

# GRADUATE OUTCOMES SURVEY QUALITY REPORT

LUCY VAN ESSEN-FISHMAN, LEAD POLICY & RESEARCH  
ANALYST  
DAN COOK, HEAD OF POLICY & DEVELOPMENT  
JUNE 2020



**Official Statistics**

HESA

95 Promenade

Cheltenham

GL50 1HZ

E [official.statistics@hesa.ac.uk](mailto:official.statistics@hesa.ac.uk)

T +44 (0) 1242 388 513

W [www.hesa.ac.uk](http://www.hesa.ac.uk)

Higher Education Statistics Agency Limited is a company limited by guarantee, registered in England at 95 Promenade, Cheltenham, GL50 1HZ. Registered No. 02766993. Registered Charity No. 1039709. Certified to ISO 27001. The members are Universities UK and GuildHE.

# CONTENTS

|  |    |
|--|----|
| About this document .....  | 4  |
| Executive summary.....   | 5  |
| 1. Introduction .....  | 8  |
| 2. Quality methodology .....   | 9  |
| 3. Quality description .....   | 10 |
| 3.1. Relevance.....  | 10 |
| 3.1.1. Users and user needs.....                                     | 10 |
| 3.1.2. Data and statistical concepts .....                           | 12 |
| 3.1.3. Assessment of gaps .....                                      | 15 |
| 3.2. Accuracy and reliability .....                                  | 16 |
| 3.2.1. The sampling frame, and how it is maintained .....            | 16 |
| 3.2.1.1. Sampling frame data based on HESA data collections.....     | 16 |
| 3.2.1.2. Sampling frame data based on other ingested data .....      | 19 |
| 3.2.2. How does the sampling frame relate to the population?.....    | 20 |
| 3.2.3. The sample.....   | 21 |
| 3.2.4. Sampling error and non-response error .....                   | 22 |
| 3.2.4.1. Unit non-response error .....                               | 23 |
| 3.2.4.2. Item non-response error.....                                | 26 |
| 3.2.5. Proxy responses.....  | 27 |
| 3.2.6. Measurement error .....                                       | 27 |
| 3.2.6.1. Respondent error .....                                      | 28 |
| 3.2.6.2. Survey instrument error.....                                | 30 |
| 3.2.6.3. Interviewer error .....                                     | 33 |
| 3.2.7. Mode effects.....   | 34 |
| 3.2.8. Reliability of sensitive data .....                           | 36 |
| 3.2.9. Processing error .....  | 38 |
| 3.2.9.1. Imputation and editing .....                                | 38 |
| 3.2.9.2. SIC and SOC coding.....                                     | 39 |
| 3.2.9.3. Handling free text responses.....                           | 40 |
| 3.2.9.4. Location of employment data (UK) – handling free text ..... | 40 |
| 3.2.9.5. Employment country data – handling free text.....           | 43 |
| 3.2.9.6. Salary data – handling free text.....                       | 43 |
| 3.2.9.7. Other free text cleaning – frequencies.....                 | 44 |
| 3.3. Timeliness and punctuality .....                                | 45 |
| 3.3.1. Timeliness and user needs.....                                | 45 |
| 3.3.2. Production timeline .....                                     | 46 |

|   |    |
|---|----|
| 3.3.3. Frequency of production .....                        | 47 |
| 3.4. Accessibility and clarity .....                        | 47 |
| 3.4.1. Confidentiality and disclosure control.....          | 47 |
| 3.4.2. Statistical products and supporting information..... | 49 |
| 3.4.3. Access and use .....                                 | 50 |
| 3.5. Coherence and comparability.....                       | 51 |
| 3.5.1. National and international data standards.....       | 51 |
| 3.5.2. Graduate Outcomes and other data on graduates .....  | 54 |
| 3.5.3. Comparability and time series.....                   | 56 |
| 4. Conclusion .....   | 58 |
| 5. References .....   | 60 |

## ABOUT THIS DOCUMENT

This report offers the most comprehensive current assessment of the strengths and weaknesses of the Graduate Outcomes data available currently, as well as providing information on any known specific quality issues. The primary audience is intended to be data analysts and other users who need more detailed information about the quality characteristics of the Graduate Outcomes data. It also acts as an advanced user's guide to further information HESA has published on Graduate Outcomes, most especially where to look for technical specifications, papers, and reports of interest to analysts. The executive summary offers an overview of the contents, including a digest of the most important points.

We would like to express thanks to our many colleagues who contributed their time, expertise, and analysis to this report.

## EXECUTIVE SUMMARY

Graduate Outcomes is a new national survey of students completing courses of higher education (HE). It is conducted across the UK and seeks to survey the entire graduate population. It is the largest annual social survey in the country and is run by the Higher Education Statistics Agency (HESA) – a producer of official statistics. In the first year of publication from the survey, we are badging our outputs as ‘experimental statistics’. Experimental statistics are newly developed or innovative official statistics undergoing evaluation. They are published with the aim of involving users and stakeholders in the assessment of their suitability and quality.

Users should exercise caution when using data from experimental statistics and evaluate the quality and coverage of any data they intend to use in the context of the intended application to ensure that it is fit for their purpose. This quality report has been written to help analysts do just that. We are keen to hear what users think of the products. Contact our Official Statistics team ([official.statistics@hesa.ac.uk](mailto:official.statistics@hesa.ac.uk) or (0)1242 388 513 [option 2]) with feedback and suggestions.

### Relevance

[For more on relevance, see section 3.1.](#)

Graduate Outcomes data has been designed to be **relevant** to a wide range of user needs. The data reflects what we know about the requirements of prospective and current students; graduate employers; the HE sector; central, devolved and local government; the press; and civil society, to have access to an independent and trusted source of information about graduates. It covers longstanding areas of interest in the activities graduates are doing, including whether they are in work or further study, and what their job or course is about. The survey also collects new data reflecting on the experience of being a graduate, their subjective wellbeing, and information about the characteristics of self-employment.

### Accuracy and reliability

[For more on accuracy and reliability, see section 3.2.](#)

The survey offers information sourced directly from graduates, and this report explores the extent to which this can be relied upon as **accurate**. While no social survey can offer the individual-level precision of an administrative data source, the scope of topics in Graduate Outcomes is much broader. Graduate Outcomes’ sample size and response rate is much larger than for other surveys. We have found the data to be representative of the population for most statistical purposes. Our decision not to weight the data was based on rigorous research reviewed by our peers.

Table 1 - overall 2017/18 survey response rates by group (full responses)

| Main target group       | Target | 2017/18 response rate |
|-------------------------|--------|-----------------------|
| UK domiciled, full-time | 60%    | 52.3%                 |
| UK domiciled, part-time | 60%    | 48.7%                 |
| Research funded         | 65%    | 58.0%                 |
| EU domiciled            | 45%    | 46.1%                 |
| Non-EU domiciled        | 25%    | 29.4%                 |

Many users wish to analyse sub-samples of data about graduates. Sample sizes are important when using disaggregated data. Analysts should consider the sample sizes, and any uncertainty that generates. HESA has published confidence intervals on key data tabulations to assist in understanding how **reliable** the data is. This quality report also explains sources of known or potential bias we have identified, to help analysts decide how they should use the data, safely. We

offer specific advice around using the data for regional or sub-regional geographic analysis. We also describe our survey instrument and processing approach in detail. One such section explains the creation of occupational and industrial classifications, and our confidence in what has been produced.

### **Timeliness and punctuality**

[For more on timeliness and punctuality, see section 3.3.](#)

Through extensive consultation with users and stakeholders, the census week at 15 months after course completion was determined as the best point at which to balance the need to generate meaningful insights into career and other outcomes with the need to deliver good rates of survey response. One implication of this is that our statistics include those who went straight on to postgraduate studies after their bachelor's degree and who may only just have finished at time of survey. Depending on onward use it may not be appropriate or **timely** to compare those who have spent 15 months in the labour market with those just graduating from a further qualification and graduate responses to the survey may be driven more by the second qualification achieved. Our publications make filtering these individuals from the data easy to achieve.

The Graduate Outcomes survey from which this year's statistical releases have been derived concluded prior to the rise of the coronavirus in the UK, so the statistical data presented is unaffected by it. Future iterations of the survey are likely to be affected by the pandemic and we will assess any implications and report on those in future editions of this quality report.

### **Accessibility and clarity**

[For more on accessibility and clarity, see section 3.4.](#)

The Graduate Outcomes data is designed to be **accessible**, and users can view the data on our website, and download our data to perform their own analysis and visualisations. Open data is released under a Creative Commons 4.0 CC-BY license. We are also making aggregate and disaggregated survey data available through our data processor, Jisc. To find out more about how Jisc data analytics can help you, see: [www.jisc.ac.uk/data-and-analytics](http://www.jisc.ac.uk/data-and-analytics)

The data release is accompanied by a comprehensive range of supporting information. Besides this quality report, users can find a methodology statement, coding manuals, reports, blogs, and detailed guidance on our website. There is a lot of information available, which can be daunting. We provide **clear** access routes to this information from the publications themselves and for visitors navigating to our website. For more advanced Graduate Outcomes users looking for a digest of the (largest single body of) information on our website, we recommend using this quality report to guide you.

### **Coherence and comparability**

[For more on coherence and comparability, see section 3.5.](#)

Graduate Outcomes forms the newest member of a family of exceptionally rich information about the HE sector. It **coheres** with the HESA Student records (and other data about HE in further education (FE) settings) to which it can be linked. It can also be used in conjunction with other data HESA collects about HE providers, their staff, finances, estates, and interactions with business and society. This survey replaces the former Destinations of Leavers from Higher Education (DLHE) survey and differs from it significantly in a number of ways that are explored later in this document. Data and statistics drawn from these two surveys are not directly comparable. HESA advises all data users against attempting to directly compare data between Graduate Outcomes and DLHE. Any such comparisons are likely to generate highly questionable results that are open to misinterpretation.

The Graduate Outcomes survey offers a rich and regular source of information collected directly from graduates themselves, offering their perceptions of their outcomes to date, as well as factual information about the kind of work they are doing, their salary and contractual status, or their further study options. This presents a breadth and level of detail about graduate experiences unparalleled in any other data source. It offers context to the tax and benefits data at the core of the Longitudinal Educational Outcomes (LEO) data from the Department for Education. It also complements the Labour Force Survey (LFS) by shining a spotlight on recent graduates and their activities. Our future plans include assessing how **comparable** our data is with similar variables in these other data sources. This will provide users further understanding of the **quality** of the Graduate Outcomes data, to increase **trust** in our data source and methods, and to demonstrate the **value** the survey offers to our understanding of society.



## 1. INTRODUCTION

As a producer of official statistics, HESA is under an obligation to demonstrate the quality of its statistical outputs. This obligation is both a formal one, inasmuch as it is specified in Section Q3 of the Code of Practice for Statistics, and a more pragmatic one, inasmuch as, by demonstrating the quality of its outputs, we can provide our stakeholders with information which will support them in the use of our statistics.

Although year one of the Graduate Outcomes survey is being released as experimental statistics, experimental statistics are a category of official statistics, and the Code of Practice therefore applies just as clearly to this first set of experimental statistics as it does to any of our other statistical outputs.<sup>1</sup> We have therefore worked to integrate the principles of the Code of Practice for Statistics into our every stage of work on the Graduate Outcomes survey.

HESA data is used by a wide variety of stakeholders, and their need for high quality data provides us with further motivation for demonstrating the quality of our statistical products. Data from the Destinations of Leavers from Higher Education (DLHE) survey, the forerunner of the Graduate Outcomes survey, was used not only by HE providers and prospective students, but also by a wide range of policy makers, researchers, and media outlets, and it will be important for Graduate Outcomes to meet the needs of this same varied group of stakeholders.

As part of the review which led to the creation of the Graduate Outcomes survey, HESA conducted a quality assurance self-assessment of DLHE.<sup>2</sup> This self-assessment found that, while some aspects of DLHE achieved the highest level of quality assurance, other aspects did not meet the same standard, and that the survey as a whole therefore did not meet the desired level of quality assurance. The self-assessment concluded with a list of quality assurance recommendations for the successor to DLHE; as well as allowing us to make a more general determination of the quality of the Graduate Outcomes outputs, this report gives us an opportunity to demonstrate the progress HESA has made towards addressing those recommendations in the design and implementation of the new Graduate Outcomes survey. A full assessment of these will be conducted during the forthcoming post-implementation review of Graduate Outcomes.

---

<sup>1</sup> See our blog post on this issue for more details <https://www.hesa.ac.uk/blog/18-03-2020/true-method-knowledge-experiment-why-graduate-outcomes-statistics-are-experimental>

<sup>2</sup> See [https://www.hesa.ac.uk/files/Quality\\_assurance\\_self\\_assessment.pdf](https://www.hesa.ac.uk/files/Quality_assurance_self_assessment.pdf)

## 2. QUALITY METHODOLOGY

HESA builds the Code of Practice for Statistics<sup>3</sup> into all aspects of its work. At HESA, quality management is an overarching practice that is prioritised in each part of the statistical business process. We operate appropriate quality regimes for each aspect of our work, and at the time of writing we are moving towards bringing these approaches together in a single overarching quality policy. For this quality report, we have taken the following approach.

First, we base our approach on the guidance offered by the National Statistician on survey quality measurement, by structuring our report around the five dimensions of quality outlined in the European Statistical System.<sup>4</sup>

Second, we have already created a methodology document,<sup>5</sup> and a range of supporting materials including a Graduate Outcomes-specific dissemination policy.<sup>6</sup> These documents are cross-referenced as needed in this document, as they form part of the evidence base for it. Sometimes, for ease of reading, there will be some repetition between this document and others we have published, though we have attempted to keep this to a minimum. Our future intention is to bring these several documents together in a single user guide for the Graduate Outcomes survey. This document is our first step towards that approach.

Third, our aim and purpose in writing this document is to offer the most up-to-date assessment of the quality characteristics of the Graduate Outcomes survey. In doing so, we have necessarily prioritised our own uses and outputs first, as these are the first national data to be presented from the survey, and these take into account the many user requirements we have already elicited. However, at this early stage in the Graduate Outcomes survey's development, a quality report cannot be as comprehensive as one that follows a period of extensive usage by other users. Notably, our own initial uses are mainly for the release of aggregated data, which is filterable by multiple characteristics, but ultimately still a summary of findings. Users of survey microdata are urged to carry out their own quality assessments of the Graduate Outcomes data, especially in areas where our own work does not provide them with the understanding they need to have confidence in the validity of their analysis. This approach will extend and enhance the work we have begun here. We urge all microdata users to publish their quality findings (and to share them with us at the earliest opportunity) so that this work can be extended over time, for the benefit of all users.

Fourth, although this is a technical report about statistics, it follows a narrative format. Our assessments and evaluations of quality characteristics are presented using a predominantly narrative approach.

---

<sup>3</sup> UK Statistics Authority. Code of Practice for Statistics. <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf>

<sup>4</sup> UK Statistics Authority. 2013. *Guidelines for Measuring Statistical Output Quality*: <https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/01/Guidelines-for-Measuring-Statistical-Outputs-Quality.pdf>

European Statistical System Committee. 2017. *European Statistics Code of Practice*. <https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7>

<sup>5</sup> In two parts, available from: <https://www.hesa.ac.uk/data-and-analysis/graduates>

<sup>6</sup> See <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

### 3. QUALITY DESCRIPTION

The following subsections detail our assessment of the quality of the survey.

We will be assessing the quality of the survey according to the five dimensions of quality specified in the European Statistical System. These are:

- Relevance
- Accuracy and reliability
- Timeliness and punctuality
- Accessibility and clarity
- Coherence and comparability.

These dimensions, which are recommended for use in measuring survey quality in the National Statistician's guidance, also map onto aspects of quality which must be assured according to section Q3 of the Code of Practice for Statistics. The following sections cover each of the dimensions in turn, exploring the quality characteristics of the Graduate Outcomes survey by utilising the relevant quality indicators and measures identified in the guidance.

#### 3.1. RELEVANCE

When considered as a dimension of statistical quality, relevance refers to the extent to which statistical outputs meet the current and potential needs of users. In order to assess relevance, it is necessary first to identify likely users of the data and their needs. The data sources and statistical concepts used in the production of a statistical output are also a factor in determining relevance; depending on user needs, different data sources and classification schemes will be appropriate. Finally, it is important to identify any gaps between the statistical output and known user needs and to assess how those gaps may be filled in future.

##### 3.1.1. USERS AND USER NEEDS

A wide variety of users, in the HE sector and beyond, work with HESA data on graduates. As Designated Data Body for England, HESA is required by law to publish 'appropriate information relating to registered HE providers and the higher education courses they provide'. According to the Higher Education and Research Act 2017 (HERA), the category of 'appropriate information' includes information which may be helpful to students in higher education, potential higher education students, and HE providers; HERA also specifies that the designated data body must provide appropriate information to the Office for Students (OfS), UK Research and Innovation (UKRI), and the Secretary of State for Education.<sup>7</sup> Comparable expectations are placed upon us by the Devolved Administrations. In addition to those users whose needs we are required by law to consider, we also wish to consider the needs of others for whom high quality data on graduates will be useful, including HE funding and regulatory bodies, local and national governmental agencies, graduate employers, and academic researchers.<sup>8</sup>

Different users have different needs for the Graduate Outcomes data. Prospective HE students may look to Graduate Outcomes in order to make informed choices about providers and courses, while HE providers may use the data for strategic planning purposes. Funding and regulatory

---

<sup>7</sup> Higher Education and Research Act 2017, sections 64 and 65.  
<http://www.legislation.gov.uk/ukpga/2017/29/section/65/enacted>

<sup>8</sup> A list of likely users of graduate outcomes data, based on known users of the DLHE survey, can be found in Part One of the Graduate Outcomes Survey Methodology Statement:  
[https://www.hesa.ac.uk/files/Graduate\\_Outcomes\\_History\\_and\\_Background\\_20200330.pdf](https://www.hesa.ac.uk/files/Graduate_Outcomes_History_and_Background_20200330.pdf)

bodies may use Graduate Outcomes data to assess the performance of providers and courses, while government agencies—both local and national—and graduate employers may look to the data to provide information both about the regional supply of graduates with different skills and about the roles played by graduates in society more generally.

