SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Report 3: Occupation coding in self-completion surveys: Evidence Review

Sebastian Kocar, Centre for Longitudinal Studies, UCL
Matt Brown, Centre for Longitudinal Studies, UCL
Lisa Calderwood, Centre for Longitudinal Studies, UCL

February 2025

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (grant ES/X014150/1) aimed at bringing about a step change in survey research to ensure that high quality social survey research can continue in the UK. The initiative brings together social survey researchers, methodologists, commissioners and other stakeholders from across academia, government, private and not-for-profit sectors. Activities include an extensive programme of research, a training and capacity-building (TCB) stream, and dissemination and promotion of good practice. The research programme aims to assess the quality implications of the most important design choices relevant to future UK surveys, with a focus on inclusivity and representativeness, while the TCB stream aims to provide understanding of capacity and skills needs in the survey sector (both interviewers and research professionals), to identify promising ways to improve both, and to take steps towards making those improvements. Survey Futures is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefitting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveyfutures.net.

Contents

Executive Summary

- This evidence review provides an overview of current practices regarding the collection and coding of occupation data, including evidence from the most relevant UK-based studies that collect occupation data. Particular focus is given to self-completion surveys, especially online surveys. The review outlines different occupation coding approaches that are typically used and presents methodological research that has been and will be conducted in the UK to further develop occupation coding solutions for survey practice.
- High-quality occupation coding requires the coding to be valid and reliable. This evidence review begins by highlighting how challenging this is to achieve, as measuring occupation in surveys can be complex for a wide range of reasons. Additionally, occupation coding in self-completion surveys presents further challenges due to the absence of interviewers who can provide assistance to respondents.
- Different approaches have been used in the UK to collect and code occupation data. These include office coding of open text after the interview (i.e., manual, software-supported, and automated coding), self-coding during the interview (the look-up), and closed-ended survey questions on occupation. Each of these approaches comes with its unique advantages and disadvantages from the perspectives of cost, complexity, time-efficiency, and data quality.
- Manual or clerical coding of open-text descriptions of jobs has been used in several UK-based studies to date, including longitudinal household panels (Understanding Society), birth cohort studies such as the 1970 British Cohort Study (BCS70), and cross-sectional surveys such as the European Social Survey (ESS UK). Some studies that have transitioned from face-to-face to online data collection have made adaptations to improve the quality of the open-text occupation data, such as adjustments to wording and instructions.
- Automated coding of open-text descriptions is a growing field, with recent developments in machine learning algorithms and language-based predictive models. In the UK, automated occupation coding has been used by the Office for National Statistics (ONS) in large-scale studies, including the 2021 Census in England and Wales and the Transformed Labour Force Survey. Some identified limitations include that around one third of occupations could not be automatically coded in the 2021 Census (and had to be manually coded), and discrepancies between office coding and automated coding were found in some other surveys.
- Self-coding is an interactive and semi-automated approach to occupation coding during the interview, often using look-up tables with a list of potential answers generated based on the respondent's open text job description. This approach has been piloted in the UK both in self-completion and interviewer-administered surveys, such as ONS's Opinions and Lifestyle Survey, ESS UK, and BCS70, with a number of challenges identified.

- Building on this research, the feasibility of the look-up approach was most thoroughly tested in Next Steps, a longitudinal study. This evidence review outlines key findings from the methodological studies using Next Steps data, including that the main challenges of the look-up approach are lower coding rates compared to the office coding approach and moderate agreement between look-up and office codes, with relatively minor difference between self-completion and interviewer administration. Nonetheless, several practical solutions are identified which could improve this approach going forward.
- Recommendations for methodological solutions for data collection, as well as for coding open descriptions of jobs, are also proposed. These include a wider application of the look-up approach in self-completion surveys, potentially supplemented by the collection of open-text job descriptions and office coding for a sub-group of respondents. Further testing of automated occupation coding and a broader application of artificial intelligence (AI) to code occupation data is also suggested in the UK context, using the UK occupational classification as the framework.
- To further advance methods and approaches to occupation coding, additional methodological research for both self-coding and manual coding is suggested. This includes examining the effect of mode on occupation coding, the role of administrative data in assessing the quality of occupation coding, and further development of the look-up function for self-completion surveys. Some of this methodological research, which is described in more detail in this evidence review, is planned as part of the Survey Futures project.

1. Introduction

Occupation is an important concept to measure in social research since it forms a significant part of an individual's identity (Tijdens, 2022). It is also an important marker of socio-economic status and has a significant impact on income, health, lifestyle, and many other aspects of life.

The traditional approach to measuring occupation in surveys has been to ask open-text questions about job title and duties which are subsequently coded manually by office-based coders to a standardised classification scheme (Lyberg & Dean, 1992). Participants are typically asked to supply descriptions of their job title, their duties or 'work tasks' (e.g., Hacking et al., 2006; Mannetje & Kromhout, 2003; Peycheva et al., 2021).

Occupation coding has a long history, with the first International Standard Classification of Occupations (ISCO) being adopted in 1957 (i.e., ISCO-58) (International Labour Organization, n.d.). Klassifikation der Berufe (KldB) Simson et al., 2023), and the U.S. Standard Occupational Classification (U.S. Bureau of Labor Statistics, n.d.) being established in 1977. The International Standard Classification of Occupations (ISCO) is often used for cross-country comparisons (Hoffmann & Thomas, 1995). Classification schemes differ but all include hundreds of specific occupation codes nested within hierarchical groups. For example, the 2020 classification of occupations in the UK, known as the Standard Occupational Classification (SOC2020), has 10 major groups (1 digit) which are sub-divided into sub-major groups (2-digit), minor groups (3-digit), and unit groups (4-digit) (Office for National Statistics, n.d.-a). Standard occupational coding classifications are periodically updated to reflect changes in job roles, industries, and labour market trends. For example, the UK SOC system is typically updated every 10 years. The most recent update was the UK SOC 2020, following the previous version, UK SOC 2010.

Other related classification schemes which are closely related to occupation include classifications of social class – such as the NS-SEC in the UK (National Statistics Socio-economic Classification). This classification uses occupation in combination with employment status, supervisory skills, organisation size and classifications of industry such as SIC (Standard Industrial Classification) which is used to categorise businesses and industries based on their primary activity.

