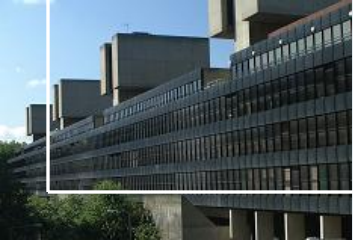




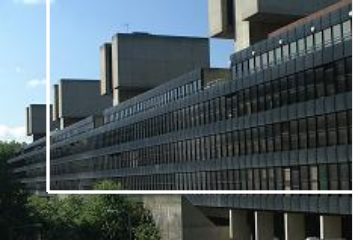
The BCS70 sample, attrition and non-response

Tarek Mostafa
Quantitative Social Science
t.mostafa@ioe.ac.uk



Outline

- Introduction – The BCS70 sample, attrition and non response.
- Plan
 - Problems of attrition in collecting longitudinal data: the case of BCS70.
 - What do we know about non-response/attrition?
 - What can we do about it?



British Cohort Study 1970

- BCS70 follows the lives of 17,000 people born in a single week in April 1970.
- Individuals were surveyed at birth and at age 5, 10, 16, 26, 30, 34, 38 and 42.
- BCS70 collects data on health, physical, educational and social development, and economic circumstances among other factors.



- Collaboration between study team and NHS – hospitals and maternity units in GB (England, Scotland and Wales)
- Geographically distributed as GB population
- No over-sampling and hence small numbers of those from ethnic minorities - supplemented by 'immigrants' identified in schools during tracing for surveys to age 16
- No weights provided to facilitate analysis



Attrition

Definition

- Attrition is the discontinued participation of some individuals in a longitudinal survey for reasons that are unknown and/or beyond the control of the researcher



Problems of non-response/attrition

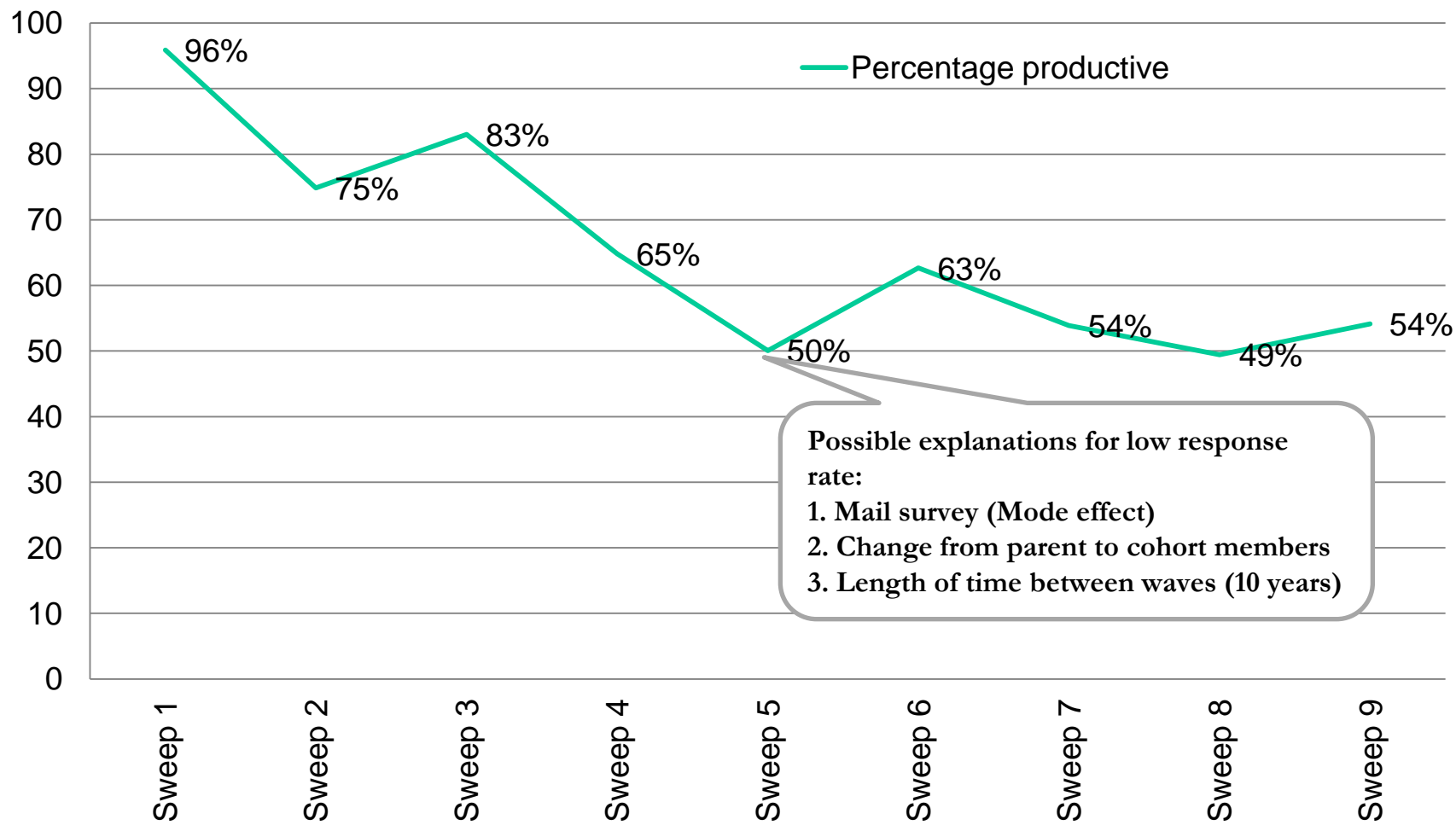
- Distinction between unit (respondents') non response and item non-response (focus on former here)
- Types of non-response (have separate reasons)
 - Non-contact Refusal Inability
 - Out of scope/ineligible
- Non-response on increase in all surveys
- Non-response may not be permanent (in panel survey)
- Effects of non-response/attrition



