

A model of unit non-response for data collection in social surveys

- To set out a central model for an analysis and counter strategy to Non-response.
- Aim 1 to raise response to 2011 levels (or above)
- Aim 2 to understand why response has fallen
- Aim 3 to prepare for alternative solutions
- Aim 4 to demonstrate we are doing all we can to users
- If nothing we can do due to constraints so be it.

Intro

- Not here to comment on current practices, to praise or otherwise.
- Happy to admit I'm probably not aware of all the work we have done so I may say something that is incorrect or already been done. That is fine.
- Did this work due to reasons I will explain later but also:
- Wanted a summer project.
- Wanted to use my OR skills after CPI has rekindled them.
- This may be useful or it may not be.
- But I'm here to get your views , not today but after you have digested this and see if there is any mileage and ways we can take it forward.

Two aspects of non-response

- Adjusting-Data analysis/processing: Sampling errors, Bias, Weights, Imputation
 - Sample design to minimise sampling errors and allow low level disaggregation
 - Non-response weighting or imputation for bias adjustment using auxillary or survey data via linear or logistic regression
 - Robust time series analysis
- Reducing: Data collection: Materials, Interviewers, Calling and Persuasion Strategies

Previous work to date

- RFU to assess bias
- Sensitivity and attrition analysis for bias
- Logistic paradata analysis for bias
- OAC analysis for bias
- Led to push for detailed CNRLS
- Calling patterns analysis-Neil Hoppers work
- ART, ACT, ILR training
- Led to less fall in non-response
- All led to following strategy

Response Strategy

- A framework for:
- 1. Keep questionnaire length and complexity to a minimum. Regularly review the questionnaire. Despite no conclusive evidence that questionnaire length and complexity effect non-response it would seem logical that there is some correlation.
- 2. Use non-response weights and impute missing data where possible.
- 3. **Consider the use of incentives especially during periods of known high non-response rates eg Summer.**
- 4. **Use a mixed calling pattern to contact responders.**
- 5. **Use ART and ACT training of interviewers.**
- 6. **Tailor contact to the responder where possible eg their nationality, socio-economic status etc.**
- 7. **Use paradata if possible to estimate non-response bias and utilize responsive data collection strategies.**
- 8. **Calculate non-response consistently and report regularly.**
- 9. **Conduct regular refusal follow up studies to ascertain reasons for non-response.**
- 10. Give responders a choice of collection modes.

- But response has still fallen so require next steps:

Why is Response Important?

Bias

Sampling error-precision

Time Series Analysis

Small Area/Group Analysis

Weighting

Need an even pattern of response

Partial response is better than no response

Response correlates to the above so is the right measure

Should produce simple leaflet for responders/interviewers

What should the targets be?

Response at the time of the 2011 Census-Why?

CNRLS either confirms:

- No bias
- Derives bias adjustment NR weights

Plus

- EUROSTAT, APS, Bulletin, Other targets acceptable met
- Sampling errors on key estimates meet precision requirements (just)

Response fall

- 75-80% pre 2005 down to 65%
- Why-maybe?
- Integrating LFS and other field forces-1-2%
- 30% now 15% Wave 1 to TO-1-2%
- Interview 1 HMO not all-1%
- Fixed Interviewer CS contracts-1-2%
- Older male interviewers/More migrants/Change of lifestyle-1-2%
- More gated communities; People out; People too busy-1-5% (ART, ACT, Calling patterns kept this up)
- Had to make the above changes for CSR efficiency savings and employment law
- Could have reduced sample size but risk response still would have fallen and allocated sample not representative as too small.
- Other EU countries with voluntary LFS achieve 72-78% W1 rate.

Response Status-BRAGW

- White 80% or above-no action required (never get 100% due to holidays, working abroad, death, moving house-95% then 20% will always decline-ala IPS)
 - White-Green 75-80%-start to monitor
 - Green 65-75%-start counter strategies
 - Green-Amber 60-65%-full scale counters and PT NR team-**Should get this or.....**
 - Amber 55-60%-full time NR team (FRU) and Internet Mode?
 - Amber-Red 50-55%-start to investigate alternative methods eg incentives (interviewer and responder), ring fenced FIF, internet mode, more ethnic interviewers, extend reissue period.
 - Red 45-50%-apply alternative methods
 - Red-Black-35-45%-start to investigate extreme methods eg Sunday collection, compulsory
 - Black under 35% apply extreme methods
-
- Above limits should apply above and below the Survey Response Rate in Census year
 - Overall quarterly targets but evenly weighted by IA and IA targets set. Monthly unweighted target just an indication.
-
- Reason why I have conducted this work is that LFS w1 response rate is now Amber and was approaching Amber-Red
-
- No investment in LFS apart from HMT uplift over 20 years.