Measuring occupation in surveys can be challenging. High-quality occupation coding requires the conducted coding to be valid and reliable. Validity relates to the accuracy with which a job is assigned to the classification system and reliability relates to the consistency of coding across repeated instances – i.e., if the same job were described twice, it would be assigned the same code on both occasions (Hoffmann & Thomas, 1995). However, achieving the required level of accuracy and consistency presents several challenges:

- Occupations can be as diverse[1] as the people participating in surveys, and different individuals might describe the same job in different ways (Simson et al., 2023).
- Respondents may provide answers that are invalid and/or cannot be converted into a codable format (Conrad et al., 2016), which can be associated with insufficient interviewer training (Belloni et al., 2016).
- The range of questions required for accurate occupation coding might be extensive (Belloni et al., 2016) and can vary greatly between occupations.
- Occupation coding is typically conducted after data collection and is based solely on the provided answers without the ability to request more information from respondents (Simson et al., 2023).
- Occupations evolve and change over time, and new or different job roles emerge, often meaning that coding schemes can lag behind the real-world developments in the labour market (Hoffmann & Thomas, 1995).

Manual coding can be impacted by coder experience and subjective interpretation of open-text job information, meaning that two coders reviewing the same information will often allocate different codes (Schierholz & Schonlau, 2021). Automated approaches to coding can increase efficiency but may not consistently capture nuances in occupation descriptions, leading to misclassification (Schierholz & Schonlau, 2021). Moreover, collecting and coding occupational data is time-consuming and costly, both in relation to the duration of questions required for collecting this information and the time and cost of manual office coding, and there is necessarily a lag before coded data is available following the survey.

Surveys are increasingly being conducted online but measuring occupation in self-completion surveys is even more challenging, as interviewers are not present to assist the respondent in providing the information required. This can have a negative impact on the quality of the collected data for coding (Conrad et al., 2016). The shift to online surveys has led to the development of new approaches to collecting and coding occupation data.

This evidence review provides an overview of current practices regarding the collection and coding of occupation data in online self-completion surveys in the UK and explores recent methodological innovations to improve both the accuracy and efficiency of coding.

We attempt to include evidence from the most relevant UK-based studies that collect occupation data. Our aim is to document various occupation coding practices and methodological research supporting the development of those practices. We worked closely with various partners and collaborators, including, but not limited to the Office for National Statistics (ONS), the National Centre for Social Research (NatCen), Ipsos, Verian and various academic institutions, to identify relevant studies and collect evidence to present the state of

[1] This could also be attributed to a modern and dynamic labour market with flexible jobs, remote and hybrid working, and people having multiple jobs at the same time. For example, the existing evidence suggests that errors are more likely to occur for those who are self-employed and certain occupation categories (Peycheva et al., 2021).

practice in occupation coding in the UK. The range of UK studies collecting occupation data is wide and we do not seek to document them all, instead we purposely focus on major surveys and those which have incorporated methodological explorations in this field. As a result, we present evidence from a range of surveys, both cross-sectional and longitudinal, as well as the census, to collate and critically discuss national evidence that would be valuable for UK (and international) survey practice. Where useful, we draw on evidence from outside of the UK and from surveys which have collected occupation data in non self-completion modes.

In Section 2 we describe the key different approaches used to collect and code occupation data in online surveys. We cover post-interview office coding of open text, using manual and automated coding, self-coding during the interview and the use of closed questions. In Section 3 we provide some conclusions, make some recommendations for best practice and suggestions for further research.

2. Approaches to collecting and coding occupation data in online surveys

In this section, we present an overview of developments and innovations in the approach to occupational coding in an online survey context in UK surveys. We focus on surveys where survey designers have implemented or trialled an approach which is different to the standard approach, defined as collecting open-text answers which are office coded after the interview. We cover surveys which have innovated by making use of partially or fully automated coding post-interview, those which have incorporated self-coding within interview by respondent and those which have used closed questions on occupation instead of, or in combination with open-text answers. Although none of these approaches are specific or exclusive to online surveys, they have often been implemented in the context of online self-completion surveys. As part of this evidence review, where relevant, we cover adaptations to standard questions about occupation made as part of online implementation, for example soft-checks on string lengths.

In each of the following sections, which describe broad approaches to occupation coding, we first present the state of the art in occupation coding worldwide, followed by a focus on examples of occupation coding in self-completion surveys in the UK. These examples include past uses of occupation coding approaches, methodological studies conducted as part of the Survey Futures project, and future research on occupation coding that is planned or already ongoing in the UK context.

2.1 Post-interview office coding of open text (manual/clerical and automated)

2.1.1 Manual or clerical coding of open text

Manual or clerical coding of occupation data has been a standard method for classifying respondents' occupations in surveys for decades. In this approach, respondents provide free-text responses to occupation-related questions, and these responses are then manually reviewed and assigned to predefined occupation categories by human coders. This process typically utilizes a standard occupational classification system, such as the Standard Occupational Classification (SOC) system, to ensure consistency and comparability across datasets. Lyberg and Dean (1992) were among the early proponents of this approach, noting its reliability in classifying occupations based on detailed job descriptions. Although time-consuming and labour-intensive, manual coding has been considered the "gold standard" in occupation coding for many years (Burstyn et al., 2014). Since the 1990s, computer-assisted office coding—utilizing software to support expert coders—has become more widely adopted (e.g., Mannetje & Kromhout, 2003). These systems use algorithms to assist coders with the aim of increasing efficiency and potentially increasing accuracy. For example, CASCOT software is widely used to facilitate coding to the UK's SOC classification scheme (Elias et al., 2014). The tool reads the open-text inputs and provides a suggested code with an associated confidence score.

Clerical coding is however not without its limitations. One significant challenge is the cost and labour involved in manually classifying large volumes of free-text responses, which can result in high administrative expenses and delays in data processing (Gweon et al., 2017). Furthermore, human coders may introduce biases or inconsistencies, especially when interpreting ambiguous or unfamiliar occupation titles. Research by Mannetje and Kromhout (2003) emphasizes the importance of coder training and the development of standardized coding protocols to mitigate these issues. Updates to coding classification schemes can also cause issues for occupation coding, particularly clerical coding (Office for National Statistics, 2022). Despite these drawbacks, clerical coding remains prevalent in surveys requiring high accuracy in capturing complex occupational data. There are many examples of online surveys which have implemented this standard approach of collecting open-text answers for post-interview coding. Such examples from the UK include major online or mixed-mode cross-sectional surveys including Understanding Society, the National Child Development Study and the 1970 British Cohort Study. It is noteworthy that the precise questions asked may vary slightly from study to study (for a list of occupation questions from different surveys and countries, see Table A1 in the Appendix A). Studies that have moved from face-to-face data collection to online data collection have often made adaptations to attempt to improve the quality of the open-text data collected for office coding. Some examples of such UK studies are provided below.

