Is the social origin pay gap bigger than we thought?

Understanding and mitigating missing class data in social surveys

Kristinn Hermannsson, Michael Vallely and Jeanette Findlay

University of Glasgow,

School of Education and Adam Smith Business School

24/11/2020

Outline

1. Social origin pay gaps

2. What is social origin and how is it operationalised in surveys?

3. Is lack of class background information non-random?

4. What does that mean for estimates of class pay gaps?

5. Who are these people and what do missing values suggest for the practice of collecting social origin information in social surveys?

Context and relevance

- Increasing income and wealth inequality in high income countries from late 20th century onwards (Picketty & Saez)
- Well established that the family you're born into matters for:
  - Educational attainment
    - HE disproportionately middle class
    - Low attainment disproportionately in particular localities, white working class and some ethnic minorities
  - Occupational attainment
    - Professional and creative occupations don't reflect population at large
    - Reproduction common, e.g. medicine law

- Policy consensus: improve attainment of disadvantaged groups to solve occupational mobility and income inequality

Unexplained pay gap identified using several UK datasets

- HESA graduate follow up survey
(Crawford & Vignoles 2014)
- BCS70 Cohort study
(Crawford & van der Erve, 2015)
- Matched university and tax data
(Britton et al, 2016, 2019)
- Labour Force Survey
(Laurison & Friedman, 2016, 2020)

- Unexplained pay gap analogous to gender pay gap
- Allows for observable characteristics like education
- Hence challenge to policy consensus(see Duta & Ianelli, 2018 for a discussion)

National Statistics Socio Economic Classification (SEC)

- Adopted by the Office for National Statistics from the 2011 census onwards
- Rests on a theoretical argument from sociology that those who share similar occupational conditions share similar material and social conditions – can be identified as a social class
- Competing approaches, but none as convenient to implement in surveys
- Uses the occupational status of previous generation to proxy social origin
  - Specifically, households main earner when respondent was age 14
- Weakness
  - 1) Information required from a sequence of questions to derive variable
  - 2) Occupational framework may not capture reality of respondents
- For details see: https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010

Data

- Quarterly Labour Force Survey

- Pool 6 waves 2014-2019
  - Use 3rd quarter as this contains social origin info

- Include individuals:
  - Earning wage income
  - Over 16
  - Reporting social origin (SOC code of parent when respondent was 14 SMSOC101 in present wave (not data brought forward)

Final response rate in achieved sample for social origin variable

Occupation of main wage earner when respondent was 14 years old (Major)	Freq.	Percent	Cum.
Does not apply	33,322	10.79	10.79
No answer	1,642	0.53	11.32
Higher managerial and professional occu	34,769	11.26	22.58
Lower managerial and professional occup	44,237	14.33	36.91
Intermediate occupations	24,956	8.08	44.99
Small employers and own account workers	15,605	5.05	50.04
Lower supervisory and technical occupat	66,949	21.68	71.73
Semi-routine occupations	9,182	2.97	74.70
Routine occupations	10,249	3.32	78.02
Never worked and long-term unemployed	37,175	12.04	90.06
Not classified	30,703	9.94	100.00
Total	308,789	100.00	

Observed traits of respondents and non-respondents

Observed characteristics	Social origin reported (n=243,122)	No answer (n=1,642)	Does not apply (n=33,322)	Not classified (n=30,703)
Age	45.01	45.17	41.95	47.66
Sex	1.52	1.52	1.52	1.53
Occupational destination - NS-SEC 1	14.16%	13.52%	7.54%	7.28%
Occupational destination - NS-SEC 2	23.94%	22.47%	15.85%	16.73%
Hourly pay	£14.85	£14.92	£12.12	£12.06
Hourly pay by Post-graduate	£21.44	£19.90	£19.77	£19.22
Hourly pay by Graduate	£18.36	£18.50	£15.91	£16.00