Response Status-BRAGW

- If we can get 60-65% pressure is off!
- Red 45-50%-apply alternative methods
 - Sunday Interviewing
 - Continuous £25-100 value incentives for interviewers and/or responders
 - Raise sample sizes
- Red-Black-35-45%-start to investigate extreme methods:
 - Interviewer pay per interview
 - Saturated Advertising in high profile media
 - Compulsory

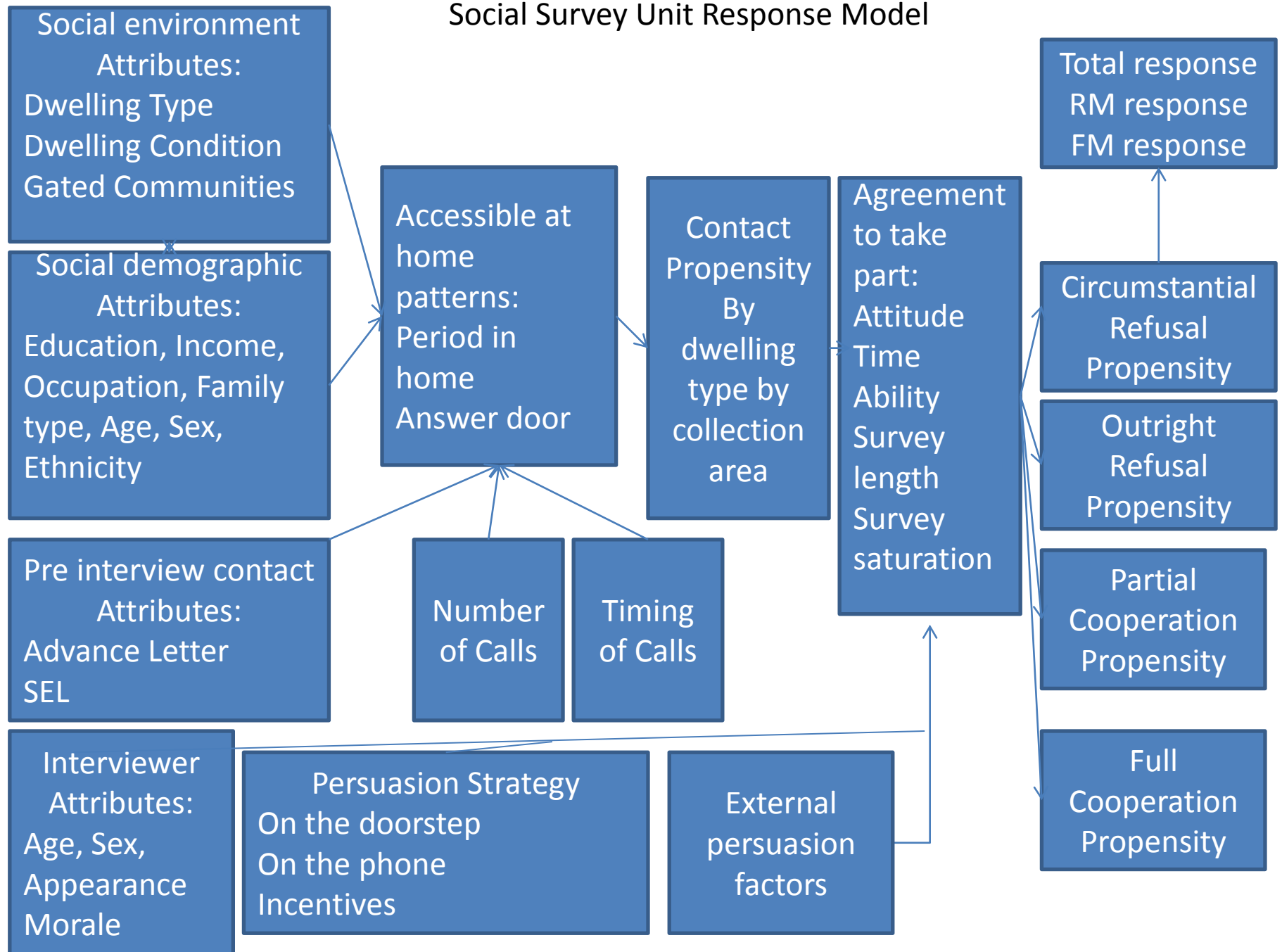
Basis of Work

- Survey Methods in Social Investigation Moser
- Non-response working group pre 2007
- All the work by Heather, Neil and others in MD
- Previous Non-response programme 2007-2010: RFU; Strategy; Calling patterns
- Non-Response on the ESS-2010-references all previous work eg Groves 2006 etc
- Hierarchical Analysis of Unit Nonresponse in Sample Surveys-2012
- Measuring Interviewer Characteristics Pertinent to Social Surveys: A conceptual framework Blom 2013
- Understanding the Respondent Stats Sweden 2013
- International Non-response papers
- CNRLS-2001 and 2011 Census
- Durrant and Steele models
- Black Belt review of the LFS

