

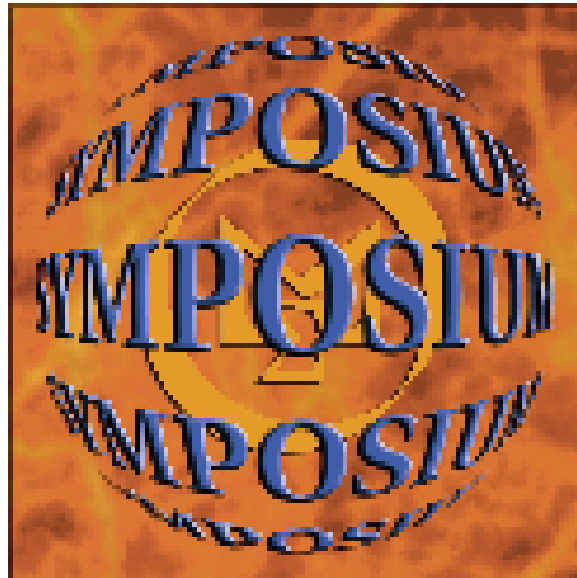
The Interplay among Censuses, Surveys and Administrative Data

John 'Mac' McDonald



Leading education
and social research
Institute of Education
University of London

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



2010 International Methodology Symposium

Statistics Canada
October 26-29, 2010
Ottawa, ON, Canada

Social Statistics: The Interplay among Censuses, Surveys and Administrative Data

Symposium 99 - Combining Data from Different Sources

The Symposium will be titled “**Social Statistics: The Interplay among Censuses, Surveys and Administrative Data**”. Members of the community, such as those from private organizations, governments, or universities, are invited to attend, particularly if they have a special interest in statistical or methodological issues resulting from the use of multiple sources of data (censuses, sample surveys or administrative data).

The first day will consist of workshops, while the following days will consist of both plenary and parallel sessions covering a variety of topics. Additional research and results will be presented via poster sessions.

We are soliciting contributed papers related to **the methodological aspects of using multiple sources of data**. Topics may include:

- Sampling Frames and Sample Design
- Coordinating Samples
- Content and Questionnaire Design
- Data Collection Methods and Acquisition of Administrative Data
- Supplementing Survey Data with Administrative Data
- Administrative Data for Direct Estimation
- Statistical Databases from Administrative Data (e.g., Population Registers)
- Imputation
- Weighting and Estimation
- Dissemination and Data Access
- Record Linkage Techniques
- Record Linkage Software
- Measurement Errors
- Response Burden
- Treatment of Nonresponse
- Confidentiality, Privacy and Ethical Issues
- Small Area Estimation

the methodological aspects of
using multiple sources of data

- Sampling Frames and Sample Design
- Coordinating Samples
- Content and Questionnaire Design
- Data Collection Methods and Acquisition of Administrative Data
- Supplementing Survey Data with Administrative Data
- Administrative Data for Direct Estimation
- Statistical Databases from Administrative Data (e.g., Population Registers)

- Imputation
- Weighting and Estimation
- Dissemination and Data Access
- Record Linkage Techniques
- Record Linkage Software
- Measurement Errors
- Response Burden
- Treatment of Nonresponse
- Confidentiality, Privacy and Ethical Issues
- Small Area Estimation

15th GSS Methodology Conference, 1 July 2010

Making data work harder

Methodology sessions on

- Introducing efficiencies
- Methodological change
- Innovation in methods
- Maintaining quality
- Use of administrative data
- Data sharing
- Survey/Analysis methods
- General statistical methods

JW ‘Mac’ McDonald – CV

EDUCATION

- 1980 Ph.D. in Biostatistics
University of Washington, Seattle
- 1972 M.A. in Demography
University of California at Berkeley
- 1970 B.A. in Mathematics & Economics
University of California at Los Angeles

EMPLOYMENT HISTORY

- 2007–present Professor of Longitudinal Social Statistics
Institute of Education, University of London
- 1984–2007 Lecturer, Senior Lecturer, Reader and
Professor of Social Statistics
University of Southampton
- 1980–1984 Research Officer, World Fertility Survey
International Statistical Institute



A diagram consisting of three blue circles with black outlines. Two large circles are positioned at the top, one on the left and one on the right. A smaller circle is centered below the space between the two larger circles. Each circle contains a white text label.