National Child Development Study (NCDS)

The National Child Development Study Age 55 Survey, conducted in 2013, was conducted with a web-first mixed mode approach. This was the first use of web interviewing in the UK birth cohort studies. Participants were asked to provide open-text descriptions of their occupation, as had been the case in the face-to-face waves that had come previously, but, with one major design adaptation due to the online context. An additional question was added where participants were asked about the qualifications required to do the job. This request had typically been provided as a prompt for interviewers in face-to-face surveys but user-testing of the web survey identified that this information was often not provided (Centre for Longitudinal Studies, 2018).

European Social Survey (ESS)

The European Social Survey (ESS), a cross-national comparative survey that explores various social issues across Europe, collects occupational data using three key questions: job title, description of work, and qualifications required. Coding is conducted by office coders.

For the first nine rounds of data collection, all data was collected face-to-face. However, due to the COVID-19 pandemic, Round 10 (2020-2022) saw a shift to a self-completion mode (both web and paper) in some countries, while others maintained face-to-face interviews (sometimes in combination with video). A significant challenge during the transition to self-completion was an increase in missing data for both occupation and industry. In Round 10, the missing data rate for occupation (ISCO-08 coding) was 15% across countries using self-completion, compared to only 2% in the previous face-to-face round (Hanson, 2021). Self-completion modes were not offered in Round 11 (2023-2024) but in Round 12 (2025-2026) half of the sample will be invited to participate via web and paper only, and this approach will be rolled out to the full sample from Round 13 onwards (2027-2028). ESS will continue to collect open-text information on occupation and industry for office coding. Minor adjustments have been made to the wording and instructions of some questions to elicit more comprehensive information and therefore, hopefully, reduce missing data.

2.1.2 Automated coding of open-text descriptions

Fully automated coding approaches have been developed, but have not yet been widely adopted as evidence suggests they are not as accurate as manual coding (Gweon et al., 2017). Automated approaches using word-matching (e.g., Ossiander & Milham, 2006) and advanced machine learning models (e.g., Schierholz & Schonlau, 2021) have been developed for occupational coding. Automated coding can be either rule-based or based on statistical or machine learning. Rule-based occupation coding employs hundreds of 'if-then' statements, written by experts, to match text to predefined rules and select the code with the highest score. On the other hand, statistical learning models are trained on already classified data and use text mining to convert text into numerical data. Using this numerical data, the code with the highest probability based on the prediction model, such as a regression model, is selected (Gweon et al., 2017). There has also been research on how to improve the accuracy of different automated coding approaches. For example, Gweon et al. (2017) tested the modified nearest neighbour method, statistical learning, and a hybrid method (with both duplicate and statistical learning approaches). Another approach that emerged recently is artificial intelligence (AI)-based or, more specifically, language-based predictive models such as GPT (e.g., Safikhani et al., 2023).

An example of a fully automated coding solution is the SOCcer algorithm, which codes provided job information into a 6-digit U.S. SOC code. The algorithm uses up to five natural language classifiers (as algorithms or models used to categorise data), including three on job titles, one on industry, and one on work tasks (depending on the availability of information). It calculates a score that represents "an estimate of the probability that an expert coder would have assigned that SOC code to the job description" (Russ et al., 2023). While the results for the most recent version of SOCcer showed that the agreement between the algorithm and a coding expert is typically about 50% at the 6-digit level, it is important to note that the agreement between two experts at the same level is often around the same figure (Russ et al., 2023).

On the other hand, some other studies have reported high agreement rates but low coding rates (as a proportion of respondents with an assigned occupation code) for automated coding software compared to manual coding by professional coders (e.g., Helppie-McFall & Sonnega, 2018). Gweon et al. (2017) also demonstrated how targeting high accuracy rates can have a negative impact on coding rate (or so-called production rates in the case of automated coding), and the right balance between these two objectives needs to be found.

In the UK, the Office for National Statistics in England and Wales (ONS) has developed and made extensive use of automated coding on the 2021 Census, the Transformed Labour Force Survey (TLFS) and other large-scale surveys. Additional information is provided in the sections below.

Population censuses in England and Wales

The ONS conduct the Census in England and Wales every 10 years. The aim is to provide a comprehensive picture of the people living in those two nations, as well as households. The most recent Census was undertaken in 2021 (Office for National Statistics, n.d.-b) which for the first time was predominantly conducted online (88.9% of responses) (Office for National Statistics, n.d.-c, n.d.-d). Open-text questions covering job-title, job duties and the main activity of the firm worked for were collected for subsequent coding to SOC 2020.

In preparation for the 2021 Census in England and Wales, ONS developed an in-house coding tool to automatically code open-text responses including occupation (and industry) data. The coding of occupation and industry variables involved a multi-step process which began with pre-processing to correct spelling errors, truncate words, and remove white spaces, followed by exact matching, fuzzy matching, and dependency-based coding. Exact matching involved direct comparisons, while fuzzy matching used weighted word and phrase comparisons to find the best fit. Uncoded data underwent additional stages including default job title matching, resulting in data being categorized as coded, residual (needing clerical coding), or not codable. Updates were made to the tool throughout the coding process in order to increase the accuracy of the algorithms. In total 65% of occupations were automatically coded, with the remainder being coded by clerical coders. Once automated and clerical coding was complete, the quality of 1% of codings were reviewed by the Census Quality Assurance Team and 89% of occupations were assessed to have met the Quality Key Performance Indicator. (The automated coding tool was also deployed to code responses to other questions including ethnicity, national identity, religion etc. A greater proportion of responses to these questions could be auto-coded and a higher proportion passed quality checks, reflecting the complexity of occupation data).