Education of respondents and non-respondents
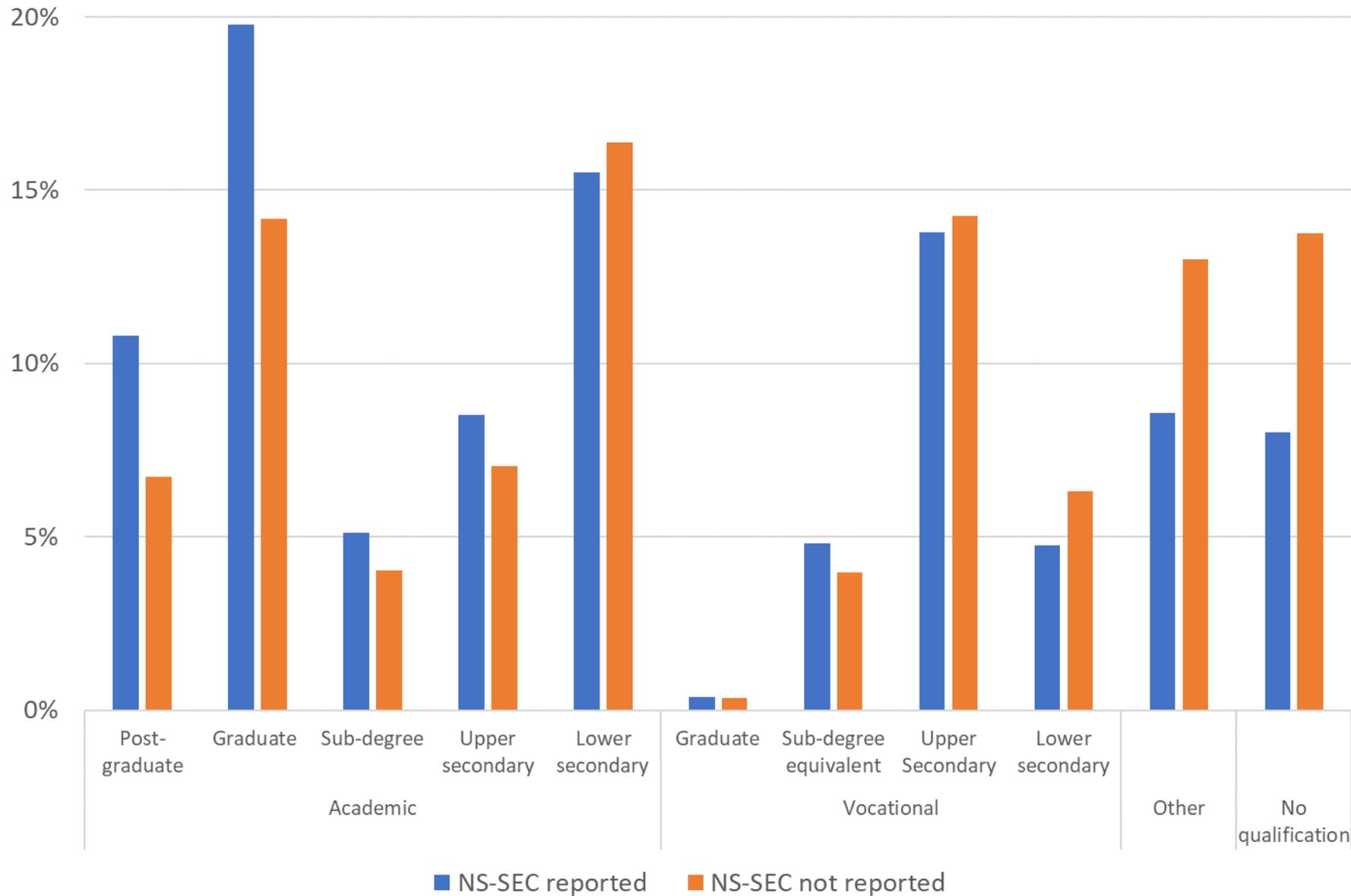
Social origin wage gaps

Cross-sectional wage equation with social origin dummies

$$\ln(w) = \alpha + \beta_i S_i + \gamma_1 X + \gamma_2 X^2 + \delta_j C_j + \varepsilon$$

- ln(w): log hourly wage
- S: Social class dummy
- X: Age
- C_k: Controls
  - Demographic: gender, disability, visible minority, birthplace Qualifications, degree class
  - Workplace: sector, firm size, location of workplace
  - Occupation