fertility

mortality

migration



census
data



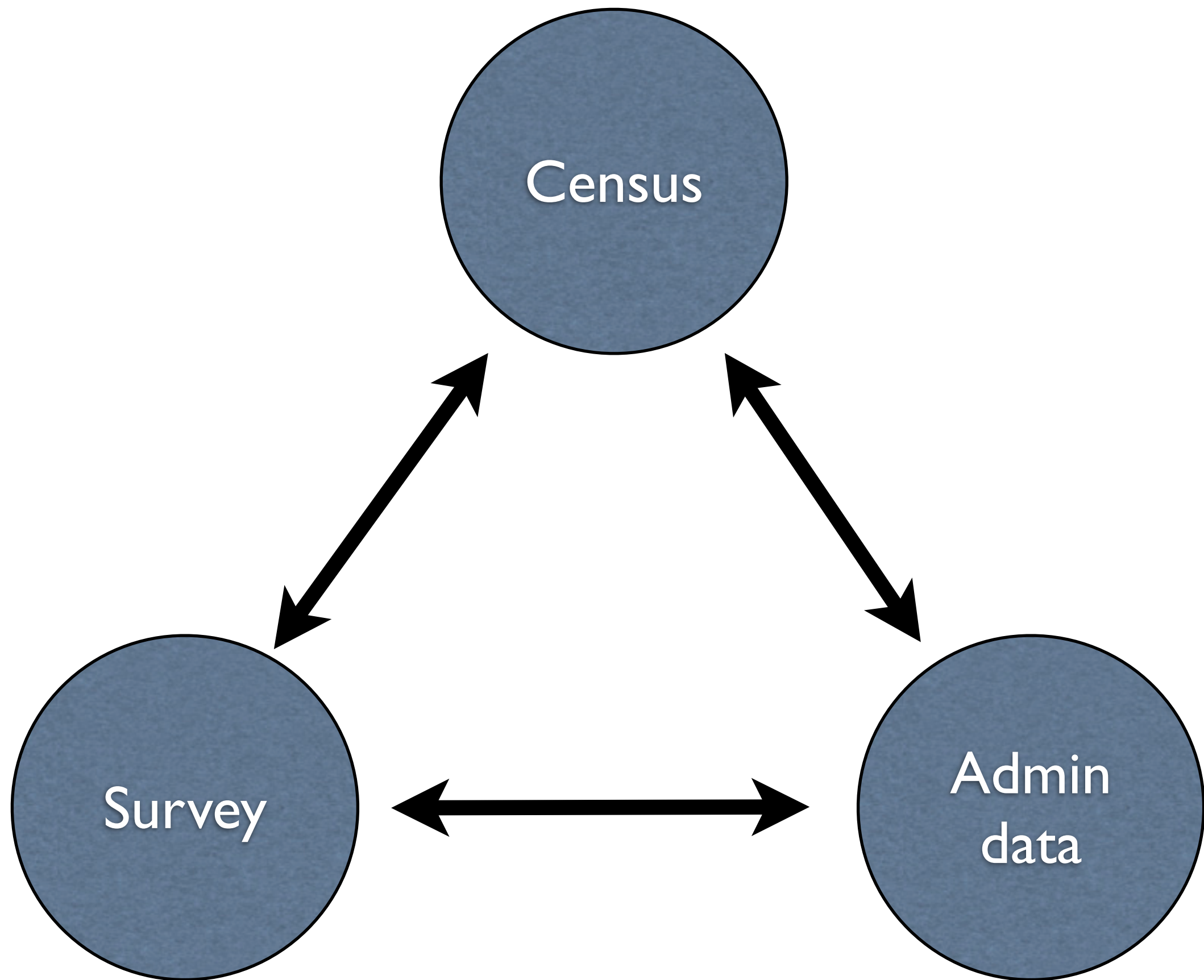
vital
registration
data



survey
data



church
records



Perspectives on Longitudinal Surveys

Randall J. Olsen

“Perhaps the greatest unexploited opportunity for survey projects lies in administrative data.”

ESRC National Centre for

Research RM Methods

A·D·M·I·N

Administrative Data -
Methods, Inference & Network



- 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 - **short courses**

- Introduction to Data Linkage
- The Value of Data Linkage for Research
- Data Linkage - Methodological and Statistical Issues
- Longitudinal Data Linkage
- Event History Analysis
- Longitudinal Data Analysis

- Introduction to the National Pupil Database
- Quantitative Analysis Using the National Pupil Database
- Using Administrative Data to Analyse the Impact of Policy Initiatives
- Enhancing Longitudinal Surveys by Linking to Administrative Data
- one new course annually

Record Linkage*

HALBERT L. DUNN, M.D., F.A.P.H.A.

*Chief, National Office of Vital Statistics, U. S. Public Health Service,
Federal Security Agency, Washington, D. C.*

EACH person in the world creates a Book of Life. This Book starts with birth and ends with death. Its pages are made up of the records of the principal events in life. Record linkage is the name given to the process of assembling the pages of this Book into a volume.

and is but a few pages in length for others. In the case of a stillbirth, the entire volume is but a single page.

The person retains the same identity throughout the Book. Except for advancing age, he is the same person. Thinking backward he can remember the important pages of his Book even though he may have forgotten some of the words. To other persons, however, his identity must be proven. "Is the John Doe who enlists today in fact the same John Doe who was born eighteen years ago?"

Events of importance worth recording in the Book of Life are frequently put on record in different places since the person moves about the world throughout his lifetime. This makes it difficult to assemble this Book into a single compact volume. Yet, sometimes it is necessary to examine all of an individual's important records simultaneously. No one would read a novel, the pages of which were not assembled. Just so, it is necessary at times to link

the various important records of a person's life.

The two most important pages in the Book of Life are the first one and the last one. Consequently, in the process of record linkage the uniting of the fact-of-death with the fact-of-birth has been given a special name, "death clearance."

IMPORTANCE OF ASSEMBLING THE BOOK OF LIFE

There are many uses for the important records of each person, brought together as a whole. At times, even now, such a collection is of sufficient value that it is made at considerable cost in time and money. Usually, it is the individual who is made to do the work since he alone knows where his records are on file. It is much more difficult for any other person or organization to assemble the records of his life since no personal cross-index exists to lead one to all of a person's records. It is important to many people and organizations to be able to assemble this type of information easily and efficiently.

IMPORTANCE TO THE INDIVIDUAL

Sooner or later most of us need to prove facts about ourselves. The most frequent facts are concerned with birth. Sometimes other facts are needed. Many of these are recorded in the vital records of marriage, divorce, adoption, legitimation, change of name, death and presumption of death. However, it is

* Modified form of paper given before the joint conference of the Vital Statistics Council for Canada and the Dominion Council of Health, held in Ottawa, Ontario, Canada, on May 10, 1946.

