SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Industry and occupation coding: A comparison of office-based coding and a closed-list approach

Cristian Domarchi[1] • Olga Maslovskaya[1] • Curtis Jessop[2] • Lisa Calderwood[3] • Matt Brown[3]

[1]University of Southampton, UK; [2]National Centre for Social Research (NatCen), UK; [3]Centre for Longitudinal Studies, University College London, UK

Survey Futures Workshop: Industry and Occupation Coding • University College London • 4 June 2026

Introduction

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- **Industry and occupation** are key measures in social surveys:
  - They are an indicator of socioeconomic status
  - They are strongly linked to income, health, and lifestyle
- The **conventional method** for industry and occupation coding involves:
  - Interviewers who ask open questions and ensure respondents provide the necessary information
  - Office-based coders manually coding this information to a standard classification
- The conventional method is:
  - Expensive and labour-intensive
  - Burdensome for respondents
  - Difficult to implement in self-administered surveys

Introduction

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Several alternative methods have been used to implement **industry and occupation coding in online surveys** (Kocar et al., 2025), including:
  - Automated post-interview office coding of open text responses (including rule-based coding, machine learning, and large language models)
  - Self-coding during the interview (including look-up codes)
- **Closed-list questions** are frequently used across surveys that only require higher-level classifications of industries and jobs
- However, there is limited methodological evidence of their effectiveness and accuracy

Aim and objectives

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- This study examines the implementation of closed-list questions for industry and occupation coding within an online probability-based panel survey in the UK
- The survey employed this method alongside conventional open-text questions, which were subsequently office-coded
- We aim to evaluate:
  - The accuracy of self-coding based on closed-list questions
  - The types of mismatches between self-selected codes and office-coded responses
  - The factors influencing the level of agreement between the two approaches

Data

- Data comes from a probability-based panel survey of adults aged 18 or over in the United Kingdom, conducted by the National Centre for Social Research (NatCen)
- We use data from the March 2022 wave of the NatCen panel survey
- This was a self-administered survey with a small fraction of interviews conducted by telephone:

	Sample size
Rough sample	+5,884
- Telephone respondents	-187
- Respondents not in work	-921
= Initial sample	**4,776**
- Missing sociodemographic information	-56
= Effective sample	**4,726**

	Sample size
= Effective sample	+4,726
- Missing industry information	-51
= Analytical sample (Industry)	**4,721**
- Missing occupation responses	-131
= Analytical sample (Occupation)	**4,589**

Closed-list: Industry

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

{IF PdWrk = 1}
CurEmpInd2
"The following questions refer to the main paid job you had in the seven days ending Sunday {date of preceding Sunday, e.g.: "6th February.

Which of the following best describes the sector you work in?

G_ReadOut_II1"

1. Agriculture, forestry and fishing [HELPLINK: "Typically involves working on farms, forests or fishing boats and often based outdoors on the land. Examples include crop farming, livestock farming, logging, silviculture, marine and freshwater fishing."]
2. Mining and quarrying [HELPLINK: "Involves removing rock and other minerals from the ground. Examples include underground mining, surface mining, petrol and natural gas extraction and support services for mining and quarrying."]
3. Manufacturing [HELPLINK: "Creating goods, by hand or machine, for sale. This includes all goods production and processing such as meat, fish, baked goods and other food products, beverages, tobacco, textiles, clothes, wood, metal, furniture, printed materials, recorded media, chemical and pharmaceutical products, computer and electrical equipment, vehicles and transport equipment."]

Closed-list: Industry

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- The options in the **industry** question correspond to the 1-digit **sections** of the 2007 UK Standard Industrial Classification of Economic Activities (**SIC**) codes:

Section	Industry
A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam & air conditioning
E	Water supply, sewerage and waste management
F	Construction
G	Trade (retail sales and wholesales) and vehicle repair
H	Transportation and storage
I	Accommodation and food service
J	Information and communication

Section	Industry
K	Finance and insurance
L	Real estate
M	Professional, scientific and technical
N	Administrative and support services
O	Public administration and defence
P	Education
Q	Human health and social work
R	Arts, entertainment and recreation
S	Other service activities
T	Employed by a household as domestic staff
U	International organisations and bodies

Closed-list: Occupation

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

{IF EconAct = 3 AND **Nov22SampSplit3 = 1**}
CurEmpOcc3
"Which of the following best describes the main paid job you had in the seven days ending last Sunday?