Response in BCS70

Response categories	Sweep 1	Sweep 2	Sweep 3	Sweep 4	Sweep 5	Sweep 6	Sweep 7	Sweep 8	Sweep 9
Participated	16,569	12,939	14,349	11,206	8,654	10,833	9,316	8,545	9,354
Contact later	715	2,899	1,277	4,004	5,979	1,791	1,497	0	0
Dead	0	565	585	597	697	748	795	824	853
No contact later	0	837	1,071	1,477	1,723	1,792	3,133	0	0
Temporary Emigrant	0	42	0	0	49	37	62	23	2,579
Permanent Emigrant	0	0	0	0	41	250	344	393	0
Refusal	0	2	0	0	141	1,292	1,334	687	145
Unproductive	0	0	0	0	0	541	803	6,812	4,353
Total	17,284	17,284	17,284	17,284	17,284	17,284	17,284	17,284	17,284

Response in BCS70

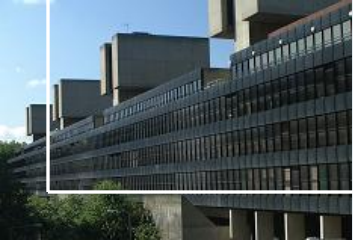




Patterns of response

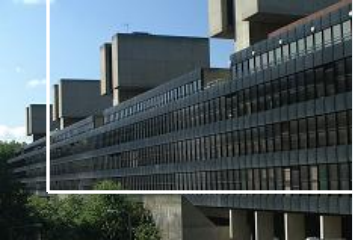
Pattern	Type of non-response
111111111	Participated in all sweeps
111111110	Monotone response
111100000	Monotone response
111101111	Non-monotone response
100000011	Non-monotone response

Pattern	Frequency	Percentage
Monotone	4,716	27
Non monotone	9,153	52.95
Participated in all sweeps	3,423	19.8
Total	17,284	100



Effects of non-response/attrition

- Biases in results (all surveys)
 - Disproportionate to some groups (mobile, disadvantaged, young, men, long working hours)
 - Problem if linked to survey topic focus/variables
- Missing data
 - lower cell sizes - fewer transitions;
 - incomplete histories



Ways to detect non-representativeness

- Compare the characteristics of the sample in any sweep to the same characteristics at birth. if sample composition is changing => Bias.
- Use binary regression models (i.e. logit and probit models) to measure the association between these characteristics and the likelihood of response.