EACH person in the world creates a Book of Life. This Book starts with birth and ends with death. Its pages are made up of the records of the principal events in life. Record linkage is the name given to the process of assembling the pages of this Book into a volume.

on record in different places since the person moves about the world throughout his lifetime. This makes it difficult to assemble this Book into a single compact volume. Yet, sometimes it is necessary to examine all of an individual's important records simultaneously. No one would read a novel, the pages of which were not assembled. Just so, it is necessary at times to link

important to many people and organizations to be able to assemble this type of information easily and efficiently.

IMPORTANCE TO THE INDIVIDUAL

Sooner or later most of us need to prove facts about ourselves. The most frequent facts are concerned with birth. Sometimes other facts are needed. Many of these are recorded in the vital records of marriage, divorce, adoption, legitimation, change of name, death and presumption of death. However, it is

* Modified form of paper given before the joint conference of the Vital Statistics Council for Canada and the Dominion Council of Health, held in Ottawa, Ontario, Canada, on May 10, 1946.

Events of importance worth recording in the Book of Life are frequently put on record in different places since the person moves about the world throughout his lifetime. This makes it difficult to assemble this Book into a single compact volume. Yet, sometimes it is necessary to examine all of an individual's important records simultaneously. No one would read a novel, the pages of which were not assembled. Just so, it is necessary at times to link

Record Linkage*

HALBERT L. DUNN, M.D., F.A.P.H.A.

*Chief, National Office of Vital Statistics, U. S. Public Health Service,
Federal Security Agency, Washington, D. C.*

EACH person in the world creates a Book of Life. This Book starts with birth and ends with death. Its pages are made up of the records of the principal events in life. Record linkage is the name given to the process of assembling the pages of this Book into a volume.

The Book has many pages for some and is but a few pages in length for others. In the case of a stillbirth, the entire volume is but a single page.

The person retains the same identity throughout the Book. Except for advancing age, he is the same person. Thinking backward he can remember the important pages of his Book even though he may have forgotten some of the words. To other persons, however, his identity must be proven. "Is the John Doe who enlists today in fact the same John Doe who was born eighteen years ago?"

Events of importance worth recording in the Book of Life are frequently put on record in different places since the person moves about the world throughout his lifetime. This makes it difficult to assemble this Book into a single compact volume. Yet, sometimes it is necessary to examine all of an individual's important records simultaneously. No one would read a novel, the pages of which were not assembled. Just so, it is necessary at times to link

the various important records of a person's life.

The two most important pages in the Book of Life are the first one and the last one. Consequently, in the process of record linkage the uniting of the fact-of-death with the fact-of-birth has been given a special name, "death clearance."

IMPORTANCE OF ASSEMBLING THE BOOK OF LIFE

There are many uses for the important records of each person, brought together as a whole. At times, even now, such a collection is of sufficient value that it is made at considerable cost in time and money. Usually, it is the individual who is made to do the work since he alone knows where his records are on file. It is much more difficult for any other person or organization to assemble the records of his life since no personal cross-index exists to lead one to all of a person's records. It is important to many people and organizations to be able to assemble this type of information easily and efficiently.

IMPORTANCE TO THE INDIVIDUAL

Sooner or later most of us need to prove facts about ourselves. The most frequent facts are concerned with birth. Sometimes other facts are needed. Many of these are recorded in the vital records of marriage, divorce, adoption, legitimation, change of name, death and presumption of death. However, it is

* Modified form of paper given before the joint conference of the Vital Statistics Council for Canada and the Dominion Council of Health, held in Ottawa, Ontario, Canada, on May 10, 1946.

Record Linkage*

HALBERT L. DUNN, M.D., F.A.P.H.A.

the various important records of a person's life.

The Book has many pages for some and is but a few pages in length for others. In the case of a stillbirth, the entire volume is but a single page.

The person retains the same identity throughout the Book. Except for advancing age, he is the same person. Thinking backward he can remember the important pages of his Book even though he may have forgotten some of the words. To other persons, however, his identity must be proven. "Is the John Doe who enlists today in fact the same John Doe who was born eighteen years ago? "

Events of importance worth recording in the Book of Life are frequently put on record in different places since the person moves about the world throughout his lifetime. This makes it difficult to assemble this Book into a single compact volume. Yet, sometimes it is necessary to examine all of an individual's important records simultaneously. No one would read a novel, the pages of which were not assembled. Just so, it is necessary at times to link

clearance."

IMPORTANCE OF ASSEMBLING THE BOOK OF LIFE

There are many uses for the important records of each person, brought together as a whole. At times, even now, such a collection is of sufficient value that it is made at considerable cost in time and money. Usually, it is the individual who is made to do the work since he alone knows where his records are on file. It is much more difficult for any other person or organization to assemble the records of his life since no personal cross-index exists to lead one to all of a person's records. It is important to many people and organizations to be able to assemble this type of information easily and efficiently.

IMPORTANCE TO THE INDIVIDUAL

Sooner or later most of us need to prove facts about ourselves. The most frequent facts are concerned with birth. Sometimes other facts are needed. Many of these are recorded in the vital records of marriage, divorce, adoption, legitimation, change of name, death and presumption of death. However, it is

* Modified form of paper given before the joint conference of the Vital Statistics Council for Canada and the Dominion Council of Health, held in Ottawa, Ontario, Canada, on May 10, 1946.

