Welcome, introduction and prospects

Ian Shuttleworth
Director of NILS-RSU

www.nils-rsu.census.ac.uk

Twitter: #NILS2011
Structure of the day

- The NILS and the NILS-RSU
- The 2011 Census Link in outline and projects using the 2011 data
  - Four brief presentations, marriage, grief, health and housing, birth month and later outcomes
- The wider context for the 2011 Link and the future
  - Five brief presentations on the UK LS context, synthetic data, the NI-ADRC, historical data and the NILS, and the future of the NILS
The NILS

• Longitudinal database of people and their major life events based on existing data sources
• Health card data linked to Census and administrative data
• Sample members selected based on birth date
• 104 birth dates
  – c.28% sample
  – c.500,000 people
• Two similar record linkage studies are available (ONS-LS, SLS) and the NILS works with these and CALLS
• NIMS 2011: 100% sample
  – Death data linked to Census data
Funded by:

Public Health Agency

Northern Ireland Statistics & Research Agency

ESRC Economic & Social Research Council
The NILS-RSU supports researchers by: supervising the safe environment; helping researchers make applications; promoting NILS-based research through seeking new users and policy engagement; building capacity.
Using the 2011 Link

• The NILS started with the 2001 Census and subsequent vital events and address changes
• The 2011 Census link means that it is now possible to look at the transitions of people and places between 2001 and 2011
• Themes that can be considered include
  – Labour market transitions (economic status, occupation)
  – Housing transitions (tenure change and health)
  – Demographic transitions (health status, religion, marital status)
Using the 2011 Link

- Other themes that can be researched include:
  - Fertility
  - Mortality (NIMS)
  - Migration – internal and moves into and out of NI
- The utility of the NILS will increase further by 2015 as by then the 1981 and 1991 Censuses will be linked to the NILS – 4 Censuses and 30 years of data
- These will mean that long-term social and demographic change can be described and analysed
Using the 2011 Link

• Thus a major research resource which has taken time and money to develop
• Challenge is therefore to capitalise on this to influence academic debates, policy debates and the public inside and outside Northern Ireland
• Value will increase further if, as seems likely, there is a 2021 Census
• Now...more on the 2011 Link and research that has already been done using it...
The 2011 Link

Catherine McLoughlin
NILS-RSU
Overview

• Development of the Census 2011 Link
• UK Longitudinal Studies
• NILS Datasets
• NILS Samples for Analysis
• NILS 2011 Linkage
• Supporting Documentation
• New Questions in the 2011 Census
Development of the Census 2011 Link (1)

- Business Case Started 2010
- Approved by NILS Steering Group 2011
- Funded by Census Office
- Development Started Summer 2011
- Assistance to Census Office
Development of the Census 2011 Link (2)

- Alpha-Testing Phase Summer 2013
- Beta-Testing Phase September 2013
- Soft Launch 9\textsuperscript{th} December 2013
- First Projects Approved January 2014
- UK Launch Event 6\textsuperscript{th} March 2014
<table>
<thead>
<tr>
<th>Study sample</th>
<th>ONS Longitudinal Study</th>
<th>Scottish Longitudinal Study</th>
<th>Northern Ireland Longitudinal Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1% of the population of England &amp; Wales</td>
<td>5% of the population of Scotland</td>
<td>28% of the population of Northern Ireland</td>
</tr>
</tbody>
</table>

| Records from 2001 Census | 540,000 | 260,000 | 452,600 |
| Records from 2011 Census | >950,000 | N/A | 477,800 |


| Academic user support | Centre for Longitudinal Study Information & User Support, UCL | Scottish Longitudinal Study Development & Support Unit, University of St. Andrews | Northern Ireland Longitudinal Study Research Support Unit, QUB |

| Census data available | Complete census data for study members and for people living in the same household as a study member | Complete census data for study members and for people living in the same household as a study member | Complete census data for study members and for people living in the same household as a study member |

| Event data available | Civil registration system • Births of sample members • Births to sample mothers • Stillbirths / Infant deaths • Deaths of sample members • Widow(er)hoods NHS Central Register • Immigration • Emigration • Minor events Cancer registries • Cancer data | Civil registration system • Births of sample members • Births to sample mothers • Births to sample fathers • Stillbirths / Infant deaths • Marriages • Deaths of sample members • Widow(er)hoods NHS Central Register • Immigration • Emigration Scottish Govt. Education Directorate • School level education data including attendance, exclusions, attainment and qualifications Other data available, subject to approval • Hospital episodes • Maternity data • Cancer data | Civil registration system • Births of sample members • Births to sample mothers • Births to sample fathers • Deaths of sample members Health card registration system • Immigration • Emigration • Internal migration Land & Property Services • Housing data Health & Social Care • Health data linked in one-off distinct linkage projects (e.g. prescription data, dental treatments) subject to approval |
NILS Samples for Analysis

<table>
<thead>
<tr>
<th></th>
<th>1991</th>
<th>2001</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>NILS Members</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Census</td>
<td>285,600</td>
<td>452,600</td>
<td>477,800</td>
</tr>
<tr>
<td>2 Censuses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Censuses</td>
<td></td>
<td></td>
<td>225,100</td>
</tr>
<tr>
<td>(so far)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vital Events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>105,500 Births of Babies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>107,000 Births to Mums</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>94,800 Births to Dads</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47,100 Deaths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migration Events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>566,400 Address Changes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NILS 2011 Census Launch - 5th June 2014
NILS 2011 Linkage

• Link Health Registration data to Census data
  - List Inflation Rate
  - Person Imputation
• High match rate 96.7%
• Consistent with 2001 rate
• Forward and Backward rates consistent with ONS LS and SLS
Supporting Documentation

• NILS Data Dictionary
• Metadata Document
• Searchable Data Dictionary
• Census 2011 Variables Overview
• Working Papers
• Imputation Rates Paper
New Questions in the 2011 Census
Voluntary Work

28 In the past year, have you helped with or carried out any voluntary work without pay?

☐ Yes  ☐ No
Nature of Health Conditions

Do you have any of the following conditions which have lasted, or are expected to last, at least 12 months?