Throughout the design and implementation phases of the Graduate Outcomes survey, HESA has been actively engaging with the various potential users of the survey data. In the early stages of the NewDLHE review, a Strategic Group and a Working Group were convened; these groups, which were comprised of representatives from a wide variety of HE providers and other sector bodies, were responsible for setting the remit for the review and developing a workplan to pursue this remit.<sup>9</sup> Later in the review, HESA carried out two consultations, the first to determine user needs for the successor to DLHE, and the second to solicit feedback on the draft model for the new survey.<sup>10</sup>

Responses to the second consultation suggested a high level of stakeholder approval of the proposed model, giving HESA a mandate from potential survey users to proceed with the implementation of the new survey. The model proposed in the second consultation called for the establishment of the Graduate Outcomes Steering Group; this group, like the earlier Strategic and Working Groups, is designed to reflect the diversity of stakeholders for the Graduate Outcomes survey and is comprised of representatives from HE providers and HESA statutory customers from across the UK. The Steering Group has been meeting quarterly throughout the implementation phase to advise HESA on all aspects of the development of the Graduate Outcomes survey. HESA values the expertise and input which has so far been contributed by the Graduate Outcomes Steering Group, and it is envisaged that the group will continue to operate in an oversight capacity and to help guide further improvements to the survey.<sup>11</sup>

In addition to the regular meetings of the Graduate Outcomes Steering Group, HESA has also solicited feedback from the sector on particular issues at different points during the implementation phase. While the charts and tables to be included in the Graduate Outcomes Statistical Bulletin and open data release were being developed from conceptual designs into logical wireframes, HESA convened a group of sector representatives to ascertain whether or not the planned outputs met with user needs. While the feedback from that session was largely positive, it did lead to some changes in the design for the data releases, including the addition of provider-level analysis of responses to the 'graduate voice' questions.<sup>12</sup>

---

<sup>9</sup> The NewDLHE review was a major review of HESA's destinations and outcomes data which ran from July 2015 to June 2017; 'NewDLHE' was the working title for replacement for DLHE, which has since become the Graduate Outcomes survey. For a complete record of the review, see

<https://www.hesa.ac.uk/innovation/records/reviews/newdlhe>

Further detailed information on the NewDLHE Working Group and Steering Group can be found on the

HESA website: <https://www.hesa.ac.uk/innovation/records/reviews/newdlhe/working-group>

<https://www.hesa.ac.uk/innovation/records/reviews/newdlhe/strategic-group>

<sup>10</sup> Syntheses of responses to the two consultations can likewise be found on the HESA website:

<https://www.hesa.ac.uk/innovation/records/reviews/newdlhe/consultation>

<https://www.hesa.ac.uk/innovation/records/reviews/newdlhe/second-consultation>

<sup>11</sup> Further information on the remit and composition of the Graduate Outcomes Steering Group:

<https://www.hesa.ac.uk/innovation/outcomes/about/steering-group>

<sup>12</sup> HESA. 2020. How to publish Graduate Outcomes data? Our consultation on open data release.

<https://www.hesa.ac.uk/blog/15-04-2020/how-publish-graduate-outcomes-data>

---

### 3.1.2. DATA AND STATISTICAL CONCEPTS

The Graduate Outcomes survey covers all graduates who obtain relevant higher education qualifications during the survey year. The list of graduates who are eligible to be surveyed is generated on the basis of data on qualifiers from the Student and Alternative provider student record along with data from the further education sector supplied by the Department for the Economy, Northern Ireland (DfENI) and the Office for Students (OfS). HE providers and Welsh and Northern Irish further education colleges are then responsible for supplying HESA with valid contact details for their graduates. While the OfS can provide HESA with contact details for graduates of English further education colleges; colleges whose contact details are provided by the OfS can subsequently amend the contact details for their graduates as required.<sup>13</sup>

Graduates are divided into four cohorts, based on the time of year at which they obtained their qualification, and they are surveyed, either online or by telephone, approximately 15 months after the completion of their studies.<sup>14</sup> Graduates are asked to respond to the survey with reference to a seven-day census week at the beginning of the sampling period; graduates in cohort A, for example, finished their qualifications in August to October 2017, and were surveyed in December 2018 to February 2019, with reference to the first week of December 2018.<sup>15</sup> Approximately 20% of graduates were surveyed as part of cohort A (having qualified from August to October 2017); 5% were surveyed as part of cohort B (having qualified from November 2017 to January 2018); 5% were surveyed as part of cohort C (having qualified from February to April 2018); and the remaining 70% were surveyed as part of cohort D (having qualified from May to July 2018). All four cohorts are analysed together to produce a single annual dataset, reflecting the fact that most UK higher education operates on a relatively standardised academic year, and the majority of graduates therefore finish their qualifications in early summer (cohort D). The division of the survey year into four cohorts primarily aids data collection and ensures a consistent 15-month gap between course completion and census week.

Graduate activities are one of the main areas of interest in the survey. Graduates are given a list of potential activities and are asked to select all activities from that list which they were undertaking during census week. The following options are available:

- Paid work for an employer
- Self-employment/freelancing
- Running my own business
- Developing a creative, artistic or professional portfolio
- Voluntary/unpaid work for an employer
- Engaged in a course of study, training or research
- Taking time out to travel – this does not include short-term holidays
- Caring for someone (unpaid)
- Retired
- Unemployed and looking for work
- Doing something else.

---

<sup>13</sup> Graduate Outcomes Survey Results record 2017/18 - Coverage of the record.

<https://www.hesa.ac.uk/collection/c17072/coverage>

<sup>14</sup> For further details on the sources of the Graduate Outcomes data, see the relevant section of Part Two of the Graduate Outcomes Methodology Statement: <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/survey-coverage>

<sup>15</sup> For further detail on cohort and census dates, see the Graduate Outcomes Definitions page on the HESA website: <https://www.hesa.ac.uk/support/definitions/graduates>

From the list of activities which they select, graduates are additionally asked to identify the activity which they consider to have been their most important activity during census week. On the basis of the activities which they select, graduates are routed to subsequent survey questions; the order in which they are routed depends on which activity they identify as most important.

Graduate employment is a key area of interest for many users of HESA data on graduates; the most recent version of the Teaching Excellence and Student Outcomes Framework (TEF), for example, makes use of DLHE data on the percentage of graduates from a given provider who are in full-time professional level employment or further study six months after finishing their qualification.<sup>16</sup> Responses to the graduate activity question in the Graduate Outcomes survey are therefore likely to be an important source of information about graduates in work or employment 15 months after course completion.<sup>17</sup> In HESA's analysis of the data, we compile tables that look at graduates in work and usually include both those working for an employer and those who are self-employed, running their own business and developing portfolios. We provide a 'work population marker' to relevant tables which allows users to view data either based on all graduates who report one or more of these activities, or alternatively to focus on those graduates who state that one of these activities is their most important activity. We also look to provide a 'work type marker' filter to relevant tables, to allow users to distinguish between respondents based on their most important activity.

Graduates who are engaged in work for an employer (whether paid or unpaid), self-employment, running their own business, are assigned both a Standard Industrial Classification (SIC) code and a Standard Occupational Classification (SOC) code. Graduates developing a portfolio are assigned a SOC code only. Accurate SIC and SOC coding makes it possible both to provide users with a clear picture of the industries and occupations in which higher education graduates are working and to allow users to compare the outcomes experienced by graduates working in different areas.

The SIC framework categorises businesses in terms of the type of economic activity in which they are engaged.<sup>18</sup> Easily comparable data on the industries in which graduates are working helps users to understand the economic contributions made by higher education graduates. Whereas SIC data provides information about the sectors of the economy in which graduates are active, the SOC framework provides a system for categorising occupations according to the skill level and type of work entailed by the jobs which graduates do.<sup>19</sup> SOC codes allow jobs to be categorised according to major groups, sub-major groups, minor groups, and unit groups; major groups are distinguished by the level of skill and experience required to perform the activities

---

<sup>16</sup> Office for Students. TEF data. <https://www.officeforstudents.org.uk/advice-and-guidance/teaching/tef-data/> For further detail, see the 'Employment metrics' section of the Year 4 TEF technical specification: [https://www.officeforstudents.org.uk/media/da96d15a-97e6-4732-a2f5-cb2473633932/ofs2018\\_45.pdf](https://www.officeforstudents.org.uk/media/da96d15a-97e6-4732-a2f5-cb2473633932/ofs2018_45.pdf) For TEF purposes, employment counts as 'professional level' if the occupation in question has a SOC code in major groups 1-3; see below for HESA's approach to SOC coding (more information is covered in this section, and further information is available below, in section 3.2.9.2. SIC and SOC coding).

<sup>17</sup> For more detail on the definitions of work and employment used by HESA, see section 3.5.1. National and international data standards below.

Future action for HESA to consider includes evaluation of the potential gap between HESA and harmonised national definitions, both through comparison with third-party linked data sources such as LEO/LFS and through evaluation of question wording in the light of this.

<sup>18</sup> Further detail about HESA's use of the SIC coding framework can be found in section 3.5.1. National and international data standards below.

<sup>19</sup> For further detail about HESA's use of the nationally recognised SOC framework, see section 3.5.1. National and international data standards below.

associated with a job, while occupations within each major group are organised according to the type of work performed. Depending on their SOC major group, occupations are classified as 'high skilled' (groups 1-3), 'medium skilled' (groups 4-6), or 'low skilled' (groups 7-9) for purposes of analysis. These classifications by SOC major group are particularly valuable to users who wish to see a broad overview of the kinds of jobs done by graduates or to compare the employment outcomes of graduates with different characteristics.

In addition to asking graduates about their activities during census week, the survey asks graduates two sets of questions about how they feel. In the first set of these questions, the 'graduate voice' questions, graduates are asked to reflect upon their activities, and to consider the extent to which those activities fit with their future plans, are meaningful, and allow them to utilise what they learned during their studies. Graduates in work are asked these questions with reference to their work, graduates in further study are asked these questions with reference to their current study, and graduates doing something else or engaged in multiple activities are asked these questions with reference to their current activities. These questions were designed by HESA in response to feedback from sector representatives, who felt that there was a need for qualitative data linking graduates' current experiences with their experiences in HE.<sup>20</sup>

The second set of questions deals with graduates' subjective wellbeing (SWB). SWB is assessed in Graduate Outcomes using a set of four questions (the ONS4), which were developed by the Office for National Statistics (ONS) for use in the Annual Population Survey and have since been used in a large number of social surveys; prior to their use in Graduate Outcomes, the ONS4 were included in the final year of the LDLHE survey.<sup>21</sup> In the ONS4, graduates are asked to think about the extent to which they:

- are satisfied with their life
- feel that the things they do are worthwhile
- feel happy
- feel anxious.

Like the 'graduate voice' questions, the section on SWB was added to the Graduate Outcomes survey as an alternative outcome measure, separate from employment and employability. Several possible alternative outcome measures were proposed during the first consultation phase, and the ONS4 SWB questions were added to the core Graduate Outcomes survey in response to feedback from HESA's statutory customers.<sup>22</sup>

Finally, the Graduate Outcomes survey includes a number of opt-in question banks, which may be asked after respondents have come to the end of the core survey. Providers are given the option to select a number of additional question banks which will be asked of their graduates. Some of the opt-in question banks are targeted at certain categories of graduates and will therefore not be asked of all graduates from a provider.

---

<sup>20</sup> For more detail about the development of the graduate reflection questions, see Part One of the Graduate Outcomes Methodology Statement:

[https://www.hesa.ac.uk/files/Graduate\\_Outcomes\\_History\\_and\\_Background\\_20200330.pdf](https://www.hesa.ac.uk/files/Graduate_Outcomes_History_and_Background_20200330.pdf)

<sup>21</sup> ONS. 2018. *Personal well-being user guidance*.

<https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide>

<sup>22</sup> See Part One of the Graduate Outcomes Methodology Statement:

[https://www.hesa.ac.uk/files/Graduate\\_Outcomes\\_History\\_and\\_Background\\_20200330.pdf](https://www.hesa.ac.uk/files/Graduate_Outcomes_History_and_Background_20200330.pdf)

---

The following opt-in question banks are available, depending on the data needs of providers:

- Finding your job
- Net promoter entity
- Graduate choice
- Research students
- Newly qualified teachers
- Careers service.

The addition of these opt-in banks gives providers some scope to tailor the survey to their particular data needs; a provider with a particular desire for data on graduate satisfaction might want its graduates to answer the 'net promoter' question bank, while a provider interested in the effectiveness of its career services provision might want to ask its graduates how they found their current jobs.

### 3.1.3. ASSESSMENT OF GAPS

The Graduate Outcomes survey was designed, so far as possible, to meet likely user needs for data on what graduates do after finishing HE and how they feel about their careers so far. We maintain a watching brief on policy issues that are of relevance to data about graduate outcomes, and report internally on emerging trends. Since the development of this survey was an extended project, however, there may be some gaps between our outputs and user needs. This could be because needs are changing faster than the survey development process, or because there are trends of which HESA is not yet aware. After the publication of the first year of Graduate Outcomes data, we intend to solicit further feedback from users of the data so that we can undertake an assessment of how well the initial publication has met user needs; we will then be able to incorporate user feedback into any adjustments we make to subsequent years of the survey and resulting statistical outputs. Our approach to evaluation is covered in the methodology statement.<sup>23</sup>

HESA is already aware of some areas in which we will aim to make improvements in subsequent years. Regional employment and skills gaps are important areas of current policy interest; in the spring budget speech, the government emphasised a policy of 'levelling up' across the UK, aimed at providing opportunities in under-served regions and reducing regional disparities, and the OfS has recently launched an initiative encouraging universities to work with local employers to develop graduate jobs.<sup>24</sup> While the Graduate Outcomes survey collects data on location of domicile, HE provider, and subsequent activity, the statistical outputs for the survey analyse graduate outcomes by region of provider, but not by region of domicile or activity in census week; we will aim to improve the regional and sub-regional coverage of our statistical outputs so as to enable more regional analysis on the part of key users of our data.

Additionally, the data collection process for year one and the initial cohorts of year two has brought to light some areas in which adjustments to the survey questionnaire have the potential to improve data quality and our ability to produce outputs which meet user needs. Significant changes to the survey questionnaire require approval from the Graduate Outcomes Steering Group; as we continue to assess the survey and integrate feedback from users, HESA will continue proposing

---

<sup>23</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/evaluation>

<sup>24</sup> Rishi Sunak. Budget Speech 2020. 11 March 2020.

<https://www.gov.uk/government/speeches/budget-speech-2020>

OfS. 2020. *No place like home – new fund to boost local graduate opportunities.*

<https://www.officeforstudents.org.uk/news-blog-and-events/press-and-media/no-place-like-home/>



changes to the Steering Group wherever it seems that modifications will enable us to produce data which better serves the needs of our stakeholders.

### 3.2. ACCURACY AND RELIABILITY

In this section we evaluate the closeness between the estimated results produced from the survey and the (unknown) true value. The design of Graduate Outcomes minimises the possibility for sampling error, due to the comprehensive approach taken to surveying all cases available to be contacted in the sampling frame. We start by describing the sampling frame, and how we maintain it, also describing the close resemblance of the sample to the sampling frame. We then go on to concentrate on various forms of non-sampling error in the subsequent subsections, including:

- coverage error
- non-response error
- measurement error
- processing error.

#### 3.2.1. THE SAMPLING FRAME, AND HOW IT IS MAINTAINED

The Graduate Outcomes survey aims to survey the population of graduates from Higher Education (HE), and the survey employs a dynamic sampling frame that is kept up to date when source data changes. The source data is a list of data about individual graduates drawn from existing administrative census datasets about students. These sources are enriched with contact details sourced from the providers where those graduates studied. Below, we cover these two separate aspects of how the sampling frame is constructed. The methodology statement section on the sampling frame also offers a good overview of this area.<sup>25</sup> We present additional information in the following paragraphs.

The sampling frame has been developed utilising the main administrative data sources for HE provision in HE settings across the UK,<sup>26</sup> and for HE in FE settings in all parts of the UK except Scotland.<sup>27</sup> These data sources all support existing official statistics publications, so our starting assumption is that they are of high quality and fit for their purposes. The sampling frame is drawn from this administrative data, according to the criteria set out in the coverage statement in the methodology statement.<sup>28</sup> The (separate) coverage statement for the Graduate Outcomes Survey Results record explains further detail of this.<sup>29</sup> The following subsection summarises this information, and provides additional commentary, starting with the main processes utilised for all data sourced from HE providers, by HESA. In the subsection after that, we cover how we derive the sampling frame related to HE in FE settings.

##### 3.2.1.1. Sampling frame data based on HESA data collections

The majority of data used to determine the sampling frame is collected by HESA. HESA collects individualised data on students in HE providers across the whole UK in its Student record and Alternative Provider Student record (referred to hereinafter as the “Student record(s)”, for brevity). Data from these records is an administrative census: their goal is to enumerate the HE student

---

<sup>25</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/sampling-frame>

<sup>26</sup> These are the HESA Student record(s) described in detail further on. See <https://www.hesa.ac.uk/data-and-analysis/students> for the data published from these records.

<sup>27</sup> See footnote 45.

<sup>28</sup> For further information about the survey coverage, see the relevant section in the Methodology statement: <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/sampling-frame>

<sup>29</sup> See <https://www.hesa.ac.uk/collection/c17072/coverage>

population and describe their personal and study characteristics. The data on qualifiers contained in the Student record is the most complete single record of graduates from HE available. The Student record(s) are the primary official record of UK HE, and are collected on behalf of the UK Government, the Devolved Administrations, and the Office for Students.<sup>30</sup> HESA collects this data annually, from a constituency of HE providers that is refreshed at least annually – referred to by HESA as ‘reporting providers’. This covers all publicly-funded and/or regulated HE providers in the UK. The HESA Student record(s) for the 2017/18 academic year were used in the creation of the sampling frame for the first year of the Graduate Outcomes survey.

The sampling frame comprises all students reported to HESA or the relevant body as obtaining relevant higher education qualifications during the reporting period 01 August to 31 July, and whose study was full-time or part-time (including sandwich students and those writing-up theses). Graduates with awards from dormant status are only included in the target population for postgraduate research students. Graduates with some qualifications are excluded from the sampling frame. These include intercalated degrees, awards to visiting students, students on post-registration health and social care courses, and professional qualifications for serving school teachers.<sup>31</sup>

Exceptionally,<sup>32</sup> issues may be found in the source administrative data, that, when corrected through the data amendments process (also termed the fixed database facility), have the effect of altering the sampling frame. Up to the dates specified in the coding manual (which overlap with the contact period substantially) changes made to the sampling frame *via* the fixed database are reflected in the “population file” that is passed to the provider through the Portal, so that additional contact details can be gathered. This would be necessary, for example, if the fixed database change increases the sampling frame data for a provider, by inserting previously missing records. Furthermore, the data that is published (including response rates in relation to targets) always reflects the most up-to-date sampling frame available from the fixed database at that point. This means that even if over-sampling has occurred (because a fixed database change removes graduates from the sampling frame, in cases where responses have already been gathered, successfully) then these results would also be discarded from the output file.

In order to derive the sample, and to obtain their contact details, information about the sampling frame is passed back to the HE providers, through an online electronic portal for providers (hereinafter, “the Portal”). The goal is to maximise the availability of usable contact details for deployment during data collection. A full data collection process exists to support this activity, and it is specified in detail in the coding manual for the Graduate Outcomes Contact Details record.<sup>33</sup> This document explains the collection schedule, the data items collected, and gives information to support interactions with graduates – an engagement strategy is defined by HESA and roles and

---

<sup>30</sup> HESA’s Collection Notice for its Student record <https://www.hesa.ac.uk/about/regulation/data-protection/notices> details the statutory background for this. The coverage statement for the Student record (2017/18) utilised in creating the sampling frame gives details on which students are included in the record: <https://www.hesa.ac.uk/collection/c17051/coverage>.

<sup>31</sup> Full details of exclusions are available at: <https://www.hesa.ac.uk/collection/c17072/coverage#contents5>

<sup>32</sup> For details of the financial impact and regulatory authorisation needed to make a change to the previously-submitted data (to amend the fixed database) see <https://www.hesa.ac.uk/support/provider-info/subscription/fees-and-charges>

<sup>33</sup> See <https://www.hesa.ac.uk/collection/c17071>

responsibilities are shared with HE providers.<sup>34</sup> The coding manual also gives details of the quality assurance regime (automated and manual) along with other guidance and training materials on the systems and processes operated *via* the Graduate Outcomes provider portal.