Ruggles S (2002) **Linking historical censuses: A new approach.** History and Computing 14, 213-224

- We will exploit new record-linkage and data-mining technology to create linked representative samples of individuals and family groups from the censuses of 1860, 1870, 1900 and 1910 to the 1880 census.
- linked samples will provide new opportunities for researchers to carry out individual-level analyses of social and geographic mobility and family transitions.

UK Longitudinal Studies

- ONS (England & Wales) Longitudinal Study
4 birthdays, 1% sample, c. 500,000
- Scottish Longitudinal Study
16 + 4 birthdays, 5% sample, c. 274,000
- Northern Ireland Longitudinal Study
104 birthdays, 28% sample, c. 500,000
- NB absolute numbers & relative numbers

File A

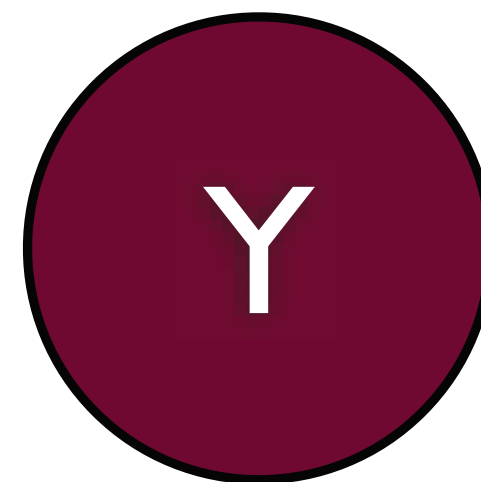
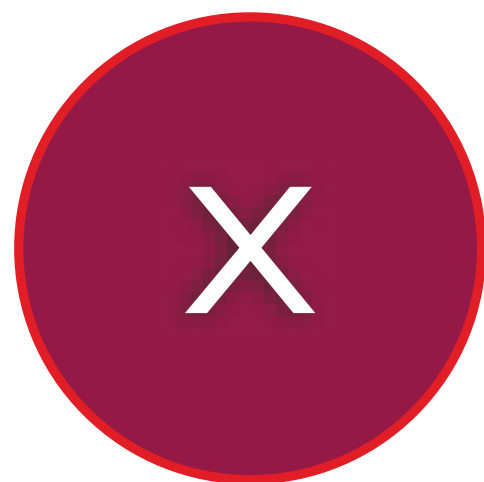
matching
variables