- Tick all that apply.
- Deafness or partial hearing loss
- Blindness or partial sight loss
- Communication difficulty (a difficulty with speaking or making yourself understood)
- A mobility or dexterity difficulty (a condition that substantially limits one or more basic physical activities such as walking, climbing stairs, lifting or carrying)
- A learning difficulty, an intellectual difficulty, or a social or behavioural difficulty
- An emotional, psychological or mental health condition (such as depression or schizophrenia)
- Long-term pain or discomfort
- Shortness of breath or difficulty breathing (such as asthma)
- Frequent periods of confusion or memory loss
- A chronic illness (such as cancer, HIV, diabetes, heart disease or epilepsy)
- Other condition
- No condition
Passports Held and National Identity

14. What passports do you hold?
- Tick all that apply.
- [ ] United Kingdom
- [ ] Ireland
- [ ] Other, write in
- [ ] None

15. How would you describe your national identity?
- Tick all that apply.
- [ ] British
- [ ] Irish
- [ ] Northern Irish
- [ ] English
- [ ] Scottish
- [ ] Welsh
- [ ] Other, write in
New Questions in the 2011 Census

- Housing Adaptations
- Intention to stay in the UK
- Lived outside Northern Ireland
- Last country lived in
- Date of most recent arrival in Northern Ireland
- Main Language
- English Proficiency
- Knowledge of Ulster-Scots
- Type of Central Heating
Contact Details

@NILSRSU

www.nils-rsu.census.ac.uk

nils-rsu@qub.ac.uk
Does exogamy (mixed marriage by religion) increase the risk of marital dissolution in Northern Ireland?

David M Wright, Michael Rosato, Dermot O’Reilly
Marital dissolution

• Profound impacts on societal and individual wellbeing

• Family history and values, age at marriage, children

• Socioeconomic factors: employment status, education

• Partner dissimilarity (e.g. by age, ethnicity, religion)
Marital dissolution in Northern Ireland

• Is there variation in dissolution risk among religious groups?

• Are mixed marriages at increased risk, especially those crossing the Catholic-Protestant boundary?

• Does residential segregation affect dissolution risk for mixed marriages?
Residential segregation by religion

Distribution of Catholic population in Northern Ireland, by Electoral Wards (as defined by Community Background (Table KS07b), Census 2001)

NILS 2011 Census Launch - 5th June 2014
Northern Ireland Longitudinal Study – 2001- 2011 Census link

• Married couples at the 2001 Census
  – Aged 16 – 74
  – At least one NILS member
  – Both survived to 2011 Census

• 115,465 couples, 23,803 dissolutions

• Marital dissolution: no longer co-resident in 2011

• Logistic regression
Dissolution risk higher among younger and dissimilar couples

- Age: 50 – 75, 35 - 50, 16 - 35
- Economic activity: Sick, Unemployed
- Religion
- First/subsequent marriage
- Education: No degree
- Housing tenure: Rented
- Country of birth
- Urban/rural
- Dependent children

Relative risk of marital dissolution (ORs)
Variation among religious groups in risk of marital dissolution

Mixed

Non-mixed

Relative risk of marital dissolution (ORs and 95% CIs)
Mixed marriages are rare and mixed couples are less likely to marry
Increase in proportion of Catholic-Protestant marriages

Proportion of marriages (%)

Census

2001

2011

Protestant-Protestant
Catholic-Catholic
Other-Other
No religion-No religion
Catholic-Protestant
Catholic-Other
No religion-Protestant
Catholic-No religion
No religion-No religion
Other-Other
Catholic-Catholic
Protestant-Protestant
Risk of marital dissolution unrelated to residential segregation by religion
Summary

• Increased risk of marital dissolution among less religious groups.

• Mixed marriages had slightly higher risks of dissolution than non-mixed marriages.

• Catholic-Protestant unions:
  – at no greater risk than other types of mixed marriage
  – rare
  – couples were more likely to cohabit
  – younger cohort
  – dissolution risk unrelated to residential segregation

• Significant barriers to cross-community relationships remain but indications of increased mixing among the young.
Acknowledgements

The help provided by the staff of the Northern Ireland Longitudinal Study (NILS) and the NILS Research Support Unit is acknowledged. The NILS is funded by the Health and Social Care Research and Development Division of the Public Health Agency (HSC R&D Division) and NISRA. The NILS-RSU is funded by the ESRC and the Northern Ireland Government. The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of NISRA/NILS.
The effect of social & economic transitions on children & young people

Dr Aideen Maguire, Dr Mark McCann & Dr John Moriarty

QUB
Children in Transition

• Importance of the early years

• Children and young people are in constant change

• All aspects of the environment are potential determinants of how children develop

• Family environment and affluence are two key candidate factors that determine development

• Gaining a better understanding of the effect that the social environment can have on children and young people is important for identifying ways to mitigate harmful influences
Outcomes of interest

• 2011 Census questions on **educational attainment**
  - education a predictor of adult success
  - higher education = higher SES
  - increase social networks and opportunity

• 2011 Census questions on **mental health**
  - mental ill health leading cause of disability
  - largest contributing factor to days of work lost
  - decreases social networks and lack of opportunity

• Using two waves of NILS data from the 2001 and 2011 censuses, this project will assess how changes in affluence or family structure can influence educational and mental health outcomes for young people
Beta Test Questions

• Do changes in household affluence affect children living in the household in terms of:
  – Academic achievement
  – Mental health

• Do changes in household structure affect children living in the household in terms of:
  – Academic achievement
  – Mental health
Method

• Health card registration data
  – NILS members selected based on 104 dates of birth
  – Identifies individuals
  – Identifies home address
  – And change of address at 21 time points between 2001 & 2011

• NILS cohort comprises c.28% of NI population

• Cohort lives in c.50% of residential households in NI
  – Non cohort member Census returns provide household characteristics and characteristics of family structure
Datasets to be linked

• **Land and property services**
  – The open market sale price of every property in Northern Ireland (as of 2005)

• **Census 2001**
  – Social/ private rented vs. owner occupier
  – Health & educational qualifications

• **Census 2011**
  – Social/ private rented vs. owner occupier
  – Health & Educational qualifications

• **Non-members 2001 & 2011**
  - Relationships within Census household/household structure
Results

• 182,545 individuals, Under the age of 18 years & enumerated in both 2001 and 2011 Censuses

• Mental ill health: 3,349 (2%)

• No qualifications: 10,887 (13%)
## Percentage with no qualifications by economic transitions

<table>
<thead>
<tr>
<th></th>
<th>House Value 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Renting</td>
</tr>
<tr>
<td><strong>House</strong></td>
<td>Renting</td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td>&lt;75K</td>
</tr>
<tr>
<td></td>
<td>75-95K</td>
</tr>
<tr>
<td><strong>2001</strong></td>
<td>95K+</td>
</tr>
</tbody>
</table>

NILS 2011 Census Launch - 5th June 2014
Percentage in poor mental health by economic transitions

<table>
<thead>
<tr>
<th></th>
<th>Renting</th>
<th>&lt;75K</th>
<th>75-95K</th>
<th>95K+</th>
</tr>
</thead>
<tbody>
<tr>
<td>House</td>
<td>Renting</td>
<td>5.5</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>Value</td>
<td>&lt;75K</td>
<td>6.5</td>
<td>1.7</td>
<td>2.2</td>
</tr>
<tr>
<td>2001</td>
<td>75-95K</td>
<td>5.6</td>
<td>3.2</td>
<td>1.3</td>
</tr>
<tr>
<td>95K+</td>
<td></td>
<td>4.8</td>
<td>4.4</td>
<td>3.3</td>
</tr>
</tbody>
</table>
## Percentage in poor mental health by family transitions