In the provider portal, providers are presented with an output file showing graduates from the sampling frame drawn from their own data (collected previously), and are asked to populate and upload an XML file with contact details. Detailed guidance and training is offered on data quality expectations and using the tools provided.<sup>35</sup> The provider portal enables HE providers to act as peers in the quality assurance process, and HESA's system logs show interaction with the data, with 1,462 separate contact details files uploaded by 190 providers, as well as extensive use of the web-based update facility (mainly used by smaller providers). On submission, checks are undertaken by HESA to identify any problems with various quality dimensions of the data:<sup>36</sup> validity,<sup>37</sup> uniqueness,<sup>38</sup> completeness,<sup>39</sup> and consistency.<sup>40</sup> Further information about the 51 automated rules applied during the first year of operation is available online.<sup>41</sup> New rules are added and existing ones amended as deemed necessary by analysis of data quality and feedback on operational survey management is received from interviewers. Version control is applied to all aspects of the coding manual and quality rules, allowing analysts to see which rules were introduced at which points.

The quality regime seeks to maximise the number of usable details available for contact. Where quality rules are triggered, providers must either update the data, or contact HESA to request that the rule be 'switched-off' for that observation. This process is managed by HESA's Liaison team who have oversight of these operational data quality issues. We do not directly assess the accuracy<sup>42</sup> of the contact details – our current checks do not determine if the contact details provided belong to the graduate. Providers must therefore warrant the accuracy of the data and fitness for purpose for use of the contact details, on submission. Providers' interactions with HESA generally form part of their internal audit and compliance mechanisms, overseen by their governing bodies.

At this point, we will summarise the quality characteristics of the contact details. During the first year of surveying, we measured the effectiveness of contact details supplied to us, in achieving contact with the graduate. During and following each cohort, we reported on challenges faced with contact details, and where manual checks revealed potential quality issues with records, we

---

<sup>34</sup> This engagement plan is detailed in the information provided on the operational management of the survey. See <https://www.hesa.ac.uk/definitions/operational-survey-information#engagement-plan>  
Communication resources are here: <https://www.hesa.ac.uk/innovation/outcomes/providers/communications>  
Roles and responsibilities are here: <https://www.hesa.ac.uk/innovation/outcomes/providers/responsibilities>

<sup>35</sup> See [https://www.hesa.ac.uk/collection/c17071/contact\\_details\\_guidance](https://www.hesa.ac.uk/collection/c17071/contact_details_guidance) for an accessible overview. For full information about types of contact details we accept and other best practice see the Portal user guide, at: [https://www.hesa.ac.uk/collection/c17071/portal\\_user\\_guide](https://www.hesa.ac.uk/collection/c17071/portal_user_guide)

<sup>36</sup> HESA's approach to data quality management during collection rests heavily on the quality dimensions specified in the DAMA DMBOK. See (DAMA UK Working Group on "Data Quality Dimensions", 2013) (For outputs, HESA uses the ESS dimensions.)

<sup>37</sup> E.g. telephone numbers consist of digits.

<sup>38</sup> E.g. identifying graduates with duplicate email addresses or telephone numbers.

<sup>39</sup> E.g. that most graduates in the sampling frame have some contact details.

<sup>40</sup> E.g. that a variety of different contact methods have been given, and they are not all, for example, comprised entirely of the provider's own 'email for life' address (where this exists) for each graduate.

<sup>41</sup> See [https://www.hesa.ac.uk/collection/c17071/quality\\_rules](https://www.hesa.ac.uk/collection/c17071/quality_rules)

<sup>42</sup> E.g. Properly-formed contact details may pass our checks, without necessarily belonging to the respondent we hope to reach.

contacted providers directly to investigate. Experience handling such issues led us to enhance our quality checks, to provide improved feedback to providers on the quality of the details they had sent us. For example, for nearly 19% of graduates in cohort A (3,283 individuals) all telephone numbers supplied were unobtainable (i.e. they did not work). For a further 514 graduates in cohort A, while the number was contactable, it was the wrong number for the graduate. However, in cohort B, the incidence of unobtainable numbers reduced by nine percentage points, to 10%. We attribute this to awareness-raising with HE providers during cohort A. During the full year, HESA received over 2.2 million individual contact details which were all received and processed in the provider portal. We gradually enhanced the quality rules applied in the portal with additional validation put in place to check mobile numbers for cohort D (the UKMOB.08 rule that utilises a check using a telephone validation service). This paid off with over 59,692 numbers being found invalid which not only saved effort, but avoided HESA paying for SMS messages that inevitably fail (reducing costs to subscribers). Similarly, over the first year, additional checks were put in place to warn providers, or to raise an error, where the characteristics of email addresses submitted for use raised issues. We also note that for some graduates, no contact details at all were provided. Quality of contact details is an important factor in achieving the high response rates our users desire. They are also a potential source of bias in the data, but at present we cannot quantify this. Investigations aimed at assessing integrity and reliability of contact details continue, therefore. We will publish further on this in due course.

During the contact details collection process, HE providers are also able to supply additional information that allows HESA to exclude graduates from the surveyable population, for example if they have become seriously ill, or have died, since graduating. In practice, some contact details prove unavailable. A few graduates do not keep in touch with their HE providers and accurate contact details held for them can become out of date. Providers are encouraged to stay in touch with their graduates through different means, enabling them to supply good quality contact details in time for the survey 15 months later.

Timeliness of the data in the sampling frame is a central consideration. The collection of contact details follows four phases, each aligned to one of the four cohorts (A, B, C, and D). Comprehensive information is provided about timescales for collection activities.<sup>43</sup> Because the survey takes place approximately 15 months following course completion, allowance has to be made for changes of circumstance following this. Contact details are therefore collected during a period when the provider has had maximum opportunity to ensure they are as up-to-date as possible, and; as mentioned earlier, providers can also provide information at this stage to tell HESA that a graduate should not be contacted, due to intervening serious illness or death.

#### 3.2.1.2. Sampling frame data based on other ingested data

A minority of HE study takes place in further education (FE) settings.<sup>44</sup> HESA collects data about these students in Wales as part of its Student record (the process for this is the same as for the other data described in the paragraphs following this one). In England, Northern Ireland, and Scotland, the equivalent data is collected by other bodies.<sup>45</sup> HESA ingests data about HE students

---

<sup>43</sup> See [https://www.hesa.ac.uk/collection/c17071/data\\_collection\\_schedule](https://www.hesa.ac.uk/collection/c17071/data_collection_schedule)

<sup>44</sup> To summarise, in 2017/18, FE providers accounted for 0.5% of the UK's total postgraduate enrolments, 1.4% of the UK's total first degree enrolments, and 47.8% of the UK's "other undergraduate" enrolments. For detailed figures and explanatory notes, see <https://www.hesa.ac.uk/data-and-analysis/sb255/figure-3b>.

<sup>45</sup> In England, the Individualised Learner Record (ILR) is collected by the Education and Skills Funding Agency (ESFA). In Northern Ireland, the Assembly mandates the collection of the Consolidated Data Return (CDR) of which an extract is supplied to HESA by the Department for the Economy (Northern Ireland). In

in FE settings from the administrative records collected in England and Northern Ireland. Graduates from HE study in FE settings in Scotland, are excluded from the survey coverage.<sup>46</sup> This data, along with, in England, contact details found within these administrative records, is provided to HESA in a timely manner by the relevant bodies, in order to permit these HE in FE graduates to be contacted during the normal operation of the survey. Where contact details are not provided, or where the FE provider is able to source improved contact details, a collection process identical to the one described in the previous section, is employed to permit this. We do not describe the quality processes followed in the construction of these administrative records here, but we do provide supporting information for Further Education Colleges (FECs) in England and Northern Ireland<sup>47</sup>. Further details should be sought from the data collectors (see footnote 45).

### 3.2.2. HOW DOES THE SAMPLING FRAME RELATE TO THE POPULATION?

This section deals with what we know about coverage error. The population of interest is graduates from HE-level courses. The exclusions from this are explicit and intentional (see footnote 31). The survey does not attempt to contact students who did not graduate – these individuals are counted elsewhere in HESA's Student data. Where students graduate with a different award than that they originally intended at the beginning of studies, they will be included in the sampling frame (except where they fall into the exclusions we list).

The administrative data described in the previous section comprises all publicly-funded and/or regulated HE providers in the UK. There are known instances of duplication of student identifiers between providers within the ILR (which does not have a globally unique identifier akin to the HESA unique student identifier, or HUSID) and between the ILR and HESA data (where 'franchise' arrangements exist). The Office for Students is expert in handling both types of duplication, and has isolated and removed these within their dataset prior to sharing data with HESA. Where other administrative data sources are concerned the separation of reporting environments militates against duplications occurring.

One legitimate question is how complete the administrative data is: could there be any undercoverage of HE graduates, because the provider they studied at is not included in the administrative data? In short, our sampling frame represents the overwhelming majority (probably in excess of 99% based on Hunt and Bolliver's figures<sup>48</sup>), but not absolutely all, UK HE students. While there is no definitive answer to how many are missing, it is known that a small amount of HE-level provision remains outside the formally-regulated sector. Research commissioned over seven years ago by the former Department of Business, Innovation and Skills identified a minimum estimate 'of 674 named privately funded HE providers operating in the UK. [...] Most providers identified [were] relatively small in scale; 217 of the 674 had fewer than 100 students. Only 35 providers had over 1000 students, with five of these having over 5000 students.'<sup>49</sup> Subsequently,

---

Scotland, the government mandates the collection of the Further Education Statistics record (FES). However, the HE in FE activity in Scotland, collected in the FES, is not within coverage for the Graduate Outcomes survey.

<sup>46</sup> See the section on HE in FE: <https://www.hesa.ac.uk/collection/c17072/coverage#contents4>

<sup>47</sup> For FECs in England, see: <https://www.hesa.ac.uk/innovation/outcomes/providers/information-english-further-education-colleges>

For FECs in Northern Ireland, see:

<https://www.hesa.ac.uk/innovation/outcomes/providers/informationnorthern-irish-further-education-colleges>

FECs in Wales are longer-standing HESA subscribers, and information for them is consistent with the general information sources, here: <https://www.hesa.ac.uk/innovation/outcomes/providers> and elsewhere.

<sup>48</sup> (Hunt & Boliver, 2019, p. 22)

<sup>49</sup> (Department for Business, Innovation and Skills, 2013, pp. 7–8)

the Higher Education and Research Act 2017<sup>50</sup> has had the effect of expanding the sphere of HE regulation in England to include a group of organisations referred to as ‘Alternative Providers’. While this terminology is no longer used officially, the providers brought into the regulated sphere under this designation are now included in administrative data returns used to create the sampling frame, and include the majority of larger organisations identified in the research.<sup>51</sup> In the Graduate Outcomes open data, we provide some information at provider-level, and users can therefore see the list of providers, where data on their graduates is included in our outputs.

Volatility in the segment of the HE marketplace comprising the very smallest providers, means that some will not have provided full data for inclusion in the sampling frame, nor would they have shouldered their share of the costs of surveying, having undergone market exit. Further research currently underway indicates that there are ‘some 813 private providers in operation in the UK – a significant increase on the 732 and 674 recorded in 2014 and 2011 respectively.’ Many of these ‘are small scale, concentrating on sub-degree or postgraduate qualification across a narrow band of subjects – often characterised as being popular but with low overheads.’<sup>52</sup> For the most up-to-date documentation on what is known about the scale and scope of this part of the HE sector, readers are directed to the Hunt and Boliver paper listed in the references. In HESA’s published data no attempt has yet been made to provide estimates that include this population, as we know too little about the characteristics of students and graduates from this part of the sector.

We therefore estimate that the list of graduates in the sampling frame comprises in excess of 99% of the population of interest, and that the impact of this slight undercoverage is therefore likely to be very slight in England, and negligible in Wales, Scotland, and Northern Ireland.<sup>53</sup> However, we remind users of the discussion about the provenance of contact details collected against the sampling frame. The practical effect of missing contact details, and those found to be unusable or ineffective during fieldwork, reduces the effective size of the sampling frame, and limits the achievable number of responses.

### 3.2.3. THE SAMPLE

Graduate Outcomes is a population-scale survey (or colloquially, a census).<sup>54</sup> Our goal is to contact the entire sampling frame. The sampling frame and the sample are therefore largely synonymous.

A marker was developed to identify the sampling frame from within the HESA Student record(s), and appropriate file(s) were extracted. Similar logic was applied by the suppliers of the HE in FE data not collected by HESA. The datasets were then combined – no matching or linking was required.

---

<sup>50</sup> HERA 2017. <http://www.legislation.gov.uk/ukpga/2017/29/contents/enacted>

<sup>51</sup> Current public laws delineate the regulatory regimes in place, but do not compel all HE providers to register with a funder or regulator.

<sup>52</sup> (Hunt & Boliver, 2019, pp. 1–3)

<sup>53</sup> Hunt and Boliver estimate 88% of private HE providers operate only in England.

<sup>54</sup> Sometimes Graduate Outcomes, and its predecessors DLHE and LDLHE are referred to as a “census”. Strictly, a census enumerates a population. This is a central function of the HESA Student record outputs. By the time students become graduates, we already know how many of them have studied in HE, and we use our pre-existing census data from the Student record(s) to construct a sampling frame for the Graduate Outcomes survey. However, Graduate Outcomes is the UK’s biggest annual social survey, and there is no standard statistical term to describe a full survey (i.e. not administrative data) of (effectively) a whole defined population. It is fine to call Graduate Outcomes a census in everyday usage, but the term “population-scale survey” probably gets the same point across in a more technically-correct manner.

Our 'base population' is the term used to refer to the dataset that comprises the entire sampling frame. This includes all graduates who fall within our coverage statement, but for whom we have inadequate, ineffective, or missing contact details.<sup>55</sup> Graduates in the 'base population' who (we learn from their provider during the collection of contact details) have since died, or become seriously ill, are excluded from our 'surveyable population' – this is the sample - graduates who we will attempt to contact during the survey. We also choose to exclude this small number from the denominator for our response rates. Hence, the survey sample is identical to the sampling frame, except for the exclusion of graduates who we know have died, or become seriously ill, based on information from their provider. The total number of graduates excluded in this way from the surveyable population for the first year was 150.<sup>56</sup> We note that a further 270 graduates were discovered to have died, or become seriously ill, during the collection of survey responses (this information was not independently verified). This latter group of graduates remain included in the denominator for response rates. Where we have achieved a usable response from them, they will also be included in the numerator. Graduates who exercise their right to opt out of the survey are also included in the denominator.

Response rate targets form part of the survey design. These rates are high, to reflect the desire among many users to evaluate smaller sub-samples as a part of their analysis, and thus to minimise the rate of unit non-response. Targets were set in October 2018, and further information on these is available in the methodology statement.<sup>57</sup> HESA's engagement strategy is the main tool for seeking high response rates.<sup>58</sup> Progress towards these targets (along with updates on the operational management of the survey) has been reported in a series of midpoint reviews and end of cohort reviews, published regularly on the HESA website.<sup>59</sup> Final response rates, by domicile and mode of study, are published in the Statistical Bulletin.

We cover issues related to non-response in the next two sections.

### 3.2.4. SAMPLING ERROR AND NON-RESPONSE ERROR

Sampling error is the difference between a population value and an estimate based on a sample, and is one of the components of total survey error. It is normal for a quality report on a sample survey to offer a caveat explaining that, in principle, many random samples could be drawn and each would give different results, due to the fact that each sample would be made up of different people, who would give different answers to the questions asked. The spread of these results is the sampling variability. However, sampling error occurs because estimates are based on a sample rather than a census. As we have previously demonstrated, Graduate Outcomes is a population scale survey<sup>60</sup> where the sample closely resembles the sampling frame, and the

---

<sup>55</sup> Our approach to collecting contact details means we may still manage to contact these graduates, if adequate contact details are supplied during the period of fieldwork.

<sup>56</sup> The prevailing HESA approach to rounding and anonymisation has been applied to these and all other specific figures about people reported in this document. See: <https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>

<sup>57</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/survey-targets>

<sup>58</sup> We do not publish the full engagement strategy. Instead, we provide an outline plan for each cohort, updated quarterly here: <https://www.hesa.ac.uk/definitions/operational-survey-information#engagement-plan> For an example of a more discursive account of the kinds of activities involved, see this blog post: <https://www.hesa.ac.uk/innovation/outcomes/providers/engagement-plan/partial-complete>

<sup>59</sup> For a full list of mid-point and end of cohort reviews from the 2017/18 cycle, with infographics, see: <https://www.hesa.ac.uk/innovation/outcomes/about/progress/progress-updates-archive>

<sup>60</sup> See footnote 54 on p.290.

sampling frame closely resembles the population. While we know that the quality and availability of contact details must affect the response rate we can achieve from the sample, we currently lack an independent and comprehensive measure for this. Therefore, the response rate achieved is itself our present best indicator of the quality of contact details. Hence, our analytical focus in this section is on the extent to which the achieved sample is representative of the population. We therefore focus on non-response error.

This section comprises two subsections, which cover the strategies HESA has followed to limit the practical effects of missing responses. In conducting a survey, one of the main types of non-sampling error that can arise is that resulting from non-response. Whilst a lower level of response causes a reduction in the precision of obtained estimates, the impact of response rates on bias is ambiguous.<sup>61</sup> The two types of error in this category are unit non-response<sup>62</sup> and item non-response.<sup>63</sup> We cover issues related to these in the next two sections.

#### 3.2.4.1. Unit non-response error

Unit non-response occurs where a graduate does not respond to the survey. A poor response rate will result in less precision in any estimates we generate. In terms of its effect on bias, this is less certain. Bias is determined by two components.<sup>64</sup> These are the response rate, as well as the variation between respondent and non-respondent values. Hence, a better response rate can be associated with greater bias, if the discrepancy between those who participate in the survey and those who do not grows ever larger. Consequently, attempting to maximise response rates will not necessarily minimise non-response bias.<sup>65</sup>

A number of elements of the survey design are intended to maximise response rates. These include:

- A website aimed at respondents to reinforce the legitimacy and credentials of the survey<sup>66</sup>
- A smartphone-optimised survey
- Allowing the survey to be completed in more than one stage, whether online, at the telephone, or using a mixture of both modes
- Bespoke email invitations and reminders that include the name of the graduate and their provider
- A dynamic engagement strategy informed by best practice and survey paradata
- Using a data collection platform that seamlessly integrates all modes together
- the adoption of a concurrent mixed-mode design (computer-assisted telephone interviewing (CATI) starts a week after the online system opens, and those who start online are not followed up until much later in the field period)
- Increasing the convenience of responding for graduates, by making appointments for telephone interviews at times that suit them
- Collecting proxy responses from half-way through the fieldwork period.

---

<sup>61</sup> As Koch and Blohm (2016) note.

<sup>62</sup> This is where we are missing all observations for a case – this would mainly happen in situations where we are unable to elicit any response from a graduate.

<sup>63</sup> This is where we are missing some observations for a case – a common situation might be a graduate who answers the survey, but does not wish to answer some questions in the survey. We explain more about how we handle this sort of issue, in the following section.

<sup>64</sup> As Groves (2004) illustrates.

<sup>65</sup> Keeter et al (2000) and Curtin et al (2000) are examples of previous studies that have demonstrated the phenomenon of achieving both higher response rates and bias.

<sup>66</sup> See <https://www.graduateoutcomes.ac.uk/>



For the rest of this section we cover the specifics of our approach where non-response bias is concerned. Root cause remediation is one of the practices HESA adopts to proactively manage data quality.<sup>67</sup> In this case, our goal was to reduce data quality issues arising during collection. Historically, organisations that have administered surveys have relied upon methods executed after collection (i.e. weighting) to deal with the challenge of non-response. Yet, over the last decade, those working in this area have increasingly looked at whether anything can also be done during the data gathering phase. Work by the Netherlands' official statistics agency<sup>68</sup> points to the advantages in attempting to do this, such as improved precision due to less variable weights. In trying to reduce non-response bias, other authors highlight the potential benefit of developing propensity models and subsequently diverting more attention to those individuals with a lower likelihood of responding in the latter stages of the collection process.<sup>69</sup> An adaptive survey design methodology was therefore designed and implemented from cohort C of the first year of the survey, onwards. Details of the practical approach to case prioritisation we took, are covered in detail in the section of the methodology statement covering data collection.<sup>70</sup> In summary, approximately halfway through a collection cycle, a logit model (consisting of student and course characteristics as independent variables) is created to generate individual response propensities. Additional resource and effort is then allocated to obtaining responses from those graduates identified as being least likely to partake in the survey. The objective of this exercise is to ensure not only higher response rates, but also to reduce possible non-response bias by aspiring to achieve a more representative sample.