Although the automated coding in the Census was successful, we are aiming, as part of the Survey Futures project, to provide evidence on measurement differences in occupational coding between the Census data, collected primarily online, and survey data, collected via interview-administered approaches. In both approaches, the occupational data are coded after the interview. We hope to do this by using data from the Census Non-Response Link Study (CNRLS), which is a data-linkage project conducted by the ONS. The study matches households and the individuals within them who participated in the 2021 Census for England and Wales with ONS survey data collected around the same time (3 months before or after). These surveys are the COVID-19 Infection Survey, Living Costs and Food Survey, Survey of

Living Conditions, Labour Force Survey, and Labour Market Survey. The project primarily aims to investigate survey nonresponse by examining differences in the distributions of key variables between participants in the population census, which provides universal coverage, and sample surveys, which typically experience higher levels of nonresponse (O'Farrell et al., 2022).

This data source enables exploration of differences between survey participants whose target variables can be matched in both census and survey data. One potential application of this matching is to analyse occupation data collected through different modes. In the 2021 Census, 4-digit SOC occupation data were collected via online self-completion (free text), while Labour Force Survey data were collected via Computer-Assisted Telephone Interviewing (CATI), with interviewers assigning the 4-digit SOC codes using a list of common occupations. These data provide an opportunity to compare the quality and consistency of occupation coding using these different approaches.

Labour Force Survey (LFS) and Transformed Labour Force Survey (TLFS)

The Labour Force Survey (LFS) is the largest household study conducted in the UK, collecting data on employment circumstances to provide official statistics for both employment and unemployment at the national level. This ongoing survey includes a quarterly sample size of 36,000 households (Office for National Statistics, n.d.-e).

The LFS has typically been conducted face-to-face with telephone follow-up, though during the COVID pandemic data collection was primarily conducted via telephone. As of October 2023, face-to-face interviewing has resumed as the primary mode for Wave 1, with telephone used for Waves 2-5.

Since February 2022, the ONS has been running a self-completion online-first survey, known as the Transformed Labour Force Survey (TLFS) (Office for National Statistics, 2024), alongside the traditional LFS. In the future, it is planned that the TLFS will replace the LFS. As a key source of national statistics on employment, maintaining quality and consistency of occupational coding has been a key element of the methodological testing and development incorporated in the TLFS development. The ONS continue to explore a range of innovative solutions to occupation coding as part of this development. While LFS is administered via Computer-Assisted Personal Interviewing (CAPI) and Computer-Assisted Telephone Interviewing (CATI), TLFS is primarily via self-administered via Computer-Assisted Web Interviewing (CAWI), with some CATI. The online questionnaires, and associated respondent guidance, have been adapted for an online, rather than an interviewer administered context.

The TLFS employs automated coding of occupation questions post-interview. During interviews, respondents are asked open questions about their job title and the main activities they perform in their role. During the data processing stage, these answers are then coded using an automated solution.

The extensive use of the automated coding on the TLFS (and also on the Census and other large ONS surveys) has been informed by previous development work. In particular, in 2010, ONS assessed the quality of the automated coding frame (ACTR) for occupation data collected in the Labour Force Survey (LFS) and Opinions Survey (OPN). They uploaded free-text descriptions of respondents' jobs, compared the codes for the same job from expert manual coders, and found discrepancies for 60% of them at the 4-digit level (Dawe & Wilson, 2021). The tool has since been improved and adapted for use at scale on the 2021 Census and on the TLFS.

We are aiming, as part of the Survey Futures, to do further analysis comparing the quality of automated and clerical coding conducted in TLFS.

2.2 Self-coding during interview

Another recent development in occupational measurement is self-coding, either by the respondent or by the interviewer, during the interview. This interactive and semi-automated approach typically uses look-up tables, where a list of potential occupation codes or categories is automatically generated based on the open-ended answers to questions (see Brugiavini et al., 2017; Peycheva et al., 2021; Tijdens, 2015). Respondents or interviewers then select the code or category from the list displayed. In online surveys, this is done by the respondent themselves, but in interviewer-administered surveys the interviewer and/or respondent can do this, or they can do this together (e.g., Hacking et al., 2006; Mannetje & Kromhout, 2003; Peycheva et al., 2021; Schierholz et al., 2018; Simson et al., 2023). The application of self-coding normally follows the following steps: (1) the respondent provides free text answers about their job, (2) a list of categories is generated by a model or look-up table and presented visually to the respondent (3) the respondent selects the most appropriate category for their occupation (Simson et al., 2023).

We outline below evidence from some studies that have trialled or implemented this approach, in particular from Next Steps, which is the only UK study we are aware of which has used this approach at scale. We include examples of surveys which have piloted this approach in interview-administered settings, as this has often been done as a precursor or test of the feasibility of this approach, prior to potential use in an online context, though sometimes as a test of whether respondent self-coding would give higher quality data in interviewer-administered surveys.

Office for National Statistics (ONS)

The ONS have trialled this approach. In the Opinions and Lifestyle Survey (OPN) 2010, respondents self-coded their answers to occupation questions in a face-to-face interview by using the interviewer's laptops and on-screen instructions for coding. Interviewers did the same, and the codes were compared to the respondents' codes, with about a 68% agreement rate. Using evidence from these two experiments, they determined that it was challenging to decide whether to encourage respondents to self-code their occupation using frames or to carry out manual expert in-house coding (Dawe & Wilson, 2021). Additionally, as part of the Census 2021 development, Verian was commissioned by ONS to test the occupation question with more focus on the online self-completion mode. A look-up approach was investigated, and certain usability issues were reported. Those associated with the look-up functionality and its specifics included respondents not being aware of the ability to overwrite the suggestions, long lists of the occupation options proposed by the function, duplicate entries, and respondents disproportionately selecting one of the first proposed options from the list of suggested occupations (also known as a primacy effect).

European Social Survey (ESS)

ESS have tested a 'job coder' tool as part of the 'Survey Codings' package developed through the SERISS project (https://www.surveycodings.org/articles/codings/industry). The tool involved a look up function where a database was loaded for each country/language. The respondent would start to type their occupation, and options would be presented for selection. This approach was not integrated into the main survey as user testing revealed a number of issues: 1) the database was incomplete or included translation errors for some countries/languages, 2) the database only included the job titles used in the ISC008 classification where the language did not always reflect the way the respondent would describe their job, and 3) many participants did not select a job in the look-up.