<table>
<thead>
<tr>
<th></th>
<th>Living Arrangements 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 parent</td>
</tr>
<tr>
<td>Living</td>
<td>2 parent</td>
</tr>
<tr>
<td>2 parent</td>
<td>1</td>
</tr>
<tr>
<td>Arrangements</td>
<td>Single parent</td>
</tr>
<tr>
<td>2001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step family</td>
</tr>
<tr>
<td>2001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Potential of New 1991 & 2011 link

1991 Census

1997

GRO Birth records

2001 Census

address changes

2008

EPD data

2011 Census

Deaths

0 yrs

4 yrs

14 yrs

0 yrs

10 yrs

20 yrs

NILS 2011 Census Launch - 5th June 2014
Potential of New 1991 & 2011 link

• Direction of effect
  – Does child health affect economic position or vice versa?

• Reason for transition
  – Parental separation, parent death, job loss

• Timing of transition
  – Critical developmental periods, economic downturn

• Reason for effect
  – Parental mental health, family conflict, financial resource

• Questions for SDAI Phase 2:
  – Families, Adolescents and Children in Transition
Acknowledgements

“The help provided by the staff of the Northern Ireland Longitudinal Study/Northern Ireland Mortality Study (NILS/NIMS) and the NILS Research Support Unit is acknowledged. The NILS/NIMS is funded by the Health and Social Care Research and Development Division of the Public Health Agency (HSC R&D Division) and NISRA. The NILS-RSU is funded by the ESRC and the Northern Ireland Government. The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of NISRA/NILS.”
Health, housing tenure & entrapment
2001-2011:
Does changing tenure and address improve health?
Myles Gould (UoL) Email: m.i.gould@leeds.ac.uk  Twitter: @Myles_Gould_UoL
Ian Shuttleworth (QUB) Email: i.shuttleworth@qub.ac.uk
Structure

• Introduction
• Project 067 Aims
• Data & analytical approach
• Population bases
• Exploratorive results
  – 2001 health/2011 health transition
• Conclusions
Introduction

• Political & policy debates about social-rented housing focus on low spatial mobility & reduced chances of upward social mobility

• Extensive literature on inter-relationships between: housing tenure, health, and wider dimensions of social wellbeing, and the measurement of these at both the individual and area level (e.g. Marmot, 2010; Macintyre et al, 2002)

• Smith & Easterlow (2005) consider concepts of housing entrapment & selective placement
Project 067 Aims

• To explore relationships between *changing* health & *housing tenure* in Northern Ireland, 2001-2011
  – plus spatial mobility (i.e. change SOA geographical area)
• To determine whether different *tenure trajectories* are associated with changes in health status e.g.
  movements from social rented to owner occupied housing & changing health status
• To explore whether these trajectories are linked to different kinds of spatial move
  – moves between different types of place/area

*n.b. only consider first two today*
Population Bases

• **Movers – changed tenure**
  – may/may not changed house/address, &/or changed SOA
  – theoretically possible to only change tenure

• **All movers changed address OAs**
  – Still to analyse this
Data and Analytical Approach

• Restricted set of variables for parsimony
• Descriptive analysis different kinds of tenure/health transition in the NILS
  – e.g. changes in individuals’ limiting long-term illness and general health, 2001-2011
  – i.e. changes tenure type
• Multilevel statistical modelling (individuals nested in SOAs)
# Health change 2011 compared to 2001

<table>
<thead>
<tr>
<th>Health 2001</th>
<th>Good health</th>
<th>Fairly good health</th>
<th>Not good health</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good health</td>
<td>143503</td>
<td>23012</td>
<td>5458</td>
<td>171973</td>
</tr>
<tr>
<td></td>
<td>83.4%</td>
<td>13.4%</td>
<td>3.2%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Fairly good health</td>
<td>29323</td>
<td>23659</td>
<td>6528</td>
<td>59510</td>
</tr>
<tr>
<td></td>
<td>49.3%</td>
<td>39.8%</td>
<td>11.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Not good health</td>
<td>6131</td>
<td>12632</td>
<td>11187</td>
<td>29950</td>
</tr>
<tr>
<td></td>
<td>20.5%</td>
<td>42.2%</td>
<td>37.4%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>178957</td>
<td>59303</td>
<td>23173</td>
<td>261433</td>
</tr>
<tr>
<td></td>
<td>68.5%</td>
<td>22.7%</td>
<td>8.9%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
• Reminds us of obvious importance of taking account of age, doing this in our statistical modelling work
## Health & Tenure Transitions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Remains Own Occup</td>
<td>113919</td>
<td>15224</td>
<td>2870</td>
<td>20803</td>
<td>14667</td>
<td>3202</td>
<td>3714</td>
<td>6912</td>
<td>5100</td>
</tr>
<tr>
<td>Own Occup &gt; Soc Rent</td>
<td>8117</td>
<td>1891</td>
<td>635</td>
<td>1682</td>
<td>1452</td>
<td>549</td>
<td>387</td>
<td>825</td>
<td>768</td>
</tr>
<tr>
<td>Own Occup &gt; Priv Rent</td>
<td>898</td>
<td>274</td>
<td>80</td>
<td>217</td>
<td>391</td>
<td>98</td>
<td>41</td>
<td>134</td>
<td>126</td>
</tr>
<tr>
<td>Soc Rent &gt; Own Occup</td>
<td>5289</td>
<td>884</td>
<td>273</td>
<td>1510</td>
<td>990</td>
<td>354</td>
<td>479</td>
<td>637</td>
<td>709</td>
</tr>
<tr>
<td>Remains Soc Rent</td>
<td>6823</td>
<td>2718</td>
<td>940</td>
<td>2744</td>
<td>3608</td>
<td>1397</td>
<td>807</td>
<td>2543</td>
<td>3149</td>
</tr>
<tr>
<td>Soc Rent &gt; Priv Rent</td>
<td>286</td>
<td>198</td>
<td>88</td>
<td>168</td>
<td>368</td>
<td>136</td>
<td>71</td>
<td>267</td>
<td>298</td>
</tr>
<tr>
<td>Priv Rent &gt; Own Occup</td>
<td>4361</td>
<td>447</td>
<td>92</td>
<td>824</td>
<td>381</td>
<td>108</td>
<td>185</td>
<td>194</td>
<td>143</td>
</tr>
<tr>
<td>Priv Rent &gt; Soc Rent</td>
<td>2643</td>
<td>704</td>
<td>224</td>
<td>873</td>
<td>798</td>
<td>292</td>
<td>247</td>
<td>549</td>
<td>515</td>
</tr>
<tr>
<td>Remains Priv Rent</td>
<td>492</td>
<td>88</td>
<td>18</td>
<td>109</td>
<td>132</td>
<td>36</td>
<td>25</td>
<td>63</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>142828</td>
<td>22428</td>
<td>5220</td>
<td>28930</td>
<td>22787</td>
<td>6172</td>
<td>5956</td>
<td>12124</td>
<td>10848</td>
</tr>
</tbody>
</table>
Health & Tenure Transitions
Modelling Approach