1. Corporate manager or Director [HELPLINK: "Such as Chief Executive, Production manager, Financial manager, marketing manager, officers in the armed forces, senior police officers, health, public health and social care managers."]
2. Other manager or owner [HELPLINK: "Such as managers and proprietors in agriculture, fishing, forestry, hotels, public houses, leisure and sports centres, travel agencies, health care practices, transport, warehousing, housing, garages, beauty salons, creative industries etc."]
3. Professional – Science, research, engineering, technology [HELPLINK: "Such as natural and social scientists, engineers, IT professionals, Web and graphic design professionals, Research & Development professionals.

Closed-list: Occupation

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- The options in the **occupation** question correspond to the 2-digit level **sub-major groups** of the 2020 Standard Occupational Classification (**SOC**) codes:

Major group		Sub-major group		Skill level
Code	Definition	Code	Definition	
1	Managers, directors and senior officials	11	Corporate managers and directors	4
		12	Other managers and proprietors	3
2	Professional occupations	21	Science, research, engineering and technology professionals	4
		22	Health professionals	4
		23	Teaching and other educational professionals	4
		24	Business, media and public service professionals	4
3	Associate professional occupations	31	Science, engineering and technology associate professionals	3
		32	Health and social care associate professionals	3
		33	Protective service occupations	3
		34	Culture, media and sports occupations	3
		35	Business and public service associate professionals	3
4	Administrative and secretarial occupations	41	Administrative occupations	2
		42	Secretarial and related occupations	2
5	Skilled trades occupations	51	Skilled agricultural and related trades	3
		52	Skilled metal, electrical and electronic trades	3
		53	Skilled construction and building trades	3
		54	Textiles, printing and other skilled trades	3
6	Caring, leisure and other service occupations	61	Caring personal service occupations	2
		62	Leisure, travel and related personal service occupations	2
		63	Community and civil enforcement occupations	2
7	Sales and customer service occupations	71	Sales occupations	2
		72	Customer service occupations	2
8	Process, plant and machine operatives	81	Process, plant and machine operatives	2
		82	Transport and mobile machine drivers and operatives	2
9	Elementary occupations	91	Elementary trades and related occupations	1
		92	Elementary administration and service occupations	1

