period of one year.

- Gastroschisis is a rare but dramatic and ultimately life threatening congenital defect where abdominal contents herniate through the abdominal wall, usually to the right of the umbilical cord, leaving the intestines and other abdominal contents exposed (Medline Plus, 2006; Torfs et al 1993).
- Produced extract of all SMR1/death records from Scottish morbidity database with date of birth 1998 - 2004.
- Analysed SCALD sweeping through all record types with date of birth/date of delivery 1998 - 2004 to find incident cases of gastroschisis (ICD-10 Q79.3).

**Univariate Logistic Regression Modelling for outcome of Gastroschisis diagnosis
Odds Ratios with 95% CI and p-values**

Variable	Unadjusted Analysis					
	births	cases	Odds Ratio	95% Conf.Int.		P-Value
				Lower	Upper	
Maternal Age Category						
19 and under	30216	36	33.261	14.014	78.943	<0.001
20-24	63756	31	13.574	5.663	32.537	<0.001
25-29	95857	17	4.951	1.952	12.558	0.001
30 and over	167501	6	*	*	*	*
Sex of baby						
Male	183173	49	*	*	*	*
Female	174111	41	0.880	0.581	1.333	0.547
Mothers Smoking Status						
No	202745	35	*	*	*	*
Yes	126885	42	1.917	1.224	3.003	0.004
Deprivation Quintile						
1 (least deprived; referent)	66549	8	*	*	*	*
2	64233	9	1.166	0.450	3.021	0.753
3	65856	15	1.895	0.803	4.469	0.144
4	70841	22	2.584	1.150	5.803	0.022
5 (most deprived)	89791	36	3.335	1.550	7.176	0.002
Gestational Age Category						
1 = premature (less than 38weeks)	40112	65	20.548	12.954	32.594	<0.001
2=not premature (38 weeks or more)	317009	25	*	*	*	*
Birth Weight Category						
1=below average bweight (less than 3000g)	75750	74	21.152	11.731	38.137	<0.001
2=average bweight and above (3000g up)	281474	13	*	*	*	*

Conclusions of Gastroschisis Study

- **No significant increase in cases of gastroschisis between 1998 and 2004.**
- **No significant geographical relationship.**
- **Much higher rates are observed both in young and in socially disadvantaged mothers.**
- **Factors significantly associated with risk of onset of gastroschisis are low maternal age, low birth weight and low gestational age.**