• Restricted set of variables for parsimony

• Multilevel statistical modelling (individuals nested in SOAs)
  – Place difference having taken account socio-demographic characteristics [bit today]
  – Cross-level interactions (individual/household & area effects) – still to do

• An number of logistical regression models treating response as binary outcomes
  – Bad 2001 to Bad 2011; V Good 2001 to V Good 2011; Moderate 2001 to Moderate 2011; V Good 2001 to Moderate 2011; V Good 2001 to Bad 2011 etc (9 suites)
  – Also modelling V Good vs Fair/Poor (2011); V Good/Fair vs Poor (2011); with 2001 health as a predictor
  – Considering Multinomial (multivariate) response specification
Health & Tenure Transitions (Model predictions)

Example Model
- Response: probability of transition from good health 2001 to good health 2011
- Having allowed for tenure change
- Also age, sex, education level, community background
- Plus allowed for rates to vary by SOA (small effect, but significant place differences)
Conclusions

• Other model results reassuring (incl. inclusion of area effects; e.g. deprivation)
  – finding similar/consistent patterns with different specifications of Y-variable
• Seeking to model considerable complexity: transitional states, compositional & contextual, & cross-level
  – Possible because of large & rich variable detail of NILS
• Results are preliminary & analysis is ongoing (this week)
• Age decreases the probability of transiting from bad to good health, and of remaining in good health in 2001 and 2011
• Moves out of owner occupation are associated with transitions to poor self-reported health
• Remaining in social renting in 2001 and 2011 is associated with poorer self-reported general health in 2001 and 2011
• Implications; tenure and spatial mobility (or its lack) linked to social residualisation
• Very keen to share more results later with interested stakeholders
Acknowledgement

The help provided by the staff of the Northern Ireland Longitudinal Study and the NILS Research Support Unit is acknowledged. The NILS is funded by the Health and Social Care Research and Development Division of the Public Health Agency (HSC R&D Division) and NISRA. The NILS-RSU is funded by the ESRC and the Northern Ireland Government. The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of NISRA/NILS.

NILS-RSU Contact Details:
Email: nils-rsu@qub.ac.uk
Does the Month of Birth affect Educational Success Later in Life?

An analysis based on Northern Ireland Longitudinal Study (NILS)-Data.

Stefanie Doebler (QUB), Ian Shuttleworth (QUB), Myles Gould (University of Leeds)
Background & Literature

There is an array of literature on the effects of the month of birth on educational and labour market outcomes in later life.

Key Problem: Because all children born before a certain cut-off date of a school-year are included in the same class, some children are almost a year younger than some of their classmates.

Research Question: Does the Month of Birth affect educational success later in life? Are students born closer before the cut-off date educationally disadvantaged?

NILS 2011 Census Launch - 5th June 2014
For England and Wales, there is evidence that those born in July and August have poorer educational and labor market outcomes (Crawford, Dearden, and Meghir 2010; Crawford, Dearden, and Meghir 2007; Sprietsma 2010; Black, Devereux, and Salvanes 2010) than those born later in the year.

In England & Wales the school year starts on 1st September.

In Northern Ireland, children born before 30th June are included in the first class of a school year. The school year starts on 1st July.

In the USA it has been suggested that those with winter births experience poorer later outcomes (Buckles et al 2010).
Motivation:

- The earlier start of the Northern Ireland school year permits a quasi-natural experiment in which Northern Ireland can be contrasted with England & Wales and the robustness of birth month effects to be assessed.

- The NILS, with its large sample size (N = c. 500,000) is an excellent data source to analyze even relatively small groups (pupils by birth-months).

- Also, the longitudinal design of the NILS allows to look at outcomes later in life.
Caveats:

The school-starting age in Northern Ireland (4 years) differs from England and Wales (5 years). This impairs the comparability of our data with the English and Welsh data.

The literature on England and Wales, and also the majority of studies from the US use educational data (data on test-scores controlling for cognitive ability). Results based on the NILS are not directly comparable to these data.

Intervals between waves: ten years is a long interval. A lot can happen between 2001 and 2011 that we cannot account for.

But Strength: Educational success later in life (‘having a degree’) is comparable to the abovementioned studies. The NILS offers a large sample to examine this.
Data & Methods

• **Data:** The Northern Ireland Longitudinal Study (NILS) with Census link for 2001 and 2011.

• **Sample:** all NILS-Members aged 12 to 18 in 2001, and 22 to 28 in 2011.

• Thus in 2001 the respondents are of school-age and living with their parents, and in 2011 can have obtained a degree.

• **N= 36,087**

• **Methods:** bivariate statistics, hierarchical linear & binary logistic Multilevel Modeling

• **Response:** ‘Having a Degree in 2011’ (binary logistic regression)
Design and Strategy

1. Bivariate and Multivariate Analyses across the Sample (everybody aged 16 and older) as a whole in 2011

2. **NILS-Sub-Sample:** Everybody aged 12-18 in 2001 (still living with their parents), the same respondents in 2011, when they are aged 22-28 (and can have obtained at least a first University degree)

3. Response Variable: ‘Having a Degree in 2011’ (binary logistic multilevel model)
<table>
<thead>
<tr>
<th>Birthmonth</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>3,117</td>
<td>8.64</td>
</tr>
<tr>
<td>February</td>
<td>2,754</td>
<td>7.63</td>
</tr>
<tr>
<td>March</td>
<td>3,249</td>
<td>9.00</td>
</tr>
<tr>
<td>April</td>
<td>2,810</td>
<td>7.79</td>
</tr>
<tr>
<td>May</td>
<td>3,221</td>
<td>8.93</td>
</tr>
<tr>
<td>June</td>
<td>3,306</td>
<td>9.16</td>
</tr>
<tr>
<td>July</td>
<td>2,875</td>
<td>7.97</td>
</tr>
<tr>
<td>August</td>
<td>3,250</td>
<td>9.01</td>
</tr>
<tr>
<td>September</td>
<td>3,170</td>
<td>8.78</td>
</tr>
<tr>
<td>October</td>
<td>2,999</td>
<td>8.31</td>
</tr>
<tr>
<td>November</td>
<td>2,679</td>
<td>7.42</td>
</tr>
<tr>
<td>December</td>
<td>2,657</td>
<td>7.36</td>
</tr>
<tr>
<td>Total</td>
<td>36,087</td>
<td>100</td>
</tr>
</tbody>
</table>
Findings:

% having a university degree

January | February | March | April | May | June | July | August | September | October | November | December | Total
32      | 33      | 33    | 36    | 37  | 36   | 35   | 34     | 34       | 33    | 32      | 33      | 33
Percent of Respondents Who have a University Degree by Birth-month and Year

- born in 1983
- born in 1984
- born in 1985
- born in 1986
- born in 1987
- born in 1988
## Binary Logistic Multilevel Model: ‘Having a Degree in 2011’ on Month of Birth and Controls

<table>
<thead>
<tr>
<th></th>
<th>M1 Coef.</th>
<th>S.E.</th>
<th>M4 Coef.</th>
<th>S.E.</th>
<th>M5 Coef.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>January</td>
<td>-0.099</td>
<td>0.058</td>
<td>-0.082</td>
<td>0.06</td>
<td>-0.079</td>
<td>0.06</td>
</tr>
<tr>
<td>February</td>
<td>-0.063</td>
<td>0.06</td>
<td>-0.051</td>
<td>0.061</td>
<td>-0.04</td>
<td>0.061</td>
</tr>
<tr>
<td>March</td>
<td>-0.069</td>
<td>0.057</td>
<td>-0.063</td>
<td>0.059</td>
<td>-0.056</td>
<td>0.059</td>
</tr>
<tr>
<td>April</td>
<td>0.01</td>
<td>0.059</td>
<td>0.051</td>
<td>0.061</td>
<td>0.053</td>
<td>0.061</td>
</tr>
<tr>
<td>May</td>
<td>0.072</td>
<td>0.057</td>
<td>0.109</td>
<td>0.059</td>
<td>0.111</td>
<td>0.059</td>
</tr>
<tr>
<td>June</td>
<td>0.064</td>
<td>0.057</td>
<td>0.096</td>
<td>0.059</td>
<td>0.099</td>
<td>0.058</td>
</tr>
<tr>
<td>July</td>
<td>-0.086</td>
<td>0.059</td>
<td>-0.064</td>
<td>0.061</td>
<td>-0.062</td>
<td>0.061</td>
</tr>
<tr>
<td>August</td>
<td>-0.038</td>
<td>0.057</td>
<td>-0.022</td>
<td>0.059</td>
<td>-0.013</td>
<td>0.059</td>
</tr>
<tr>
<td>September</td>
<td>-0.032</td>
<td>0.058</td>
<td>-0.003</td>
<td>0.059</td>
<td>-0.002</td>
<td>0.059</td>
</tr>
<tr>
<td>October</td>
<td>-0.028</td>
<td>0.058</td>
<td>0.007</td>
<td>0.06</td>
<td>0.016</td>
<td>0.06</td>
</tr>
<tr>
<td>November</td>
<td>-0.09</td>
<td>0.06</td>
<td>-0.08</td>
<td>0.062</td>
<td>-0.069</td>
<td>0.062</td>
</tr>
<tr>
<td>Sex: female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.090***</td>
<td>0.006</td>
<td>0.091***</td>
<td>0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents divorced/separated</td>
<td>-0.613***</td>
<td>0.036</td>
<td>-0.560***</td>
<td>0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>House-owner</td>
<td>0.310***</td>
<td>0.031</td>
<td>0.312***</td>
<td>0.031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parental Deprivation: Education</td>
<td>-0.764***</td>
<td>0.034</td>
<td>-0.729***</td>
<td>0.033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parental Deprivation: Employment</td>
<td>-0.557***</td>
<td>0.03</td>
<td>-0.489***</td>
<td>0.030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area: Multiple Deprivation Score</td>
<td></td>
<td></td>
<td>-0.421***</td>
<td>0.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.677***</td>
<td>0.046</td>
<td>-3.590***</td>
<td>0.169</td>
<td>-2.466***</td>
<td>0.175</td>
</tr>
<tr>
<td>Level2-Variance</td>
<td>0.529***</td>
<td>0.046</td>
<td>0.403**</td>
<td>0.047</td>
<td>0.266***</td>
<td>0.069</td>
</tr>
<tr>
<td>N</td>
<td>36087</td>
<td></td>
<td>36087</td>
<td></td>
<td>36087</td>
<td></td>
</tr>
</tbody>
</table>
Some Cross-Level Interactions:

Predictive Margins of Parental Educational Deprivation with 95% CIs

'Having a degree in 2011', predicted probabilities vs. Multiple Area Deprivation Score (log)

Educational Deprivation
- no
- yes
Some Cross-Level Interactions:

Predictive Margins of parental employment deprivation with 95% CIs

"Having a degree in 2011", predicted probabilities

Multiple Area Deprivation Score (log)

parents not deprived

parents deprived
Some Cross-Level Interactions:

Predictive Margins of homeowner with 95% CIs

'having a degree in 2011', predicted probabilities

Multiple Area Deprivation Score (log)

not a homeowner  
homeowner
Conclusions

- In Northern Ireland the month of birth does not have a (substantially or statistically) significant effect on the likelihood of having obtained a degree later in life.
- Thus, our data do not confirm the hypothesis of a long-term educational disadvantage.
- The same analysis was carried out for ‘having obtained A-levels’, with the same result.
- What matters for tertiary educational outcomes are parental (material, educational and employment) deprivation, having separated/divorced parents, and area-deprivation.
- Future analyses should focus on these factors.
Acknowledgement

The help provided by the staff of the Northern Ireland Longitudinal Study and the NILS Research Support Unit is acknowledged. The NILS is funded by the Health and Social Care Research and Development Division of the Public Health Agency (HSC R&D Division) and NISRA. The NILS-RSU is funded by the ESRC and the Northern Ireland Government. The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of NISRA/NILS.

NILS-RSU Contact Details:
Email: nils-rsu@qub.ac.uk
References


CALLS Hub and the SPiNe

Professor Allan Findlay
University of St Andrews
Introduction

• Thinking longitudinally
• The purpose of CALLS Hub – (Census and Administrative Data Longitudinal Studies Hub)
• Introduction to SPiNe (Science and Policy Network)
The key feature of longitudinal studies is that one uses data for the same individual for multiple points in time. The events may be from different data sets. It is a good way to investigate features such as social change or health change over time.
Longitudinal Effects

- Age
- Period
- Cohort

Temporary

Lasting

Lagged

Anticipated

Effect:  
Event:  X  
Observations:  

NILS 2011 Census Launch - 5th June 2014
Aims of Census and Administrative data Longitudinal Studies Hub – CALLS Hub

• One stop shop – information and advice
• Data dictionary
• Examples of uses of the LS’s – impact
• Establish and engage with a Science and Policy Network
• Help with running UK level analysis
Aims of Census and Administrative data
Longitudinal Studies Hub – CALLS Hub

• Data dictionary – Harmonization of variables

• Help with running UK level analysis
High impact Publications

THE LANCET

Socioeconomic inequalities in morbidity and mortality in western Europe


American Journal of PUBLIC HEALTH

BMJ

Higher mortality in deprived areas: community or personal disadvantage?