Open-text questions

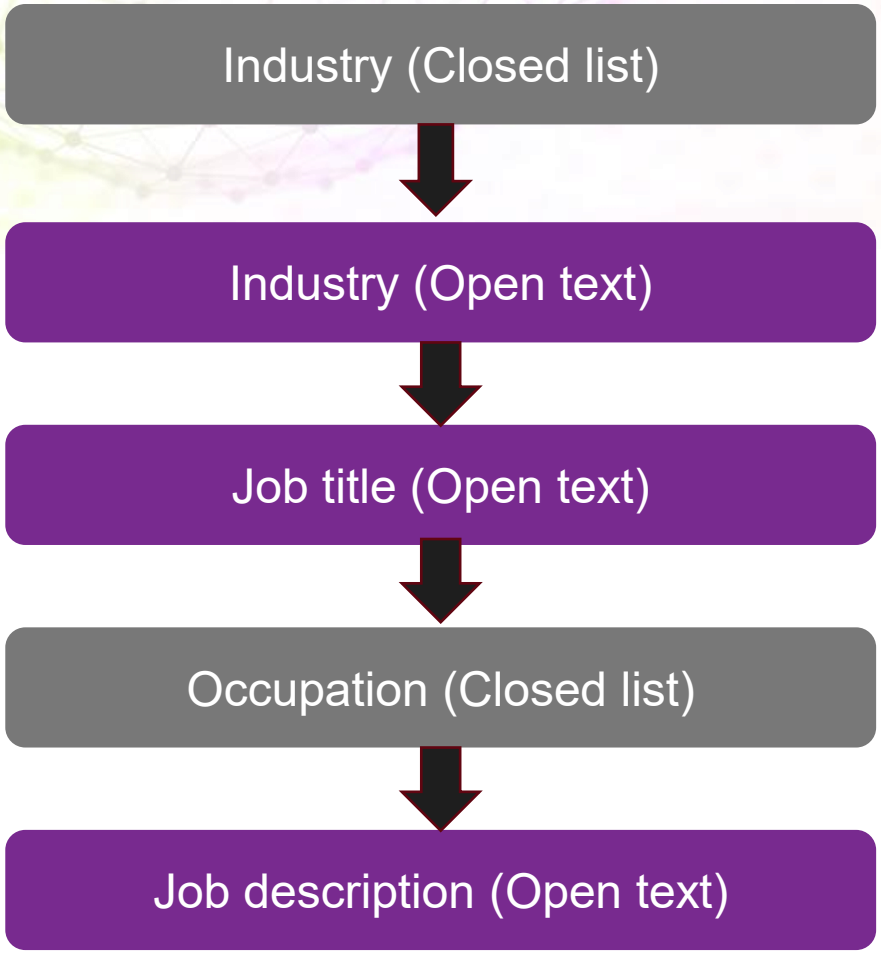
Question	Text
Industry	"*What does the firm or organisation you work for mainly make or do at the place where you work? Please give a full description. For example, do you work in 'manufacturing', 'processing' or 'distributing'? What are the main goods you produce and materials you use? Is it wholesale or retail?*"
Occupation – Job title	"*What is the name of your job or your job title?*"
Occupation – Job description	"*And in your own words, what do you mainly do in your job? Please give a full description*"

- The **industry** and **job description** questions also included a **soft check** when respondents entered 15 or fewer characters in the open-text response:

  "*It is important that you enter enough information for us to clearly understand [where you work OR your job]. Please ensure you have fully described what you do*".

Industry and occupation questions

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION



Paradata: Industry

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- **SIC sample:** 60.3% respondents used mobile phone, 39.7% used other devices:

Category	Indicator	Open-text question (OT)				Closed-list question (CL)				OT vs. CL Comparison (p<0.05)
		Mobile	Others	Sig. (Mob vs. Oth.)	All	Mobile	Others	Sig. (Mob vs. Oth.)	All	
Completion time	Elapsed time (seconds)	51.2	60.3	-	**54.5**	41.8	46.1	-	**43.4**	OT ≠ CL
Errors and timeouts	Error messages (n)	0.27	0.19	***	**0.24**	0.003	0.004	-	**0.003**	OT ≠ CL
	Timeouts (n)	0.04	0.06	**	**0.045**	0.03	0.05	**	**0.035**	OT ≠ CL
Text length and complexity	Number of characters	43.5	50.2	***	**45.9**	-	-	-	-	-
	Number of syllables	12.4	14.4	***	**12.9**	-	-	-	-	-
	Number of words	6.4	7.1	***	**6.6**	-	-	-	-	-
	SMOG Index	9.4	10.3	***	**9.7**	-	-	-	-	-

Paradata: Occupation

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- **SOC sample:** 63.6% respondents used mobile phone, 36.4% used other devices:

Category	Indicator	Open-text question (OT)				Closed-list question (CL)				OT vs. CL Comparison (p<0.05)
		Mobile	Others	Sig. (Mob vs. Oth.)	All	Mobile	Others	Sig. (Mob vs. Oth.)	All	
Completion time	Elapsed time (seconds)	71.2	82.9	**	**75.5**	40.9	39.7	-	**40.5**	OT ≠ CL
Errors and timeouts	Error messages (n)	0.08	0.06	**	**0.075**	0.006	0.01	-	**0.007**	OT ≠ CL
	Timeouts (n)	0.09	0.15	***	**0.108**	0.04	0.07	**	**0.06**	OT ≠ CL
Text length and complexity (*Job title*)	Number of characters	19.4	20.5	**	**19.8**	-	-	-	-	-
	Number of syllables	6.1	6.4	**	**6.2**	-	-	-	-	-
	Number of words	2.5	2.6	-	**2.5**	-	-	-	-	-
	SMOG Index	8.6	9.0	***	**8.8**	-	-	-	-	-
Text length and complexity (*Job description*)	Number of characters	68.8	78.1	***	**72.2**	-	-	-	-	-
	Number of syllables	18.8	21.7	**	**19.9**	-	-	-	-	-
	Number of words	10.2	11.5	**	**10.7**	-	-	-	-	-
	SMOG Index	10.5	11.7	***	**11.0**	-	-	-	-	-

Results: Industry

Section	Description	Selected (%)	Coded (%)	Agreement rate (%)	N (coded)
A	Agriculture, forestry and fishing	0.7	0.3	*	*
B	Mining and quarrying	0.2	0.1	*	*
C	Manufacturing	5.7	9.3	51.3	433
D	Electricity, gas, steam & air conditioning	1.2	0.6	*	*
E	Water supply, sewerage and waste management	0.7	0.8	72.2	36
F	Construction	3.2	3.6	58.7	167
G	Trade (retail sales and wholesales) and vehicle repair	7.0	9.8	57.1	461
H	Transportation and storage	3.6	3.3	62.8	156
I	Accommodation and food service activities	2.8	2.8	68.5	127
J	Information and communication	4.6	5.3	48.8	248
K	Finance and insurance	5.9	4.4	86.5	207
L	Real estate	0.8	1.3	41.0	61
M	Professional, scientific and technical	10.9	8.2	44.0	382
N	Administrative and support services	7.4	3.0	16.2	142
O	Public administration and defence	6.6	9.2	53.8	433
P	Education	14.0	14.6	85.0	687
Q	Human health and social work	14.9	18.4	71.1	855
R	Arts, entertainment and recreation	2.6	2.6	57.1	119
S	Other service activities	6.8	2.0	38.8	93
T	Employed by a household as domestic staff	0.4	0.3	*	*
U	International organisations and bodies	0.4	0.1	*	*

*Only groups with 30 or more observations are displayed

- Distribution of industries differs across coding method ($p < 0.001$)
- Dissimilarity index: 15.0%

- Overall agreement rates:

Device	Rate (%)
Mobile	63.5%
Others	58.4%
Total	**61.7%**

*Difference across devices is significant ($p < 0.001$)

Results: Industry

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Most typical mismatches between industries, among those with mismatched industry codes ($N = 1{,}790$):

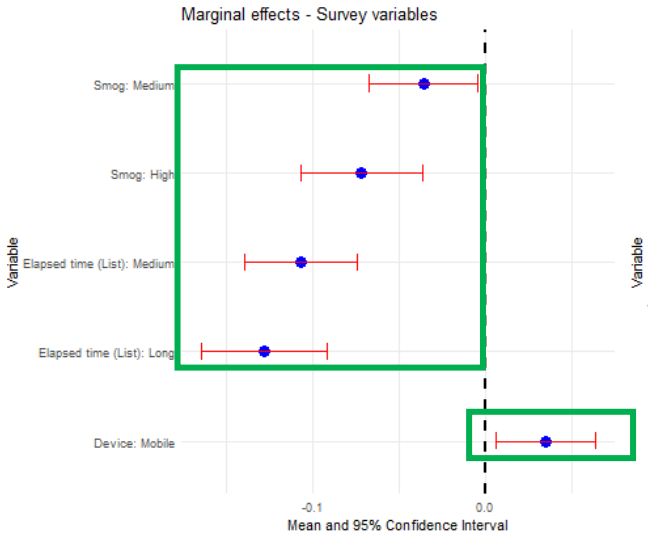
Category	Code	Industry description	% of responses
Most frequently selected	M	Professional, scientific and technical	18.9
	N	Administrative and support services	18.1
	S	Other service activities	15.5
Most frequently coded	Q	Human health and social work	14.3
	C	Manufacturing	12.4
	M	Professional, scientific and technical	12.4
	O	Public administration and defence	11.8
	G	Trade (retail sales and wholesales) and vehicle repair	11.3