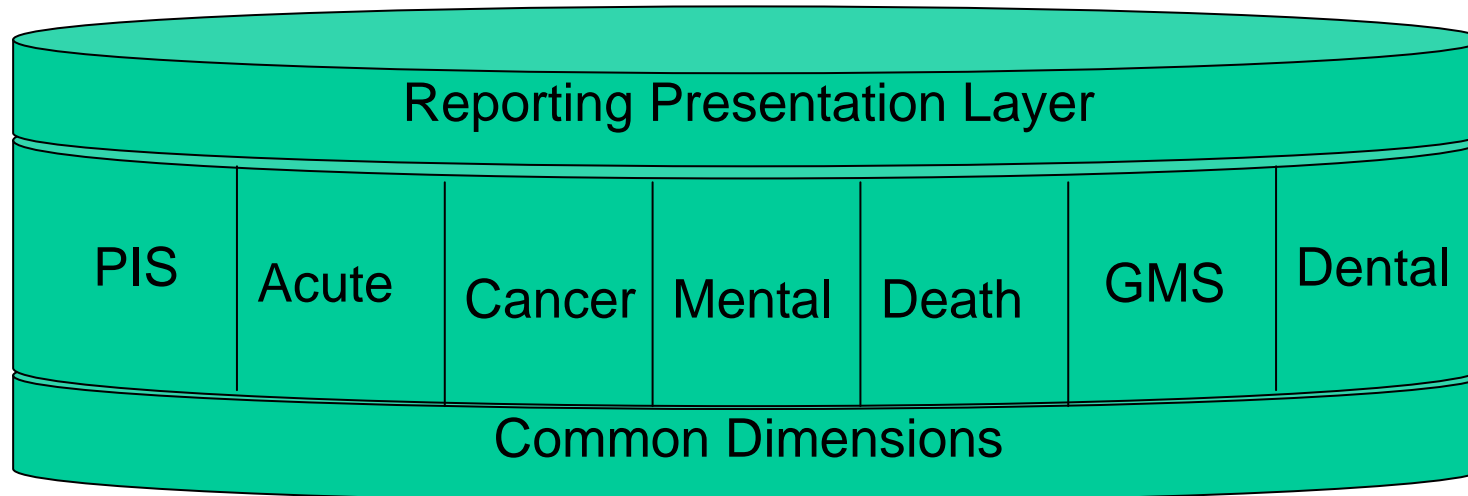
Record linkage in Scotland – the future

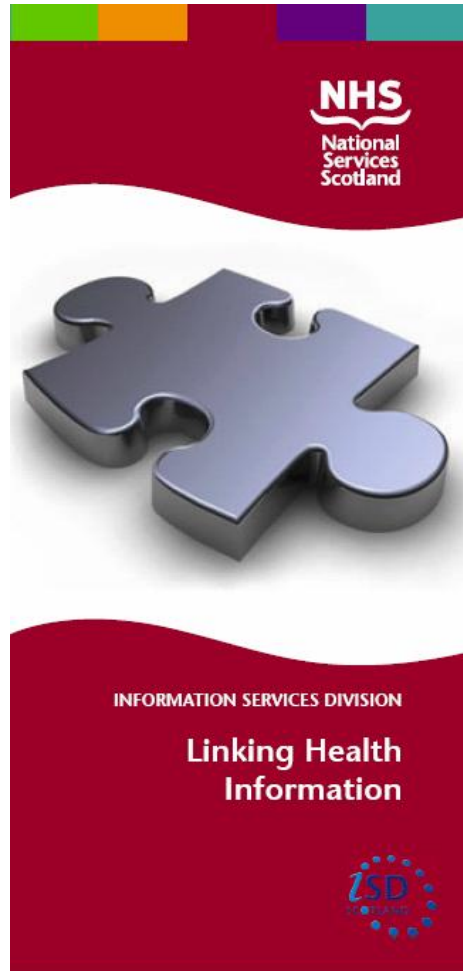


- **Unique Patient Identifier**
 - Otherwise known as CHI (Community Health Index)
- **ISD Linked Databases matched to the UPI**
 - easier linkages between external datasets
 - Data Quality improvements for both ISD and CHI
 - quicker turnaround
 - additional resources available for analyses
- **NHS National Data Warehouse**
 - ‘Patient-based’ data analyses
 - Provide ‘open’ access to NHS Scotland
 - All available national data sets ‘linked’
 - e.g. Prescribing, Cancer Screening, Dental etc
 - Easy to use front-end tools for access

