

Main biases in mobile phone data

Susan Williams

24 May 2017

Overview

- 95% of adults have a mobile (Ofcom)
- Research on population densities/flows requires mobile phone = an individual
- Must weight to official population estimates
- Need to consider and adjust for the biases
- Different methods and accuracy depending on format of data (individual level vs counts)
- Opinions Survey 2015 identified some key information to help reduce bias (similar survey work currently being proposed in Eurostat).
- Need survey data to help correct biases (such as Labour force survey to inform on employment/work patterns)

Multiple phones

Opinions survey Autumn 2015

Issue: People can have more than one mobile

Number of phones	Totals	Percentage
0	2583697	5%
1	41614933	82%
2	5991779	12%
More than 2	329879	1%
Refuse to say or unrecorded	94284	0%
Total	50614572	100%

Individual Level data

- Identify same individual /multiple phones by same movement patterns over a period (MNO indicates 'this can be difficult')
- Ethical/privacy implications

Counts (e.g. # connections to masts. NB not # active call events*)

- Prepare counts having identified multiple phones from prior processing with individual level data
- Assuming homogeneity: Use survey data on distribution of # mobiles (as in opinions survey above) to adjust counts so that individuals are represented.
- For relative densities: assume homogenous distribution across areas (so if area A has twice the # of mobiles as area B – it contains double the population).

*People have widely varying propensities to use their mobiles. Academic research indicates that density of call events does not = density of population.

Population Definition*

- **Residence** is generally inferred as where the mobile (proxy for an individual) is mainly located (and possibly timed to during the night/evening).
- **Workplace** is generally inferred as where the mobile is found mainly in standard work hours (and repeat journeys during the working week seen)
- **Workday** population can also be inferred (different methods re individual vs count data)

Individual level data

Data spanning multiple weeks can reveal the repeated patterns needed to infer home location and work location.

Official residential population estimates needed to weight (to account for variation in each MNO's market share).

Individuals allocated a weight based on residence. This weight is used for journeys made by the individual etc.

Count data

Average # connected mobiles/call volumes in evening/night time hours typically used to infer residence.

Official residential population estimates needed to weight (to account for variation in each MNO's market share).

Relationship between # subscribers 'resident' in an area and official residence counts is assumed to hold for other times of the day/year.

*Official Statistics on residential population requires a period of residence in the UK. Individual level data covering this period might identify appropriate individuals; count data would need to have been prior processed at individual level to identify the appropriate subscribers – or need to use survey data to adjust counts

Population Definition Issues

Potential biases may arise from subscribers who have non-typical movement patterns such that home/work location methods are not suited. These include:

- shift/night workers;
- workers who work far from home and stay away mid week;
- workers with varying work locations;
- part-time / zero hours workers;
- students who might have similar movement patterns to workers;
- individuals who spend a lot of time at an alternate or varying addresses;
- mobile users who do not generate enough geocoded information (e.g. switch phone off a lot). These are typically removed and weighted as per non-responders in surveys (see age bias as older people tend to have different mobile phone use behaviours);
- Children: it is illegal for an MNO to process children's data for statistical products. In practice this is difficult as many children are on parental contracts or pay as you go subscriptions.

Age bias (NB. 18 years and over only)

- Nearly 30% of people aged 75+ do not have a phone
- 60% of 75+ year olds switch phone off (mainly at night)
- Ofcom suggests that older people are more likely to use PAYG.

MNOs use the age/sex information on contracts within their weighting to official residential population.

They assume that non-registered subscribers, individuals without phones or who do not generate enough information have the same demographic mix. This needs testing.

Figure 1: Number of phones owned by age-group

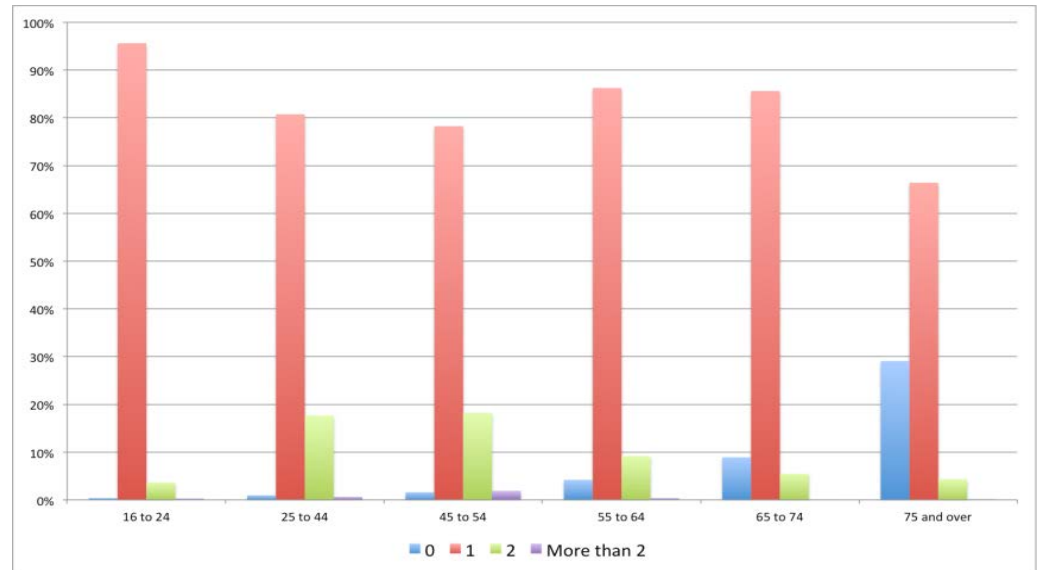
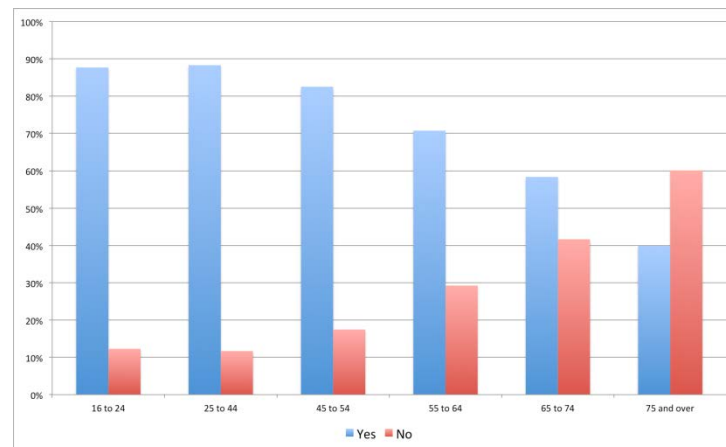