v_1 \dots v_K X

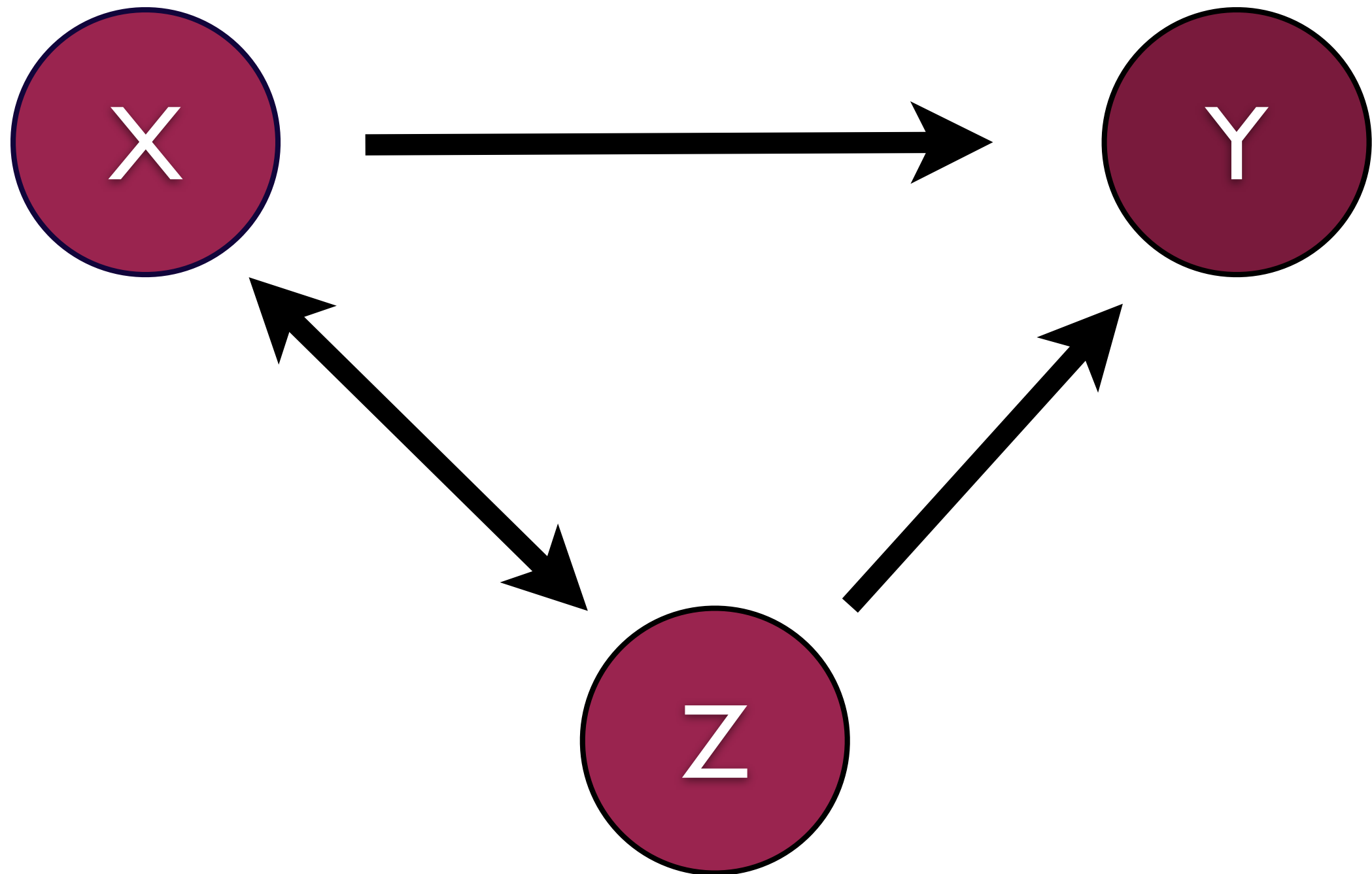
File B

matching
variables

Y w_1 \dots w_K



Confounder Z



What variables to match on?

- Jenkins S et al (2006) The feasibility of linking household survey and administrative record data: New evidence for Britain. International Journal of Social Research Methodology 11, 29-43

What variables to match on?

- IDs are subject to problems of survey item non-response and measurement error
- 5 linkages: respondent-supplied NINo and 4 linkages using different combinations of sex, name, address and DOB
- as many linkages were made using non-NINo matches as were made using NINo matches
- former also relatively accurate in terms of false-positive and false-negative linkage rates

Analysis of linked data

- Scheuren F & Winkler WE (1993)
Regression analysis of data files
that are computer matched - part 1.
Survey Methodology 19, 39-58

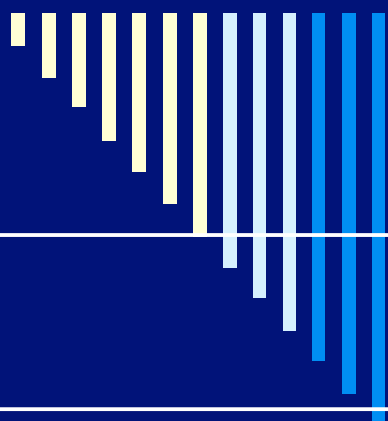
Analysis of Linked Data

- What should the linker do to help the analyst?
- What should the analyst know about the linkage and how should that information be used?
- In our opinion it is important to conceptualize the linkage and analysis steps as part of a single statistical system and to devise appropriate strategies accordingly. Obviously the quality of the linkage effort may directly impact on any analysis done.

Quality of linkage

- O'Reilly D, Rosato M & Connolly S (2008)
Unlinked vital events in census-based longitudinal studies can bias subsequent analysis.
Journal of Clinical Epidemiology, 61, 380-385
- 6.0% of deaths of residents of Northern Ireland could not be linked to a census record
- for those < 65 years at census, this linkage would exclude from analysis 20.2% of suicides and 19.7% of deaths by external causes





	E&W LS	SLS	NILS
Births to Sample Fathers			✓
Stillbirths/ Infant Deaths	✓		✓
Internal Migration			✓
Cancer Registrations	✓	✓	
Hospital Episodes		✓	
Education		✓	
Marriages		✓	✓
Claimant Count	✓		

Migration flows

- increasingly important
- difficult to measure
- proxy admin data sources often used
- small area estimates of local pop change
- differentials by age, sex, LTLI, ...
- new and improved methods needed, e.g.
Bayesian methods for estimates of precision,
missing data, inconsistent data, calibration

References

- Brierley MJ, Forster JJ, McDonald JW & Smith P WF (2008) Chapter 7: [Bayesian estimation of migration flows](#). In Raymer J & Willekens F, editors. International Migration in Europe: Data, Models and Estimates. Wiley: Chichester, pp. 149-174.
- McDonald JW, Forster JJ, & Smith P WF (2007), Invited Paper '[Bayesian Estimation of Migration Flows](#)', 56th Session of the ISI, International Statistical Institute, 22-29 August 2007, Lisboa, Portugal.

British residents born in New Commonwealth

Residence in 1966	Residence in 1971				Total
	CC	L&Y	WM	GL	
CC	118	12	7	23	160
L&Y	14	2 127	86	130	2 357
WM	8	69	2 548	107	2 732
GL	12	110	88	7 712	7 922
Total	152	2 318	2 729	7 972	13 171

CC: Central Clydeside

L&Y: Lancashire & Yorkshire

WM: West Midlands

GL: Greater London

Migration from 1966 to 1971 (flows)