1970 British Cohort Study (BCS70)

In 2016, as part of the pilot phase for the BCS70 age 46 Survey, the potential of using a look-up tool to enhance the accuracy of occupational coding was explored. Nurse interviewers inputted respondent-provided keywords and selected corresponding occupation codes from a list of offered options. These look-up codes were then compared to those assigned by office-based coders to assess their accuracy. The results showed that the look-up method used by nurses was less precise than the open-text coding performed by office coders, leading to the decision not to adopt the look-up approach for subsequent survey waves (Morgan & Taylor, 2018).

Next Steps

Next Steps is a longitudinal study following participants in England born in 1989/1990 which commenced in 2004 when participants were aged 13 or 14. There have been nine waves of data collection. Parents were interviewed in the first waves and occupation data was collected via open text for office coding. The last five waves have been conducted using a web-first mixed mode approach. Upon reaching adulthood, Next Steps participants were initially also asked to provide open-text descriptions of their occupations but a look-up approach was introduced in the Age 25 Survey. The questionnaire included a text-based search and a coding system based on the 4-digit SOC 2010 code frame. Respondents participating via web or interviewers in the interviewer-administered modes (F2F and CATI) entered keywords which described their job and which generated a list of potential SOC codes for them to select (see

Table A1 for the exact wording of the open-ended questions). Importantly, interviewers could probe for more information after entering keywords and reviewing the generated list of occupations, potentially adding more information that could result in a correct code. Following an initial pilot study the look-up approach to collecting occupation data was deemed feasible and implemented for the main stage of fieldwork. A look-up approach was used again in the Next Steps Age 32 Survey, though amendments were made to its functionality which intended to increase the accuracy of the matching process. Office coding was conducted when no look-up code was selected. We provide detailed information below about how these look-ups were implemented at age 25 and 32, and the findings from this in relation to occupational measurement. Additionally, we outline methodological research in this area which is planned for the future using data from a recently collected mixed-mode 'measurement lab'.

Next Steps Age 25 (Wave 8). As noted above, the Next Steps Age 25 Survey, a web-first mixed mode survey (with telephone then face-to-face follow-up of non-respondents), implemented a look-up based approach to collecting occupation. The approach was trialled primarily due to concern that online participants would provide insufficient information to allow accurate coding but also to increase cost-effectiveness through reducing the need for office coding. Peycheva et al. (2021) evaluated the approach and found the following key findings:

- During the interview, 82% of respondents were successfully assigned an occupation code, with only 18% requiring office-based coding.
- There were notable differences across modes: the highest coding rate (90%) was achieved through web and telephone interviews, while the F2F mode had a 20% lower rate. The reasons for the mode difference were not explained in the paper.
- The quality of occupational coding was consistent across modes, with comparable proportions being allocated generic codes (those with 0 or 9 as the last digit at the SOC 4-digit level).
- Web respondents spent more time identifying an appropriate code and responding to questions than those interviewed by an interviewer. They also provided longer answers, but without a positive effect on coding rates.
- Socio-demographic characteristics in the web mode and interviewer characteristics in the F2F and CATI modes affected whether an occupation code was assigned. White study members, those who attended university and respondents living with a partner were more likely to be assigned an occupation code. Also, female, younger and more experienced interviewers were more likely to successfully assign an occupation code using the look-up method.

Next Steps Age 32 (wave 9). In the subsequent wave of the Next Steps survey, an experiment was conducted which sought to further explore the use of a look-up approach by comparing the method with office-based coding (Kocar et al., 2025 for further information). As per the Age 25 Survey, participants were first asked to enter their job-title and keywords which described their job into a look-up which suggested a list of SOC codes. Respondents were asked to rate the accuracy of the SOC code they selected and then to provide a full open-text description of their job which was office coded by two independent coders. The main findings of that study were the following:

- The proportion of respondents who were able to select an occupation code from the look-up was lower in the web mode compared to the interviewer-administered modes (82% Web, and 88% F2F and 93% CATI[2]). This was different to the finding in the Next Steps Age 25 Survey.
- Office coders successfully assigned an occupation code for more than 99% of all respondents, with no observable differences between the modes, including for almost 97% of those respondents who could not select a code during the interview using the look-up.
- In comparison to the interviewer-administered modes (F2F, CATI), four reasons were identified for why coding rates in the Web mode were lower: higher nonresponse to job title and job description open-ended questions, a higher propensity for having to edit job information in the lookup function (associated with lower coding rates), and a higher refusal rate. These sources of missingness explain a proportion of the differences in coding rates between the modes, but a notable gap of about five percentage points remains.
- Consistently with evidence presented in Peycheva et al. (2021), the data revealed that Web respondents provided longer job titles, job descriptions (keywords), and open-descriptions (as a separate request after the look-up). While longer job title text entries increased coding rates, longer job descriptions (i.e., more than 25 characters) in the look-up negatively affected coding rates.
- The 'look up-office coding' agreement rates were between 62% (Web) and 70% (CATI) at the 4-digit level and 78% (Web) and 81% (CATI) at the 1-digit level. These differences between the modes were not statistically significant. Longer job descriptions in the look-up negatively affected agreement rates as well.
- On average, although consistency across modes was generally similar, F2F respondents tended to feel that the selected occupation code more accurately represented their job compared to web respondents. This difference may be attributed to the additional assistance provided by interviewers in determining the most appropriate code. As expected, agreement rates were higher when respondents felt that the occupation code more closely reflected their job.
- The look up-office coding agreement rate was significantly lower than the agreement between two office coders.

[2] The LVI subsample was very small (n=8) and was combined with the in-person sample into the F2F subsample sample.