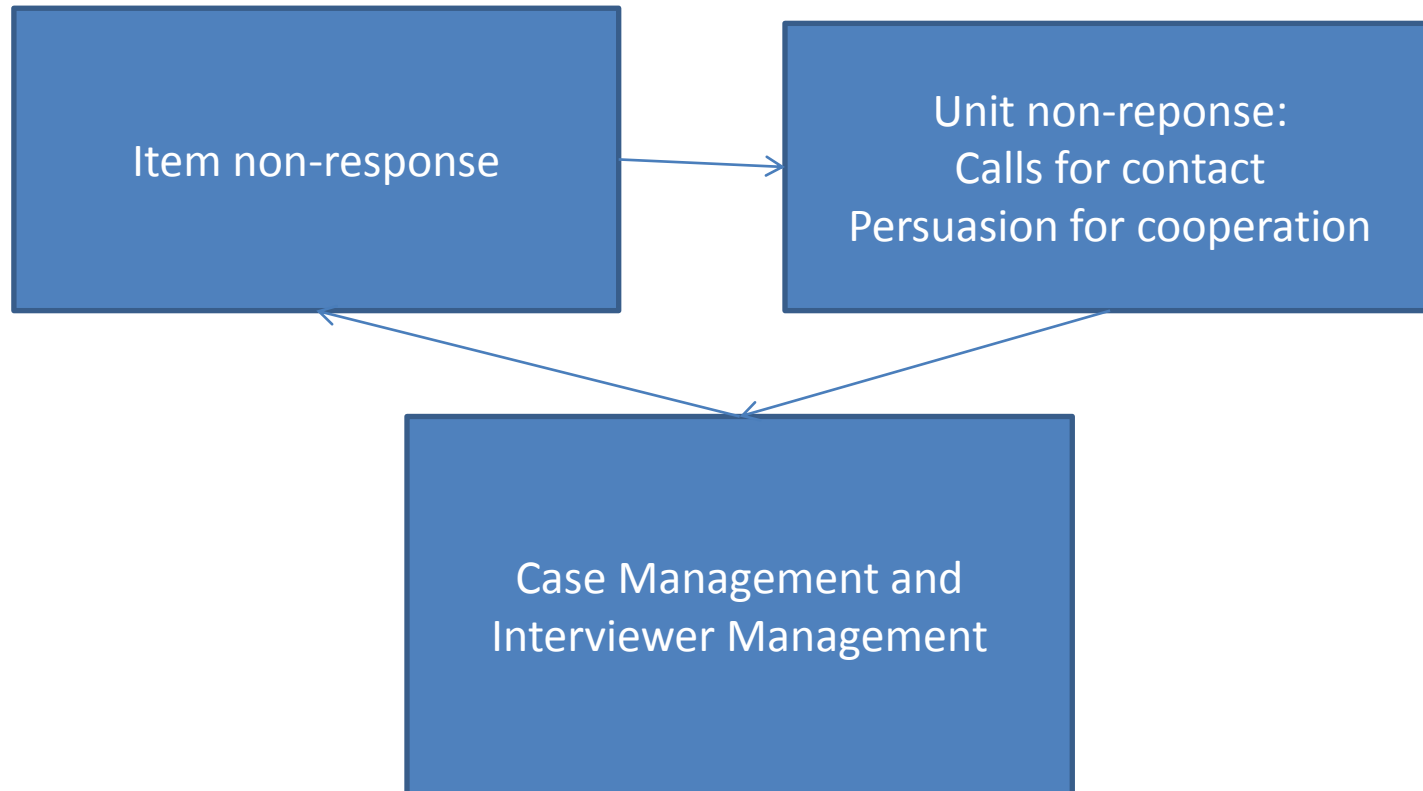
Hierarchical Analysis of Unit Nonresponse in Sample Surveys-2012 Gabriele Durrant /Skinner

- The project team has already been able to publish 6 articles in international academic journals; another manuscript has been conditionally accepted, one is under review and one is in preparation. The findings on nonresponse analysis indicate a systematic correlation between different types of nonresponse and socio-economic and demographic characteristics. (Durrant and Steele, 2009, published in Journal of the Royal Statistical Society, Series A (JRSSA)). A joint analysis of refusal and noncontact, allowing for residual correlation between them, revealed two distinct processes (Steele and Durrant, 2011, in International Statistical Review). The work by Durrant, Groves, Staetsky and Steele (2010, in Public Opinion Quarterly) found that interviewers play a crucial role in gaining response and interviewer confidence and attitudes explain part of the between-interviewer variation. Optimal calling times differ for contact and cooperation and depend on household and interviewer characteristics. Call record data and interviewer observation variables are predictive of nonresponse outcomes at future calls to a household (Durrant, D'Arrigo and Steele, 2011, published in JRSSA; Durrant, D'Arrigo and Steele, 2012, forthcoming in JRSSA). Work on nonresponse bias indicated that response rates alone are not a good guide to data quality and the exploration of different weighting methods led to quantification of response biases in the PISA survey for England (Micklewright, Schnepf and Skinner, 2011, forthcoming in JRSSA). Weighting methods were derived to adjust for clustered nonresponse (Skinner and D'Arrigo, 2011, in Biometrika). The findings have wide ranging implications for survey practice and have informed methods to reduce nonresponse rates and bias.
- Durrant at Southampton-Have we spoken to her-probably?

Social Survey Unit Response Model



A model of data collection



Mathematical model

- Weighted by $1/\text{number of IA's}$ to overall Response Rate
- Binomial at IA aggregate level
- Multinomial at dwelling type level.
- Can be modelled by a probit hierarchical model
- We are less about modelling more about analysing changes. Use the binomial model for the filtering only.
- Once the filters flag a fail then there probably is an underlying cause-change. No need to model this or prove it is significant just need to identify differences and derive strategies to counter the largest ones.
- More detailed modelling may help to identify key causes but small sample sizes will make this problematic.