We cannot, however, assume that the adaptive survey design will achieve its objective. The resulting data must be assessed and if necessary, action taken to address bias. This is referred to as "weighting" the survey. The overarching objective of weighting is to enable the sample to be adjusted such that it is more representative of the population.<sup>71</sup> Most surveys are weighted following collection. However, the Graduate Outcomes survey has some unusual features, such as a large sample size, an adaptive survey design, and an concurrent mixed-mode data collection approach. We therefore undertook a study to determine whether year one of the Graduate Outcomes survey should be weighted. The recommendation of this study was that weighting will not be applied to all statistics published by HESA for this first year (17/18) of survey data. Our analysis of the survey data did not identify any evidence of bias relating to mis-match between the achieved sample and graduate population characteristics in any direction at sector level. Indeed, when analysing across a range of demographic and course variables, we find a high level of similarity between the sample and population distributions. We trialled various weighting methods, and these did not improve the quality of our estimates. Unweighted and weighted estimates were generated at the overall level, as well as by key subgroups for each of three different weighting

---

<sup>67</sup> Addressing quality issues closest to their source is generally the most efficient approach, and follows established data quality management principles (Data Management Association, 2017, p. 453).

<sup>68</sup> (Schouten & Shlomo, 2017)

<sup>69</sup> See Rosen et al. (2014) for details. The use of this approach has also been applied in a similar fashion by Peytchev et al (2010) and Wagner (2013).

<sup>70</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/data-collection> (particularly the section on case prioritisation).

<sup>71</sup> The creation of weights can comprise of several components. First, the base weight refers to the probability that an individual is selected into the sample given the design of the survey. In Graduate Outcomes, we aim to send the survey to everyone in the sampling frame. We have not quantified how many people actually receive the survey. Second, a (unit) non-response weight may be generated, which seeks to account for the fact that participation may vary among different groups. In instances where information is available on the entire population, a final step would be to ensure that the weights can allow the sample data to match known population totals for a chosen set of categories.

methods.<sup>72</sup> Overall, across the breadth of HESA variables analysed, we generally observe close resemblance between the sample and the population, reducing concerns over potential bias. For a summary of our research and the findings, see the methodology statement section on data analysis.<sup>73</sup> Technical details of the study we undertook are also available in our research paper titled '*Should we weight?*'.<sup>74</sup> This paper offers a detailed account of how we reached the decision not to apply weighting. It describes the research methodology and illustrates the results that were found from the analysis. The paper is mainly aimed at academics, statisticians, and other interested parties wishing to understand the weighting research and its conclusions. Included in appendices A and B of the research paper, are a series of tables and graphs that illustrate our findings in detail.

The position regarding use of weighting in future years of the survey remains under review. HESA is planning some additional exploration of more nuanced approaches to weighting, which will utilise data from the second year of the survey. If an approach to weighting can be identified at that time that can be shown to improve the quality of statistics derived from the surveys, then this will be applied from 2018/19 Graduate Outcomes data onwards and will also be applied retrospectively to key data outputs from the 2017/18 survey, to enable comparisons across years.

Some statistics published from the Graduate Outcomes survey are at a very granular level, e.g. activity by provider, domicile, level of qualification and mode of qualification. In some cases, the sample size for such statistics may be small. In these cases, the statistics may be subject to high levels of variability and a lack of statistical precision. Confidence intervals on these statistics (ranges within which we have a high level of confidence that the equivalent whole-population parameter would fall, where a narrow range indicates greater precision and a wide range indicates less precision) are, for key tables, published alongside the data.

In addition, for some statistics, it may be necessary to introduce publication thresholds whereby statistics based on very small sample sizes and/or lower response rates are suppressed – this will be explained in any statistical releases where this decision is taken.<sup>75</sup>

The risk of non-response bias appears to have been minimised by the combination of relatively high response rates, and the adaptive survey design. We are fortunate to be able to link to good data on population characteristics to support our assessment. Despite this, it is not easy to quantify the extent to which non-response bias remains a problem. As noted in '*Should we weight?*' the Longitudinal Educational Outcomes data offers a suitable source for analysis of bias. Users of Graduate Outcomes microdata may wish to conduct their own analyses to ensure the Graduate Outcomes data supports their analytical objectives. We welcome future work aimed at further

---

<sup>72</sup> This included subject area, provider and subject area within providers, which tend to be groups of interest for different stakeholders across the sector (e.g. to help providers evaluate their performance and for prospective students considering what course to study). As policy matters in this area are devolved across the four nations, estimates were also produced by country of provider. Additionally, the Equality Act of 2010 requires public sector bodies to promote equal opportunity among individuals from all types of backgrounds. Consequently, we have also produced estimates by some of the key protected characteristics, such as age, ethnicity, disability and gender. Others such as marital status and gender reassignment were not covered, due to insufficient coverage in the data.

<sup>73</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/data-analysis>

<sup>74</sup> See <https://www.hesa.ac.uk/files/Graduate-Outcomes-Should-we-weight-20200521.pdf>

<sup>75</sup> Where suppression is applied, this will be done in line with the prevailing HESA statistical confidentiality policy (see <https://www.hesa.ac.uk/about/regulation/official-statistics/confidentiality>) and the associated rounding and suppression approach: <https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>

illuminating the quality characteristics of our data. However, there is currently no evidence to suggest that non-response bias presents a particular issue for users of the Graduate Outcomes survey data.

#### 3.2.4.2. Item non-response error

Item non-response occurs where a value for a particular variable is missing for a graduate, in a case where this observation was expected. In our survey, this typically occurs when respondents decline to answer particular questions. Non-respondents may differ from respondents, which can lead to non-response bias. Higher non-response is often understood to mean that the presence of bias may be more likely, and also increases the variance of estimates. However, as we noted in the previous section, there may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders for the value in question. The impact of non-response on bias is therefore ambiguous.

It is important to understand that no single graduate is expected to answer all available survey questions. A routing structure directs respondents to particular sets of questions that are most relevant to their circumstances.<sup>76</sup> Furthermore, optional questions will not be presented to all respondents. So, some data will not be present, but this does not mean it is missing – it may never have been sought, as it was not relevant to be asked in that case. In HESA's publications, these issues will be made clear in the data and the notes, for example by indicating the sample used to produce a table or chart in its title, and by enumerating the unknown values. Researchers and other microdata users in particular will need to note this feature of the survey.

A core set of mandatory questions<sup>77</sup> are required to be completed for a response to be marked as completed.<sup>78</sup> The terms 'complete' and 'full response'<sup>79</sup> are used interchangeably to refer to those cases where all the questions requiring a response have been completed and are populated with an answer. A derived field (ZRESPSTATUS<sup>80</sup>) describes the status of response to the Graduate Outcomes survey for each graduate for whom some (however minimal) results data has been received.

In addition to responses classified as 'survey completed', a status of 'partially completed'<sup>81</sup> has been assigned where some of the core questions are missing but the first two questions have been answered.<sup>82</sup> Although partially completed responses do not contribute to the survey's response rate targets, partially complete responses are used alongside 'survey completed' responses in statistical outputs. Again, data from such responses will appear in published statistics in the following ways: in tables with numbers, unknown values are shown for questions that were not answered. Wherever we display % values, we exclude unknowns from the calculations. The sample used will be clear in the title or accompanying text.

---

<sup>76</sup> A flow diagram showing the survey response record fields produced given each survey routing, is available in the coding manual:

[https://www.hesa.ac.uk/collection/c17072/download/Overall\\_Survey\\_Routing\\_Structure.pdf](https://www.hesa.ac.uk/collection/c17072/download/Overall_Survey_Routing_Structure.pdf)

<sup>77</sup> Details of mandatory questions can be found as a PDF download from:

<https://www.hesa.ac.uk/innovation/outcomes/survey>

<sup>78</sup> ZRESPSTATUS=04

<sup>79</sup> See <https://www.hesa.ac.uk/definitions/glossary#F>

<sup>80</sup> See the derived field specification at: <https://www.hesa.ac.uk/collection/c17072/derived/zrespstatus>

<sup>81</sup> ZRESPSTATUS=03

<sup>82</sup> The observations gathered from the first two survey questions permit the derived field XACTIVITY to be produced – see <https://www.hesa.ac.uk/collection/c17072/derived/xactivity>. Since 'activity' is the Graduate Outcomes survey's central concept, these responses are often partly usable.

Additional value could be provided for users of microdata by developing an appropriate formal logic to apply a description of the type of nullability in responses where there is no observation, although we note this is not yet on the future development work plan.

As this is a new survey, we are keen to undertake further quality assurance work to investigate the extent of non-response error, and indications of the directions to be followed by the programme of research work we are developing, can be found in the evaluation section of the methodology statement.<sup>83</sup> We intend to provide a summary of item non-response for key variables as a part of that work.

### 3.2.5. PROXY RESPONSES

A proxy response is a response made on behalf of the sampled graduate by someone other than the graduate. It is an indicator of accuracy as information given by a proxy may be less accurate than information given by the desired respondent. However, if the respondent is unavailable, someone in their household or family (who is therefore likely to know them well) may be able to offer some useful information about their activity. Since our users value high levels of completeness, we viewed the risks to accuracy and reliability as acceptable, if we could seek to minimise them.

Our survey therefore uses the following strategy to minimise proxy responses. During the first half of the field period for each cohort (approximately six or seven weeks) proxy responses are not sought by telephone interviewers. During the second half of the field period, interviewers are advised to collect responses from third parties, where possible, and where a suitable proxy respondent (defined as a partner, relative, carer or close friend) is available. Only the mandatory questions are asked of proxies, and subjective questions are excluded. We did not collect proxy responses from the graduates of English Further Education Colleges as a matter of policy<sup>84</sup>. We also make sure that responses collected from third parties do not exceed 10% of a provider's target population; limiting the impact on data quality.

In the first year of surveying, we introduced proxy surveying for the first time at the halfway point of cohort B, but no proxy responses were collected. We activated proxy surveying functionality again halfway through surveying cohort C, and we received 10 proxy responses between this point and the end of cohort C. We took the same approach by introducing proxy surveying as an option at the mid-way point during cohort D. Between that point and the close of cohort D, we surveyed 1,755 proxy respondents. This equates to 0.5% of the total number of responses.

Other than the data in the previous paragraph, we have not provided information on proxy responses in either the published outputs, or the microdata. To further support users' understanding of the likely accuracy or reliability of this data, we would ideally like to provide this information to all users in the main outputs. Since we possess the survey paradata required, adding this variable to the derived fields is a target for future improvement.

### 3.2.6. MEASUREMENT ERROR

Measurement error occurs from failing to collect the true data values from respondents. Sources of

---

<sup>83</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/evaluation>

<sup>84</sup> Information on who can answer the survey is available under the privacy notice <https://www.graduateoutcomes.ac.uk/privacy-info>

measurement error in Graduate Outcomes are: the survey instrument(s); the telephone interviewers, and the respondents themselves. This section of the report covers these aspects, in turn. The mode of data collection is also a source of measurement error, and we cover this in more detail in the next section.

#### 3.2.6.1. Respondent error

The survey takes the following measures to minimise respondent error. We cognitively tested the survey questions prior to launch, and adapted our questionnaire design in the light of the research findings. Information on cognitive testing is available in a technical report<sup>85</sup> and an outcomes report.<sup>86</sup> The implementation of the survey questions in the survey instrument was undertaken with expert input and testing from HESA and our suppliers, in order to pro-actively identify and overcome potential respondent error issues.

The survey instrument is available in both English and Welsh languages. This allows respondents graduating from providers in Wales to use whichever language they prefer. This should reduce respondent error due to language issues.

The instrument is deployed online, and over the telephone, which offers respondents some choice over how to engage. Details about the implementation of the instrument can be found in the methodology statement sections dealing with the online<sup>87</sup> and telephone<sup>88</sup> based aspects of our approach, and these materials also contain further information about how we seek to minimise respondent error. Online, we use a series of prompts to encourage the respondent to check the accuracy of their responses. Over the telephone, our interviewers' script similarly prompts operatives to elicit accurate responses through checking understanding back with the respondent. (We will from now on refer to the computer-assisted telephone interviewing by its standard acronym – CATI.)

Some examples of respondent error we believe may occur are:

- Information retrieval may be difficult for those respondents reporting several jobs. They may not remember precisely, or may not have access to, information about, for example, their previous earnings for a job they left months beforehand.
- Brevity or lack of response to free text questions could lead to differences in SOC codes for graduates in similar jobs. This equally applies to other coded free-text data. However, the SOC coding process would be more sensitive to this sort of issue, than, for example, free text country data, as the input data is more extensive, and there is some degree of semantic overlap between the output codes.
- Cases where respondents select unemployed and paid work simultaneously. (Of the respondents in paid work for an employer, 950 had also indicated they are unemployed. Of these, 270 had said that being unemployed was their most important activity).<sup>89</sup>
- Acquiescence bias (sometimes called agreement bias, 'straight-lining', or alternatively referred to as 'yea-saying/nay-saying') is where there is a tendency on the part of respondents to indicate positive (or negative) responses in a routine fashion, perhaps not

---

<sup>85</sup> See <https://www.hesa.ac.uk/files/Cognitive%20Testing%20Technical%20report.pdf>

<sup>86</sup> See <https://www.hesa.ac.uk/files/Cognitive%20Testing%20Outcomes%20report.pdf>

<sup>87</sup> For online aspects, see: <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/online-survey-design>

<sup>88</sup> For telephone and contact centre aspects of the instrument, see <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/telephone-survey-design>

<sup>89</sup> For details of how HESA reflects this contradictory information in published outputs, see the dissemination policy and XACTIVITY spec

reflecting their 'true' feelings. The design of the survey mitigates this by avoiding questions where this kind of response is easy to offer and HESA is continuously reviewing the impact of survey design on response distribution.

- Social desirability bias occurs where respondents tend to give socially desirable responses instead of choosing responses that are reflective of their 'true' situation. Examples where this could occur might include reporting a higher salary, or a greater sense of subjective wellbeing (SWB). Other studies have indicated that this kind of bias may vary by mode of response.<sup>90</sup>

For details of our investigations into these forms of respondent error, readers are directed to the section 3.2.8. Reliability of Sensitive Data, where we discuss our analysis of the data. While further work is required to investigate the extent of these forms of bias on the survey, we are able to show the current extent of our understanding of their effect.

In the Graduate Outcomes dissemination policy, details are given about how HESA interprets and publishes responses.<sup>91</sup> In the section of that document covering key data concepts and standards, explanations are given around the analysis that has been carried out on a number of key data items. In the section on salary, there is specific information about the approach HESA has taken to handling any potential respondent error. This includes trimming the salaries to exclude outliers, and future corrective actions, including improvements to the instrument to reduce the risk of misunderstanding that leads to respondent error.

One limitation on the respondent's ability to correct their own errors was HESA's decision to remove the 'back-button' functionality on the online deployment of the instrument mid-way through cohort A. Respondents are therefore unable to go back and change their answers to previous questions. This is done for data protection reasons (this is covered at greater length in the section of the methodology statement on the online survey design;<sup>92</sup> we also cover our estimates of the impact of processing error prior to the removal of the back button functionality, below, in section 3.2.9. Processing error).

During the first few cohorts, we noticed that some respondents indicated they believed they should not be in the sample because they had not graduated. This sometimes occurred when they had gone on to further study or had only completed part of their qualification but were still eligible to take part based on that component. We amended the introduction to the survey to allow interviewers more time to explain the eligibility, if needed. We also made necessary amendments to our emails and other communication highlighting the eligibility criterion as having completed a sufficient component of an HE course. This amendment was implemented in cohort C.

We are aware that more evidence needs to be gathered on whether respondent error represents a significant issue in the survey. For instance, for those who stated in the survey that they were undertaking further study in the UK HE sector, there is the potential to link their response to the HESA student record. This would offer the opportunity to evaluate the extent of measurement error in this part of the survey.

---

<sup>90</sup> For some further discussion of acquiescence and social desirability bias, see section 3.2.8. Reliability of Sensitive Data. We intend to report further on our investigations into these matters in the coming months.

<sup>91</sup> See <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf> pp.13-19.

<sup>92</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/online-survey-design>

### 3.2.6.2. Survey instrument error

Significant effort is invested in reducing opportunities for instrument error, and the first element of this is the choices of platforms, partners, and personnel involved. HESA manages the survey and appoints the suppliers.<sup>93</sup> HESA's procurement and supplier management approaches ensure that suppliers deliver on process quality requirements imposed by HESA. Conformat is HESA's current feedback management solution supplier. Conformat's technology is widely used to conduct surveys by leading sector bodies, including the Office for National Statistics, and also in market research contexts. It includes a smartphone compatible online system. HESA's current contact centre provider is IFF research. IFF has worked with many individual providers, previously, in their delivery of Graduate Outcomes predecessor DLHE. IFF was also the survey contractor for all six iterations of the Longitudinal DLHE survey.

The survey instrument is ultimately HESA's responsibility, and HESA is an official statistics producer with a track record in delivering the DLHE and LDLHE surveys for over twenty years. HESA's staff are skilled across the range of statistical business processes, including developing the methodologies, procuring survey and coding services, developing and commissioning software systems, data processing and enrichment, quality assurance, commissioning research, analysis, dissemination, and undertaking reviews. Users can therefore trust that the survey is being delivered by an organisation with experience and skill in appropriate professional domains.

The instrument was tested thoroughly by staff from HESA, IFF, and Conformat prior to deployment. However, the complexity of the survey routing meant that some less likely routing combinations were only tested to a limited extent. All problems discovered during testing were fixed prior to launch.

HESA and the Graduate Outcomes Steering Group considered whether to pilot the Graduate Outcomes survey or run it and the DLHE in tandem for a time in September 2017. The decision was that while a pilot was desirable, it was deemed infeasible due to time and cost constraints. The combination of strong management, effective procurement, and cognitive testing gave confidence that alternatives to a pilot were feasible, and might yield answers to the questions a pilot might answer, as effectively. The relatively few graduates in the first three cohorts also reduced the risks accruing from instrument error. The impact on graduates (who would be surveyed twice) and the costs involved were the main reasons for not pursuing dual-running of DLHE and Graduate Outcomes.

HESA demonstrates an evidence-based approach to operational data quality management, backed up by a clear governance approach. The Head of Research and Insight was able to provide a log that confirmed that all instances of potential instrument error that have been identified during data collection were investigated and assessed for the level of impact. The log showed that no matter how small the impact, the instrument has been fixed whenever necessary, showing assignment of tasks, description, and closure. This was substantiated by the regular progress updates, which explained these same issues to stakeholders<sup>94</sup>. Fixes were usually applied rapidly, but some issues were amended in time for the start of the second year of surveying. Higher-impact issues typically concerned aspects of operational survey management, survey routing, or missing or problematic quality rules on specific questions, and a subset of these will present some instrument error, especially in the first three cohorts. Other issues related to flaws

---

<sup>93</sup> See our press release: <https://www.hesa.ac.uk/news/14-11-2018/complete-graduate-outcomes-line-up>

<sup>94</sup> Readers wishing to understand these issues in detail, and in chronological order, are recommended to read the mid-point and end of cohort reviews, which are published at: <https://www.hesa.ac.uk/innovation/outcomes/about/progress/progress-updates-archive>

that generated excessive respondent queries, and where addressing these could improve respondent experience and/or improve operational efficiency. A few issues reflected design choices that were queried by users and will be evaluated for future action. Medium and lower impact sources of potential instrument error tend to focus on applicant experience improvement, and minor comprehensibility improvement, especially of the CATI script.