Next Steps Measurement Lab, 2024. To expand on the findings of the initial studies conducted by CLS (Kocar et al., 2025; Peycheva et al., 2021), a two-wave mixed-mode survey experiment was designed as part of the Next Steps Measurement Lab project. Nonprobability sampling, incorporating elements of convenience sampling and quota sampling, was employed to roughly balance the recruited sample across various characteristics: region, age (ranging from 20 to 40 years), working status, social grade, and education. Face-to-face recruitment was carried out across different regions to enrol 1500 participants for the experiment. These participants were randomly assigned to nine experimental groups, categorised by the mode of completion (Web, live video interviewing, in-person) in each wave. Specifically, three experimental groups were interviewed in Wave 2 using the same mode as in Wave 1, while the remaining six groups switched modes between the waves. In contrast to the prior Next Steps studies focusing on occupation coding (Kocar et al., 2025; Peycheva et al., 2021), this survey design facilitates the examination of occupation coding effects associated with the mode of administration, namely measurement mode effects. This entails exploring the impact of mode switching on the collection of occupation data and subsequent coding consistency. The primary objectives of this experiment are as follows:

- To identify any differences between the modes in Wave 1 regarding the proportion of respondents able to select an occupation code.
- To detect any changes in the selected occupation over time influenced by the mode change (i.e., between Waves 1 and 2).
- To identify any shifts in the consistency between look-up and office coding, conditional on the mode change between Waves 1 and 2. To determine any differences in the length of the look-up keywords and open-ended descriptions for office coding, also conditional on the mode change between Waves 1 and 2.

2.3 Closed questions

A number of online surveys have used closed questions on occupation based on the SOC groups as an alternative to or in combination with open-text questions. For example, surveys may offer broad occupation categories, such as the ten 1-digit classifications of the ISCO system (Tijdens, 2014). While this approach provides less detailed information than open-ended coding, it can be sufficient for many types of analysis, and is less expensive and easier to implement. Some examples of surveys which have used this approach in the UK are the Community Life Survey, Participation Survey, Public Attitudes Tracker Survey of Attitudes to the Environment.

Government surveys conducted on the NatCen Opinion Panel

The panel managed by NatCen is a general population probability-based online panel. It is primarily used for cross-sectional data collection for commissioned academic and government research on different topics. Survey data are typically collected using a sequential mixed-mode (CAWI/CATI) fieldwork design (National Centre for Social Research, n.d.). A number of innovations in occupational coding using closed questions have been trialed on the panel. This has been both as part of the core questionnaire and as part of data collected for specific clients.

An example of such a survey is the 2023 Skills and Employment Survey, funded by the Advisory, Conciliation and Arbitration Service, Department for Education, and the Economic and Social Research Council (Wales Institute of Social and Economic Research and Data, n.d.). It was conducted using the NatCen panel, with interviews conducted via web or telephone and occupation data collected using an open-ended question, followed by a closed question with top-level SOC categories (26 categories in total). The same approach was used in a survey conducted on the panel for the Health and Safety Executive in March 2022. In research conducted in November 2022 for His Majesty's Revenue & Customs, only the closed SOC question including 26 categories was used.

In the two studies which used both approaches, the open-text descriptions were office-coded. The NatCen panel has also experimented with the use of character length checks. As part of Survey Futures, we intend to investigate the consistency of the office codes and the self-administered top-level coding, in order to assess the accuracy of this approach and to evaluate the impact of character length checks on office coding quality.

3. Conclusions

3.1 Summary of key findings and insights from the evidence review

This evidence review reveals various practices in occupation coding, in the UK and worldwide. A general finding of the review is that occupation coding has been developing hand in hand with the development of survey practice, with an introduction of new statistical and machine learning techniques and algorithms, as well as the emergence of AI. Some of these were purposely introduced with self-completion in mind.

On the one hand, collection of open-ended descriptions of jobs, which might include collection of some other associated questions and information, and office coding by professional coders, remains considered the gold standard in occupation coding. This should generally apply to self-completion surveys as well. On the other hand, the evidence suggests that even the gold standard approach is not without error, evidenced in a moderate inter-coding reliability in a number of studies. It is also less time- or cost-efficient, and collecting such data in self-completion surveys without the presence and assistance of an interviewer can be challenging. For those reasons, innovative approaches for self-coding (e.g., a look-up) and automated coding (e.g., machine learning-supported coding) have been tested and methodologically assessed. Some of those relatively recently introduced approaches are still combined with open-descriptions and manual coding, which can lead to better accuracy and can also provide the insights required to improve the quality of self-coding and automated coding. Comparing occupation data collected in surveys with accurate administrative records would be a hugely useful way of assessing the quality of occupation data collected using different approaches, including the "gold standard" office coding approach. However, it is unfortunate that, at least in the UK, it is not thought that such an administrative dataset exists.

We can argue that the development of occupation coding in the UK followed a similar path as in other countries, including the U.S. (e.g., Russ et al., 2023), Germany (e.g., Schierholz et al., 2018), and the Netherlands (e.g., Tijdens, 2014). In practice, it means that new solutions have been introduced and tested in the UK context – examples are an introduction of a look-up method in longitudinal studies, including in the self-completion mode, and (somewhat limited attempts) for automated coding in official statistics (e.g., Transformed Labour Force Survey, Census). Both of those solutions have been compared to more traditional collection of open-ended descriptions and office-coding.

Importantly, methodological research has also been conducted and these new approaches tested and evaluated in the UK context by the ONS, fieldwork agencies, and academia (e.g., CLS at UCL). The results have been reported at events such as GenPopWeb2 workshops, in methodological and technical reports (see the References), and in scientific publications (e.g., Peycheva et al., 2021). The recommendations for best practices in occupation coding in the UK, which are presented in the subsequent section of this Conclusion, are predominantly based on these findings.

Additionally, there are several UK studies and projects either in progress or planned in the near future, also as part of the Survey Futures project (n.d.), that will provide further evidence on some of the most suitable occupation coding solutions.

3.2 Recommendations for best practices in occupation coding in the UK

The existing evidence from the UK and more broadly suggests three groups of general recommendations for occupation coding practice in social surveys:

- Recommendations for methodological solutions for data collection (including self-coding).
- Recommendations for methodological solutions for coding open-descriptions of jobs.
- Recommendations for further methodological research, which are addressed in the last section of this evidence review.

For methodological solutions for data collection, we generally recommend a wider application of the look-up approach in self-completion surveys. The existing evidence indicates that between 80% and 90% of respondents participating via the Web can successfully select a code with no assistance from an interviewer, with more than 90% of respondents with a missing look-up code later being assigned a code by professional coders. The most suitable solution would therefore be to ask respondents whose search does not result in a suitable code being listed after entering search terms, to provide longer descriptions of their jobs with as much detail as possible for subsequent office coding. Evidence from the Next Steps study suggests that there are certain sub-groups where agreement rates between look-up coding and office-coding are particularly low – for example, those who do not feel the code they selected accurately described their job, those who edited their text entries, those who selected occupations lower down the list of presented occupations and those in particular forms of work. Supplementing the look-up approach with the collection of open-text job descriptions and office-coding for this sub-group only (perhaps 15-20%) could significantly reduce manual coding costs without significant impact on coding accuracy (Kocar et al., 2025).