Models cont

Multinomial:

$$\text{pmf} \quad \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

Grouped to give binomial:

$$\text{pmf} \quad \binom{n}{k} p^k (1-p)^{n-k}$$

Filter 1

You estimate the difference between two population proportions, $p_1 - p_2$, by taking a sample from each population and using the difference of the two sample proportions,

$$\hat{p}_1 - \hat{p}_2,$$

plus or minus a margin of error. The result is called a *confidence interval for the difference of two population proportions*, $p_1 - p_2$.

The formula for a confidence interval (CI) for the difference between two population proportions is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \text{ where } \hat{p}_1$$

and n_1 are the sample proportion and sample size of the first sample, and

$$\hat{p}_2$$

and n_2 are the sample proportion and sample size of the second sample. The value z^* is the appropriate value from the standard normal distribution for your desired confidence level. (Refer to the following table for z^* -values.)

z*-values for Various Confidence Levels	
Confidence Level	z*-value
80%	1.28
90%	1.645 (by convention)
95%	1.96
98%	2.33
99%	2.58

Filter 2

Modified Sign Test

Let $p = \Pr(X > Y)$, and then test the [null hypothesis](#) $H_0: p = 0.50$. In other words, the null hypothesis states that given a [random](#) pair of measurements (x_i, y_i) , then x_i and y_i are equally likely to be larger than the other.

To test the null hypothesis, independent pairs of sample data are collected from the populations $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. Pairs are omitted for which there is no difference so that there is a possibility of a reduced sample of m pairs.^[1]

Then let W be the number of pairs for which $y_i - x_i > 0$. Assuming that H_0 is true, then W follows a [binomial distribution](#) $W \sim b(m, 0.5)$.

1 tail $b(m, 0.66)$

Models Cont

- Why 2 filters?
- To pick up IAs that change by $>10\%$ very quickly but also to pick up IA's that change by $>3.33\%$ over a few years or sooner
- Without picking up too many that haven't changed or only ones that have changed a lot.
- Compared to 2011 averages out seasonal effects; is census year where CNRLS gives bias estimation etc. Interview attitudes survey year.
- Yearly averages for filter 1 and yearly/quarterly for filter 2 provide large enough sample sizes and account for variation in both periods
- So there can be no counter arguments eg 2011 was a 'good year'

Non-response Monitoring and Analysis

- The filter approach-Microanalysis
- Take Callrecords-merge to quarters tabulate week by area by is-inelg is-elg, is-coop, is-noncont, is-refusal in SPSS output to Excel
- Need to merge on other metrics eg int age,

37				37				37				37
38	430	0.697674	0.650943	56	112	0.723214	0.723214	67	122	0.639344	0.650943	56
39				56				55				52
40	264	0.549242	0.470588	31	66	0.530303	0.530303	51	87	0.563218	0.470588	34
41				35				36				26
42	358	0.550279	0.622222	42	87	0.517241	0.517241	32	77	0.597403	0.622222	48
43				45				45				41
44	602	0.609635	0.585987	77	143	0.587413	0.587413	45	126	0.579365	0.585987	76
45				66				81				73
46	432	0.55787	0.52381	58	105	0.542857	0.542857	76	127	0.629921	0.52381	47
47				47				51				58
48	415	0.561446	0.542056	45	103	0.582524	0.582524	54	104	0.442308	0.542056	51
49				58				50				56
50	259	0.57529	0.566667	3	64	0.578125	0.578125	57	59	0.559322	0.566667	4
51				61				2				67
52	227	0.550661	0.614035	59	59	0.525424	0.525424	58	58	0.586207	0.614035	53
53	393	0.541985	0.515464	50	100	0.56	0.56	57	104	0.625	0.515464	52
54				50				47				50
55	403	0.64268	0.653465	55	99	0.575758	0.575758	50	103	0.553398	0.653465	52
56				44				53				49
57	352	0.607955	0.597701	46	87	0.54023	0.54023	48	83	0.614458	0.597701	44
58				41				35				46
59	271	0.586716	0.550725	39	68	0.705882	0.705882	52	94	0.595745	0.550725	31
60				29				42				36
61	398	0.592965	0.646465	51	100	0.56	0.56	27	75	0.613333	0.646465	46
62				49				48				55
63	402	0.614428	0.652632	51	102	0.666667	0.666667	51	96	0.604167	0.652632	48
64				51				45				55
65	320	0.6125	0.573171	38	78	0.602564	0.602564	50	90	0.611111	0.573171	41

77				1 Refusec	year surv	2011	1	2
78						2013	2	1
79				Total			3	3
80								
81				year survey stage	year identifier * area	is_noncon		
82				Count				
83				is_noncon		area		
84						921	922	
85				0 Contact	year surv	2011	8	5
86						2013	4	2
87				Total			12	7
88				1 Not Cor	year surv	2011	0	1
89						2013	1	4
90				Total			1	5
91								
92							4	
93								
94								
95								
96								
97								
98								
99								
100								

Filters

- 3 filters for each IA-the Field Data Tests
- 99% confidence limits for response rate on previous rolling year compared to 2011. So probability of less than $1/100^{\text{th}}$ if no change.
- 4 failures for last 4 quarterly response rates compared to 2011 rate each with a probability of less than $1/3^{\text{rd}}$ if no change.
- Priority order-response rate below 50%

Filter(cont.)