Non-response bias

$$\text{Population mean} = W_R (\text{Respondent Mean}) + W_{NR} (\text{Non-respondent mean})$$

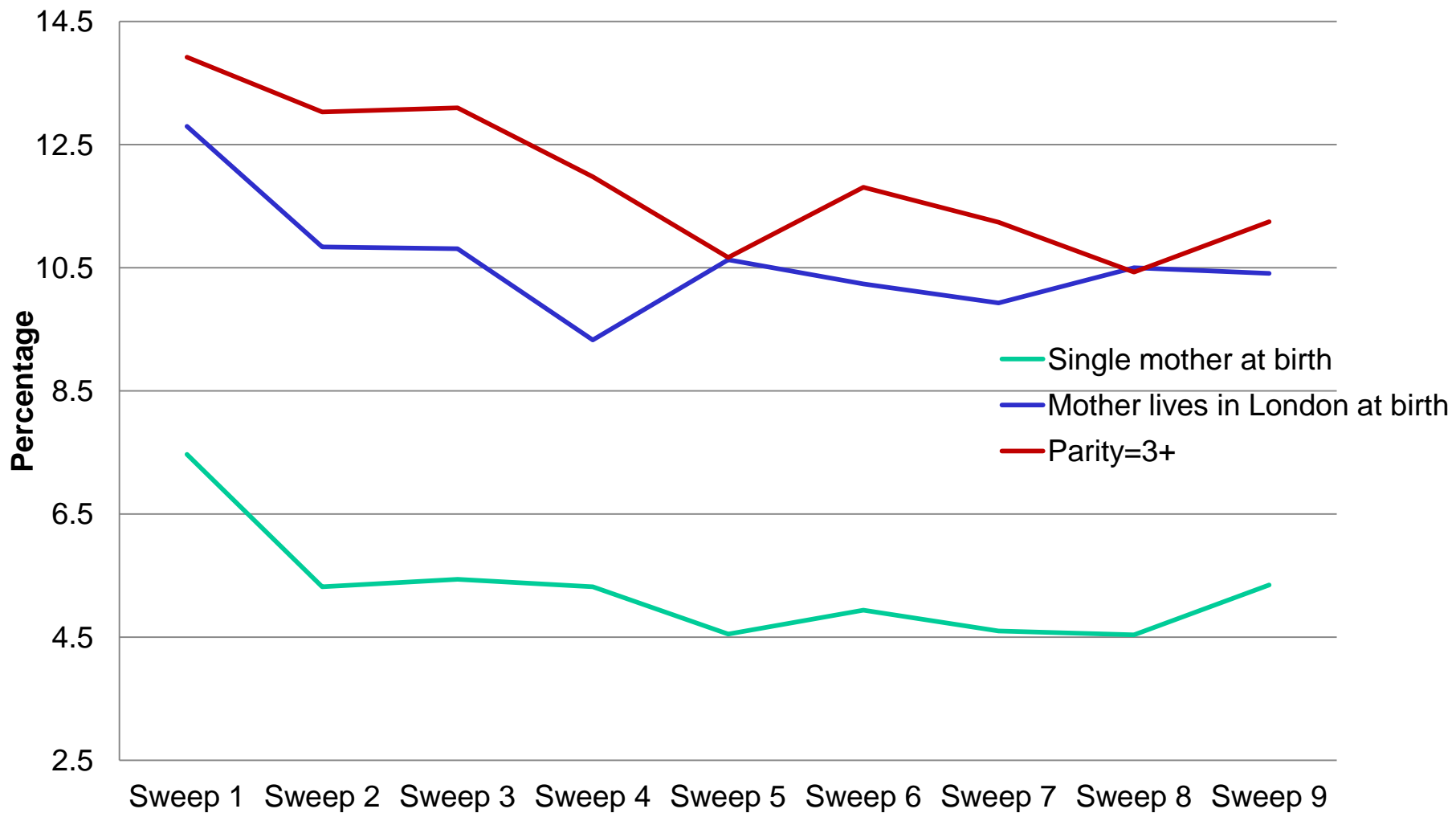