Regarding methodological solutions for coding open-descriptions, this evidence review showed that automated coding algorithms have, with the exception of by ONS, not been extensively used and tested in the UK. This is an area where in other countries, the field has advanced, but there have been limited evaluation of different machine learning techniques as yet (at least not documented in the UK). With a wider application of AI in social research in general, there are opportunities for using AI to code occupation data, which has been tested in other contexts. Although the evidence from other countries (as well as from an ONS study) showed that automated coding is less accurate, it could be combined with manual coding, especially for codes with a lower predicted accuracy score. With further development of coding algorithms and AI, a wider use of automated coding will likely become more feasible. As different countries use their own national standard occupational classifications, evidence from other countries can only be partially extended to other contexts. Consequently, such methodological research would ideally be carried out in the UK with UK survey or census data and the national SOC classification as the frame.

3.3 Suggestions for future research directions and advancements in occupation coding methodologies in the UK

Regarding the methodological research of occupation coding solutions, we strongly recommend further methodological research to identify and test the best existing practices, and develop and refine methods in both self-classification and manual coding. As previously noted, there are a number of UK studies already either in progress or planned, and some additional research is required to answer other relevant questions. These include the effect of mode on occupation coding at an individual level in longitudinal studies, the role of administrative data in assessing (and potentially improving) the quality of occupation coding, the role of AI in automated coding of occupation, and the further improvement of the look-up function for self-completion surveys. The latter suggestion includes investigation into what information and instruction should be provided to web respondents to increase the quality of information they provide and increase the likelihood that they will successfully (and accurately) select the most suitable code. There might also be opportunities to assess to what extent AI could be included in the look-up approach, if the self-coding process could be somewhat interactive with an addition of language-based models, and how including AI affects both the coding rate and the accuracy relative to a “standard” look-up approach, manual or automated coding.

4. References

Belloni, M., Brugiavini, A., Meschi, E., & Tijdens, K. (2016). Measuring and detecting errors in occupational coding: an analysis of share data. *Journal of Official Statistics, 32*(4), 917-945.

Brugiavini, A., Belloni, M., Martens, M., & Buia, R. E. (2017). The "Job Coder". In *SHARE Wave 6: Panel innovations and collecting Dried Blood Spots* (pp. 51-70). MEA, Max Planck Institute for Social Law and Social Policy.

Burstyn, I., Slutsky, A., Lee, D. G., Singer, A. B., An, Y., & Michael, Y. L. (2014). Beyond crosswalks: reliability of exposure assessment following automated coding of free-text job descriptions for occupational epidemiology. *Annals of occupational hygiene*, *58*(4), 482-492.

Centre for Longitudinal Studies. (2018). *BCS70 2016-18 survey CAPI Questionnaire Documentation*. Available at: https://cls.ucl.ac.uk/wp-content/uploads/2018/10/BCS70_collated-CAPI-specs_FINAL.pdf

Conrad, F. G., Couper, M. P., & Sakshaug, J. W. (2016). Classifying open-ended reports: Factors affecting the reliability of occupation codes. *Journal of Official Statistics*, *32*(1), 75-92.

Dawe, F., & Wilson, L. (2021, March 12). Building Web Collection Capability for ONS Social Surveys [Workshop presentation]. *GenPop2 Event, March 2021*.

Elias, P., Birch, M., & Ellison, R. (2014). *CASCOT international version 5 user guide*. Institute for Employment Research, University of Warwick, Coventry. Available at: https://warwick.ac.uk/fac/soc/ier/software/cascot/internat/cascot_international_user_guide.pptx

Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M., & Steiner, S. (2017). Three methods for occupation coding based on statistical learning. *Journal of Official Statistics*, *33*(1), 101-122.

Hacking, W., Michiels, J., & Janssen-Jansen, S. (2006). Computer assisted coding by interviewers. In *Proc. 10th Int. Blaise Users Conf* (pp. 283-296).

Hanson, T. (2021, July 14). Developing self-completion instruments for the European Social Survey [Workshop presentation]. *Designing and delivering online surveys - Transitioning long questionnaires to online workshop*, virtual.

Helppie-McFall, B., & Sonnega, A. (2018). Feasibility and reliability of automated coding of occupation in the health and retirement study. *Michigan Retirement Research Center Research Paper*, (2018-392).

Hoffmann, E., & Thomas, R. (1995). *What kind of work do you do? Data collection and processing strategies when measuring" occupation" for statistical surveys and administrative records*. International Labour Organization.

International Labour Organization. (n.d.). *International Standard Classification of Occupations (ISCO)*. Retrieved June 24, 2024, from https://www.ilo.org/public/english/bureau/stat/isco/

Ipsos & University College London. (2023). *Children of the 2020s Wave 1 Technical Report*. https://assets.publishing.service.gov.uk/media/65671ebdd6ad75000d02fc89/Children_of_the_2020s_technical_report.pdf

Kocar, S., Peycheva, D., Brown, M., Sakshaug, J., Bhaumik, C., & Calderwood, L. (2025). *Navigating Challenges in Occupation Data Collection via Web Interviews: Insights into the Look-Up Approach* [Manuscript submitted for publication]. Centre for Longitudinal Studies, University College London.

Lyberg, L., & Dean, P. (1992). *Automated Coding of Survey Responses: An International Review. R&D Reports (1992–2)*. Statistics Sweden, Stockholm, Sweden. Available at: https://www.scb.se/contentassets/7c4edb581f8745e3a081e1ba9b332eb4/rnd-report-1992-02-green.pdf.

Mannetje, A. T., & Kromhout, H. (2003). The use of occupation and industry classifications in general population studies. *International Journal of Epidemiology*, *32*(3), 419-428.

Morgan, H., & Taylor, K. (2018). *BCS70 technical report Age 46 Survey*. Available at: https://cls.ucl.ac.uk/wp-content/uploads/2019/03/BCS46_Technical-Report_FINAL.pdf

National Centre for Social Research. (n.d.). *NatCen Opinion Panel.* Retrieved May 23, 2024, from https://natcen.ac.uk/centres/natcen-panel

O'Farrell, K., Higham-Lloyd, S., Kilner, H., Nelson, D., & Pike, O. (2022). *CNRLS 2021 Quality Report*. Office for National Statistics.