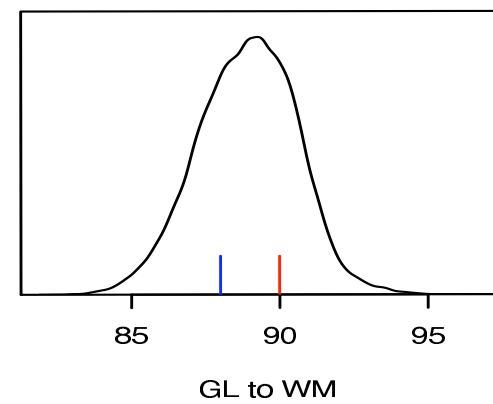
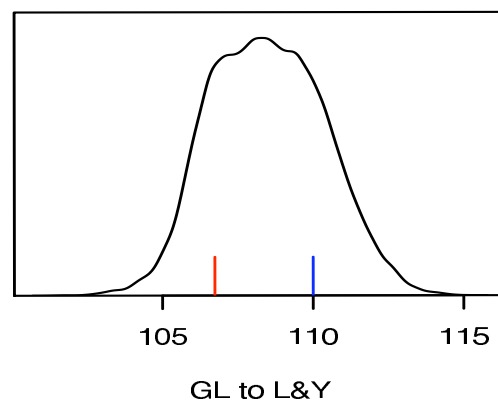
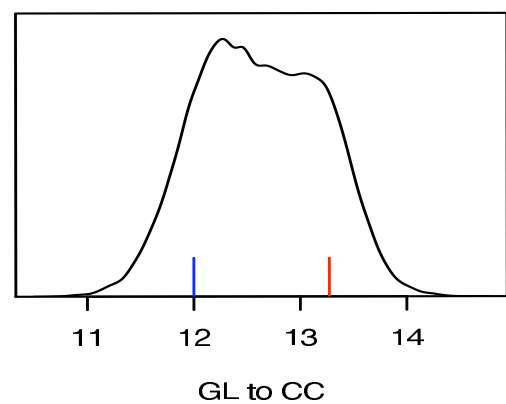
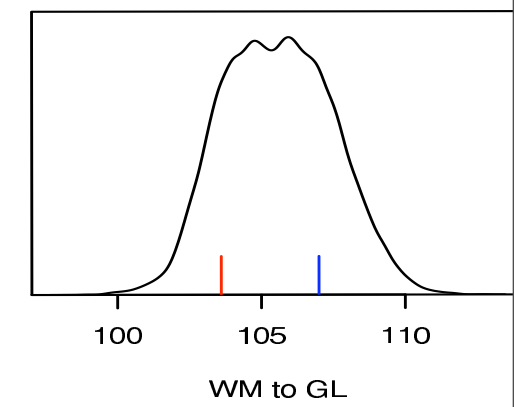
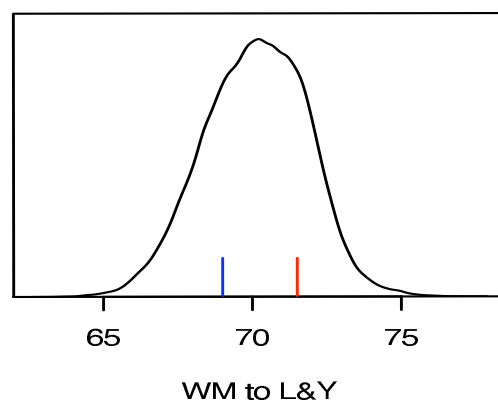
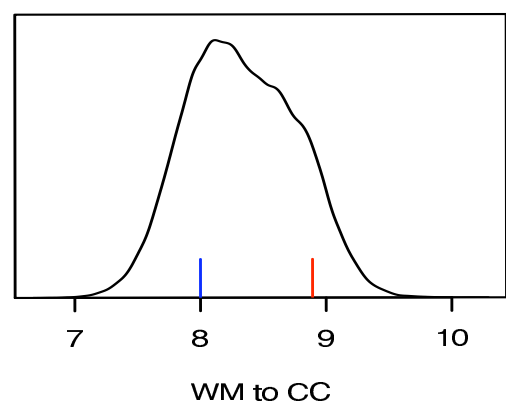
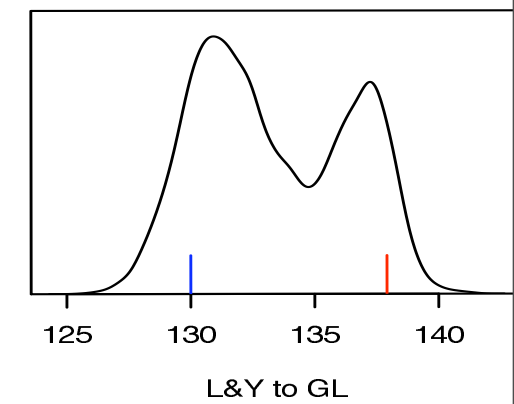
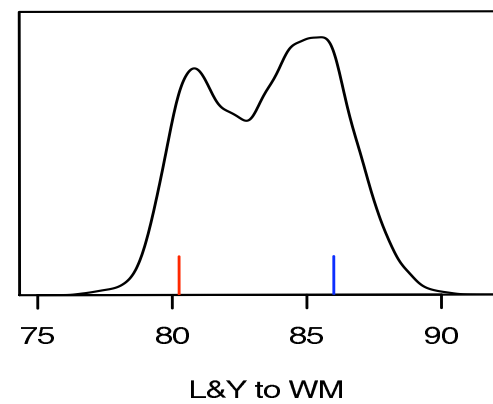
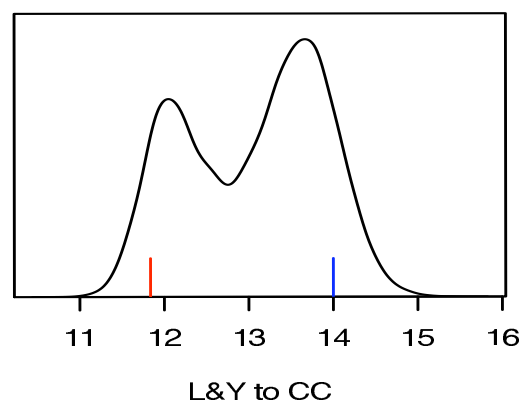
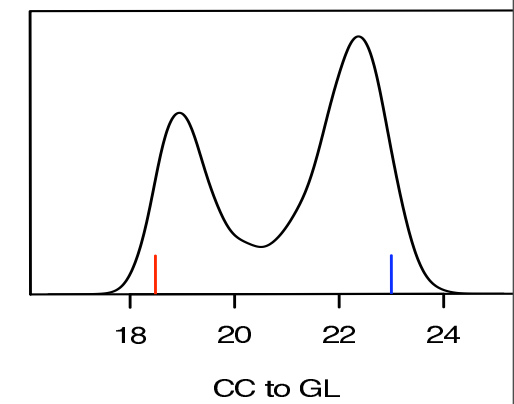
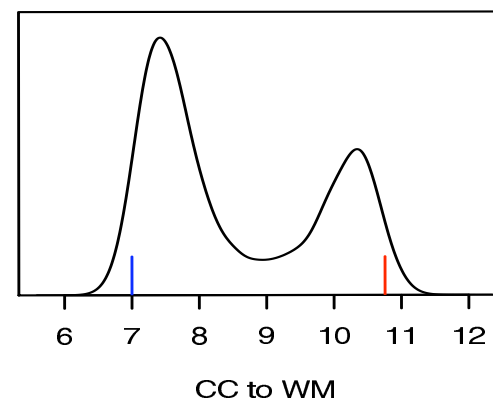
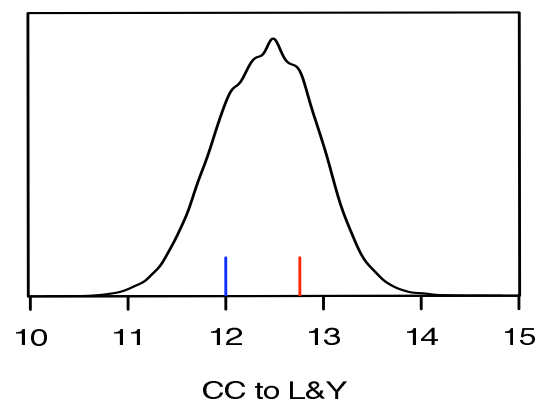
Residence in 1966	Residence in 1971				Total
	CC	L&Y	WM	GL	
CC		12	7	23	42
L&Y	14		86	130	230
WM	8	69		107	184
GL	12	110	88		210
Total	34	191	181	260	666