- Filters pick up has where response has dropped and it is unlikely it is by chance.
- Filter 1 should detect changes over 10%.
- Filter 2 should detect changes over 3.33%
- Filter 3 flags IA's that are in the Red

Filters

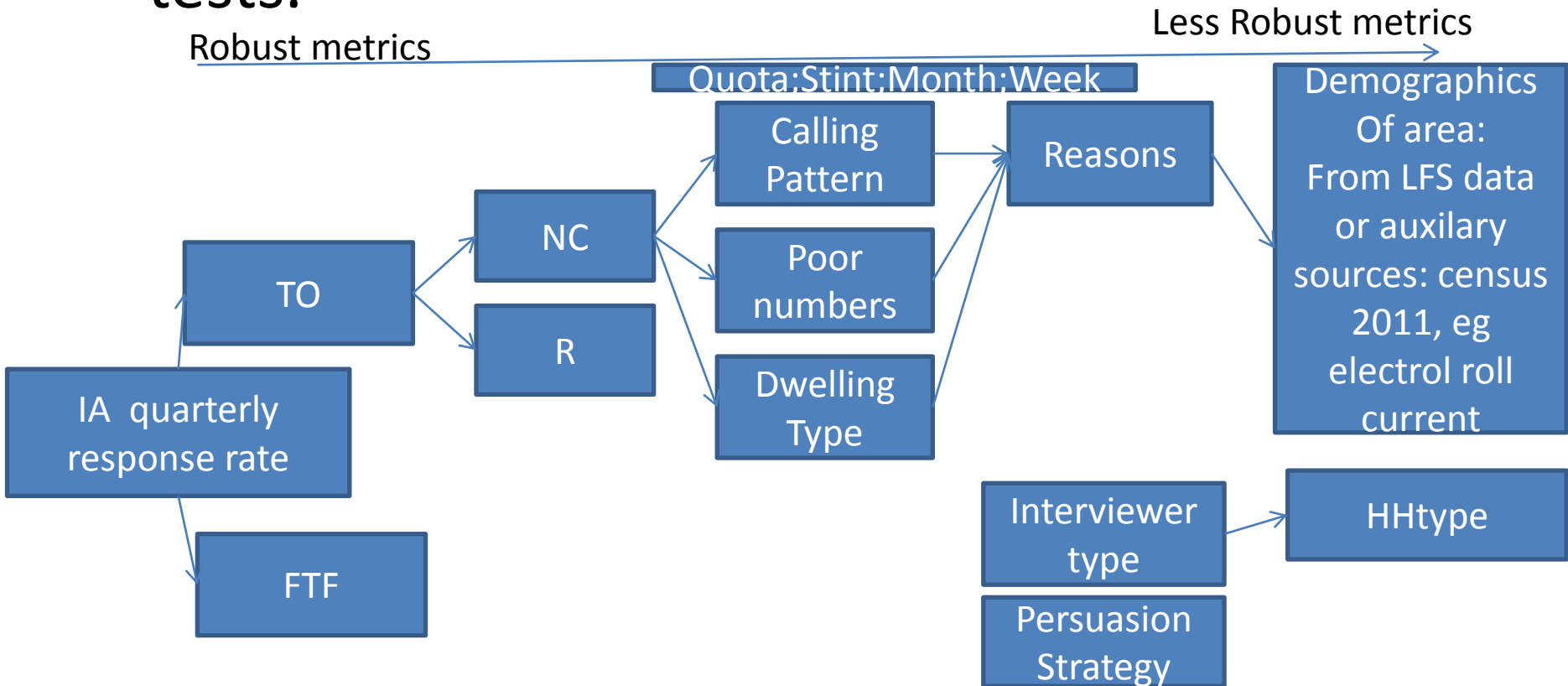
- Filters applied each quarter on a rolling basis.
- If a IA fails a filter then it is investigated.
- Investigation compares the failed quarter/year with 2011-what has changed?
- If there is no change uncovered then it is just chance or we don't have enough information
- Change can also be investigated on the ground by visits

Filters

- Interview areas can still have a common target
- Also this is not to criticise interviewers and not to replace audits and probity checks
- This is to see what has changed and then:
 - Make suggestions to counter the change.
 - Lower the 2011 target and/or reset it.
 - Understand the issues we cannot change unless we implement alternative costly strategies
 - Show to users we are doing all that we can.

Analysis Tree

- Look for changes from 2011-order in size, all significant up to sum that would have failed tests.