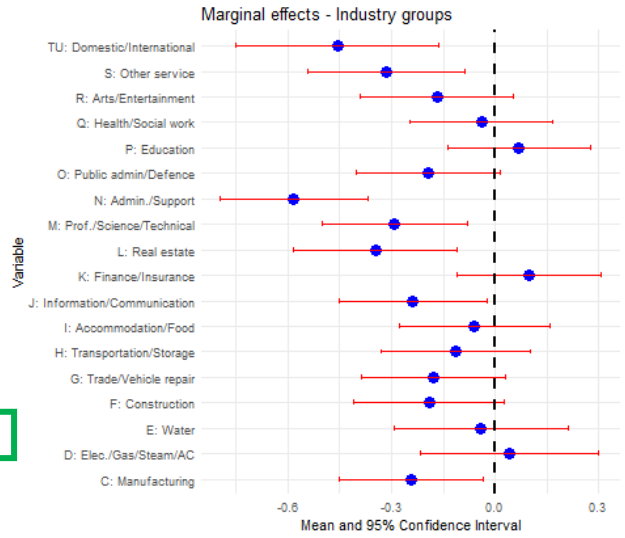
Results: Industry

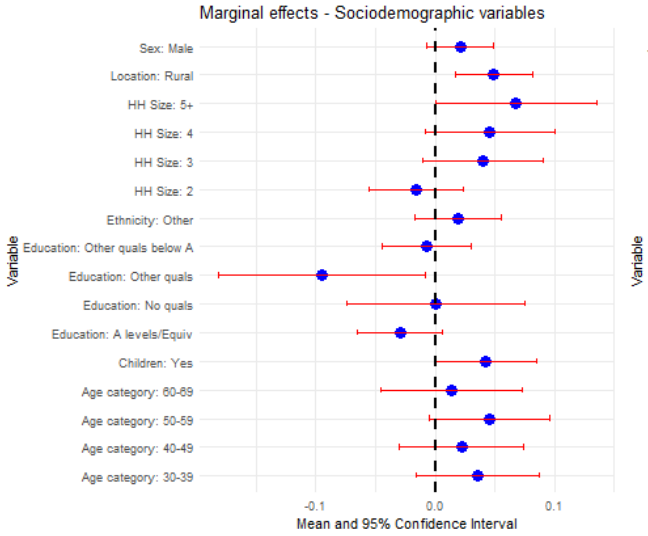
SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

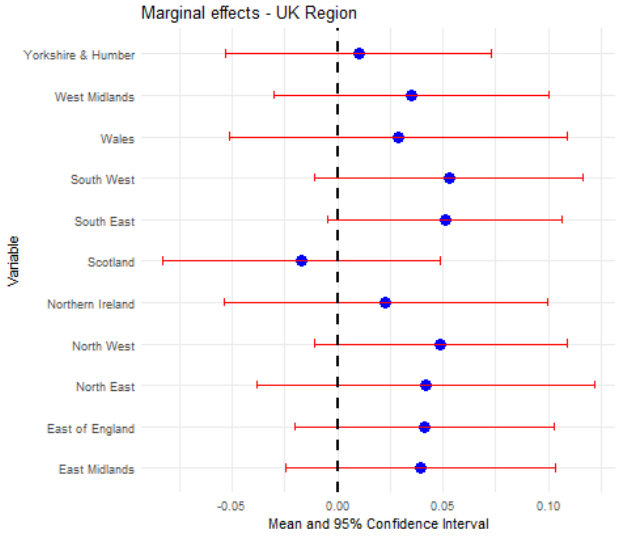
- Logistic regression model of agreement between coded and selected industries.