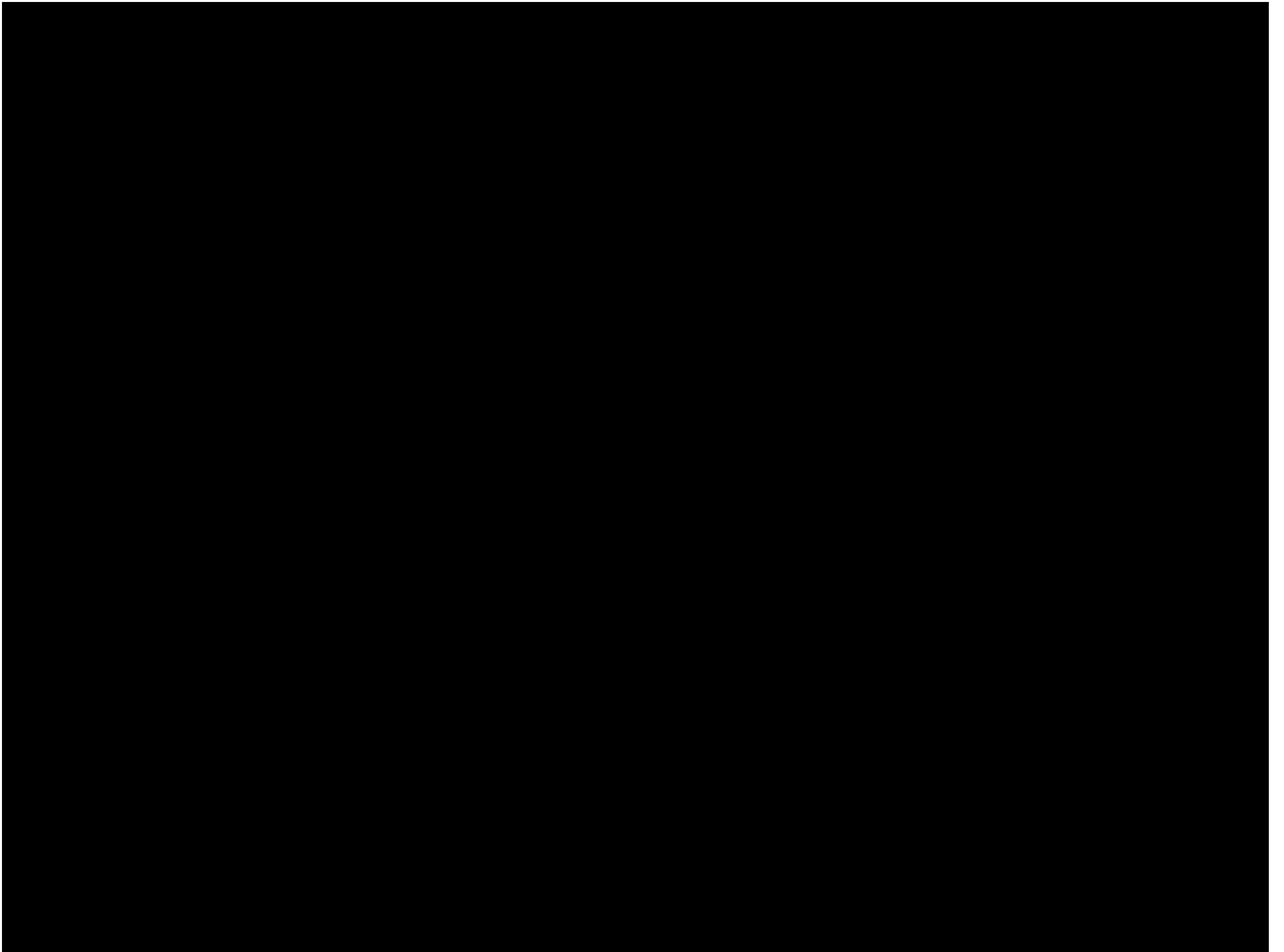
Modernising our IT (SHIS)

Business Objects Universe





- Information Services Division (ISD)
 - www.isdscotland.org
- Linking Health Information
 - www.isdscotland.org/record-linkage
 - Carole.Morris@isd.csa.scot.nhs.uk
- Data Confidentiality/ PAC
 - www.isdscotland.org/pac
- Chief Scientist Office (CSO)
 - www.show.scot.nhs.uk/cso/



Record linkage using probability matching and the challenges of limited identifying information

Catherine Storey

Linking records

- **Need to be confident records belong to the same person**
- **Ideally need:**
 - Surname and soundex code
 - First name
 - Sex
 - Date of birth
 - Post code
 - Others eg unique CHI number, case reference number

Linking records – Examples (1)

Surname	Soundex	Forename	Date of birth	Post code	CHI number
Duck	D200	Donald	1 January 1955	DL1 2BY	0101551234
Dukk	D200	Donald	1 January 1955	DL2 3SL	0101551234
✘	✔	✔	✔	✘	✔

Surname	Soundex	Forename	Date of birth	Post code	CHI number
Swan	S500	Daisy	24 December 1959	AZ1 2BY	2412591234
Duck	D200	Daisy	24 December 1959	DL2 3SL	2412591234
✘	✘	✔	✔	✘	✔

Linking Records – Examples (2)

- **Both the above would be considered good links**
- **Change of details such as postcode, surname taken into account**
- **Misspelling of surname can be compensated by use of soundex**
- **Unusual surnames given more weight**
- **Constants such as date of birth can be subject to error – allowances made**

A more challenging linkage

- **Request to link Scottish Morbidity Record (SMR) to the Scottish Drugs Misuse Database (SDMD)**
- **SDMD data highly confidential – limited person identifying information:**
 - No surname available – 1st and 4th characters only; as a result, no soundex
 - Forename reduced to initial only, no second initial
 - Postcode sector only
 - No ‘extras’ such as maiden name