Occupational class and cause specific mortality in middle aged men in 11 European countries: comparison of population based studiesCommentary: Unequal inequalities across Europe

Transactions of the Institute of British Geographers

Does Britain have ghettos? Peach,

C 1996 Volume: 21 Issue: 1 Page: 216 -235
Citations of ONS LS research

- Mortality, 1135
- Heart diseases, 437
- Socio-economic status, 931
- Cancer, 391
- Social, 755
- Inequalities, 708
- Occupation/work, 694
- Household/Family, 567
- Aging, 535
- Gender, 391
- Migration, 427
- Economics/labour/pension, 103
- Time trends, 153
- Fertility, 49
Impact from the LSs – E&W LS

• Data from the ONS LS has a long history of use in Gov reviews and enquiries, e.g.:
  – Dilnot Commission (2011)
  – Marmot Review (2010)

• DH and DWP: LS-based statistics are key to their work on pension projections
Impact from the LSs – NILS

• O’Reilly, QUB with Director of Screening Services (NI Public Health Agency) - Breast Screening uptake project
  – PI invited to join the NI PHA Screening Planning and Review Group

• Shuttleworth, QUB work on internal migration, social deprivation and religion in Northern Ireland:
  – Led to advice to Community Relations Council, NI Assembly and NI Office
Potential developments and data linkages in the LSs
A message to everyone - act now!

Everyone should be included in the census. Households and overnight visitors are included in the census to help plan and fund services, community - services like transport, education and health.

Please complete your census questionnaire by 27 March 2011, or as soon as possible afterwards. You can fill it in online or on paper. If you are taking part in the census it is very important that it is also compulsory for you. If you do not participate or if you supply false information, you could be fined.

St Andrews

Local Services Network

Opening Hours

<table>
<thead>
<tr>
<th>Day</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>9.00 to 4.30</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9.00 to 4.30</td>
</tr>
<tr>
<td>Wednesday</td>
<td>9.30 to 4.30</td>
</tr>
<tr>
<td>Thursday</td>
<td>9.00 to 4.30</td>
</tr>
<tr>
<td>Friday</td>
<td>9.00 to 4.30</td>
</tr>
</tbody>
</table>

Registration of Births, Deaths and Marriages available here.

Authourised for Civil Marriages
Hospital data

• In patient stays
• Mental health
• Violence – knife injury
• Drug misuse
• Child health surveillance systems
• Prescribing data
Education data in Scotland

- Attendance/Absence
- Exclusions
- Attainment

- School census
  - Main difficulty in learning
  - Free school meals registered
  - Individual educational programme (IEP)
  - Records of needs (RON)
  - English as a second language
Ecological data

- **Pollution data**
  - CO, NO, SO$_2$, O$_3$
  - Particulate matter

- **Weather data**
  - Temp, frost, sunshine, precipitation, clouds
SPiNe

• CALLS Hub Science & Policy Network

• Aims
  – to expand the traditional body of LS data users, to include a wider range of interest groups
  – to expand the range of research themes being explored using LS data
  – to help guide future data developments
SPiNe

• We aim to include
  – Existing and additional academic users
  – Local and national government users
  – Think tank analysts
  – Voluntary sector researchers

• We see mutual benefits
  – New users for LS RSUs
  – Advice and expertise to members
SPiNe

• Research Working Groups
  – Clusters of researchers interested in particular issues
  – Support for data access
  – Production of impartial reports
Thank you

For more information visit:
calls.ac.uk
Outline

• What is SYLLS and why is it important?
• Our approach
  – The National Synthetic Data Spine
  – Individual Bespoke Synthetic Data
What is SYLLS?

• A project to generate synthetic versions of the national LSs which look and behave like the real thing, but are not subject to the same access restrictions
Why do we need synthetic data?

• ONS LS, Scottish LS and Northern Ireland LS are an unparalleled resource for social science research in the UK.

• But compared to other Census data products, we have a small user base (although larger teams than number of researchers).

<table>
<thead>
<tr>
<th>Census Data product</th>
<th>Unique users 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONS LS</td>
<td>62 (open projects) 46 (active)</td>
</tr>
<tr>
<td>Flow data*</td>
<td>616</td>
</tr>
<tr>
<td>Aggregate data (Casweb)*</td>
<td>5781</td>
</tr>
<tr>
<td>Boundary data*</td>
<td>2873</td>
</tr>
</tbody>
</table>

*data from Q4 2012, Q3 2013 Census Support Service
Why a relatively small user base?

<table>
<thead>
<tr>
<th>Route to accessing flow data</th>
<th>Route to accessing ONS LS data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Formulate research question</td>
<td>1. Formulate research question</td>
</tr>
<tr>
<td>2. Turn on computer</td>
<td>2. Turn on computer</td>
</tr>
<tr>
<td>3. Go to cider.census.ac.uk</td>
<td>3. Go to ucl.ac.uk/celsius</td>
</tr>
<tr>
<td>4. Log on to WICID (now open access)</td>
<td>3. Download customer request form, data access agreement and approved researcher form</td>
</tr>
<tr>
<td>5. Choose your data</td>
<td>4. Fill out forms and submit for approval</td>
</tr>
<tr>
<td>6. Download to your own computer and analyse with preferred software</td>
<td>5. Wait for approval from LS research board</td>
</tr>
<tr>
<td>7. Repeat as necessary</td>
<td>6. Attend safe researcher certification course</td>
</tr>
<tr>
<td></td>
<td>7. Ask research support officer to build your dataset from LS database</td>
</tr>
<tr>
<td></td>
<td>8. Hop on train to London, Newport or Titchfield to attend VML</td>
</tr>
<tr>
<td></td>
<td>9. Carry out analysis on VML terminal with old, slow software</td>
</tr>
<tr>
<td></td>
<td>10. Ask for intermediate outputs to be cleared</td>
</tr>
<tr>
<td></td>
<td>11. Seek final output clearance from LS research board</td>
</tr>
<tr>
<td></td>
<td>12. Repeat as necessary</td>
</tr>
</tbody>
</table>

NILS 2011 Census Launch - 5th June 2014
Why such a small user base?

- Complex data (compared to other cross-sectional Census data products)
- Lack of exposure early in academic careers
- These are not ‘hands on’ data
Why do we need synthetic data?