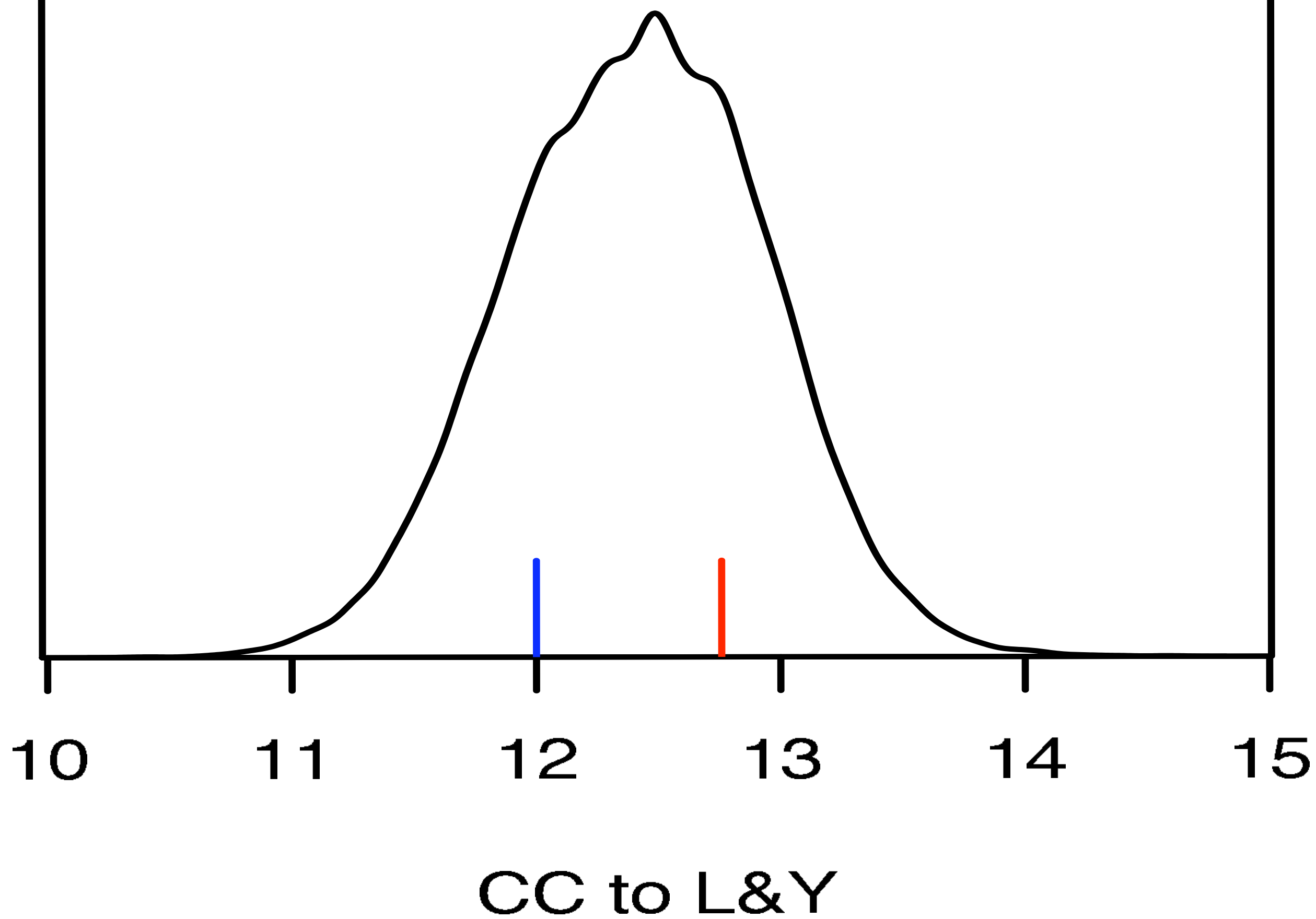
CC: Central Clydeside L&Y: Lancashire & Yorkshire
WM: West Midlands GL: Greater London