The survey instrument is generally of high quality, and any initial problems were not serious in effect, and were overcome rapidly. There will therefore be some modest instrument errors in data for cohorts A, B and C, and microdata users should be mindful of the potential impacts in those cohorts, especially on salary and to some extent, location data. (In HESA's published outputs, these issues are handled as noted in the Graduate Outcomes dissemination policy.<sup>95</sup>) Cohort D (which accounts for the overwhelming majority of graduates in the sample) was therefore conducted on a stable and well-tested instrument, and there is no evidence to suggest that instrument error is likely to be a significant factor in the survey data for this cohort.

We summarise the main sources of potential instrument error in the following subsections.

#### **3.2.6.2.1. Survey routing issues**

At the outset, the online version of the survey instrument incorporated a 'back button'. No issues with this functionality were identified in testing. Early analysis during cohort A identified <50 cases with unexpected instances of item non-response accompanied by additional responses that defied the survey logic. These showed characteristics that indicated random variance, rather than bias. Root cause analysis demonstrated that the back button could be used to generate these effects. Respondents appeared to have progressed along the survey routing, then retraced their steps and followed other routing paths, leaving a trail of 'orphaned' data from the abandoned route. The back button was removed from the online survey mid-way through cohort A. Further information about this is in the online survey design section of the methodology statement<sup>96</sup>.

CATI operatives retain access to a back button (to maintain a good interviewer-respondent relationship). This means that there is still a small risk of processing error arising in the same way, however this risk is ameliorated considerably through CATI operative awareness and training, and by increasing the validation checks undertaken either automatically, or through analysis.

A second issue was identified in the routing of the study section of the survey relating to two variables: UCNAME and STUCOUNTRY. This issue was fixed on 2019-03-22. In this instance, HESA decided to impute missing data from a trusted source. This issue is explained in detail in section 3.2.9.1. Imputation and editing.

A third survey routing issue (unrelated to the second issue) was also identified in the further study section of the survey. This routing issue affected the whole first year of the survey. It has affected 5,040 respondents in year one. This is likely to result in a higher number of records with missing country information, where the provider was 'other', and the name of the provider has not been supplied. This was caused due to an error in survey routing that did not consider a scenario where a person might refuse to provide an answer to a free-text question, when prompted, which should still lead them to the next immediate question and treat them the same as someone who provided information in the free-text box. While the survey was tested extensively before launch, the focus was largely on making sure that the completion logic is working as expected i.e. response to one question leads to the presentation of the following question. In the present scenario, however,

---

<sup>95</sup> See <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

<sup>96</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/online-survey-design>



absence of response should have also been treated as a 'valid response'. The statistical release for year one is not affected by this as the outputs are not disaggregated by the country of further study provider. However, microdata users will wish to be aware of this data quality issue.

In response to these issues, a wider review of the implementation of the questionnaire is currently being planned and is scheduled to start in autumn 2020. Included in this review is further testing of survey routing with a specific emphasis on the handling of unexpected or less expected responses. We will announce any further issues and corrective actions identified by this review once it has completed.

#### **3.2.6.2.2. Survey alterations to increase retention**

We monitored respondent progress through the survey throughout. During the first cohort we noticed a higher drop-out rate than we expected, and isolated the step from the introduction screen to the first question as implicated in 89% of the total drop-out rate. To address this, we merged the first two pages (screens) of the survey and reduced the amount of descriptive text without sacrificing key messaging or fair processing information. This now means that when graduates click the link to the survey, they will go straight into question one. This change was implemented from the start of cohort B. We noticed the drop-out rate decrease by nine percentage points in cohort B, compared to cohort A.

During the implementation of the first two cohorts, we identified areas of the survey that could be improved to enhance user experience, which in turn would impact non-response and data quality. Modifications were only made where they were necessary and would enable us to collect better quality data. We ensured they had no impact on the integrity of the survey itself and were not likely to introduce bias. Therefore, most changes made for cohort C were minor wording amendments to questions and instruction text to aid usability and accessibility for graduates.

One significant change in the survey (made for cohort C), involved adding new routing functionality. This enables graduates (who have said their two working activities are the same) to only be asked one set of questions regarding their activities, instead of being asked two sets about the same activity. This was a necessary change to help mitigate survey fatigue and improve the flow of the survey. This amendment does alter the structure of data available for 11,110 graduates, but does not lose any information.

The main modifications introduced at the beginning of cohort D aimed to decrease the length of the survey by reducing the number of pages in the online survey that graduates would have to click through. In some cases, we were able to remove a few redundant pages that contained contextual information and merge them with questions on one page (whilst keeping the question clear and accessible). In other cases, we were able to remove a page altogether as the information had already been covered on a previous page or in the question itself.

Other substantive changes made to the survey included the removal of 'N/A' for employment intensity questions (only full-time and part-time), tweaking of instruction text to better define the question criteria (what is required in the question) and to improve the flow of the survey.

#### **3.2.6.2.3. Email delivery**

During cohort B we experienced some issues with the delivery of emails to graduates, where they were mistaken as unsolicited email and therefore blocked by internet service providers. This problem saw key providers (e.g. Gmail, Office 365) blocking our email invitations (by blacklisting our IP address). In the short term, we made changes to the setup of our IP address and staggered

the deployment of email invitations over a longer period in the first week of cohort C. This worked well and we saw an improvement in the health of our IP address as a result. In the longer term, HESA determined that the best course of action was to engage with a third-party that provides expertise in managing the sending of emails.

Accordingly, at the end of cohort C, we implemented a new system called MailJet to help us with the issues related to internet service providers (e.g. Gmail, Office 365). At the same time, we had also been looking to trial new activity that aimed to improve our response rates and a pre-notification email (a warm up) was one strategy that we had been considering. To successfully warm up our new IP addresses and keep them warm, we needed to move to an almost continual delivery of emails. We determined that this warmup activity should commence within cohort C. The proportion of responses achieved online improved significantly from 30% in cohort B to 42% in cohort C (as a proportion of total number of completed responses for each cohort). In mid-August, we staggered the delivery of a pre-notification email to approved cohort D graduates that shared key information about the survey and let them know they would receive their unique survey link in early September 2019.

A pre-notification email<sup>97</sup> was sent to the first email address of every graduate in cohort D with an approved email address. According to MailJet, 96% of these emails were delivered. Of the emails that were delivered, 36.3% were opened and of these, respondents clicked a link in 1.5% of them. Note that pre-notification emails did not contain a link to the survey, but respondents could access the Graduate Outcomes website (for more information on its purpose, privacy and data collection methods) as well as our social media channels. As a result we saw an increase in survey engagement online.

#### **3.2.6.2.4. Call handling**

Managing the call queue to ensure all graduates have an equal chance of being called is an intricate planning process. This is particularly the case given that we may have several telephone numbers for each graduate, some in different time zones. Prioritisation is key, whilst maintaining the principle of randomisation (and acknowledging that there are multiple ways to randomise). For cohort A, we gave equal priority to all telephone numbers supplied for a graduate and a higher priority to UK domiciled graduates. Randomisation is designed to give all graduates at each provider an equal chance of being called during the fieldwork period.

In cohort C we introduced a new call management system which enabled more systematic calling of all graduates, using all numbers available.

Using pre-determined rules, we ensure that the time between calls to each graduate is appropriately spaced out and that all graduates receive a similar number of call attempts, unless we are deliberately attempting to boost the response rate of a sub group that is under represented, as a part of our adaptive survey design.

#### **3.2.6.3. Interviewer error**

Interviewer error is the effect of a human interviewer on the data gathering process. Graduate Outcomes uses many interviewers concurrently. CATI interviewers undergo training developed especially for the Graduate Outcomes survey, and which focuses on the contextual knowledge interviewers need to perform their roles effectively. They are recruited and trained by IFF according to closely-monitored quality criteria. Quality assurance by monitoring calls is also a part of the

---

<sup>97</sup> See <https://www.hesa.ac.uk/definitions/operational-survey-information#survey-materials> for sample communications.

standard practice. All interviews are recorded digitally to keep an accurate record of interviews. A minimum of 5% of each interviewers' calls are reviewed in full by a team leader. Quality control reviews are all documented using a series of scores. Should an interviewer have below acceptable scores, this will be discussed with them along with the issue raised, an action plan agreed and signed, and their work further quality controlled. Information about this is covered in the data collection section of the methodology statement.<sup>98</sup> Further details are given in the operational survey information section on the contact centre.<sup>99</sup>

CATI operatives utilise an adapted version of the same instrument as online respondents. This allows a further level of data quality checks to be performed, as CATI operatives get similar feedback from the online instrument to online respondents, in addition to having their own quality processes built into the script. This also prevents any 'clash' or data problems occurring due to respondent mode switches. One difference is that a 'back button' is available to CATI operatives, which allows adjustments to be made, if a respondent wishes to change an earlier answer in the light of a later question. This kind of anecdotal feedback could help identify potential sources of respondent error, and HESA and IFF evaluate feedback from CATI operatives regularly, to determine if instrument improvements could offer marginal enhancements to data collection. While human error is always a potential factor, this is likely to be a matter of random variance in keying errors. There is no evidence to suggest that interviewer error has had any significant impact on the conduct of the survey. Rather, CATI operatives are a useful source of quality improvement suggestions, and regular fortnightly meetings occur where performance and survey issues are discussed, and recommendations logged for further assessment and action.

### 3.2.7. MODE EFFECTS

A mode effect is a systematic difference that is attributable to the mode of data collection. Analysing the effect of mode on item responses (and aspects of response propensity) is part of our future programme of work. Prior to designing a research project, we need to do some preparatory work to define what we mean by 'survey mode', rigorously. We explain this situation below.

Mixed-mode surveys are increasingly common. A typical research survey operated in a mixed-mode fashion might survey a sample electronically, and then follow-up with a telephone survey later on, either to provide a more qualitative set of insights into a sub-sample, or to address non-response issues during the initial survey period. There are many possible such designs. The design of the Graduate Outcomes survey was a collaborative exercise that took into account knowledge developed by HESA and the HE sector during the operation of DLHE and LDLHE, its predecessor surveys. One important factor we took into account was the widely-held perception that telephone surveying from an early stage, combined with online surveying, was likely to be necessary in order to meet user needs for both high response rates and efficiencies generated through an online mode. We therefore sought to retain the best aspects of the previous practices, and these are reflected in making a concurrent mixed-mode design our adopted approach.<sup>100</sup>

---

<sup>98</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/data-collection>

<sup>99</sup> See <https://www.hesa.ac.uk/definitions/operational-survey-information#contact-centre-methodology>

<sup>100</sup> For completeness, we must explain that a separate, paper-based approach is used in a minority of cases where respondents are known not to have access to a telephone or computer. This mode asks the mandatory questions required for a complete response. Only 25 postal responses were received during the first year of surveying. Because these responses are so few, we do not discuss the paper-based mode further in this report.

Our approach is described in detail in the section of the methodology statement covering data collection,<sup>101</sup> and in the associated operational survey information.<sup>102</sup> It is underpinned by a single technology solution (Confirmit) that links online (mobile and desktop) and telephone-based modes together seamlessly. Survey responses can be saved and picked-up later, in either mode. In practice, this means that respondents may begin the survey in one mode, and end it in another, or even, potentially, change mode several times during the period of time during which they are engaging with the survey. The system logs all events, and these system logs form the basis of HESA's paradata, including modal information. The paradata, which also includes timing and duration information, is very rich, and, as we learn more about the capabilities of this system, we are extending the formally-catalogued paradata we wish to extract from the system. This system-generated logging data is, in its own way, as rich as the collected survey data itself, and offers us insights into the behavioural characteristics of respondents. When combined with our data on the population characteristics, it also yields potential insights into non-respondents. Our first task is therefore to define more precisely the characteristics of the various survey engagement modes.

Our current paradata dictionary<sup>103</sup> includes variables for the start mode, partial completion mode, completion mode, various status markers, last question viewed, number of calls made, and a range of variables relating to the sending of emails and SMS messages. Over the last few cohorts we have been using some of this paradata to inform our data collection processes such as identifying the most suitable time for sending emails and SMSs based on completion times, changing subject lines to encourage higher email open and click rates, monitoring interviewer performance using average number of calls, to name a few.

However, we are aware of the additional potential hidden within the various markers in the system, which could yield additional formally catalogued paradata. We are keen to use it to support operational improvements, as well as to investigate mode effects.

The current focus of our work is therefore on better understanding and cataloguing the paradata available to us, to determine the likely limits of our analysis. Once this work is complete, we intend to develop deeper understanding of respondent behaviours and characteristics, and non-respondent characteristics. We will adopt an appropriate analytical approach to yield two statistical benefits, both with practical utility.

The first benefit is that we will be in a position to characterise more fully the modal patterns of survey engagement. There will be simple categories for those respondents who complete in one go by either mobile, desktop, or CATI. We will also be in a position to observe and categorise the various mixed-mode patterns, and take into account factors like numbers of engagements, and various timescale factors. This study will permit us to cluster and characterise the observed engagement approaches of respondents, into an emergent set of modal categories. We can then utilise these emergent modes in our analysis of mode effects. Given the highly integrated management of respondent engagement, we believe this approach to be a better strategy than attempting a necessarily flawed analysis of mode effects utilizing the current paradata only. This analysis will also yield results that permit us to plan a research project. Ideally, this would be a randomised controlled trial, but we are still at a very early stage with this work.

The second benefit of this approach is that it will yield useful insights into potential determinants of respondent behaviour. By linking our emergent modal categories to both survey data and

---

<sup>101</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/data-collection>

<sup>102</sup> See <https://www.hesa.ac.uk/definitions/operational-survey-information>

<sup>103</sup> This is an unpublished internal document at the time of writing.

population data from the Student record, we hope to learn more about the characteristics of the respondents in the different modal categories. We can also compare these characteristics to known characteristics of non-respondents. Our goal in this respect will be to generate behavioural insights, which may help us in developing improved engagement strategies. This could address non-response error, raise overall response rates, and pursue operational efficiencies that benefit our stakeholders.

### 3.2.8. RELIABILITY OF SENSITIVE DATA

The Graduate Outcomes survey collects data on a number of topics which might be considered sensitive. In the case of both salary data and subjective wellbeing (SWB) data, there is a risk of social desirability bias, inasmuch as respondents might be expected to assume that some responses (such as higher salaries and generally higher SWB scores) are more favourable than others. Given that data on both of these topics is reported back to HE providers – although SWB data is only returned to providers in aggregate – there is some possibility that respondents will feel an incentive to answer questions about salary and SWB in such a way as to create a more favourable impression of how they are doing 15 months after course completion. In order to ascertain whether these potentially sensitive areas are subject to an elevated risk of misreporting, HESA has therefore undertaken a range of quality checks.<sup>104</sup>

As part of the quality assurance process for the Graduate Outcomes salary data, HESA conducted a series of comparisons between Graduate Outcomes and Longitudinal Educational Outcomes (LEO) salary data in order to check that the Graduate Outcomes salary data was not showing any unexpected patterns.<sup>105</sup> Although there are some key differences between the two datasets, including the fact that LEO does not distinguish between full- and part-time earnings, the comparisons still allowed us to see that the Graduate Outcomes salary data exhibited the trends which we would expect. Looking at the 2017/18 tax year, LEO data showed that median graduate earnings a year after graduation were £19,700 for females and £21,000 for males. Median Graduate Outcomes salaries for holders of first degrees were slightly higher, with females earning £22,000 and males earning £24,000, but the difference in earnings between male and female graduate in the two surveys is roughly equivalent. A comparison between Graduate Outcomes salary data by JACS subject and equivalent LEO data likewise showed broad similarities, with both datasets showing graduates with degrees in medicine and dentistry earning the highest salaries and graduates with degrees in creative arts and design earning the lowest salaries. A further comparison between these data sources cross-tabulated by SOC major groups again showed similar patterns.

We are also aware of a potential issue with the wording of salary question that required respondents to round salary to the nearest thousands. One theory is that some respondents may interpret this as a request to provide leading whole numbers without the zeros; for example, 23 instead of 23,000. To counter this misinterpretation, we have removed the requirement for respondents to round the data as it can easily be handled in outputs. This was implemented in the

---

<sup>104</sup> For details of the various kinds of respondent error that we consider salient to Graduate Outcomes, see section 3.2.6.1. Respondent error.

<sup>105</sup> For the LEO data used in these comparisons, see Department for Education (2020), *Graduate Outcomes (LEO): employment and earnings outcomes for higher education graduates by subject studied and graduate characteristics in 2017/18*.

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/874410/2020\\_03\\_HE\\_LEO\\_main\\_text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/874410/2020_03_HE_LEO_main_text.pdf)

second cohort of year two, and we will only be able to confirm whether this change has resolved some of the quality issues experienced thus far after we have collected sufficient data for analysis.

As a first quality check of the SWB data, Graduate Outcomes SWB data was compared to data on SWB from the Annual Population Survey (APS), looking at the percentage of respondents whose answers fell into each ONS individual response option for each of the SWB questions.<sup>106</sup> While the responses to the happiness, life satisfaction, and life worthwhile questions were relatively consistent between the two surveys, Graduate Outcomes showed a higher percentage of respondents in the 'very high' anxiety band and a considerably lower percentage of respondents in the 'very low' anxiety band. The differences were present in both online and CATI responses, but were more striking in the case of online responses.

Several possible explanations for this discrepancy between Graduate Outcomes and APS SWB data were explored. One possible explanation was that some respondents were 'straight-lining', that is giving the same answer to a series of successive questions, whether or not that answer is the most appropriate for each question in the series. Of the 312,810 respondents who completed all four SWB questions, only 1.59% apparently straight-lined by giving the same numerical answer for all four questions; the straight-line responses were, however, more common at the higher end of the numerical scale, and online respondents were more likely to straight-line than CATI respondents. Although the overall incidence of straight-lining in the Graduate Outcomes survey appeared to be very low, it could therefore account for some of the differences between the anxiety data collected by the Graduate Outcomes survey and that collected by the APS.

The possibility that the reverse coding of the anxiety question led to confusion was also explored. While for the questions about happiness, life satisfaction, and whether respondents feel their lives are worthwhile, high numerical answers correlate with high SWB, the opposite is generally true for anxiety. That is, a score of 10 on the happiness question indicates feeling extremely happy, while a score of 10 on the anxiety question indicates feeling extremely anxious, which is generally a negative emotional state. While it is possible to imagine a scenario in which a respondent would feel highly anxious but also happy, satisfied with life, and that their life is worthwhile, high levels of anxiety are not generally associated with high levels of personal wellbeing.<sup>107</sup>

Of respondents who completed both the anxiety and the happiness questions, 11.16% answered both questions in the 7-10 range; if straight-lining response are removed, 10.34% of respondents still report both happiness and anxiety scores in the 7-10 range. Of those respondents who listed an anxiety score of 7 or above, over half also selected a happiness score in the same range, which suggests that the reverse coding of the anxiety question may have been a source of confusion. Another possible explanation for this pattern is a degree of social desirability bias, according to which candidates are reluctant to provide a negative answer to the happiness question, even when they are experiencing high levels of anxiety.<sup>108</sup> Social desirability bias could also be compatible with straight-line responses at the high end of the numerical scale, if respondents, feeling some pressure to answer the SWB questions in such a way as to give an impression of high all-around

---

<sup>106</sup> For HESA's use of the ONS banding scheme in our analysis of the SWB data, see section 3.5.1. National and international data standards below.