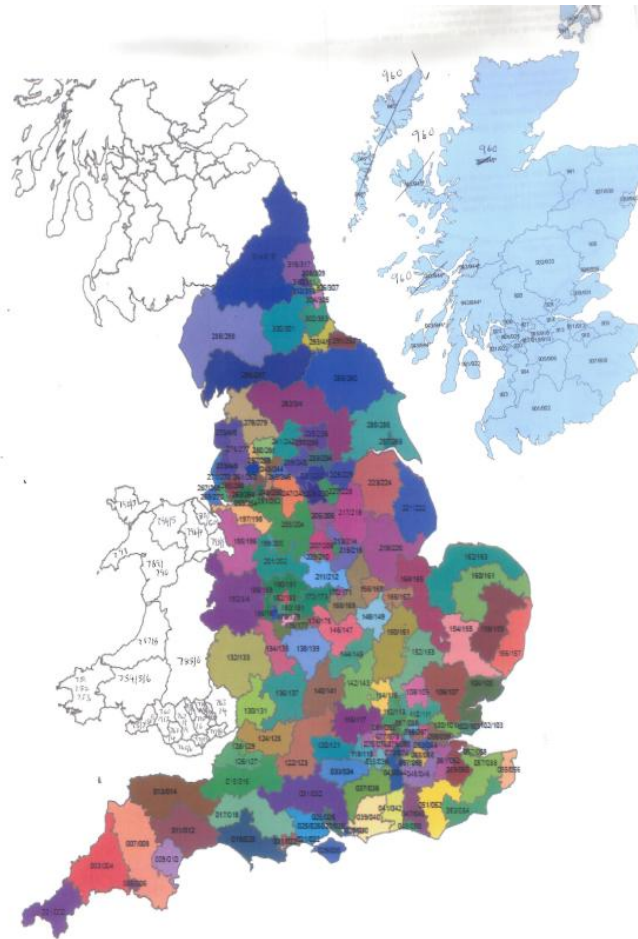


MoE

- Response rate but by IA compared to 2011 (sample size, se, MSE, bias all related to rate)
- Quarterly targets as weekly/monthly too small sizes
- Calling patterns by dwelling type
- Persuasion strategy by hhtype
- Quality of para and survey data
- Dwelling type calling patterns by IA and interview type, persuasion strategy
- Want even response rates across GB. A response is better than no response but full better than partial.
- Key is calling patterns and persuasion
- Link output datasets-callrecords and interviewer records
- Plus other sources eg electoral roll, Census data, google maps/trends


Applied to LFS Wave 1

LFS IA's:



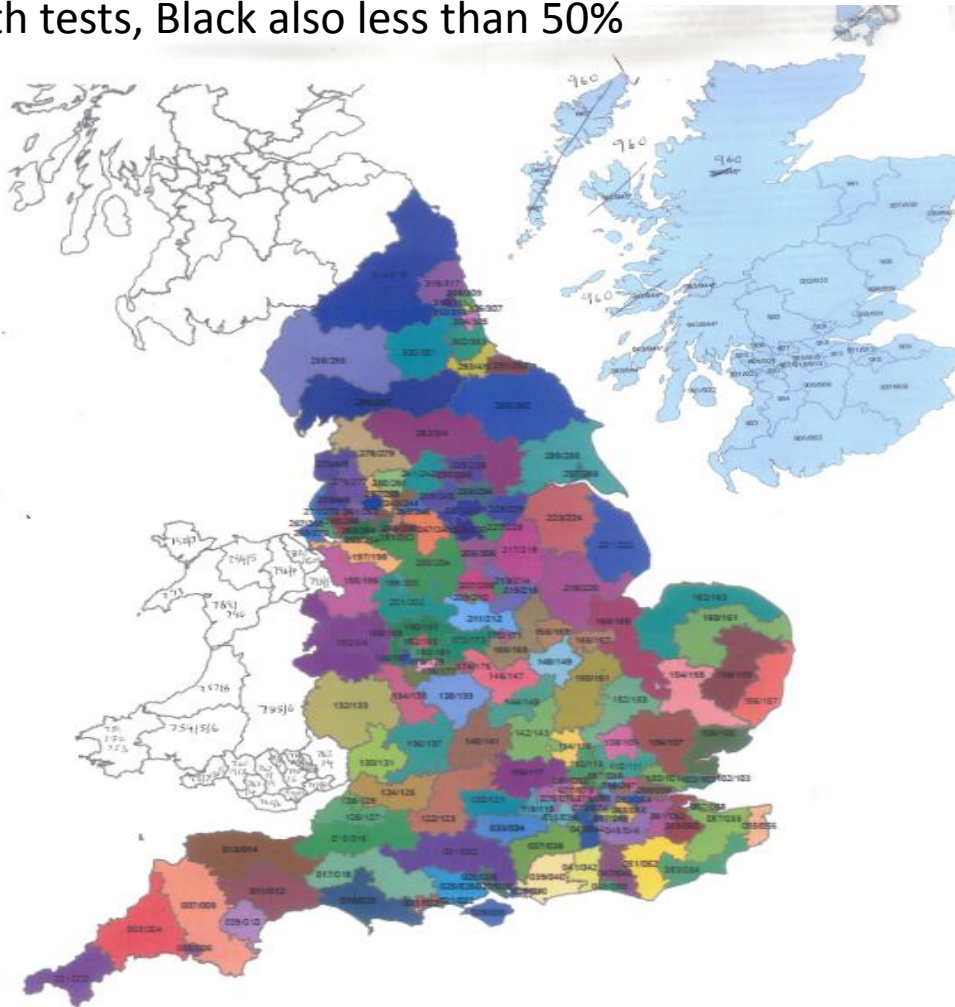
LFS Filters



- 208 IAs Spreadsheet
- 59 fail filter 1 or 2 or both for q2 2013.
Response fall over 10% compared to others
less than 2%.  NRdetailanalysis.xlsx
- 33 drop below 50% by q2 2013
- By chance only 6 should fail.
- Lower level investigation:Spreadsheet

Applied to LFS Wave 1

LFS IA's: amber fail test 2 (3%), Red fail Test 1(10%), Red-Black fail both tests, Black also less than 50%



100/101



- 100/101-Chelmsford compare last quarter with same quarter in 2011
- 101 drops over 10%, both TO and ftf.
- Interviewers-6496 only has 1 out of 9, 1652 has 4 out of 13 and 5413 has 1 out of 5, others OK. 6496 has large number of calls average 4 per household but none before 10am less than average in the evening. 2011 calling pattern more am and evening. If normal response to these 3 then overall face to face response OK
- Dwelling type a mix.
- Change calling patterns
- Else Sunday Follow Up required.