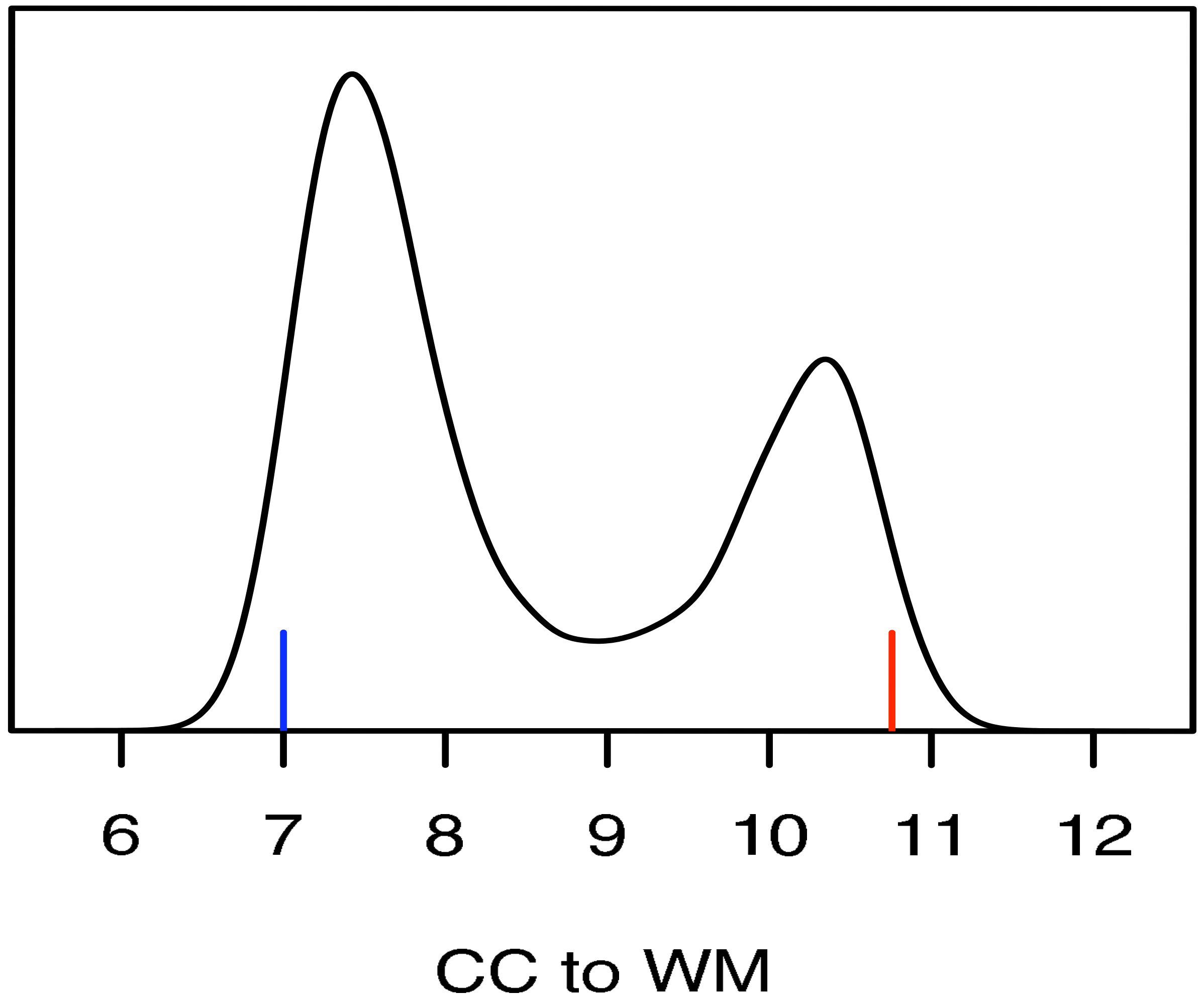
Only marginal totals of flows known

Residence in 1966		Residence in 1971				Total
		CC	L&Y	WM	GL	
CC						42
L&Y						230
WM						184
GL						210
Total		34	191	181	260	666

CC: Central Clydeside L&Y: Lancashire & Yorkshire
WM: West Midlands GL: Greater London







Take home points

- use NLS to study the 'book of life' (lifecourse)
- linkages with other data sources extremely valuable, e.g. to study/estimate migration flows
- 28% sample permits small area estimation at a low level of granularity
- conceptualize the linkage and analysis steps as part of a single statistical system
- linkage methodology should be transparent and quality of linkage studied