Probability matching: making the best use of available data

- Full date of birth and sex present
- Part postcode still useful
- Weight combinations of surname 1st and 4th letter
- Add weight for patients with a history of drugs misuse in SMR – psychiatric records added for this purpose

Managing without a surname

- **Use similar principle to soundex weighting**
- **Assign binit weights according to frequency of character combination in Scottish population – Community Health Index used**
- **Program written to read same character combinations from SMR**
- **Disadvantages**
 - does not compensate for misspelling as well as soundex
 - Difficult to judge at checking stage

Calculating binit weights

Character combination	Frequency of combination in population	Percentage of combination in population	Odds ratio - estimated agreement % (97) divided by percentage in population	Binit weight = log to base 2 of odds ratio
A	5,201	0.096	1012.427289	9.98
AA	25,887	0.477	203.4084417	7.67
AB	862	0.016	6108.624513	12.58
AC	2,998	0.055	1756.382365	10.78
AD	3,280	0.060	1605.37632	10.65
AE	50,103	0.923	105.0961885	6.72
AF	788	0.015	6682.277069	12.71
AG	333	0.006	15812.71571	13.95
AH	8,423	0.155	625.1495109	9.29
AI	12,033	0.222	437.5994623	8.77
AJ	178	0.003	29582.21534	14.85
AK	8,199	0.151	642.2288486	9.33
AL	3,133	0.058	1680.700393	10.71

Incorporating patient history

- Intention to increase probability of good match, compensate for lack of identifiers
- Psychiatric records included in linkage to capture drugs misuse
- ICD-9 codes 292, 304, 305
- ICD-10 codes F10-F19, T40
- Bonus of 3 points given to add weight to patients with history of drugs misuse

Initial results – checking the paired records

- **Vital stage in setting a threshold for the linkage**
- **Judgement more difficult in absence of full surnames and postcodes**
- **More attention given to potential misspelling of surname**
- **Ensured that ‘drug’ bonus was not falsely enhancing score by being too high, or eliminating good matches with no history**

Checking the paired records – examples (1)

Example 1: Strong all-round score without ‘drug’ bonus

Type	Surname	Forename	Date of birth	Sex	Post code	Drug
SMR1	POTTER	J	19750101	M	G1 8BZ	
SDMD	PT	J	19750101	M	G1 8	No

Scores:

Total	Surname	Forename	Date of birth	Sex	Post code	Drug
34.90	7.71	2.28	14.76	1.00	9.15	0.00

Example 2: Lower score enhanced by ‘drug’ bonus

Type	Surname	Forename	Date of birth	Sex	Post code	Drug
SMR1	PIERCE	M	19701231	M	EH1 9RU	
SDMD	PR	M	19701231	M	EH	Yes

Scores:

Total	Surname	Forename	Date of birth	Sex	Post code	Drug
34.81	8.37	4.27	14.76	1.00	3.41	3.00

Checking the paired records – examples (2)

Example 3: Possible misspelling of surname

Type	Surname	Forename	Date of birth	Sex	Post code	Drug
SMR1	MACDONALD	G	19501231	M	G6 6PS	
SDMD	MO	G	19501231	M	G6 6	Yes

Scores:

Total	Surname	Forename	Date of birth	Sex	Post code	Drug
26.80	-5.00	3.89	14.76	1.00	9.15	3.00

Example 4: Possible incorrect character in date of birth

Type	Surname	Forename	Date of birth	Sex	Post code	Drug
SMR1	SMITH	D	19800831	M	EH165JY	
SDMD	ST	D	19800331	M	EH16	Yes

Scores:

Total	Surname	Forename	Date of birth	Sex	Post code	Drug
28.79	8.82	3.56	5.20	1.00	7.21	3.00

Setting the threshold

- **Around 200 sampled pairs checked at scores from 24 to 31**
- **Threshold set at score at which it was more likely that pairs matched than did not match – 26**
- **Linkage program re-run with this threshold**

Results of the linkage

- **Of 146,993 SDMD records, 115,287 linked to SMR – 78.4%**
- **Final data set included**
 - 292,516 hospital discharges
 - 39,748 psychiatric discharges
 - 1,450 cancer registrations
 - 3,193 deaths