771/772



- 771/772-Cardiff
- Actually Q2 response 2011-2013 similar. Low compared to other quarters so look at q1:
- Big drop in face to face
- 2011 less eligible higher coop.
- Across all interviewers loss of 7227. Refusals much higher 5 to 14. Can't be bothered and personnel problems 0 to 8. More blank reasons.
- Most afternoon call times. 1 HQ refusal. Mix of dwellings.
- Some interviewers travelling from Swansea etc no calls before 10 am.

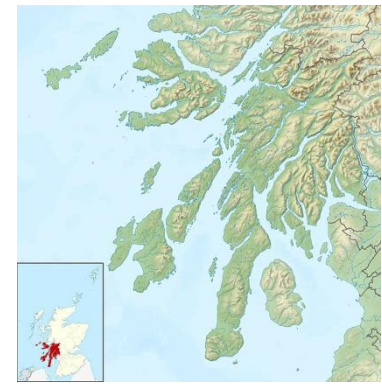
771/772

- Further drill down
- More face to face up 7 down 7 in TO
- TO coop from 6 to 1. Mainly in 771 in refusals and non-contacts
- Can't be bothered across North and South Cardiff
- Look at Dwelling type-semi, terrace, flats
- More retired people in 2011
- Refusal follow up required-use video and hooks

771/772

- Q3 2013 now available
- NB 16 different failures, 10 drop outs, 6 swaps, 18 both fail, 16 test 1, 19 test 2 only.
- 771/772 a drop out why?
- 2011 response low due to low Q2 in 772-only 47%.
- 2013 Q1-high refusals, Q2-non-contacts, Q3-bit of both but less than before so now OK within limits. Drop in people being too busy. Calls made after 9 pm. Double pre 10 am calls. New interviewers calling earlier and later.
- No action but monitor.

911/922



- 921/922-Isle of Arran
- Big drop in telephone cases especially in 922 but also 921.
- If response same as in TO 2011 then overall response is the same. One quarter response fell to 43%.
- 922 non-contacts especially rose. O numbers. But also no calls before 10am in 2013.
- Change calling patterns
- Else Sunday Follow Up required.

215/216



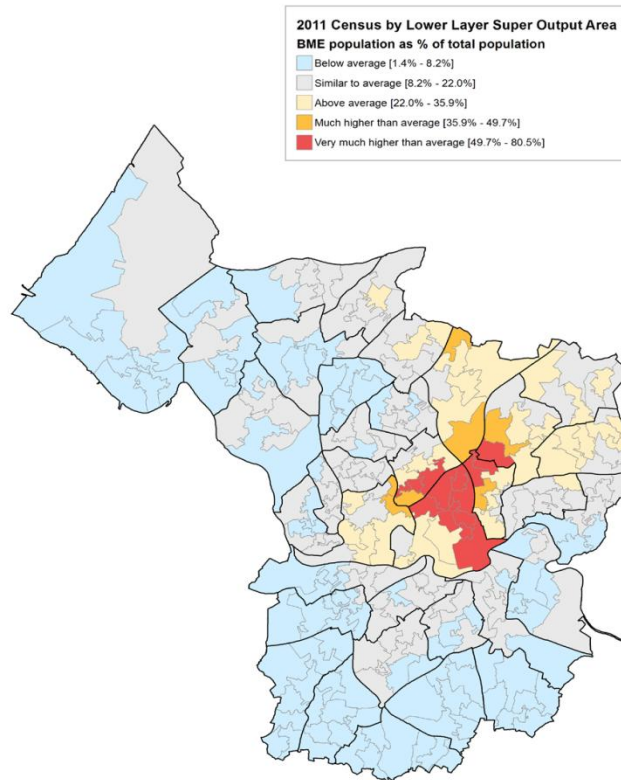
- 215/216-Rushcliffe etc by Nottingham
- Fall in face to face in both 215 and 216 especially 215 14% fall. Much higher not contacted in 215 in 2013. Again no calls before 10 and none after 8 pm in 2013 but many more calls-258 compared to 125.
- Earlier calling patterns
- Else Sunday Follow up required.

128/129



- 128/129-Bristol
- Both ftf and Tel rates have crashed in both areas ftf 80/63% to 50/47%, tel 58/69% to 27/28%.
- Much higher refusal rate than in 2011. Much higher too busy 1-6 and other 0-4.
- Why? What are the demographics. Too busy 3 interviewers. No early calls but late calls. A lot of flats. All bar 1 in South Bristol near Ed's.
- Other most not answer door. All N Bristol. 2 interviewers Houses.
- All experienced interviewers but again no early contacts.