<sup>107</sup> Arab, M., Rafiei, H., Safarizadeh, M., Ahmadi, J. and Safarizadeh, M. (2016). 'Stress, anxiety and depression among medical university students and its relationship with their level of happiness'. *IOSR Journal of Nursing and Health Science (IOSR-JNHS)*, 5(1):44-47.

<sup>108</sup> On the likelihood of social desirability bias in SWB data, see Duffy, B., Smith, K., Terhanian, G. and Bremer, J. (2005). 'Comparing data from online and face-to-face surveys,' *International Journal of Market Research*, 47(6), 615-639.

wellbeing, were to miss the reverse coding of the anxiety question and give answers of 10 for all four SWB questions.

In order to reduce the risk of confusion about the anxiety question and to discourage straight-lining, several changes have been implemented from year two of the online survey. For the first three, positively worded questions (happiness, life satisfaction, and life worthwhile), the question is asked, followed by the scale of possible responses; for the negatively worded anxiety question, however, additional wording clarifying the scale has been added before the question to highlight the change in direction of coding. The layout of the questions and possible answers has also been adjusted to avoid a grid-like layout, which might otherwise encourage straight-lining. These changes have been made with reference to other national surveys which use the ONS4, bringing HESA's use of the ONS4 into closer alignment with that of the ONS.

While we cannot yet identify the precise effect of survey mode on SWB responses, the quality checks of SWB data did suggest that there are some differences in responses to the anxiety question depending on the mode in which the survey was completed. These differences may be attributable in part to social desirability bias, which has been shown to be a factor in telephone interviews.<sup>109</sup> On the one hand, online responses may reflect a lower degree of social desirability bias than CATI responses, with online respondents feeling comfortable reporting higher levels of anxiety. On the other hand, otherwise unexpected responses (such as straight-lining or reporting both very high anxiety and very high happiness) were more common when the survey was completed online. Graduate Outcomes is one of the very few online implementations of the ONS4 known to us, and as such there is a prospect of further work to ascertain the impact of online delivery (see also the previous section). Further work is therefore still to be done on the effect of survey mode on the reliability of SWB data.

### 3.2.9. PROCESSING ERROR

Processing error includes processing-related errors in data capture, coding, editing and tabulation of the data. This section describes the processes used and the quality assurance apparatus that is employed to avoid bias in processing, and to limit the incidence of variance. We cover the issues that have arisen, and our estimates of their impact.

HESA's processing practices and quality assurance approach are explained in the methodology statement section on data processing.<sup>110</sup> It covers data capture, data quality checking, SIC/SOC data coding (where HESA employs a specialist contractor), free text field 'cleaning', and derived fields.

#### 3.2.9.1. Imputation and editing

In the first year of Graduate Outcomes data processing, HESA has applied imputation in one variable that records which country the graduate was studying in on the census week.<sup>111</sup> This variable (STUCOUNTRY) is required to be answered when the previous question, which identifies the university or college the graduate studied at (UCNAME) has an answer that is either not in the pre-defined list of providers, or has not been answered. However, the issue has arisen in the routing where if UCNAME has not been answered, then STUCOUNTRY does not display and cannot be answered. This issued affected respondents meeting the above conditions prior to a fix

---

<sup>109</sup> Dolan, P., & Kavetsos, G. (2016) Happy talk: mode of administration effects on subjective well-being. *Journal of Happiness Studies*, 17(3), 273-1291. 10.1007/s10902-015-9642-8

<sup>110</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/data-processing>

<sup>111</sup> See the field specification at: <https://www.hesa.ac.uk/collection/c17072/a/stucountry>

being applied on 2019-03-22. Of the 2,260 graduates that have missing data, we successfully imputed observations for 625 of them.

Our solution is, where possible, to use imputation to fill the gap of missing data by utilising linked data from the 2018/19 Student record(s) to identify graduates who studied at a UK Higher Education provider whose data is collected by HESA. A process of fuzzy matching was carried out to attempt to link these graduates to the HESA Student and AP records and for the appropriate country code of further study to be picked up (England = XF, Wales = XI, Scotland = XH, Northern Ireland = XG).

HESA has not determined any other area in which editing or imputing missing or unclear values would unambiguously offer enhanced value during the timeframe prior to the first published outputs. However, we will keep this determination under review, pending feedback from users.

### 3.2.9.2. SIC and SOC coding

SIC and SOC codes are applied wherever we have sufficient data to allow this. The data processing section of the methodology statement<sup>112</sup> explains this further. An experienced external supplier (Oblong<sup>113</sup>) undertakes this coding, and the quality checks they apply are explained in the methodology statement. Established SIC-coding methodology has proved stable over the long term.<sup>114</sup> A new method had to be developed for SOC coding. Provisional SOC codes were processed using an agreed method by Oblong. These were then supplied to HE providers (through the Portal) which were invited to quality assure the data for themselves. During this phase, more than 90 providers undertook peer review. This was a semi-structured quality assurance process and relied on the varying resource that providers were able to bring. Although we received feedback from only a sub-set of providers, any changes to SOC coding resulting from this feedback were applied consistently across the entire collection.

All the provider feedback received was individually reviewed and placed into one of the following four categories: Systemic (where the error is widespread and there is a clear pattern of miscoding); Non-systemic (isolated cases); Inconsistent (where multiple records in an occupation group are coded inconsistently with no obvious pattern) or Not actionable (no basis or evidence exists for coding to be changed).

This helped us identify potential processing issues that affected a large number of records. Non-systemic issues could not be used to improve individual-level data, as this would have been inequitable, and introduce bias through inconsistent application. This exercise revealed some systemic errors in SOC coding, as well as scrutinising some areas where the coding ultimately met our quality standards. An overview of this process can be found in the data processing section of the operational survey information.<sup>115</sup> Detailed information on the exercise undertaken to review feedback and improve the data processing approach is also available in a detailed briefing<sup>116</sup>, which identifies the impact of the issues identified.

---

<sup>112</sup> See <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology/data-processing#data-coding>

<sup>113</sup> Information on our suppliers is here: <https://www.hesa.ac.uk/innovation/outcomes/about/our-suppliers>

<sup>114</sup> HESA has commissioned Oblong as a SIC-code supplier in the past, using DLHE data that was similar to the structure of the relevant parts of Graduate Outcomes data. This longstanding methodology continued to prove robust.

<sup>115</sup> See <https://www.hesa.ac.uk/definitions/operational-survey-information#data-classification-sicsoc>

<sup>116</sup> See <https://www.hesa.ac.uk/innovation/outcomes/providers/assessment-1718-soc-coding>

---



As a result of the exercise described above, 8% of records from the first year of surveying have had SOC codes changed. Many users group all codes within SOC major groups 1-3 to identify 'professional and managerial', 'highly skilled', or 'graduate' jobs. Of all the records that changed at the major group level as a result of identified coding problems, 28% moved from major groups 4-9 to 1-3; 13% from major groups 1-3 to 4-9 and the remaining 59% continued to be coded within these two groupings. We also amended the logic to remove the impact of qualification requirements on coding, and to allow many partially-completed responses to also be coded, increasing the usefulness of the data.<sup>117</sup> As a result of this comprehensive checking exercise, we believe the sources of systematic processing error identified by HE provider manual quality checks have been removed, and the processing system fixed. There is no evidence that there is any remaining bias in the coding strategy for SOC, and any remaining processing error is likely to be minimal, and the product of random variation only. We continue to review and refine our future approach to coding and quality assurance, working in partnership with our supplier and other stakeholders.

### 3.2.9.3. Handling free text responses

Most questions in Graduate Outcomes map directly to established lists of values, and details of these are available in the coding manual.<sup>118</sup> However, there is often an "Other" option that permits a free text response. In this subsection and the subsequent ones, we cover the most important issues relating to free text processing, and explain the risks around processing error, giving our estimates for this.

At the end of the collection process, data returned for questions that permit a free-text response goes through a cleansing process, in order to improve data quality. This is usually where the respondent has not chosen a value from the drop-down list provided but has instead selected "other" and typed their own answer. This process also runs for questions seeking postcode, city/area and country of employment, or self-employment / running own business; country in which graduate is living and of further study; provider of further study, and salary currency. Where possible, the free text is mapped to an appropriate value from a dictionary published within the appropriate derived field specification.

We have encountered some specific issues in the processing of UK-based location information, which we turn to next. Later subsections offer comparable quality descriptions of cleansing of country and salary data.

### 3.2.9.4. Location of employment data (UK) – handling free text

Respondents in employment are asked to tell us where they worked during the census week.<sup>119</sup> The majority of respondents supplied data that we could process into a structured format, such as

---

<sup>117</sup> We undertook an investigation to determine whether an improved methodology could yield more complete SOC code data, without reducing consistency. We learned that accurate coding could be achieved for some records that were not previously coded. As a reminder, our previous methodology required four fields (Company Name; SIC code; Job title; Job Duties Description) to be completed by the graduate for SOC to be coded. However, on reflection, where we have found that responses of sufficient quality have been provided in job title and job duties, even where the employer's name and/or duties are missing, we can derive a code, satisfactorily.

<sup>118</sup> The Graduate Outcomes survey results coding manual is available here:

<https://www.hesa.ac.uk/collection/c17072>

<sup>119</sup> This data is gathered through various survey questions (dependent on routing) and stored in the fields: EMPPLC; EMPPCODE; EMPPCODE\_UNKNOWN; EMPCOUNTRY; EMPCOUNTRY\_OTHER, and; EMPCITY. We also collect parallel data on self-employed graduates, using the fields: BUSEMPPLC;

their employer's postcode.<sup>120</sup> However, not all respondents know the information, or have it to hand, and in these cases a free text response is possible. Where respondents offered free text responses, we parsed these using an algorithm that attempts to make an exact match between our dictionary of text strings aligned to geolocation data, with other text strings (in this case, the free text responses) to generate location data. Location data is then mapped to as many levels of a standard geography as we can: a basic UK country/EU/overseas split; (detailed) country; UK government region, and UK counties/unitary authorities.

Our exact matching process is specified in detail in the two processing fields ZEMPAREA<sup>121</sup> and ZBUSAREA.<sup>122</sup> Predictably, we encountered issues with certain place names having different labels in the look-up files preventing an exact match on the string. Examples include Chester and Hull which are listed as Cheshire West and Chester and Kingston upon Hull respectively. Newcastle is another example which cannot readily be mapped to either Newcastle upon Tyne (North East) or Newcastle-under-Lyme (West Midlands) or even the Newcastle in Wales. Searching for containment of place names within the look-up file also posed problems with one place name being a common components of other place names. Chester is a prominent example as it is contained in the place names Manchester and Chesterfield, among others. Whilst we were able to pick out some of the larger towns and cities within the string, for the reasons given above, it was not possible for all and therefore introduces a degree of bias in the data. We explore below the extent to which these processing errors can be considered to be random.

Of the 245,315 responses where graduates were in employment or voluntary work in the UK, 42,285 (17.2%) proved to be unmatchable<sup>123</sup> below the level of the UK country. Within this unmatchable data, there were 18,129 separate unmatchable text strings. These are not distributed evenly: this is a dataset with a 'long tail'. A single string ('Newcastle') accounts for 1.6% of the unmatchable total (or 0.3% of the total in employment or voluntary work in the UK). The next most prevalent string ('Hull') accounts for 0.8% of the unmatchable total (or 0.1% of the total in employment or voluntary work in the UK). The 100 most common unmatchable responses account for 21.3% of the total of unmatchable responses. The 1,000 most common unmatchable responses account for 47.7% of the total of unmatchable responses.

Some of these strings would not be codable by any means, as searching the responses selectively reveals, for example, 300 instances of 'N/A', 165 of 'REFUSED', 30 examples of 'RATHER NOT SAY', and ten examples of 'Refused to answer'. There appear to be others, such as the enigmatic 'ref' (25 examples), and in total these and similar indications of likely intended item non-response

---

BUSEMPPCODE; BUSEMPPCODE\_UNKNOWN; BUSEMPCOUNTRY; BUSEMPCOUNTRY\_OTHER, and; BUSEMPCITY. Results for these fields are similar in proportion to those in employment, though the prevalence of self-employment is much lower, and hence we do not offer a detailed analysis on this much smaller group. Detailed metadata on all these fields can be viewed by following links from the data items index in the Graduate Outcomes survey results record coding manual, here:

<https://www.hesa.ac.uk/collection/c17072/index>

<sup>120</sup> Post-processing, location data can be found in the following derived fields: XMLOCGR; XMLOCN; XMLOCUC; XSTULOCGR; XSTULOCN; XSTULOCUC; XEMPLOCGR; XEMPLOCN; XEMPLOCUC; XBUSLOCGR; XBUSLOCN; XBUSLOCUC, and; XCURRLOC. Details of the processing involved in production is described by following the relevant links available from the derived fields specification contents page in the Graduate Outcomes survey results record coding manual, here:

<https://www.hesa.ac.uk/collection/c17072/derived/contents>

<sup>121</sup> See the derived field specification: <https://www.hesa.ac.uk/collection/c17072/derived/zemparea>

<sup>122</sup> See the derived field specification: <https://www.hesa.ac.uk/collection/c17072/derived/zbusarea>

<sup>123</sup> We use 'unmatchable' to mean 'currently unmatchable using our current algorithm'.

probably account for an estimate of between 2% and 3% of the unmatchable total (perhaps 0.3% of the total in employment or voluntary work in the UK).

We evaluated the data manually for evidence of potential bias. The main issue we identified relates to derivatives of the word 'newcastle'. We identified 225 separate unmatched free text uses based on derivative spellings of 'newcastle'. The most prevalent of these is the spelling 'Newcastle', (as noted above, this accounts for 685 responses, or 1.6% of the total unmatched responses. Manual investigation also revealed an additional unknown number of likely misspelled examples like "Newcatle" and "Newcasrle". We estimate that various spellings and intended spellings of 'newcastle' account for around 2.1% of the unmatchable total. We also note that there are at least five separate places in the UK spelled in a similar way to 'newcastle' (not to mention six in the Republic of Ireland and more further afield). It is therefore not possible to unambiguously identify which of these places is indicated by the free text response. Although the substantial majority of these responses are likely to refer to 'the Toon'<sup>124</sup> (as it is the largest conurbation identified by a variation of this spelling, rather than any of the various smaller 'newcastles') we cannot be definitive about how many. This equates to a probable overall slight unfavourable bias of location of employment/self-employment against being mapped as located in the North East of England (or Tyne and Wear, or Newcastle-upon-Tyne) in our data, to the extent that we estimate that around 0.2% of missing data on the location of employment or voluntary work, nationally, might have been expected to refer to that place.

The preceding narrative gives a sense of the scale of the challenge when applied to other place names. Other free text spellings that we could not map were less numerous, as the distribution described in this section indicates. Given the volume and heterogeneity of unmatchable text, we had to determine an appropriate course of action, with the options to either 1) edit the values manually, or 2) to accept the values as effectively missing for the time being. Resources were insufficient to pursue the first option, and there is a risk of introducing additional bias, unnecessarily. We therefore accepted a lower quality of data than we would prefer for use in our publications. Regional analyses are, as a result, more limited than we intended in the Statistical Bulletin and open data.

For analytical purposes dealing with the UK as a whole, or the entirety of each Devolved Administration, the resolution of our data on location is reliable. We obtain a high degree of completeness, because we collect additional data at country level. For users performing regional and sub-regional analysis, we suggest an appropriate allowance for error based on missing data, could be made. The precise form of this would depend on the analysis being undertaken. Regional (and lower-level) geographical analysis is becoming increasingly important in understanding graduate labour markets, and the economic and civic impact of HE providers on places. We therefore do not regard the level of data quality achieved in year one as likely to meet increasing user needs, and we are focussing efforts in two areas. The first is in improvements to the survey instrument. We believe it is likely that an auto-suggest mechanism could radically improve data quality. We are currently evaluating two potential options for this and hope to introduce an improvement in due course. The second focus for improvement is on post-publication data processing. We would like to improve our approach to matching free text response data, and

---

<sup>124</sup> We are advised this is the correct way to refer to the City of Newcastle on Tyne, on the authority of the Evening Chronicle: <https://www.chroniclelive.co.uk/news/history/call-newcastle-toon-12917622> (We also believe their interviewee "Dr Adams Means" *sic* is most likely to be Dr Adam Mearns of the University of Newcastle - <https://www.ncl.ac.uk/eill/staff/profile/adammearns.html#publications> whom we take this opportunity to thank).

we are currently considering what additional post-publication processing can be undertaken to offer enhanced user value from the existing data, prior to the next annual release.

### 3.2.9.5. Employment country data – handling free text

HESA has developed an algorithm<sup>125</sup> for the processing field ZEMPCOUNTRY, which cleans up the data provided by a graduate responding to the Graduate Outcomes survey question “In which country is your place of work?”. It does this by combining the data provided in EMPCOUNTRY (which is based on a restricted list of values available to the respondent as a drop-down menu) with that from the free text field EMPCOUNTRY\_OTHER. From a sample of 32,940 total responses from graduates in employment outside the UK, 185 free text responses were not able to be processed (0.6%). Manual checking revealed that answers were split between three emergent categories. First, intended item non-responses (refusals) of whom around half indicated they could not share their work location for security reasons, and a handful who had a geopolitical opinion to share on the available drop-down options. Second, responses that seemed to indicate probable intended geographical locations included those that could be matched with a better dictionary (e.g. ‘Dubai’; ‘Ibiza’; ‘Holland’) and some that couldn’t (‘djabalta’; ‘gersny’; ‘fgn’). Third, and most significantly was a category of globally mobile workers. Many of these individuals gave variations on the theme of ‘working on a ship’ but a few seemed to indicate a nomadic lifestyle following work from country to country, or regularly being posted to remote locations. Details of the analysis are presented in Table 2. We judge the small number of effective item non-responses that go beyond the limits of our processing approach to not be material for most analyses, despite being missing not at random (MNAR). Based on these findings, in cohort C of year two we have removed the option ‘other’ from the list of countries.

Table 2 – Frequency analysis of unmatchable text responses to country of employment or voluntary work

| Manually coded category                 | Number     |
|---|------------|
| Globally mobile worker                  | 35         |
| Probable intended geographical location | 30         |
| Refusal (intended item non-response)    | 10         |
| NULL                                    | 110        |
| <b>Total</b>                            | <b>185</b> |

### 3.2.9.6. Salary data – handling free text

HESA has developed an algorithm<sup>126</sup> for the processing field ZCURRENCY, which cleans up the currency data provided by a graduate responding to the Graduate Outcomes survey question(s) “What currency were you paid in?” It does this by combining the data provided in EMPCURRENCY (which is based on a restricted list of values available to the respondent as a drop-down menu) with that from the free text field EMPCURRENCY\_OTHER. From a sample of 306,670 total responses of graduates in employment, 36,110 could not be processed (11.8%). Most of these were nulls – instances of item non-response. 155 free text responses were not able to be processed (0.05%). Manual checking revealed that free text responses were split between various forms of intended item non-responses (refusals); indications of payment currencies not resolvable by the algorithm (or in most cases, by the analyst), and; in a handful of cases, respondents who were paid in multiple currencies. Details of the analysis are presented in Table 3. We judge the small number of effective item non-responses that go beyond the limits of our processing approach to not be material for most analyses. These responses appear to be missing at random (MAR). Based on these findings, in cohort C of year two we have removed the option ‘other’ from the list of

<sup>125</sup> See the spec. for ZEMPCOUNTRY: <https://www.hesa.ac.uk/collection/c17072/derived/zempcountry>

<sup>126</sup> See the specification for ZCURRENCY: <https://www.hesa.ac.uk/collection/c17072/derived/zcurrency>

currencies. We are currently exploring user requirement for collecting data on non-British currencies.