Figure 1: Responses to "Switch On/Off" question by age-group



Gender

There is little suggestion of differences by gender in UK via Opinion Survey 2015 (although there may be on subscriber bases – contracts vs PAYG?). MNOs do use gender within their weighting procedures (if they have legal access to it – NB. This information may not be passed onto intermediate analytical organisations).

Figure 1: Number of phones owned by gender

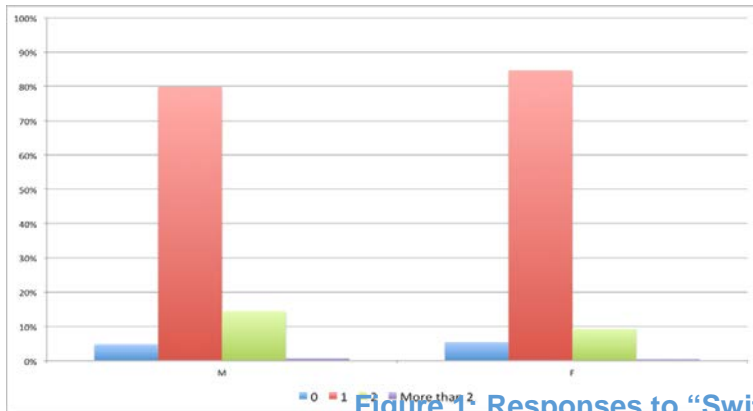
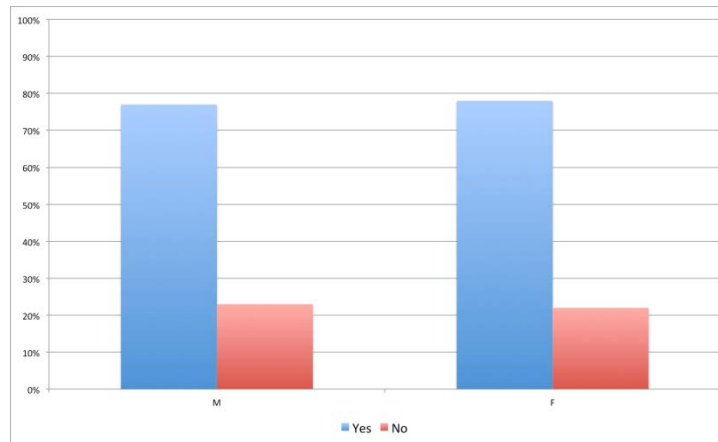


Figure 1: Responses to “Switch On/Off” question by gender



Socio-Economic level

Ofcom consider socio-economic levels to be a main source of bias

- Lower income Socio-Economic groups are less likely to own a smart-phone / have a 3G or 4G subscription and will produce less geocoded information for inference.
- Currently unaware of processing in an MNO to adjust for Socio-economic level although Ofcom survey information may be used.

Ethnicity

Research studies have been conducted – but high ethical concerns

Research (in Estonia) has indicated that it may be possible to use language setting on phone for some ethnic groups (who speak different languages)

Potential to use a tool such as ONOmap to profile names on contracts (can only be done in the MNO due to data protection).

Other studies consider the number of calls overseas as an indicator for nationality.

MNOs do not weight according to ethnicity.

Contracts

Pay As You Go is popular amongst mobile phone users who do not use their phone much, need a simplified package or do not want to give personal data to an MNO. Ofcom research suggests these subscribers include the elderly or children.

In MNO processing, unregistered Pay As You Go subscribers are assumed to have the same age/gender distribution as for contracted subscribers who will have given age and gender. Also, the population who does not have a mobile phone or do not generate sufficient data for analysis are assumed to be the same demographic mix as for contracted subscribers.

Ofcom research reveals that 4G contracts has differential uptake according to socio-economic levels (greater prevalence for people with higher incomes). The location data generated from 4G is much more than for 2/3G and would enable more accurate modelling of movement patterns.

NB. There may be ethical considerations if PAYG subscribers have not opted in to their data being used [REDACTED]

References

- Mobile Phone Report (internal ONS report)
- Literature Review on statistical uses of mobile phone data
<https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseriesno8statisticalusesformobilephonedataliteraturereview>
- Opinion Survey research paper: internal ONS working paper.
- Ofcom Communications Market Reports