128/129



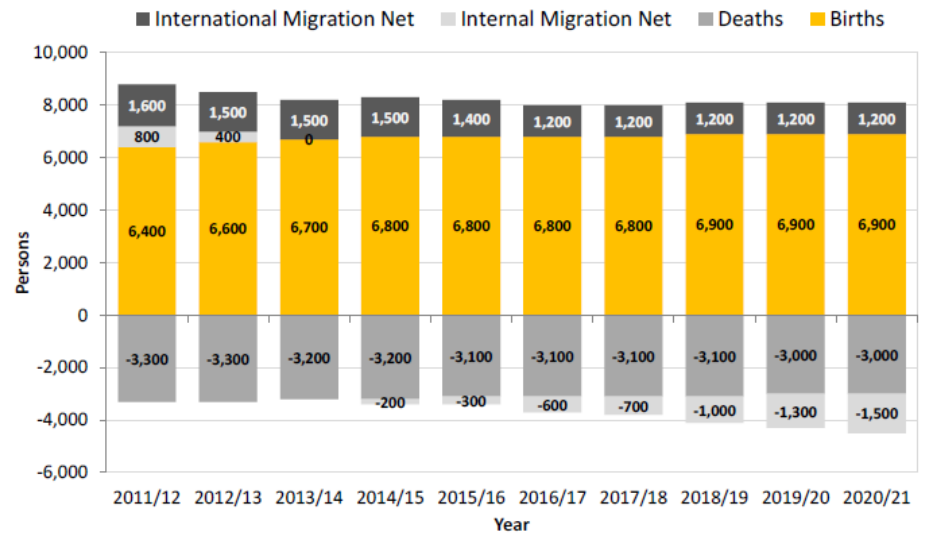
Produced by Strategic Planning, Bristol City Council
Source: Office for National Statistics © Crown Copyright 2013.
© Crown Copyright and database rights 2013. Ordnance Survey 100023406.

Students

5.22 The number of full time students aged 18 and over living in Bristol during term time has increased by just over 10,000 from 25,573 to 35,638. Students now make up 8.3% of the total population of Bristol.

Figure 14. Projected components of population change for Bristol 2011-2021

Source: Interim 2011-based Sub-national Population Projections, Office for National Statistics



128/129

- Further drill down
- 5474 and 6631 made 200 more calls. Had a lot more cases. 5474 only 43% success rate. 5416 both in 2011 and 2013 only 33% success rate. As opposed to 6693 and 7174 in 2011, 100% and 83% success rate respectively with large allocations.
- Investigate 5474 and 5416 persuasion strategy- a lot of can't be bothered/doesn't believe
- 5474 Calling patterns OK, 5416 no calls before 10.50 am.
- Employment has dropped in Bristol 128/129 down 63% to 59%
- Non UK up from 8 to 12%. UK down in numbers.
- After above possible refusal follow up required-use video and hooks

128/129

- 2013 Q3 up to 62%, still fails test1
- Why is it up? Looks as if 128 non-contact and refusals are down so response up to 67%. 129 also higher but q3 2011 quite low.
- Less too busy. Several different interviewers between Q2 and Q3. 6901/2/3 have good cooperation rates. No refusal reason of too busy. 1 non-contact out of 41 cases and 9 refusals! Rest 49 cases 10 non-contacts and 13 refusals.
- Monitor 6901/2/3 and others-more training others?
- Linked FMIS: 6901 and 2 women?
- Linked survey data: Many more 1 person hh in Q3 coop.

Recommendations

- Visit area-go out with interviewer, assess area characteristics use auxiliary data eg electric usage
- Change calling patterns
- More persuasion training
- Assign interviewer type to dwelling type by area
- Sunday analysis
- Refusal analysis-incentives, compulsory, int training

Outputs

- FRU-RO/HEO
- 4 reports per month
- Overseen bt FRT, RT and SSPG
- 1 report per 6-12 months with overview and longer term recommendations
- Changes to calling patterns
- Changes to persuasion strategies in IA's
- Possible strategies tailored to dwelling type and IA.

The Generic Approach-Macroanalysis

- To use feedback from the field, deductive reasoning and work by external bodies to investigate generic strategies to combat non-response.
- To analyse non-response at the national, regional, UALAD and OAC level.
- Interview attitude surveys
- Logistic regression and other MD work
- This approach is still required.
- Advanced Letter analysis
- Improved MIS reporting
- Review of incentives
- Video
- Keep questionnaires to current length
- Suggest persuasion hooks

Additional work

- Additional work
- Follow up conversions
- APS and other surveys-any correlation with response and LFS IA
- Attrition analysis
- Paradata use in TO and question review

Combined approach

- To monitor field work using the filters for indepth analysis with a dedicated FRU.
- To continue the generic approach.
- To monitor the generic approach against the IA's that have failed/passed the filters.
- Plus:
- Liaise with non-response experts and other research
- Review literature, maintain NR document domain

The End

- Concluding remarks. The difference between this and other strategies
- A defined model of non-response for LFS/SS
- A combined strategy of detailed micro analysis and strategy with generic macro analysis and strategy
- Clear targets and a base period to compare against