• Access LS-like data on own computer
  — Iteratively refine research ideas, update analysis code etc.
• Use data in teaching and expose social science students to longitudinal data early in their research careers
• A UK longitudinal study dataset
• Methodological innovation for UK Census microdata – beyond 2011 agenda
Our approach

• Two project streams:
  – National Synthetic LS Data Spine
    • Adam Dennett, Belinda Wu, Nicola Shelton, Mike Batty and Rachel Stuchbury (UCL)
  – Bespoke Synthetic Datasets
    • Chris Dibben, Gillian Raab and Beata Nowok (Edinburgh)

• Ian Shuttleworth and Tony Gallagher also project partners (Queen’s Belfast)
National Synthetic Data Spine

• Aims:
  – To create a core ‘spine’ dataset which:
    • Contains the same number of individuals (E&W 500,000 people + Scot 274,000 people + NI 500,000 people) as are in the LSs across 1991 > 2001 Censuses
    • Has variable distributions which match those in the LS data for Age, Sex, Ethnicity, Limiting Long Term Illness, Marital Status, Births and Deaths
    • And has accurate spatial distributions of these individuals and their characteristics at the 1991 county district level
    • Has accurate state transitions of individuals from 1991 to 2001 at ‘91CD level
National Synthetic Data Spine

• Our method: Spatial Microsimulation

1. Take sample population from (publicly available) 1991 Individual SAR
2. Update values for SAR individuals according to LS distributions at county district level
4. Finish with a full set of microdata records for all individuals in UK, with accurate transitions between 1991 and 2001 and accurate spatial distributions for 8 core variables
### Synthetic Spine Data

#### 1991

<table>
<thead>
<tr>
<th>UID</th>
<th>AREAP</th>
<th>ETHGROUP</th>
<th>LTILL</th>
<th>MSTATUS</th>
<th>SEX</th>
<th>agegroup</th>
<th>age</th>
<th>death</th>
<th>birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1871</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1872</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1873</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1874</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1875</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1876</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1877</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 2001

<table>
<thead>
<tr>
<th>UID</th>
<th>AREAP</th>
<th>ETHGROUP</th>
<th>LTILL</th>
<th>MSTATUS</th>
<th>SEX</th>
<th>agegroup</th>
<th>age</th>
<th>death</th>
<th>birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-9</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1871</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1872</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1873</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1874</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1875</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1876</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1877</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
National Synthetic Data Spine

• National Synthetic Spine almost complete:
  – Bespoke Spatial Microsimulation Software finished
  – E&W 1991-2001 static data complete, longitudinal linkage almost there
  – Scotland 1991-2001 almost complete
  – NI in progress
Synthetic Spine Release Plans

- Currently in conversation with ONS, NRS and NISRA, but plans are for:
  - Open Access
  - Available through CALLS Hub and national research support units

- Completed software means potential for 2011 linkage in the future
Bespoke Synthetic Datasets

• Aims:
  – To develop a methodology and accompanying software which will allow the swift generation of statistically representative, but completely synthetic, versions of data requests submitted to the national LS Research Support Units
  – To make some bespoke synthetic datasets available for teaching, subject to disclosure control.
Bespoke Synthetic Datasets

• Our method: Conditional Simulation Models
  1. Take a data extract from one of the national LS datasets
  2. Sequentially generate synthetic data from fitted conditional models
  3. Final result is a completely synthetic representation of the joint distribution (if the models are true)
synthpop

• synthpop package developed in R
• Structure is based on the ‘mice’ multiple imputation package
• Range of parametric and non-parametric (classification and regression trees) options for data synthesis
• Allows for data rules, e.g. no married children
• Models missing data to produce missing data patterns like the real data
<table>
<thead>
<tr>
<th>sex</th>
<th>age</th>
<th>edu</th>
<th>marital</th>
<th>incomemn</th>
<th>ls</th>
<th>wkabint</th>
<th>wkabintdur</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMAN</td>
<td>57</td>
<td>VOCATIONAL/GRAMMAR</td>
<td>MARRIED</td>
<td>800</td>
<td>PLEASED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>20</td>
<td>VOCATIONAL/GRAMMAR</td>
<td>MARRIED</td>
<td>350</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>18</td>
<td>VOCATIONAL/GRAMMAR</td>
<td>UNMARRIED</td>
<td>NA</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>78</td>
<td>PRIMARY/NO EDUCATION</td>
<td>MARRIED</td>
<td>900</td>
<td>PLEASED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>20</td>
<td>VOCATIONAL/GRAMMAR</td>
<td>MARRIED</td>
<td>1500</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>39</td>
<td>SECONDARY</td>
<td>MARRIED</td>
<td>2000</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>39</td>
<td>SECONDARY</td>
<td>MARRIED</td>
<td>1197</td>
<td>MIXED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>43</td>
<td>SECONDARY</td>
<td>MARRIED</td>
<td>580</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>63</td>
<td>SECONDARY</td>
<td>MARRIED</td>
<td>1400</td>
<td>PLEASED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>38</td>
<td>PRIMARY/NO EDUCATION</td>
<td>WIDOW(ER)</td>
<td>1350</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>73</td>
<td>PRIMARY/NO EDUCATION</td>
<td>WIDOW(ER)</td>
<td>1313</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>70</td>
<td>SECONDARY</td>
<td>MARRIED</td>
<td>800</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>30</td>
<td>PRIMARY/NO EDUCATION</td>
<td>UNMARRIED</td>
<td>1300</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>68</td>
<td>PRIMARY/NO EDUCATION</td>
<td>MARRIED</td>
<td>1500</td>
<td>MOSTLY SATISFIED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>61</td>
<td>PRIMARY/NO EDUCATION</td>
<td>MARRIED</td>
<td>-8</td>
<td>MIXED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>MAN</td>
<td>84</td>
<td>PRIMARY/NO EDUCATION</td>
<td>UNMARRIED</td>
<td>2000</td>
<td>PLEASED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
<tr>
<td>WOMAN</td>
<td>87</td>
<td>SECONDARY</td>
<td>DIVORCED</td>
<td>1400</td>
<td>MIXED</td>
<td>NO</td>
<td>MISS/NA</td>
</tr>
</tbody>
</table>
R code to synthesise: `test <- syn(data)`
And compare to real data: `compare.synds(test, data)`
Produces the plots below
R code to synthesise:
```
test <- syn(data, m=10)
```

Fit to synthetic data:
```
fit.test <- glm.synds(wkabint~sex+age+edu+log(incomenm), object=test, family="binomial")
```

And compare to fit for real data:
```
compare.fit.syn(fit.test, data, plot="Z")
```

Produces plot on RHS

Young men more likely to intend to work abroad – other factors don’t matter
Same conclusion from synthetic data
synthpop

- synthpop produces fully synthetic datasets which closely resemble the real longitudinal microdata
- Users who submit project proposals will be able to request synthetic datasets for personal research purposes
When can we access SYLLS data?