Table 3 – Frequency analysis of unprocessed free text responses to the currency question(s) <sup>127</sup>

| Manually coded category                 | Number        |
|---|---------------|
| Probably intended identifiable currency | 55            |
| Refusal (intended item non-response)    | 95            |
| Multiple currencies                     | 5             |
| NULL                                    | 35,960        |
| <b>Total</b>                            | <b>36,110</b> |

### 3.2.9.7. Other free text cleaning – frequencies

To avoid further repetitious analysis, we offer the following data on frequency counts of the combined figure for nulls and unmatchable free text responses for a number of fields. In each case we have reviewed, we have only looked at a relevant subpopulation of graduates in either employment or study and not those due to start employment/study who are also asked these questions. These are described below, with tabular data, also showing the proportion of the relevant population affected, in Table 4.

The analysis for the data on UK area where a business is located <sup>128</sup> is very similar to that given in the previous section on location of employment data.

Analysis of business country <sup>129</sup> and further study country data follow the analysis given in the previous section on employment country, closely.

Home country data <sup>130</sup> was almost complete. The 20 unmatchable response offered a familiar pattern of refusals, probable intended locations, and indications of nomadic existence.

Further study data on the provider attended <sup>131</sup> reflects the very large number of HE providers that UK graduates go on to study at. Manual review of a random sample of these records reveals a mix of overseas providers, smaller learning providers that appear to be in the UK, a variety of course-level details, a range of refusals (with and without explanations), and a proportion of unclassifiable and uncodable results.

<sup>127</sup> The apparent error in the total is merely an effect of applying the HESA rounding method to all respondent data.

<sup>128</sup> See derived field specifications XBUSLOCGR and XBUSLOCUC (navigating from <https://www.hesa.ac.uk/collection/c17072/derived/contents>)

<sup>129</sup> See derived field specification for ZBUSCOUNTRY at: <https://www.hesa.ac.uk/collection/c17072/derived/zbuscountry>

<sup>130</sup> See derived field specification at <https://www.hesa.ac.uk/collection/c17072/derived/ZHOMECOUNTRY>

<sup>131</sup> See the specification for free text 'other' responses at [https://www.hesa.ac.uk/collection/c17072/a/ucname\\_other](https://www.hesa.ac.uk/collection/c17072/a/ucname_other). This is returned where a respondent does not locate suitable option from the list of values at: <https://www.hesa.ac.uk/collection/c17072/a/ucname>

Table 4 - unmapped free text responses for various fields.

| Fields with 'other' free text responses       | Business Area | Business Country | Study Country <sup>132</sup> | Study Provider | Home country |
|---|---------------|------------------|------------------------------|----------------|--------------|
| <b>Total relevant sample</b>                  | 33,475        | 7,750            | 26,935                       | 74,470         | 35,745       |
| <b>Responses not mapped</b>                   | 5,360         | 40               | 65                           | 11,745         | 20           |
| <b>% of relevant sample not mapped</b>        | 16.0%         | 0.5%             | 0.2%                         | 15.8%          | 0.1%         |
| <b>Responses not mapped (including NULLs)</b> | 5,505         | 70               | 7,825                        | 26,480         | 690          |
| <b>% of not mapped (including NULLs)</b>      | 16.4%         | 0.9%             | 29.0%                        | 35.6%          | 1.9%         |

We also undertook a word group frequency analysis to gain insight on the substantial minority of unmatchable study provider data. Identifying those word groups with a frequency  $\geq 100$  (i.e. words [or groups of similar words] appearing within unmatchable free text responses 100 or more times) we identified possible avenues for further investigation, with a view to instrument or processing improvements. These could include supporting respondents by offering autofill suggestions for common responses, utilising a wider range of lookups based on ingested lists of study providers (e.g. FE colleges throughout the UK) or improving our text matching algorithm.

### 3.3. TIMELINESS AND PUNCTUALITY

In this section, we assess the timeliness and punctuality of the collection, analysis, and publication of Graduate Outcomes data.

Timeliness here refers to the gap between the publication of data and the period to which the data refer. Timeliness of data is an important aspect of meeting user needs; where data is going to be used to guide decisions on the part of users, it is important both that users have access to the most current data and that the gap between collection and publication is reduced as much as is compatible with the production of high quality statistical outputs.

Punctuality refers to the publication of statistical outputs according to a pre-announced timetable. In the interests of transparency and fair access to data, it is a requirement of the Code of Practice for Statistics that official statistics outputs should be pre-announced as part of a 12-month release calendar, and that any deviations from planned publication dates should be announced and explained as soon as possible.<sup>133</sup>

#### 3.3.1. TIMELINESS AND USER NEEDS

As discussed in section 3.1.1 above, HESA data on graduates is of interest to a wide variety of users. For many users, HESA data provides important support for decision making processes; prospective students may use information about what graduates do after completing their qualifications to inform their choices of course and provider, while graduate employers may target their efforts on the basis of outcomes data. For all of these users, their ability to make good

<sup>132</sup> The study country population is respondents who did not pick a provider from the drop-down list, as they are (uniquely) routed to the country question. For details, see section 3.2.6.2.1. Survey routing issues.

<sup>133</sup> Code of Practice for Statistics, Sections T3.1 and T3.2.

decisions will depend in part on access to timely data; given the rapidly evolving nature of the graduate labour market, for example, policies which aim to attract graduates with certain skills to a city or region will be less effective if it is based on out-of-date information about where graduates are more likely to do certain kinds of jobs.

The timescale for collecting and publishing Graduate Outcomes data was carefully considered in the design of the new survey. The DLHE survey collected information about graduates six months after the completion of their qualification, and a stratified sample of DLHE respondents were surveyed again three years later for the LDLHE survey; the majority of stakeholders who took part in the first NewDLHE consultation believed that the new survey should be carried out at some point between six and thirty-six months after the completion of qualifications. Six months was seen as too early, inasmuch as graduates would not yet have had time to make much progress in their post-HE careers; on the other hand, there was seen to be a risk that thirty-six months after graduation it would be difficult to contact enough graduates to provide a suitable dataset.<sup>134</sup> The 15-month interval between graduation and data collection used in the Graduate Outcomes survey was therefore selected so as to strike a balance between the availability of more useful careers data and the ability to obtain a high response rate.

Once all four cohorts for any given year have been surveyed, HESA aims to move swiftly towards publication, delivering final provider-level data to back to the providers it concerns about three months after the close of the data collection for the final cohort and releasing the Statistical Bulletin and open data about two months later. This timeline ensures that users of the survey have access to data on what graduates are doing 15 months after graduation while that data is still current.

### 3.3.2. PRODUCTION TIMELINE

In accordance with the Code of Practice for Statistics, HESA announces its planned data releases in advance. Upcoming data releases are announced on the HESA website, with their month of publication, at least six months before the planned publication date; National Statistics data releases are also pre-announced on the National Statistics hub.<sup>135</sup> Exact dates for publication are confirmed at least four weeks before each data release.

The first release of Graduate Outcomes data was initially scheduled to take place in spring 2020; this date was established once it was decided that Graduate Outcomes would be implemented for the 2017/18 academic year, with data collection for the final cohort of graduates set to close at the end of November 2019, 15 months after the last 2017/18 graduates completed their qualifications. In the autumn of 2019, it was announced that both the Statistical Bulletin and the open data release would take place in April 2020.

Before a precise date for publication could be released, the COVID-19 pandemic disrupted patterns of working at HESA; the agency's Cheltenham office closed on 19 March, with all staff moving to home-based working, and on 26 March, HESA announced that disruption caused by the coronavirus crisis might affect the publication schedule for official statistics outputs. On 1 April, HESA announced that the Graduate Outcomes data releases were to be postponed until the second half of May; this date was pushed back further due to issues of capacity caused by the

---

<sup>134</sup> HESA. 2016. 'Synthesis of Consultation Responses.'

[https://www.hesa.ac.uk/files/NewDLHE\\_consultation-synthesis.pdf](https://www.hesa.ac.uk/files/NewDLHE_consultation-synthesis.pdf)

<sup>135</sup> HESA. 2020. *Upcoming data releases*. <https://www.hesa.ac.uk/data-and-analysis/upcoming>  
For upcoming National Statistics releases, see also [https://www.gov.uk/search/research-and-statistics?content\\_store\\_document\\_type=upcoming\\_statistics](https://www.gov.uk/search/research-and-statistics?content_store_document_type=upcoming_statistics)

pandemic, and the publication of the Statistical Bulletin was scheduled for 18 June 2020.<sup>136</sup> Although it has not proved possible to adhere to the timeline which we published before the pandemic, HESA has followed the guidance issued by the UK Statistics Authority on the production of statistics during the coronavirus crisis, and we have announced any changes to our publication timeline as far as possible in advance.<sup>137</sup>

### 3.3.3. FREQUENCY OF PRODUCTION

From its inception, the Graduate Outcomes survey was designed to be published, like DLHE, as an annual data release. The Higher Education and Research Act 2017 specifies that data relating to HE providers and their courses must be published at least once a year, and an annual timetable reflects the fact that most UK higher education activities are organised around the academic year, which runs from early autumn to early summer.<sup>138</sup>

Not all graduates, however, complete their qualification at the same point in the academic year, and collecting data in quarterly cohorts allows us to make sure that we obtain data from all graduates about their activities 15 months after completion, regardless of when they completed their qualification. If all graduates finishing their degrees in the 2017/18 academic year (August 2017 to July 2018) had been surveyed with reference to a single census week in September 2019, for example, we would have data from twenty five months after completion for those students who had completed their qualifications in August 2017, but only fourteen months after completion for those who had finished in July 2018; such a discrepancy in timescale would make it difficult to compare outcomes for graduates finishing their qualifications at different points in the academic year.

From year two of Graduate Outcomes, it is hoped that HESA will be able to return to the collection and publication timetable initially established for year one, with data collection for cohort D closing at the end of November and statistical releases being published annually in the late spring or early summer.

## 3.4. ACCESSIBILITY AND CLARITY

In this section, we discuss issues of accessibility and clarity relating to the Graduate Outcomes dataset and the statistical outputs which are based upon it. In assessing statistical quality, accessibility refers to the ease with which users are able to obtain the data, including the format or formats in which the data is available and any supporting information which may be needed. Clarity refers to the availability and comprehensibility of any metadata which its users may need to understand the statistical data fully.

### 3.4.1. CONFIDENTIALITY AND DISCLOSURE CONTROL

Given that the Graduate Outcomes survey requires the collection of contact details and other personal information about respondents, issues of data protection, confidentiality, and disclosure control have been important throughout the design and implementation phases of the survey.

---

<sup>136</sup> HESA. 2020. *Coronavirus update*. <https://www.hesa.ac.uk/news/coronavirus>

For the final publication timetable for Graduate Outcomes, see <https://www.hesa.ac.uk/data-and-analysis/upcoming>

<sup>137</sup> UK Statistics Authority. 2020. 'Regulatory Guidance.' <https://www.statisticsauthority.gov.uk/wp-content/uploads/2020/03/Regulatory-guidance-changing-methods-Coronavirus.pdf>

<sup>138</sup> Higher Education and Research Act 2017, section 65.



HESA receives contact details for most graduates from providers.<sup>139</sup> Students are informed that their contact details will be passed on to HESA via HESA's Student Collection Notice, which informs students that, after graduation, providers will pass graduate contact details on to HESA and any organisations contracted by HESA to enable the collection of Graduate Outcomes data. The Student Collection Notice further informs students of the legal basis for the processing of their contact details for use in Graduate Outcomes, stating that contact details obtained from providers will be processed by HESA on the grounds that such processing is necessary for the performance of a task carried out in the public interest and for research and statistical purposes.<sup>140</sup>

Data protection policies available both on the HESA website and on the separate Graduate Outcomes website inform both providers and respondents of the uses which will be made of graduate data. Providers and respondents are informed that graduates may opt out of completing the survey, but that responses to the survey will be processed in accordance with GDPR on the basis of public interest, not consent.<sup>141</sup> Respondents are informed that their survey responses will be passed on to their HE provider, but that, unless they explicitly agree to be contacted by their provider about their survey responses, providers will only use survey responses for statistical and research purposes; while providers receive SWB data, they do not receive SWB responses for individual graduates, but instead receive aggregated statistical information about all their graduates' responses. Similarly, while Graduate Outcomes responses are passed on to a variety of other public and private bodies (including HE funding and regulatory bodies, public authorities, and others who have a legitimate interest in using the data for research and statistical purposes), survey responses are not used to make decisions about individuals. Where Graduate Outcomes data is passed on to third parties for use in research about higher education and the student population, the data is supplied under contracts which ensure that individuals cannot be identified from the data.<sup>142</sup>

When Graduate Outcomes data – or any other HESA data about people – is used in statistics published by HESA or any other users of HESA data, the data is subject to HESA's rounding and suppression strategy, which aims to reduce the risk of identifying individuals from published statistics. There are three main aspects to HESA's rounding strategy, each of which contributes to the protection of individual data: first, all counts of people are rounded to the nearest multiple of five; second, percentages based on fewer than 22.5 people are suppressed, and, third, averages are not published if they are based on seven or fewer individuals. Rounding counts of people prevents the use of multiple tables to identify small numbers of individuals, while the suppression of percentages based on fewer than 22.5 individuals and averages based on fewer than seven individuals prevents users from working back from an average or a percentage in order to obtain individual data.<sup>143</sup>

---

<sup>139</sup> As discussed in section 3.1.2. Data and statistical concepts, above, contact details for graduates of English further education colleges can be supplied to HESA by the OfS.

<sup>140</sup> HESA. 2020. *Student Collection Notice*. Available at: <https://www.hesa.ac.uk/about/regulation/data-protection/notices>

The legal basis for processing contact details for the collection of Graduate Outcomes data refers to GDPR Articles 6(1)(e) and 89.

<sup>141</sup> Information for providers: HESA. *Data protection guidance: Lawfulness of processing*.

<https://www.hesa.ac.uk/innovation/outcomes/providers/data-protection>

Information for graduates: HESA. *Graduate Outcomes: Privacy Information*.

<https://www.graduateoutcomes.ac.uk/privacy-info>

<sup>142</sup> HESA. *Graduate Outcomes: Privacy Information*. <https://www.graduateoutcomes.ac.uk/privacy-info>

<sup>143</sup> The full rounding methodology and rationale for the rounding strategy is available on the HESA website: <https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>

HESA's rounding strategy is designed to protect personal data, while still enabling HESA and other users of HESA data concerning individuals to publish useful statistics. In this vein, in order to prevent the compounding of inaccuracy which would occur if calculations were based on rounded figures, the rounding strategy is applied to the data only after any calculations have been carried out. Likewise, the specific thresholds applied in the rounding strategy represent an attempt to strike a balance between disclosure control and the production of detailed statistics; while rounding to multiples of 50, for example, would make it even harder to identify individuals, such a strategy would reduce the usefulness of the statistics which could be published.

### 3.4.2. STATISTICAL PRODUCTS AND SUPPORTING INFORMATION

As has been discussed in sections 3.3.2 and 3.3.3 above, data for the first year of the Graduate Outcomes survey will be published in June 2020, with subsequent data releases occurring annually in the spring. Like other HESA statistical releases, Graduate Outcomes data is not subject to scheduled revision; revisions to statistical releases are only carried out in the event of errors in HESA's data collection and production processes.<sup>144</sup>

HESA produces two main statistical outputs based on the Graduate Outcomes data. The first is a Statistical Bulletin, which contains a range of tables, charts, and summary analysis of headline figures drawn from the data; the second is the release of open data, published about a week after the Statistical Bulletin, containing a wider range of tables and charts, including provider-level for some variables. Both the Statistical Bulletin and the open data are available for free on the HESA website, and each chart is accompanied by a freely available data download, allowing users to conduct their own analysis of the data.<sup>145</sup>

In addition to the Statistical Bulletin and the open data, HESA has also published a variety of outputs designed to help users understand the Graduate Outcomes survey and the statistical outputs derived from it. In March 2020, HESA published a two-part methodology statement concerning the Graduate Outcomes survey, along with an accompanying blog post explaining the main points covered in each part. The first part of the methodology statement outlines the predecessors to the Graduate Outcomes survey, DLHE and LDLHE, the need for a new survey, and the process by which the new survey was developed; the second part discusses in detail the most important aspects of the design and implementation of the survey, with sections on survey coverage, survey design, data collection, data processing and analysis, data dissemination, sector engagement, and the evaluation of the survey.<sup>146</sup> In May 2020, HESA published a dissemination policy for the Graduate Outcomes survey, setting out HESA's policy, approaches, and standards for the dissemination of Graduate Outcomes data. The dissemination policy, which applies both to HESA publications and to those which may be produced by other users, includes sections on key users and uses of the data, legal and ethical considerations, and HESA's policy on misrepresentation of data; it also contains sections on HESA's statistical outputs based on the

---

<sup>144</sup> HESA. *Revisions policy*. <https://www.hesa.ac.uk/about/regulation/official-statistics/revisions>

<sup>145</sup> Further detail about the format and contents of the Statistical Bulletin and the open data release can be found in the Graduate Outcomes Dissemination Policy: <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

<sup>146</sup> Blog post: 'The ultimate guide to Graduate Outcomes.' <https://www.hesa.ac.uk/blog/16-04-2020/ultimate-guide-graduate-outcomes-hesa-publishes-survey-methodology-statement>

Methodology statement part one:

[https://www.hesa.ac.uk/files/Graduate\\_Outcomes\\_History\\_and\\_Background\\_20200330.pdf](https://www.hesa.ac.uk/files/Graduate_Outcomes_History_and_Background_20200330.pdf)

Methodology statement part two: <https://www.hesa.ac.uk/data-and-analysis/graduates/methodology>

---

survey and supporting information for users of the data, including key data concepts and standards.<sup>147</sup>

HESA also makes a range of other metadata available to users of the survey. The Graduate Outcomes section of the HESA website includes general information about the project and the survey, a link to the information page for students and graduates, a link to the information page for providers, and links to the Graduate Outcomes coding manuals; the survey results coding manual contains a variety of detailed metadata, including information on survey coverage, survey routing, and the variables used in the dataset.<sup>148</sup> The information page for providers includes a variety of resources, including detailed operational survey information; the operational survey information page includes detail on how the survey is being carried out, as well as answers to FAQs about survey operation, response rates, and the delivery of data to providers.<sup>149</sup> In addition, the operational survey information page includes a link to the data quality assessment which was carried out concerning SOC coding for the 2017/18 survey.<sup>150</sup>

Further information about HESA's data can also be found on the 'Definitions and data standards' page of the HESA website. This page includes a glossary which defines terms and acronyms frequently used in HESA outputs; information about the coding of subjects, disciplines, industries, and occupations; data intelligence notes which describe specific issues in the HESA data; and lists of definitions relevant to each HESA data stream.<sup>151</sup> The 'Definitions and data standards' page also includes answers to a number of FAQs which are relevant to multiple HESA collections, including a specific page covering Graduate Outcomes.