Office for National Statistics. (n.d.-a). *Standard Occupational Classification (SOC)*. Retrieved May 23, 2024, from https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc

Office for National Statistics. (n.d.-b). *About census*. Retrieved May 23, 2024, from https://www.ons.gov.uk/census/aboutcensus

Office for National Statistics. (n.d.-c). *The National Statistics Socio-economic classification (NS-SEC)*. Retrieved May 23, 2024, from https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010#deriving-the-ns-sec-self-coded-method

Office for National Statistics. (n.d.-d). *Occupation (current) variable: Census 2021*. Retrieved May 23, 2024, from https://www.ons.gov.uk/census/census2021dictionary/variablesbytopic/labourmarketvariablescensus2021/occupationcurrent

Office for National Statistics. (n.d.-e). *Labour Force Survey*. Retrieved May 23, 2024, from https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/labourforcesurvey

Office for National Statistics. (2024, May 23). *Transformed Labour Force Survey – user guidance* https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/transformedlabourforcesurveyuserguidance

Office for National Statistics (2022). *Occupational data in ONS Surveys,* from https://www.ons.gov.uk/news/statementsandletters/occupationaldatainonssurveys

Ossiander, E. M., & Milham, S. (2006). A computer system for coding occupation. *American journal of industrial medicine*, *49*(10), 854-857.

Peycheva, D. N., Sakshaug, J. W., & Calderwood, L. (2021). Occupation coding during the interview in a web-first sequential mixed-mode survey. *Journal of Official Statistics*, *37*(4), 981-1007.

Russ, D. E., Josse, P., Remen, T., Hofmann, J. N., Purdue, M. P., Siemiatycki, J., Silverman, D. T., Zhang, Y., Lavoué, J., & Friesen, M. C. (2023). Evaluation of the updated SOCcer v2 algorithm for coding free-text job descriptions in three epidemiologic studies. *Annals of Work Exposures and Health*, *67*(6), 772-783.

Safikhani, P., Avetisyan, H., Föste-Eggers, D., & Broneske, D. (2023). Automated occupation coding with hierarchical features: A data-centric approach to classification with pre-trained language models. *Discover Artificial Intelligence*, *3*(1), 6.

Schierholz, M., Gensicke, M., Tschersich, N., & Kreuter, F. (2018). Occupation coding during the interview. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *181*(2), 379-407.

Schierholz, M., & Schonlau, M. (2021). Machine learning for occupation coding—A comparison study. *Journal of Survey Statistics and Methodology*, *9*(5), 1013-1034.

Simson, J., Kononykhina, O., & Schierholz, M. (2023). Measurement: A Comprehensive Toolbox for Interactive Occupation Coding in Surveys. *Journal of Open Source Software*, *8*(88), 5505.

SurveyCodings. (n.d.). *SurveyCodings – Occupation*. Retrieved May 23, 2024, from https://www.surveycodings.org/articles/codings/occupation

Survey Futures. (n.d.). *Survey Futures - Survey data collection methods collaboration*. Retrieved May 23, 2024, from https://surveyfutures.net

Tijdens, K. G. (2014). Reviewing the measurement and comparison of occupations across Europe. *AIAS Working Paper.*

Tijdens, K. (2015). Self-identification of occupation in web surveys: requirements for search trees and look-up tables. *Survey Methods: Insights from the Field*, 1-11. https://doi.org/10.13094/SMIF-2015-00008.

Tijdens, K. (2022). The importance of occupation coding quality: lessons for EU-SILC from SHARE and other international surveys. In P. Lynn & L. Lyberg (Eds.) *Improving the measurement of poverty and social exclusion in Europe: reducing non-sampling errors*, 413.

Understanding Society. (n.d.). *About the Study*. Retrieved May 23, 2024, from https://www.understandingsociety.ac.uk/about/about-the-study/

Understanding Society. (2023). *Understanding Society Innovation Panel wave 15: results from methodological experiments*. https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2023-10

U.S. Bureau of Labor Statistics. (n.d.). *Standard Occupational Classification*. https://www.bls.gov/soc/

Wales Institute of Social and Economic Research and Data. (n.d.). *Skills and Employment Survey 2023*. Retrieved May 23, 2024, from https://wiserd.ac.uk/project/ses/ses2023/

Appendices

Appendix A

Table A1: Examples of open-ended occupation coding questions

Survey	Question wording
International Social Survey Programme (ISSP) 2008 Survey[a]	And in your current job, what is your main occupation? (If you are not working now, please tell us about your last job.)
Higher Education as a Generator of Strategic Competences (HEGESCO)2008 Survey[a]	What is your current occupation or job title?
European Social Survey 2012/2013[a]	What is/was the name or title of your main job?
International Social Survey Programme (ISSP) 2010 Survey[a]	What kind of work (do you/did you) normally do? That is, what (is/was) your job called?
British Household Panel Survey 2013 (UK) [a]	What was your (main) job last week? Please tell me the exact job title and describe fully the sort of work you do.
Next Steps Age 25 Survey, 2015-2016 (UK)[b]	What is your current job title? & Please enter keywords which describe what you mainly do in your job into the box below and then select the most appropriate option. *(for a look up)* Please describe in your own words what you mainly do in your job. *(open-description for office coding)*
Understanding Society Innovation Panel wave 15, 2022 (UK)[c]	What was your job? Please tell us your job title and describe fully the work that you did. If you had more than one job, please describe the job that was the highest paid. If equal earnings, then describe the job that was the most hours.
occupationMeasurement (R package)[d]	What is your main occupation at the moment? *(for a look-up)*
Next Steps Age 32 Survey, 2022-2023 (UK)[e]	What is your job title? & Please tell us keywords which describe what you do in your job. *(for a look up)* Could you also describe in your own words what you mainly do in your job? Please describe in detail (for example, the type of work, the department you are in, and what level you work at). *(open-description for office coding)*

Sources: [a]Tijdens (2014), [b]Peycheva et al. (2021), [c]Understanding Society (2023), [d]Simson et al., 2023, [e]Kocar et al. (2025)

www.surveyfutures.net