Where W_R = Proportion of respondents

And W_{NR} = Proportion of non-respondents

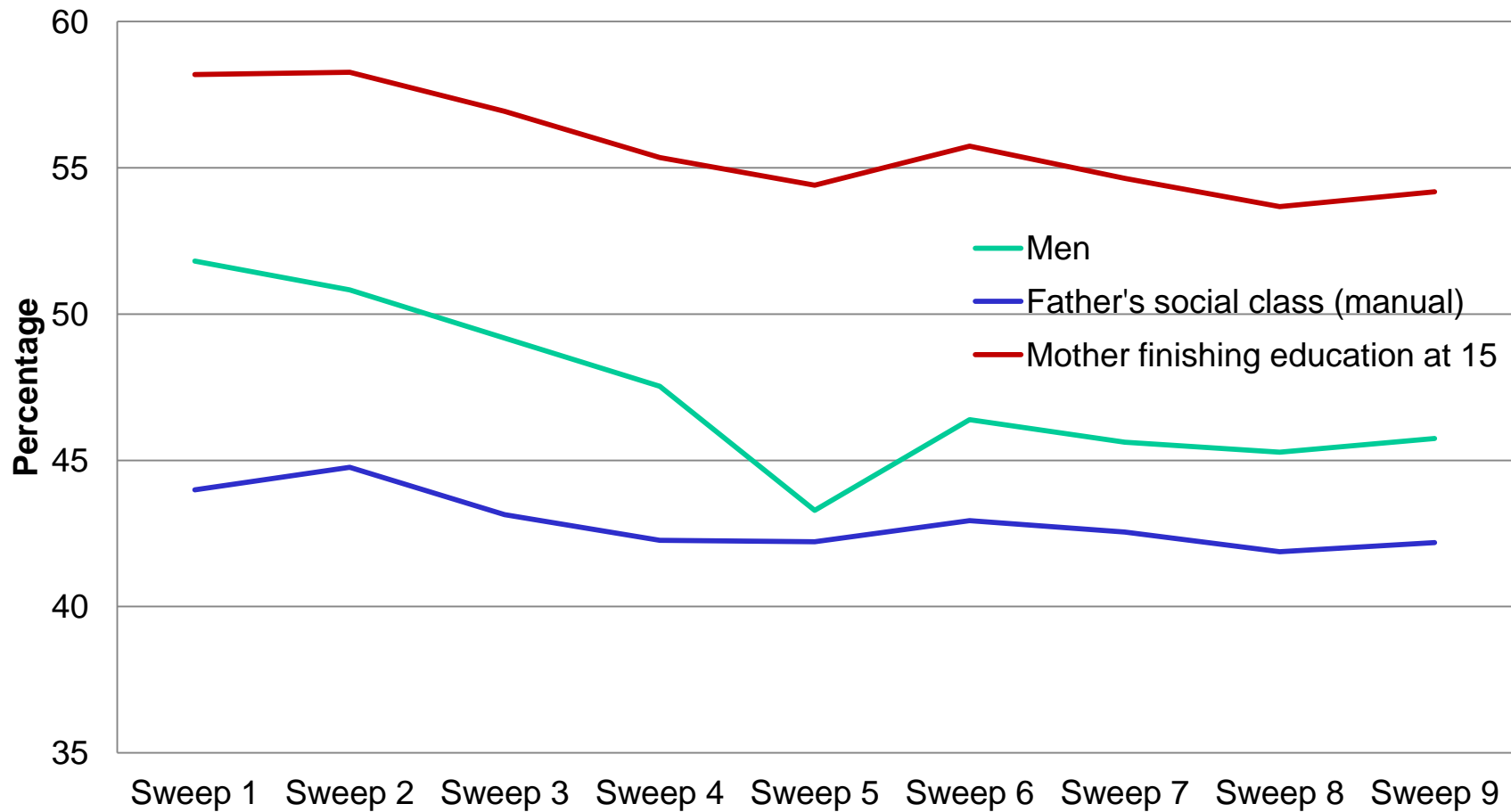
$$\text{So, } W_R + W_{NR} = 1$$

Define non-response bias ?