Code	(More likely to agree)	(Less likely to agree)
Survey variables	Response via mobile	More elapsed time in list question More complex text (longer words) in open-text question
Industry	-	C, J, L, M, N, S, T/U
Sociodemographic Variables	Rural respondents	Respondents with lower qualifications
UK region		









University of Essex
University of Southampton
UKRI Economic and Social Research Council

Results: Occupation

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Major group	Description	Selected (%)	Coded (%)	Agreement rate (%)	N (coded)
1	Managers, directors, and senior officials	13.9	10.2	56.5	467
2	Professional occupations	37.0	31.7	74.6	1,454
3	Associate professional and technical occupations	10.5	16.2	23.2	745
4	Administrative and secretarial occupations	12.4	13.1	61.2	603
5	Skilled trade occupations	5.4	5.9	57.9	271
6	Caring, leisure and other service occupations	5.7	8.5	38.7	390
7	Sales and customer service occupations	9.5	5.8	77.7	265
8	Process, plant and machine operatives	4.4	4.2	63.9	194
9	Elementary occupations	1.2	4.4	15.5	200

- Distribution of occupations differs across coding methods ($p < 0.001$)
- Dissimilarity index: 12.9%
- Overall agreement rates:

Device	1-digit (%)	2-digit (%)
Mobile	56.7%	46.7%
Others	54.2%	44.3%
Total	**55.8%**	**45.8%**

**Differences across devices are not statistically different (p = 0.1 in both cases)*

University of Essex | University of Southampton | UKRI Economic and Social Research Council

Results: Occupation

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

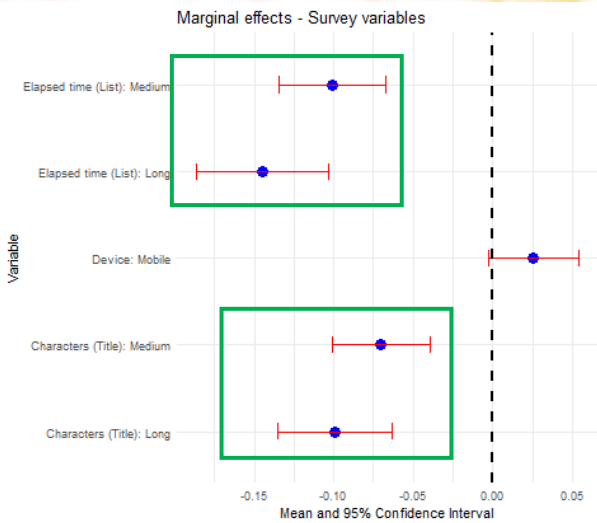
- Among respondents with mismatched occupations at the sub-major group (2-digit) level ($N = 2{,}486$):
  - **82%** are also mismatched at the major group (1-digit) level:
    - **Most occupation mismatches are observed between major groups (rather than within major groups)**
  - We did not find any specific pattern in skill levels for occupation mismatches:
    - **40%** select an occupation that requires **higher skill levels** than the occupation they were coded into
    - **30%** select an occupation that requires **lower skill levels** than the occupation they were coded into