| | log(hourly wage) | | | | | | | |
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	0.066***	0.062***	0.062***	0.062***	0.060***	0.058***	0.056***	0.046***
Age^2	-0.001***	-0.001**	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***
Sex	-0.197***	-0.213***	-0.212***	-0.209***	-0.155***	-0.157***	-0.135***	-0.120***
Disability	-0.122***	-0.092***	-0.095***	-0.092***	-0.082***	-0.083***	-0.080***	-0.062***
Non-white ethnicity	-0.065***	-0.106***	-0.088***	-0.132***	-0.123***	-0.123***	-0.112***	-0.080***
NS-SEC 2 Lower managerial and professional	0.082***	0.007	0.002	0.001	-0.000	-0.000	-0.003	-0.004
NS-SEC 3 Intermediate occupations	-0.021**	-0.024**	-0.030***	-0.031***	-0.031***	-0.031***	-0.032***	-0.032***
NS-SEC 4 Small employers and own account workers	-0.043***	-0.031***	-0.031***	-0.030***	-0.035***	-0.037***	-0.040***	-0.039***
NS-SEC 5 Lower supervisory and technical	-0.162***	-0.069***	-0.069***	-0.069***	-0.064***	-0.064***	-0.064***	-0.060***
NS-SEC 6 Semi-routine occupations	-0.194***	-0.076***	-0.080***	-0.079***	-0.078***	-0.072***	-0.070***	-0.065***
NS-SEC 7 Routine occupations	-0.178***	-0.087***	-0.083***	-0.083***	-0.082***	-0.078***	-0.080***	-0.073***
NS-SEC 8 Never worked and long-term unemployed	-0.265***	-0.120***	-0.120***	-0.120***	-0.107***	-0.108***	-0.107***	-0.102***
NS-SEC information not reported	-0.275***	-0.126***	-0.125***	-0.124***	-0.118***	-0.115***	-0.114***	-0.108***
Qualifications		√	√	√	√	√	√	√
Country of birth			√	√	√	√	√	√
Regional controls (location of workplace)				√	√	√	√	√
Part-time					√	√	√	√
Firm size					√	√	√	√
Sector of employment							√	√
Occupational status								√
Observations	47,834	47,834	47,834	47,834	47,834	47,834	47,834	47,834
R-squared	0.125	0.271	0.278	0.297	0.309	0.331	0.350	0.440

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Class coefficients in wage equation controlling for qualifications
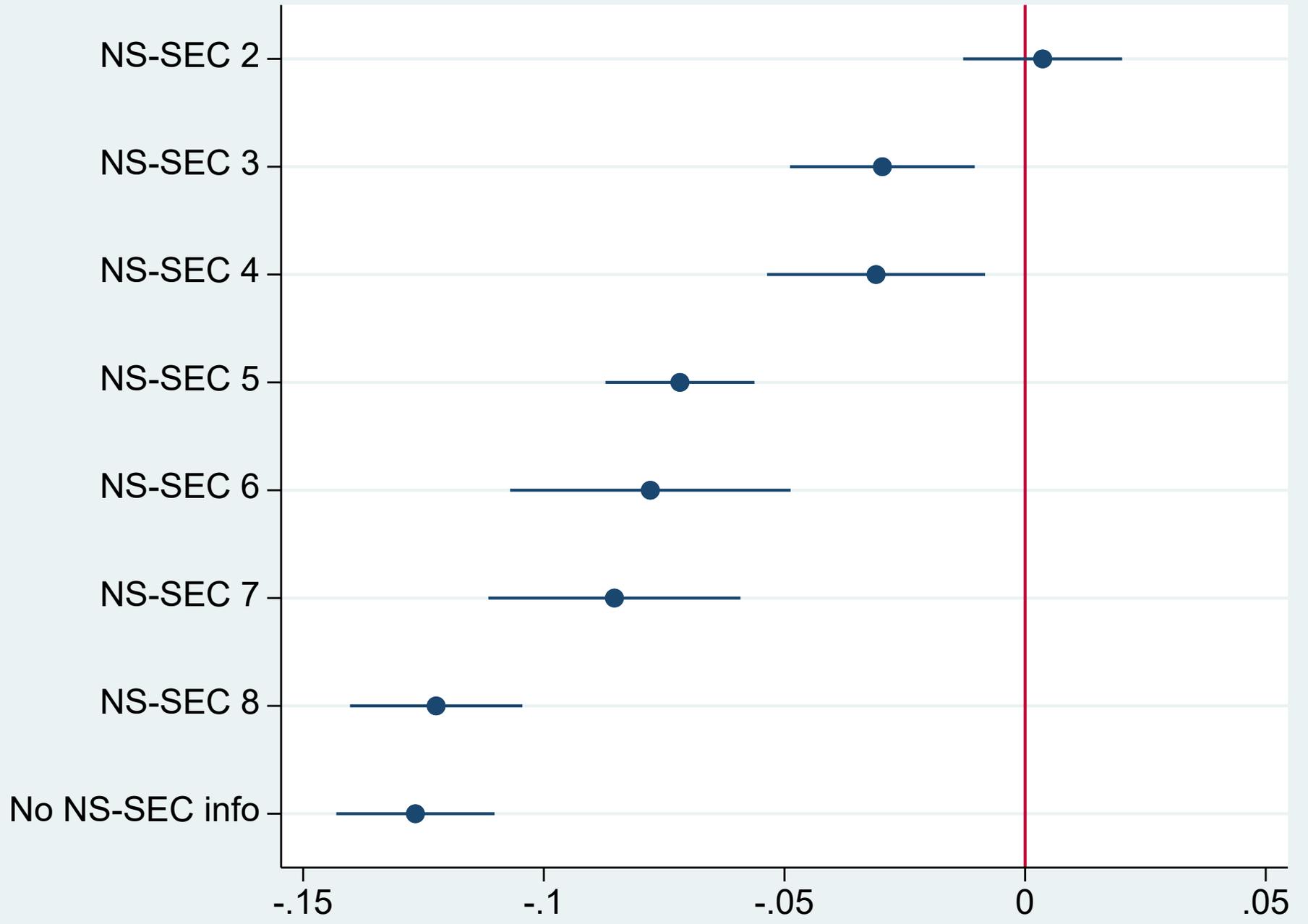
Interpreting wage gap results

