

# The ONS Longitudinal Study

Quality issues from 30 years of data linkage

Jillian Smith, Louisa Blackwell and Kevin Lynch

# The presentation will cover:

- Characteristics of the Longitudinal Study data
- Emerging quality models
- How the LS meets quality requirements

## What is the ONS Longitudinal Study?

- Record linkage study of England and Wales
- Data from censuses and vital registration systems
- Initial sample from 1971 Census - all people with 4 birth dates in any calendar year
- LS members = 500,000 + people (1% of pop.)
- Also includes other household members
- For each census, sample is drawn on same basis
- Confidentiality is of paramount importance in preparation and use

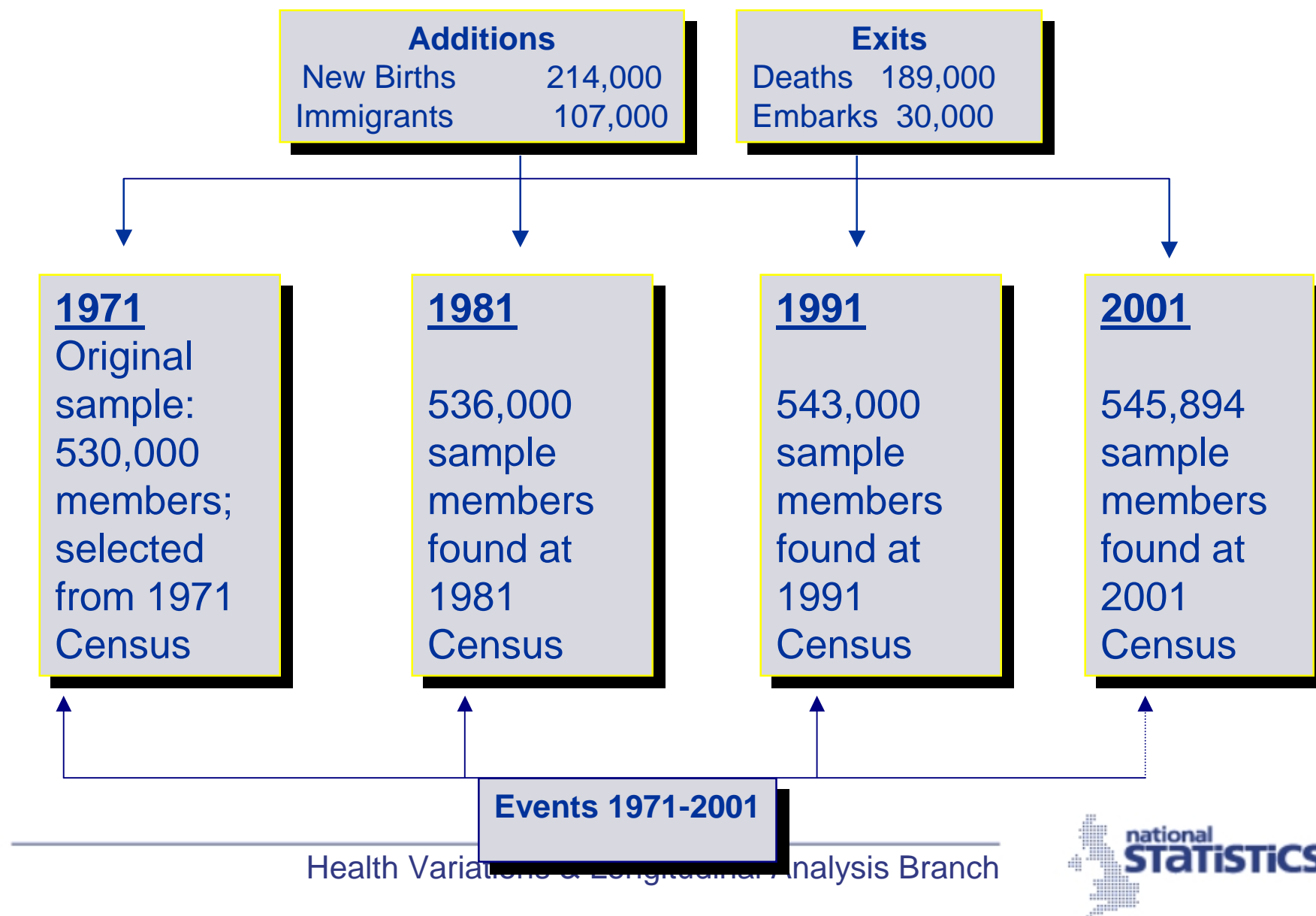
# ONS Longitudinal Study

- **Census Data:**
  - Census 1971 - sample: 530,000
  - Census 1981 - sample: 536,000
  - Census 1991 - sample: 544,000
  - Census 2001: - sample 546,000
- **Entry Events:**
  - Births on LS dates
  - Immigrants with LS birthdays

# ONS Longitudinal Study

- **Other Linked Events:**
  - Births to LS Members
  - Infant Deaths to LS Members
  - Embarkations
  - Cancer Registrations
  - Widow(er) hoods
  - Entry into Armed Forces
  - Re-entrants to Sample
  - Deaths

# Structure of the ONS Longitudinal Study (LS)



## Events 1971-2001

- Births to sample women 201,000
- Births to sample men 49,500
- Infant Deaths 2,000
- Widow(er)hoods 66,000
- Cancer registrations 70,000

# The Longitudinal Study data

- The data contains the entire, unadjusted census and event information
- Information for the LS member and all other members of the household
- No records are deleted: exits are entered into the data as events



# Linkage

- The LS data is anonymised for use
- Linked using an intermediary register (the National Health Service Central Register )
- LS flags are set on the NHSCR for future linkage
- Currently linking the 2001 UK Census into the LS

# Quality Model: 7 dimensions

- Relevance
- Accuracy
- Timeliness
- Access and clarity
- Comparability
- Coherence
- Completeness

# Relevance

- LS Review 1999
- Publications
- Wide range of users
- User consultation:
  - during Review
  - in relation to issues such as double coding

# Accuracy

- Intensive linkage exercise
- Attention to detail in coding and processing
- Attention to quality evaluation of the data
- Preparation of
  - tracing rates
  - overall linkage rates
  - sampling fractions
- Implications of edit and imputation

# Timeliness

- Data preparation takes time
- This is minimised by project management
- The trade-off between quality and timeliness is actively managed
  - for example, tracing activities for each record are limited
  - the advantage of additional quality investigations is evaluated against the impact on the delivery timetable

# Accessibility and Clarity

- ONS aims to ensure maximum use of the LS data
- The data is therefore provided free at the point of use
- Documentation has a high priority, assisting understanding of:
  - data and accompanying classifications
  - collection and processing which influences the data
- User involvement at all stages
- Improving computer systems

# Types of Comparability

- Comparability
  - in data collection methods
  - in data outputs over time
  - with other datasets
  - internationally

# Comparability

- Comparability over time is achieved by
  - consistent derived variables
  - double classifications (eg occupations and socio-economic position)
- International comparability is often possible with the detail of the LS data



# Coherence

- The LS is a secondary dataset, relying upon other primary sources for its content
- Subtle differences in definitions across sources occur
- Additionally definitions change across time
- Good documentation is required
  - adhering to modern standards where possible
  - adapting to maintain consistence with older data where necessary

# Completeness

- New data linkage
- Use in conjunction with other datasets
  - with General Household Survey
  - with national mortality data

# Conclusions

- High data quality is difficult to achieve, requiring dedication
- Data quality involves complex trade-offs of quality, cost and timeliness