Sample composition over time



Sample composition over time



Modelling response in BCS70

	Sweep 2	Sweep 3	Sweep 4	Sweep 5	Sweep 6	Sweep 7	Sweep 8	Sweep 9
Gender (reference: Men)								
Women	1.00 (0.040)	1.08 (0.049)	1.26*** (0.044)	1.80*** (0.060)	1.49*** (0.052)	1.48*** (0.049)	1.48*** (0.049)	1.44*** (0.048)
Marital status (reference: Single)								
Married	1.47*** (0.140)	2.18*** (0.218)	1.67*** (0.151)	1.85*** (0.174)	1.89*** (0.171)	1.89*** (0.174)	1.79*** (0.169)	1.42*** (0.128)
Mother lives in London in 1970 (reference: not in London)								
In London	0.57*** (0.032)	0.55*** (0.034)	0.47*** (0.024)	0.71*** (0.037)	0.61*** (0.031)	0.62*** (0.032)	0.70*** (0.036)	0.67*** (0.034)
Parity (reference: 0)								
1	0.97 (0.050)	1.02 (0.059)	0.87** (0.039)	0.92 (0.039)	0.94 (0.042)	0.89** (0.038)	0.93 (0.039)	0.92* (0.039)
2	0.82** (0.053)	0.89 (0.065)	0.81*** (0.046)	0.79*** (0.042)	0.84** (0.047)	0.74*** (0.040)	0.75*** (0.040)	0.81*** (0.044)
3+	0.72*** (0.053)	0.90 (0.076)	0.70*** (0.045)	0.58*** (0.036)	0.65*** (0.041)	0.58*** (0.036)	0.54*** (0.033)	0.61*** (0.038)

Modelling response in BCS70

	Sweep 2	Sweep 3	Sweep 4	Sweep 5	Sweep 6	Sweep 7	Sweep 8	Sweep 9
Breastfeeding (reference: attempted)								
Not attempted	0.82*** (0.036)	0.84*** (0.041)	0.85*** (0.032)	0.85*** (0.031)	0.92* (0.034)	0.87*** (0.031)	0.87*** (0.031)	0.80*** (0.029)
Mother's age at Delivery (reference: less than 20)								
[20-24]	1.42*** (0.105)	1.17 (0.098)	1.20** (0.080)	1.31*** (0.085)	1.23** (0.080)	1.33*** (0.085)	1.28*** (0.083)	1.26*** (0.081)
[25-29]	1.51*** (0.121)	1.27** (0.115)	1.28*** (0.092)	1.46*** (0.102)	1.35*** (0.096)	1.50*** (0.103)	1.45*** (0.101)	1.35*** (0.093)
[30-34]	1.63*** (0.151)	1.36** (0.143)	1.30** (0.106)	1.62*** (0.129)	1.44*** (0.117)	1.66*** (0.131)	1.59*** (0.125)	1.39*** (0.109)
35 or more	1.81*** (0.204)	1.56*** (0.198)	1.40*** (0.140)	1.69*** (0.164)	1.51*** (0.149)	1.81*** (0.175)	1.73*** (0.167)	1.45*** (0.139)
Mother's age at completion of education (reference: 14 or less)								
15	1.56*** (0.141)	1.81*** (0.179)	1.29** (0.106)	1.38*** (0.114)	1.20* (0.098)	1.32*** (0.107)	1.15 (0.094)	1.04 (0.084)
16	1.63*** (0.164)	1.73*** (0.190)	1.50*** (0.137)	1.50*** (0.135)	1.37*** (0.124)	1.51*** (0.134)	1.34*** (0.119)	1.21* (0.107)
17	1.47*** (0.172)	1.42** (0.180)	1.32** (0.138)	1.56*** (0.160)	1.26* (0.131)	1.45*** (0.148)	1.32** (0.134)	1.18 (0.120)
18 or more	1.31* (0.147)	1.34* (0.164)	1.30* (0.133)	1.48*** (0.149)	1.14 (0.116)	1.33** (0.134)	1.24* (0.124)	1.05 (0.105)

Modelling response in BCS70

	Sweep 2	Sweep 3	Sweep 4	Sweep 5	Sweep 6	Sweep 7	Sweep 8	Sweep 9
Father's social class (reference: SC 1)								
Professional	0.94 (0.102)	0.98 (0.116)	0.85 (0.084)	0.94 (0.087)	0.93 (0.090)	0.99 (0.092)	0.95 (0.088)	0.97 (0.090)
Clerical, non-manual	1.06 (0.122)	1.20 (0.151)	1.04 (0.107)	1.07 (0.102)	1.10 (0.111)	1.13 (0.108)	0.99 (0.094)	1.00 (0.095)
Skilled manual	0.90 (0.097)	0.94 (0.109)	0.79* (0.076)	0.79* (0.071)	0.83* (0.078)	0.84 (0.076)	0.74*** (0.067)	0.77** (0.070)
Unskilled manual	0.87 (0.101)	0.85 (0.108)	0.75** (0.079)	0.70*** (0.068)	0.76** (0.077)	0.75** (0.073)	0.68*** (0.066)	0.69*** (0.068)
Lowest grade workers	0.70** (0.091)	0.77 (0.111)	0.69** (0.081)	0.56*** (0.063)	0.64*** (0.074)	0.65*** (0.072)	0.56*** (0.063)	0.59*** (0.065)
Other	0.34*** (0.044)	0.60*** (0.085)	0.70** (0.086)	0.70** (0.082)	0.69** (0.083)	0.76* (0.088)	0.65*** (0.075)	0.70** (0.081)