- Outcomes of non-respondents similar to NS-SEC 8
- Respondent with no NS-SEC information combination of random and non-random item non-response
  - Including them as a separate category likely to underestimate impact of non-response

- How do respondents "drop out" of NS-SEC when variable derived?
  - Background in non-traditional households: NS-SEC does not apply
  - Occupation of parent's didn't fit occupational scheme: not-classified

- Plausible that lack of NS-SEC information is picking up non-traditional working class respondents ie social origin information is missing because of social origin

Final response rate in achieved sample for social origin variable

Occupation of main wage earner when respondent was 14 years old (Major)	Freq.	Percent	Cum.
Does not apply	33,322	10.79	10.79
No answer	1,642	0.53	11.32
Higher managerial and professional occu	34,769	11.26	22.58
Lower managerial and professional occup	44,237	14.33	36.91
Intermediate occupations	24,956	8.08	44.99
Small employers and own account workers	15,605	5.05	50.04
Lower supervisory and technical occupat	66,949	21.68	71.73
Semi-routine occupations	9,182	2.97	74.70
Routine occupations	10,249	3.32	78.02
Never worked and long-term unemployed	37,175	12.04	90.06
Not classified	30,703	9.94	100.00
Total	308,789	100.00	

Data route to social origin outcome: intersection of two key variables for deriving class

| Key |
| --- |
| *frequency*
row percentage |

Household composition when respondent was 14 years old	Main wage earner when respondent was 14 years old							Total
	Does not	No answer	Mother	Father	Other fam	Joint ear	No-one wa	
Does not apply	207 100.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	207 100.00
No answer	2,325 100.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	2,325 100.00
Living with one or bo	37 0.01	2,604 0.87	44,651 14.86	225,750 75.12	3,059 1.02	11,681 3.89	12,736 4.24	300,518 100.00
Living with other fam	0 0.00	92 2.57	41 1.14	155 4.32	2,247 62.68	59 1.65	991 27.64	3,585 100.00
Not living with famil	2,838 100.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	2,838 100.00
Total	5,407 1.75	2,696 0.87	44,692 14.44	225,905 73.00	5,306 1.71	11,740 3.79	13,727 4.44	309,473 100.00

Missing cases:

Drop out stage		
Family Composition (SMHCOMP)	5407	Does not apply, no answer not living with parents plus 37 who lived with both parents but was coded as 'does not apply'
Main earner (SMEARNER)	21830	Above plus no answer plus no one was earning
Occupation of Main Earner (SMSOC101)	65667	Above plus not classified, no answer and 11492 additional 'do not apply'