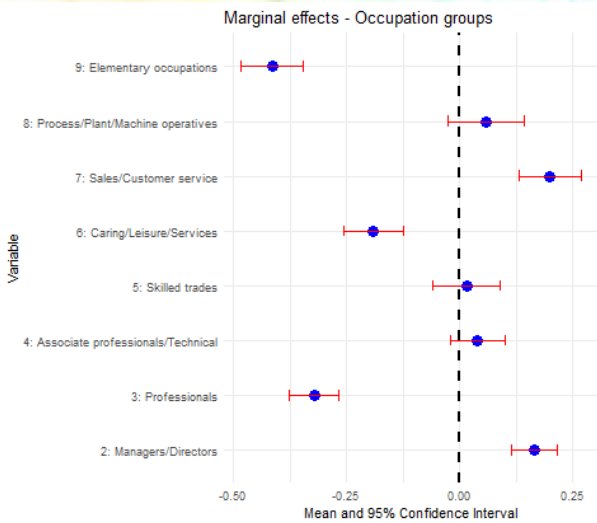
Results: Occupation

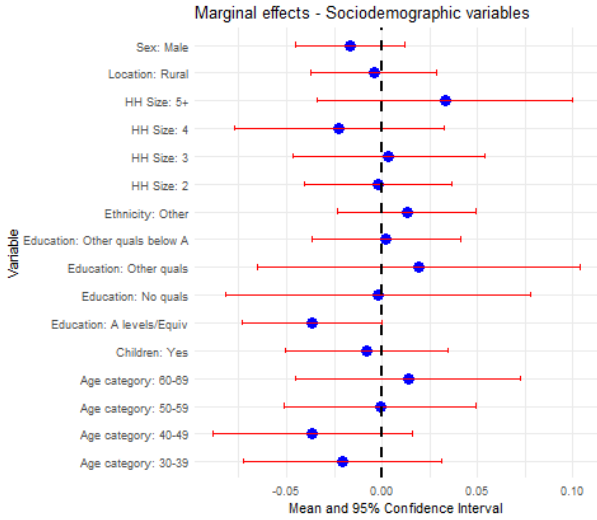
SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

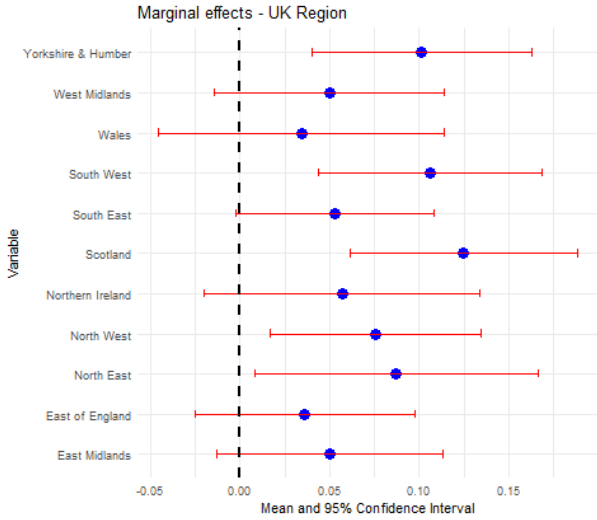
- Logistic regression model of agreement between coded and selected (major group level) occupations.

Code	(More likely to agree)	(Less likely to agree)
Survey variables	-	More elapsed time in list question More complex text (longer words) in open-text question
Occupation	2, 7	3, 6, 9
Sociodemographic Variables	No effects	No effects
UK region	5 regions	-









University of Essex
University of Southampton
UKRI Economic and Social Research Council

Conclusions

- Industry codes at the 1-digit SIC level, and occupation codes at the 1- and 2-digit SOC levels collected via closed-list questions show relatively low agreement with office codes derived from open-text responses:
  - Respondents are more likely to self-code into industries involving administration, support, and services than they are to be office-coded into these categories
  - Respondents are also more likely to select professional or managerial occupations than they are to be office-coded into these categories. However, no clear trends appear for mismatched occupations and skill levels
- Elapsed time in the closed-list question and complexity of open-text responses are associated with mismatches in both industry and occupation

Conclusions

- This approach is unlikely to be suitable if a high degree of accuracy is required for these variables
- However, it may still be helpful for surveys not requiring these accuracy levels, as respondent burden and processing costs are reduced
- The approach may be more effective if code frames were developed which described occupations and industries using language that participants themselves would use to describe their work

SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Industry and occupation coding: A comparison of office-based coding and a closed-list approach

Cristian Domarchi[1] • Olga Maslovskaya[1] • Curtis Jessop[2] • Lisa Calderwood[3] • Matt Brown[3]

[1]University of Southampton, UK; [2]National Centre for Social Research (NatCen), UK; [3]Centre for Longitudinal Studies, University College London, UK

Survey Futures Workshop: Industry and Occupation Coding • University College London • 4 June 2026