Improving contact rates in the field through analysis of linked Census survey data

Folasade Ariyibi
Overview

• Introduction
• Matching Process
• Characteristics of Non-contacts
• Paradata
• Multivariate Analysis
• Examples of Profiles
• Conclusion
Introduction

• Response rates have been declining over time

• Constraints used in weighting may account for some non-response bias but not all

• Weighting based on incorrect assumptions may lead to bias results

• Better to try and get more representative responses as oppose to compensating for non-response
Introduction

• The ONS carry out a Census every 10 years
• 2011 Census response rate – 94% overall
• At the same period as the Census, other social surveys are run in field.
• Six social surveys were selected for inclusion in this study.
• Census Non-Response Link Study (CNRLS)
Introduction

CNRLS involves matching Census data with social survey data to analyse patterns of non-response. This allows for:

• Analysis of potential bias in social surveys

• Opportunity to inform field force of better collection strategies (contact and co-operation)
Matching Process

- Matching carried out by Census-matching staff in ONS
- Matching at household and individual level
- Links survey-census addresses around 2011 Census date
- Match rates 94%
- Analysis restricted to Wave 1 for panel/longitudinal surveys
Characteristics of Non-contacts

Non-contact Rates by Age

Non-contact Rate (%)

Age Band

18-24
25-34
35-44
45-54
55-64
65-74
Characteristics of Non-contacts

Non-contact Rates by Marital Status

Non-contact Rate (%)

Marital Status
- Single
- Married/Civil Partnership
- Separated
- Widowed
Characteristics of Non-contacts

Non-contact Rate by Household Structure

Non-contact Rate (%)

Household Structure

<table>
<thead>
<tr>
<th>Household Structure</th>
<th>Non-contact Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Ad</td>
<td></td>
</tr>
<tr>
<td>1 Ad With Ch</td>
<td></td>
</tr>
<tr>
<td>2 Ad No Ch</td>
<td></td>
</tr>
<tr>
<td>2 Ad With Ch</td>
<td></td>
</tr>
<tr>
<td>2+ Ad No Ch (related)</td>
<td></td>
</tr>
<tr>
<td>2+ Ad With Ch (related)</td>
<td></td>
</tr>
<tr>
<td>2+ Ad No Ch (unrelated)</td>
<td></td>
</tr>
<tr>
<td>2+ Ad With Ch (unrelated)</td>
<td></td>
</tr>
</tbody>
</table>
Characteristics of Non-contacts

Non-contact Rate by Economic Activity Status

- Economically active: Employed
- Economically active: Unemployed
- Economically inactive
Characteristics of Non-contacts

Non-contact Rate by Ethnicity

- White
- Asian
- Black
- Other

Ethnicity

Non-contact (%)
But how will this help the field force?
Field interviewers tasked with collecting additional data about sampled addresses. The following variables were considered for use in the paradata analysis:

- Region (frame data)
- Calling Day
- Calling Time
- Type of Accommodation
- Call Duration
- Total Number of Attempts
- Number of Rooms
Non-contact Rates by Region

- LDN: 14%
- NE: 10%
- NW: 8%
- SE: 11%
- SW: 10%
- WM: 9%
- Wa: 9%
- YH: 7%
Non-contact Rate by No. of Rooms

- 1 room: 16%
- 2 rooms: 12%
- 3 rooms: 9%
- 4 rooms: 8%
- 5+ rooms: 7%
Non-contact Rate by Type of Accommodation

Type of Accommodation

Detached
Semi-Detached
Terrace
Purpose Built Flat
Converted Flat/Commercial Building/Caravan or Temporary Structure

Non-contact Rate (%)

0 2 4 6 8 10 12 14 16 18
Paradata

Non-contact Rate by Calling Time

Calling Time

Non-contact Rate (%)
Multivariate Analysis

- Logistic regression used to consider the joint predictive power of a set of covariates on non-contact
- Binary outcome (contact/non-contact)
- Modelled using household and person characteristics
- Some classes were collapsed to simplify model
- Model fitted using stepwise regression
Logistic Regression Findings