Class coefficients in wage equation controlling for qualifications
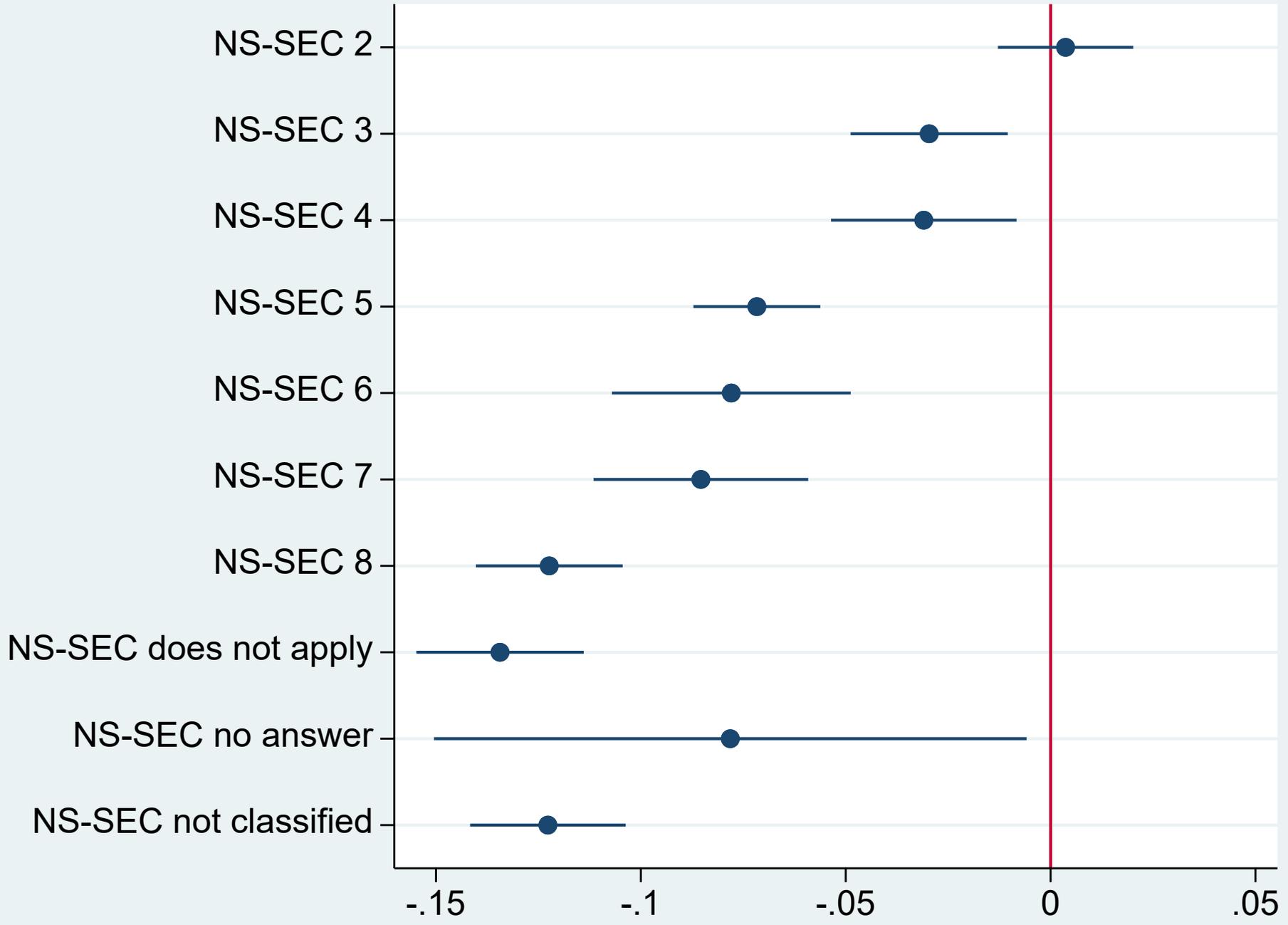
Does it matter?

- Likely to be underestimating size of social origin wage gaps

- Risk of underestimating the number of individuals affected

- Not understanding the nature of the disadvantage of the sub-populations not identified in the statistics

What can be done?

- Examine non-respondents more closely to see if we can *recover* or at least understand the data

- Separate out different categories of non-response in the wage equation

- Find an alternative/complementary variable to capture social class which has fewer missing cases

- Weight the NS-SEC8 respondents to account for the non-respondents

- Important to establish an alternative approach to omitting respondents without NS-SEC information

Recurring weakness in NS-SEC data

- Similarly high rates of missing NS-SEC data in other social surveys and administrative data
- We checked
  - Understanding Society
  - BCS70
  - HESA Students in Higher Education

HESA Missing Social Class Data 2018-19

	Private/Public School	NS-SEC codes of parents	Parental Education	Low Participation neighbourhood	Scottish/English /Welsh/NI IMD
Number	71775	234240	181705	6395	2365
Percentage	5.6%	18.4%	14.2%	0.5%	0.18%