Modelling response in BCS70

	Sweep 2	Sweep 3	Sweep 4	Sweep 5	Sweep 6	Sweep 7	Sweep 8	Sweep 9
Father's age at completion of education (reference: 14 or less)								
15	1.20*	1.24*	1.11	1.02	1.19*	1.03	1.11	1.03
	(0.102)	(0.119)	(0.083)	(0.076)	(0.089)	(0.076)	(0.082)	(0.076)
16	1.09	1.00	1.14	1.07	1.13	1.00	1.10	0.99
	(0.107)	(0.108)	(0.098)	(0.090)	(0.096)	(0.084)	(0.092)	(0.082)
17	0.92	1.04	1.25*	1.21	1.27*	1.10	1.29*	1.08
	(0.107)	(0.136)	(0.131)	(0.122)	(0.132)	(0.111)	(0.130)	(0.108)
18 or more	0.79*	0.82	0.98	0.96	1.05	1.00	1.06	0.92
	(0.083)	(0.094)	(0.092)	(0.088)	(0.097)	(0.091)	(0.097)	(0.083)
<i>N</i>	15270	15270	15270	15270	15270	15270	15270	15270
pseudo <i>R</i> ²	0.036	0.034	0.026	0.040	0.028	0.031	0.033	0.025



Ways to deal with non-response

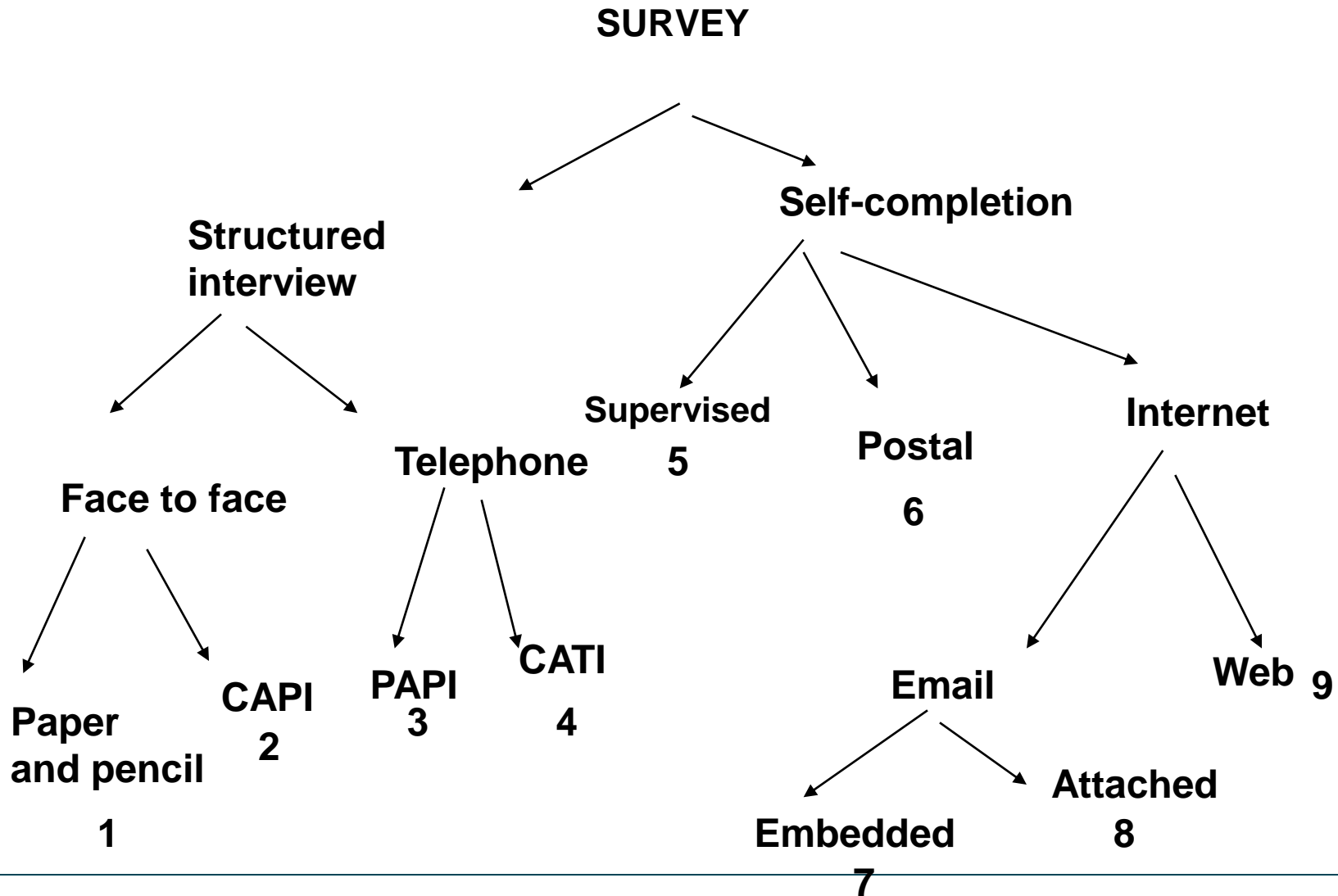
- Two ways:
- Avoid attrition from the outset.
- Adjust for unit non-response bias when it happens.