- Spine dataset(s) available soon (end of June) via CALLS and RSUs
SYLLS

• When can we access SYLLS data?
  • Users will shortly be able to request bespoke datasets from synthpop to accompany data requests *although a few software and disclosure control hurdles to jump first*
Thank you
Better Knowledge
Better Society

Administrative Data
Research Centre –
Northern Ireland

Dermot O’Reilly
The UK Administrative Data Research Network:
Improving Access for Research and Policy

Report from the Administrative Data Taskforce
December 2012
The key recommendations from the BIS report are:

- A UK Administrative Data Research Network (ADRN) will be responsible for linking data between government departments. The proposed network will provide a single governance structure that will allow for consistent and robust decision-making.

- An Administrative Data Research Centre (ADRC) should be established in each of the four countries in the UK.

- Legislation should be enacted to facilitate research access to administrative data and to allow linkage between departments to take place more efficiently.

- A single UK-wide researcher accreditation process, built on national and international best practice should be established.

- A strategy for engaging with the public should be instituted.

- Sufficient funds should be put in place to support improved research access to and linkage between administrative data.
Big Data Network

ESRC Big Data Network

Governance Structures

UK Data Service and Administrative Data Service

ADRC

ADRC

ADRC

Data Research Centre

Data Research Centre

Data Research Centre

Data Research Centre

Third Sector Data Research Centre

Social Media Data Research Centre

International Centre for SM Research

Administrative Data

Phase 1

Business and Local Government Data

Phase 2

Third Sector and Social Media Data

Phase 3

NILS 2011 Census Launch - 5th June 2014
Some benefits...

• Realising the opportunity to exploit the richness of administrative data for social and economic research and policy, and for better targeting of interventions.

• A step-change in procedures for access to, and linking between, such data across Government Departments

• Ensuring the UK remains at the forefront of research based on linked administrative data and is competitive on the international market

• Allowing cost effective re-use of existing data by avoiding the duplication of data collection and capitalising on past investments

• Enabling new methodologies and cutting-edge research

• Developing skills and capacity
Farr Institute and ADRN

The Farr institute
- London
- Manchester
- Swansea
- Dundee

ADRN
- Southampton
- Belfast
- Swansea
- Edinburgh
- Essex (ADS)
Steering Committee

Support function

Operational Group

Administrative Data Forum

Data Sub-Group

Substantive Research Group

User Forum

Methodological Research Group

Training and Capacity Building Group

Public Engagement and Communications Group

NILS 2011 Census Launch - 5th June 2014
Lead roles and responsibilities

- Linkage service
- Safe setting
- Data acquisition

- Public engagement
- Training, cap’ building
- Research

Research support
5 Safety Principles...

- Safe Projects (Valid research purpose)
- Safe People (Trusted Researchers)
- Safe Data (Anonymisation of data)
- Safe Setting (Security controls)
- Safe Outputs (Disclosure control of outputs)

=Safe use
Data

What:
- De-identified
- Administrative (non-commercial)
- Linked and non-linked

How accessed:
- Secure setting
- Syntax
- Developmental: synthetic or remote access
Public Engagement and Communications

- Awareness-raising
- Consultation
- Empowerment
Training and capacity building

- Accredited researcher
- Training related to ...
  - Data protection, disclosure control, etc
  - Record linkage
  - Administrative data
  - Statistical methods
An ADRN Project - Definition

- The project is for non-commercial research purposes
- The project has evident potential public benefit
- There is a demonstrable value from using unit level administrative data to answer the research question.
- The project would not be more appropriately served by other existing services (e.g. FARR, UKDS Secure Lab, Longitudinal Studies, HMRC Data Lab, etc.)
- The project does not constitute normal operational functions undertaken by government departments or their agencies.
An ADRN Project - Eligible people

- The researcher must be a ‘fit and proper person’ i.e.
  - They must be capable of carrying out the research either independently or under the direction of an appropriate supervisor or lead investigator;
  - They must have completed the ADRN Accreditation Training.
- The researcher must be from academia, the public sector or a research organisation on the Research Councils UK list of eligible independent research organisations (http://www.rcuk.ac.uk/funding/eligibilityforrcs/).
Types of Projects

• Using micro-data
  – Unlinked
  – Linked

• Others ‘Branded’ as ADRC projects
  – Aggregated Data
  – Methodological
  – Others

• Geography
  – NI only
  – Inter country
NILS Developments

Andrew Kerr
Data Integration And Linkage Service
NILS Developments

- Current Data
- 1991 Project
- Other Developments
NILS Developments – Current Data

• Routine Bi-Annual Updates

<table>
<thead>
<tr>
<th>January</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health (October Download)</td>
<td>Health (April Download)</td>
</tr>
<tr>
<td>GRO - Births</td>
<td></td>
</tr>
<tr>
<td>GRO - Deaths</td>
<td></td>
</tr>
<tr>
<td>NIMS11 Update</td>
<td></td>
</tr>
</tbody>
</table>

• NILS Metadata Update with each data release
NILS Developments – 1991 Project

1991 Census

2001 Census

2011 Census

Health Data
27 Downloads
Apr 2001 – Apr 2014
+ Pre-Apr 2001 Data

LPS Data

NILS Births
1991-2012

Births to NILS Mothers
1991-2012

Births to NILS Fathers
1991-2012

Deaths
1991-2012

NIMS01

Deaths
2001-2011

Census 2001

NIMS11

Deaths
2011-2012

Census 2011

NILS 2011 Census Launch - 5th June 2014
NILS Developments – 1991 Project

Also:

• 2001 Census Occupations for NILS members and associated HRPs recoded to SOC90
• Release by end September 2014
• Phase 2 - Further enhancements due to efficiencies from original 1991 Census work plan
• Release by end Spring 2015
NILS Developments – 1991
Project Phase 2

- 1981 Census
- 1991 Census
- 2001 Census
- 2011 Census

Health Data
- 29 Downloads Apr 2001 – Apr 2015 + Pre-Apr 2001 Data

NILS Data
- NILS Births 1974-2013
- Births to NILS Mothers 1974-2013
- Births to NILS Fathers 1974-2013
- Deaths 1981-2013

LPS Data

Census Data
- Census 1991
- Census 2001
- Census 2011

Deaths
- Deaths 1991-2001
- Deaths 2001-2011
- Deaths 2011-2013
NILS Developments – Other

- GRO Marriage Data – 2005 onwards
- Infant Mortality
- Widowerhoods
- Online Data Dictionary
Beyond 2011: Future Population and Social Statistics

Dr David Marshall
NISRA
Contact Details

@NILSRSU

www.nils-rsu.census.ac.uk

nils-rsu@qub.ac.uk