Main significant variables
Logistic Regression Findings

Main significant variables

Marital status
Logistic Regression Findings

Main significant variables:

- Pension Age
- Marital status
Logistic Regression Findings

Main significant variables:
- Pension Age
- Calling Day
- Marital status
Logistic Regression Findings

Main significant variables:
- Pension Age
- Calling Day
- Marital status
- Type of Accommodation
Logistic Regression Findings

Main significant variables:
- Pension Age
- Calling Day
- Marital status
- Household structure
- Type of Accommodation
Logistic Regression Findings

Main significant variables

- Pension Age
- Calling Day
- Marital status
- Household structure
- Type of Accommodation
- Age
Logistic Regression Findings

Main significant variables:
- Pension Age
- Calling Day
- Marital status
- Household structure
- Type of Accommodation
- Age
- Hours Worked
Logistic Regression Findings

Main significant variables:

- Pension Age
- Calling Day
- Marital status
- Region
- Household structure
- Type of Accommodation
- Age
- Hours Worked
Logistic Regression Findings

Main significant variables:
- Pension Age
- Calling Day
- Marital status
- Region
- Household structure
- Type of Accommodation
- Age
- Hours Worked
- Central Heating
Logistic Regression Findings

Main significant variables

- Pension Age
- Calling Day
- Marital status
- Region
- Household structure
- Type of Accommodation
- Age
- Hours Worked
- Central Heating
- Highest Qualification
Logistic Regression Findings

Main significant variables:

- Pension Age
- Calling Day
- Marital status
- Region
- Household structure
- Highest Qualification
- Type of Accommodation
- Disability Status
- Age
- Hours Worked
- Central Heating
- Region
- Disability Status
Logistic Regression Findings

Main significant variables:
- Pension Age
- Calling Day
- Marital status
- Ethnicity
- Highest Qualification
- Household structure
- Region
- Type of Accommodation
- Disability Status
- Age
- Hours Worked
- Central Heating
- Status
Logistic Regression Findings

Main significant variables:

- Pension Age
- Calling Day
- Marital status
- Region
- Ethnicity
- Highest Qualification
- Economic Activity
- Household structure
- Type of Accommodation
- Disability Status
- Age
- Hours Worked
- Central Heating
- Economic Activity
Example of Profiles

• Single Working Mothers (SWM)
  – Non-contact Rate = 13%

• Young Working Single Males (YWSM)
  – Non-contact Rate = 16%

• Employed in London Flats (ELT)
  – Non-contact Rate = 20%

• All
  – Non-contact Rate = 10%
Profiles - Calling Time

Non-contact Rate by Calling Time

- SWM
- YWSM
- ELF
- ALL

Calling Time:
- 8am
- 9am
- 10am
- 11am
- 12pm
- 1pm
- 2pm
- 3pm
- 4pm
- 5pm
- 6pm
- 7pm
- 8pm

Non-contact Rate (%): 0-35

Graph shows the non-contact rate percentage by calling time for different groups over the course of a day.
Profiles - Calling Time

Refusal Rate by Calling Time

Refusal Rate (%)

Calling Time

SWM

YWSM

ELF

ALL

8am 9am 10am 11am 12pm 1pm 2pm 3pm 4pm 5pm 6pm 7pm 8pm

Calling Time
Profiles - Calling Day

Non-response Rate by Calling Day

- SWM
- YESM
- ELF
- ALL

Non-response Rate (%)

Calling Day:
- Monday
- Tuesday
- Wednesday
- Thursday
- Friday
- Saturday
Example of Profiles

• Contact rates vary when looking at different subgroups

• Higher contact rates possible when contact strategies take account of the different subgroups

• Still difficult to implement in the field! How do we target these people??

• Use of additional variables or admin data
Conclusion

• CNRLS – a great but rare opportunity to assess response patterns and effectiveness of non-response weighting strategies

• Difficult to identify non-contacts from paradata alone

• Challenging to modify data collection strategy

• Need to look at other sources of information
Thank you!