Avoid attrition from the outset.

- Survey mode (face to face, telephone, postal).
- Interviewee characteristics. Inability: (Language-ethnicity/immigration; disabilities).
- Interviewers characteristics and experience.
- Cost of survey.
- Non-contact: (Flats, gatekeepers, assisted living; answer machines; caller ID, mobile phones; time of day).
- Non-cooperation-refusal. (people too busy).

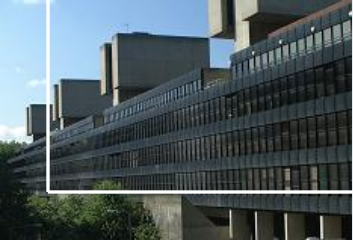
Mode Effects.





Impact on Costs per Case

- Costs of surveys are on the rise.
- More extensive use of advance letters (especially in telephone surveys where they weren't used before).
- More extensive use of incentives (revisit issue of cost impact).
- More callbacks
 - Steeh et al. (2001) present evidence that Michigan's Survey of Consumer Attitudes used to take around 6 calls per complete (mid-1990's).
 - By 1999, it took 12 calls on average.



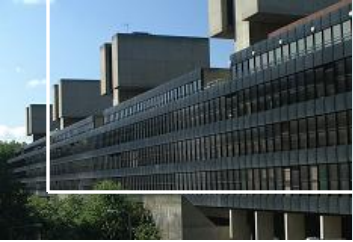
What can we do? Analysis solutions

- Weighting to adjust for unit non-response.
- Imputing by modelling is possible for some missing data.
- In BCS70 the solution depends on the analysis of interest.



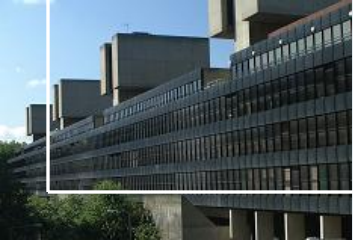
Weights

- Unit non-response weights are constructed using Logit models of binary response outcomes. Outcomes in wave M are modelled using respondents characteristics from previous waves.
- Weights are constructed as inverse probability weights.



References

- Singer, E. (2006). *Non-response bias in household surveys*. Special Issue on non-response. *Public Opinion Quarterly* 70(5), 637-645.
- De Leeuw, E & de Wim. Trends in household survey non-response: a longitudinal and international comparison. In Groves et al., (eds). Survey Non-response. Wiley, New York.



References

- PLEWIS, I. (2007) Non-Response in a Birth Cohort Study: The Case of the Millennium Cohort Study. [International Journal of Social Research Methodology](#), 10(5), 325-334.
- PLEWIS, I, CALDERWOOD, L, HAWKES, D and NATHAN, G. (2004) [Changes in the NCDS and BCS70 populations and samples over time](#). CLS Technical Report. London: CLS, Institute of Education.
- For further reports/publications – see: www.cls.ioe.ac.uk/publications