### 3.4.3. ACCESS AND USE

The Graduate Outcomes Statistical Bulletin and open data are freely available and downloadable on the HESA website under a Creative Commons Attribution 4.0 (CC BY 4.0) license.<sup>152</sup> Users of the data are free to copy, use, share, or adapt it for any purpose, provided that they give appropriate credit to HESA, provide a link to the Creative Commons license, and indicate if any changes have been made to the data.<sup>153</sup>

Since Graduate Outcomes data is freely available for public use, HESA cannot be responsible for the uses made of its data by external parties; HESA neither has the resources to police external uses of its data nor desires to be an arbiter of truth in the domains in which it publishes data. At the same time, HESA is aware that use of its data to support invalid conclusions or interpretations

---

<sup>147</sup> Graduate Outcomes dissemination policy: <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

<sup>148</sup> The main HESA Graduate Outcomes site: <https://www.hesa.ac.uk/innovation/outcomes>;  
Graduate Outcomes survey results coding manual: <https://www.hesa.ac.uk/collection/c17072>  
Index of data items: <https://www.hesa.ac.uk/collection/c17072/index>.

<sup>149</sup> HESA. 2020. *Operational survey information*. <https://www.hesa.ac.uk/definitions/operational-survey-information>

<sup>150</sup> HESA. 2020. *Operational survey information: Data classification (SIC/SOC)*:  
<https://www.hesa.ac.uk/definitions/operational-survey-information#data-classification-sicsoc>

The full results of the SOC coding assessment can be found here:

<https://www.hesa.ac.uk/innovation/outcomes/providers/assessment-1718-soc-coding>

<sup>151</sup> HESA. General Definitions and data standards are here: <https://www.hesa.ac.uk/support/definitions>

The list of definitions relevant specifically to the Graduate Outcomes survey can be found here:

<https://www.hesa.ac.uk/support/definitions/graduates>

<sup>152</sup> Creative Commons. <https://creativecommons.org/licenses/by/4.0/>

<sup>153</sup> HESA. *Open data and official statistics*. <https://www.hesa.ac.uk/data-and-analysis>

could entail a risk to the perceived trustworthiness, quality, and value of HESA's statistical outputs. With this risk in mind, the Graduate Outcomes dissemination policy includes HESA's policy on potential misrepresentations of the Graduate Outcomes data, outlining the steps which HESA may take in the event that a factual misrepresentation is perceived to have taken place.<sup>154</sup>

In addition to the Graduate Outcomes data which is available on the HESA website as open data, other datasets relating to the Graduate Outcomes survey are available to certain categories of users. HESA's statutory customers receive quality-assured microdata covering HE providers in their constituencies and a range of data fields aligned with their statutory powers and public functions; individual providers also receive microdata for their own graduates on an individual basis, with the exception of the SWB data, which is released to providers only in aggregated form.

Tailored datasets are also available for users who have data needs which are not met by the Graduate Outcomes open data. These datasets are available for a fee and can be commissioned through Jisc, HESA's data analytics partner.<sup>155</sup> Graduate outcomes data will be available for use in tailored datasets as soon as possible after the release of the open data.<sup>156</sup>

Further information about Graduate Outcomes data and publications is available from HESA's Official Statistics team ([official.statistics@hesa.ac.uk](mailto:official.statistics@hesa.ac.uk) or (0)1242 388 513 [option 2]).

### 3.5. COHERENCE AND COMPARABILITY

In this section, we discuss the coherence and comparability of the Graduate Outcomes data. Coherence here refers to the degree to which the Graduate Outcomes survey uses the same processes and harmonised methods which are used in other investigations of the same or similar domains; under the category of coherence we will be discussing both the uses of and deviations from national and international standard definitions in the Graduate Outcomes data and the relationship between Graduate Outcomes data and other datasets which may be available on the post-university careers of graduates. Comparability refers to the degree to which data can be compared over time; under this heading, we will be discussing the status of Graduate Outcomes as a new survey, and, in particular, its relationship with the DLHE survey.

#### 3.5.1. NATIONAL AND INTERNATIONAL DATA STANDARDS

Several of the domains covered in the Graduate Outcomes survey are domains to which established data standards apply. Work and employment, occupation, industry, and subjective wellbeing have all been the subject of considerable previous study, and, as a result of that study, standardised conceptual frameworks and definitions have been developed to facilitate their discussion and analysis. Where possible, HESA aims to conform to these accepted data standards so as to enable comparisons between HESA data and other studies relating to the same concepts, but it is important to discuss any areas in which we adapt internationally recognised standards to suit our particular analytical needs.

---

<sup>154</sup> For more detail on HESA's policy concerning misrepresentation of the Graduate Outcomes data, see the Graduate Outcomes dissemination policy: <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

<sup>155</sup> For further information on tailored datasets, see the Jisc website: <https://www.jisc.ac.uk/tailored-datasets>

<sup>156</sup> Graduate Outcomes dissemination policy: <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

Where Graduate Outcomes data refers to work or employment, HESA aims to conform to standard definitions wherever practical. The UK Office for National Statistics (ONS) has developed a standard framework, based on the concepts of labour supply and demand, for labour market statistics, which includes definitions for important concepts such as employment. This approach to labour market statistics is broadly compatible with the approaches taken by other international bodies, and the ONS definitions of key terms align closely with those used by the International Labour Organization (ILO).<sup>157</sup>

HESA for the most part follows the definitions of work and employment used by the ONS and the ILO.<sup>158</sup> The ILO defines work as ‘an activity performed by persons of any age and sex to produce goods or to provide services for use by others or for own use’, while employment is a sub-category of work referring to those who are ‘engaged in any activity to produce goods or provide services for pay or profit’.<sup>159</sup> Thus graduates who identify their most important activity as being engaged in unpaid or voluntary work for an employer are classified by HESA as in work, but not in employment. Although caring for someone else meets the ILO definition of work, however, graduates whose main activity is caring for someone else on an unpaid basis are classified as neither in work nor employment, and are included for analysis in the group of graduates undertaking ‘any other activity’.

On the basis of Graduate Outcomes data, it is possible to identify those graduates who fit the ILO definitions of work or employment. Identifying those who are unemployed according to the ILO definition, however, is less straightforward. The ILO defines ‘persons in unemployment’ as ‘those of working age who were not in employment, carried out activities to seek employment during a specified recent period and were currently available to take up employment given a job opportunity’; the ONS further specifies that, in order to be classified as in unemployment, people must be available to start a job within the next two weeks.<sup>160</sup> While the list of possible activities offered to respondents includes ‘unemployed and looking for work’, graduates who select this option are not asked how soon they would be able to take up work, and it is therefore not possible to identify them as unemployed according to national or international standards. Users wishing to compare the percentage of graduates who are not in work or further study with the unemployment rate in the wider population – a figure derived using the ONS definition of unemployment – should therefore use caution, since the relevant concepts are not directly comparable.

Graduates who are engaged in work for an employer (whether paid or unpaid), self-employment, running their own business, or developing a portfolio, are assigned both a Standard Industrial Classification (SIC) code and a Standard Occupational Classification (SOC) code. SIC codes for Graduate Outcomes are assigned using the SIC 2007 framework, which is the current industrial classification system maintained by the ONS; SIC 2007 is based on NACE (originally an acronym for *Nomenclature générale des activités économiques dans les Communautés européennes*), the

---

<sup>157</sup> Office for National Statistics. *A guide to labour market statistics*.

<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/aguidetolabourmarketstatistics#introduction>

<sup>158</sup> Graduate Outcomes dissemination policy: <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

<sup>159</sup> ILO Department of Statistics (ILOSTAT). *Glossary*.

<https://ilostat.ilo.org/glossary/work/>

<https://ilostat.ilo.org/glossary/employment/>

<sup>160</sup> ILO Department of Statistics (ILOSTAT). *Glossary*. <https://ilostat.ilo.org/glossary/unemployment/>

Office for National Statistics. *A guide to labour market statistics*.

<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/aguidetolabourmarketstatistics#unemployment>

European Community classification of economic activities, but with the addition of a fifth digit where it has been found necessary.<sup>161</sup> While SOC is a UK-based classification system administered by the ONS, it is broadly aligned with the International Standard Classification of Occupations 2008 (ISCO-08) so as to allow for comparison between UK and international employment roles.<sup>162</sup> SOC codes for Graduate Outcomes are assigned using SOC 2010 (DLHE), a fifth-digit expansion of the four-digit ONS SOC 2010 framework. SOC 2010 (DLHE) was developed for use with the DLHE survey in order to provide more detail about certain jobs often favoured by graduates, particularly those in areas where graduates were closely associated with a proliferation of new roles in rapidly-developing parts of the economy.<sup>163</sup> Although SOC 2010 (DLHE) is a bespoke framework, the first four digits of any SOC 2010 (DLHE) code map directly onto the appropriate four-digit SOC 2010 unit group, which enables comparisons with SOC data from other national datasets. A new UK SOC coding framework, SOC 2020, was published in February 2020, and it has been decided that this framework will be adopted for use in Graduate Outcomes from year two of the survey.

The use of nationally and internationally recognised standards to classify the industries and occupations in which graduates work enables comparison between HESA data on graduates in the workforce and other studies of employment which include data on industry and occupation. The move to SOC 2020 from year two of Graduate Outcomes will further facilitate such comparisons by ensuring that graduates are classified according to the system which most closely reflects the current state of the labour market. The training requirements for occupations can change over time, and occupations may therefore move between SOC major groups when the SOC framework is revised; thus some occupations, including higher level teaching assistants and veterinary nurses, have moved from major group 6 ('Caring, leisure and other service occupations') in SOC 2010 to major group 3 ('Associate professional occupations') in SOC 2020. Although the shift to SOC 2020 will introduce a level of discontinuity between the SOC data for years one and two of the survey, we are currently exploring options for enabling comparisons between the years.

Nationally accepted data standards are also relevant to the Graduate Outcomes SWB data. Graduate Outcomes measures SWB using a set of four questions (the ONS4) which were originally designed for the ONS as a harmonised standard of personal wellbeing; the ONS4 were first added by the ONS to the 2011 Annual Population Survey, and they have since been included in a range of other social surveys, including the Labour Force Survey (LFS).<sup>164</sup> HESA follows the ONS guidance on use of the SWB questions; the four questions are used verbatim in the Graduate Outcomes survey, and respondents are asked to give their answers to each question on a scale of 0 to 10, as specified by the ONS. HESA has also adopted the ONS' bracketing methodology in outputs based on the SWB data. The adoption of a widely used set of SWB measures in Graduate Outcomes enables comparisons between graduate wellbeing data and wellbeing data collected in

---

<sup>161</sup> Office for National Statistics. 2009. *UK Standard Industrial Classification of Economic Activities 2007 (SIC 2007): Structure and explanatory notes*. Available from <https://www.ons.gov.uk/methodology/classificationsandstandards/ukstandardindustrialclassificationofeconomicactivities/uksic2007>

<sup>162</sup> Office for National Statistics. *SOC 2010*. <https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2010>

<sup>163</sup> Elias, P. and R. Ellison. 2012. 'Standard Occupational Classification (2010) for the Destinations of Leavers from Higher Education: SOC 2010 (DLHE)'.

<https://www.hesa.ac.uk/collection/c14018/download/soc2010dlhe.pdf>

<sup>164</sup> Office for National Statistics. *Personal well-being user guidance*. <https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide>

other social surveys; although it will be important to take potentially confounding factors into account in any analysis, the SWB measures themselves will be comparable.

### 3.5.2. GRADUATE OUTCOMES AND OTHER DATA ON GRADUATES

While Graduate Outcomes is the only national survey designed specifically to provide insight into the experiences of higher education graduates, the domains of several other studies overlap to an extent with the domain of the Graduate Outcomes survey. In particular, both the Longitudinal Educational Outcomes (LEO) study and the Labour Force Survey (LFS) collect data on education and salary, with the LFS also including detailed information on employment and occupation. While the three data sources can provide complementary views of graduates in the workforce, it is important to understand key differences between the studies.

The LEO dataset, which was first published in 2017, brings together education data from the Department for Education (DfE) along with employment, earnings, and benefits data from the Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC). Using these sources, LEO provides earnings and benefits information for graduates one, three, five, and ten years after completion of their qualifications; it also includes data on personal characteristics (gender, ethnicity, and age), university attended, subject studied, and qualification achieved.<sup>165</sup>

Unlike Graduate Outcomes, which, as a survey, depends on the individual responses of graduates, the LEO dataset is drawn from administrative data and includes information on all graduates in paid work in the UK; since LEO earnings data comes directly from HMRC, it is free of some of the risks of inaccuracy inherent in self-reported salary data. LEO does not, however, include data on hours worked, so it is not possible to distinguish between graduates who are in full-time work and those who are working part-time. LEO also does not include data on graduates doing voluntary or unpaid work, and, because the LEO earnings data does not include self-assessment earnings, LEO data on graduates in self-employment cannot be entirely representative.<sup>166</sup> LEO data, moreover, does not include information about occupation; the LEO record tells us what graduates earn, but it does not give us any further information about what graduates do.<sup>167</sup>

Graduate Outcomes and LEO thus provide different pictures of the graduate population in the UK. One of the goals in the design of the Graduate Outcomes survey was to provide statistical outputs which could contextualise data on graduates from other sources, such as LEO, and this goal is reflected in the breadth of information collected in the survey.<sup>168</sup> While the LEO dataset provides data on a small number of variables for the most graduates in the UK, and while it, moreover, tracks changes in earnings over time, the Graduate Outcomes survey provides a more detailed picture of each annual cohort at a single point in their post-university careers. The LEO dataset measures graduate outcomes only in terms of whether graduates are in paid employment and, if

---

<sup>165</sup> Department for Education. 2020. *Statistics: higher education graduate employment and earnings*.

<https://www.gov.uk/government/collections/statistics-higher-education-graduate-employment-and-earnings>

<sup>166</sup> Department for Education. 2017. *Employment and earnings outcomes for higher education graduates by subject and institution: experimental statistics using the Longitudinal Educational Outcomes (LEO) data*.

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/718225/SFR\\_18\\_2017\\_LEO\\_mainText.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/718225/SFR_18_2017_LEO_mainText.pdf)

<sup>167</sup> Office for Statistical Regulation. 2019. *Exploring the public value of statistics about post-16 education and skills in England*. Office for Statistical Regulation Systematic Review Programme.

<sup>168</sup> Graduate Outcomes methodology statement, part one:

[https://www.hesa.ac.uk/files/Graduate\\_Outcomes\\_History\\_and\\_Background\\_20200330.pdf](https://www.hesa.ac.uk/files/Graduate_Outcomes_History_and_Background_20200330.pdf)

so, how much they are earning, while the Graduate Outcomes survey collects a broader range of information about what graduates are doing and how they feel about it.

While LEO is specifically geared towards collecting data about employment outcomes for higher education graduates, the LFS is a household survey designed to collect data about the employment circumstances of the UK population as a whole. It was first run in 1973 as a biennial survey and shifted to an annual survey in 1984; since 1992, the LFS has been collected quarterly, with a switch from seasonal to calendar quarters in 2006. Households participating in the LFS are surveyed for five consecutive quarters, with a fifth of the overall sample being replaced each quarter. Where LEO collects administrative data on all graduates in employment in the UK, the LFS is administered to a systematic sample of approximately 35,000 households in Great Britain, plus approximately 2,500 households from Northern Ireland; conclusions about overall patterns in employment circumstances are thus drawn from a relatively small portion of the UK population.<sup>169</sup> Unlike the LFS, which is concerned with the entire UK labour force, Graduate Outcomes is concerned only with those who have completed HE qualifications in a given year, and, while there will inevitably be some level of non-response, Graduate Outcomes aims to collect data from the entire target population. With 361,260 responses in the first year, the Graduate Outcomes sample is thus much larger than the annual sample collected by the LFS, despite the narrower focus of the Graduate Outcomes survey.<sup>170</sup>

Although both Graduate Outcomes and the LFS include questions about employment and education, the focuses of the two surveys are different. The LFS is primarily focused on employment, but participants are also asked to respond to the ONS4 SWB questions and a series of questions about their educational attainment.<sup>171</sup> Since not all LFS respondents have the same educational qualifications, the educational information collected in the survey allows for some comparison of outcomes between people with different educational histories. All Graduate Outcomes respondents, on the other hand, are higher education graduates, so different comparisons are possible; rather than encouraging comparisons between graduates and non-graduates, Graduate Outcomes encourages comparisons between different categories of graduates.

Respondents to the LFS can be at any stage in their careers; for those who have higher education qualifications, this means that they may be selected to participate in the LFS shortly after finishing their qualifications, or they may be selected many years later. Even within the subset of LFS respondents with higher education qualifications, there will therefore be a wider variation in experiences and possible outcomes than is likely to be visible in Graduate Outcomes, where graduates are deliberately surveyed at the same point in their post-university careers. While Graduate Outcomes provides a cross-section of the experiences of higher education graduates 15 months after finishing their qualifications, the LFS can provide glimpses into what their lives may be like at a variety of different points.

---

<sup>169</sup> Office for National Statistics. 2018. Labour Force Survey User Guide, Volume 1: LFS Background and Methodology. Available from: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance>

<sup>170</sup> HESA. 2020. *Graduate Outcomes Cohort D Review*. <https://www.hesa.ac.uk/files/End%20of%20cohort%20D%20report.pdf>

<sup>171</sup> Office for National Statistics. 2020. Labour Force Survey User Guide, Volume 2: User guide to the LFS questionnaire. Available from: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance>



If we are looking for a complete picture of what happens to higher education graduates in the UK, Graduate Outcomes, LEO, and the LFS all fill in different pieces of the puzzle. Although the datasets could fruitfully be used in conjunction with each other – the use of the same set of SWB questions in Graduate Outcomes and the LFS might, for example, allow for some research into the comparative SWB of graduates and non-graduates – in making any comparison between the three data sources, it will be important to recognise the differences in methodology and coverage between the sources. To return to the example of SWB comparisons, although LFS and Graduate Outcomes respondents answer the same four questions about SWB, they are faced with those questions at different points in their careers, and differences in SWB may depend on a range of factors not necessarily connected to education.

In addition to enabling careful comparisons between graduates and the population as a whole or between different stages in graduates' careers, the existence of other datasets with overlapping domains is likely to be important in the future development of Graduate Outcomes. When LEO data was first published, the DfE conducted a comparison between the LEO and DLHE datasets; HESA has in the past carried out similar comparisons in order to check the quality of DLHE salary data, and a further, detailed comparison of LEO and Graduate Outcomes would provide useful information about the respective strengths and weaknesses of the two datasets.<sup>172</sup> HESA also hopes in future years to explore the possibility of linking the Graduate Outcomes record with other relevant datasets, including LEO salary data.<sup>173</sup> Doing so will not only allow us to streamline our collection processes, but also, and perhaps more importantly, it will allow us to provide a fuller view of the trajectories of graduates after they leave higher education.

### 3.5.3. COMPARABILITY AND TIME SERIES

Graduate Outcomes is a new survey, and as such there are no time series available. When the new Graduate Outcomes survey was being designed, the outputs developed from the DLHE data were seen to have value, and it was therefore decided to begin iterating from the DLHE approach in designing the new survey.<sup>174</sup> Nevertheless, Graduate Outcomes is an entirely new survey, and important differences in timescale, methodology, and survey questions between Graduate Outcomes and DLHE make it impossible for direct comparisons to be made between data from the two surveys.

For the DLHE survey, graduates were contacted six months after the completion of their qualifications; Graduate Outcomes surveys graduates 15 months after the completion of their qualifications, that is, nine months later than they would have been surveyed for DLHE. Graduates surveyed for Graduate Outcomes are therefore at a very different stage in their post-HE careers than those who were surveyed for DLHE, which means that comparing the outcomes of respondents to the two surveys will not be a like-for-like comparison.

Methodological differences between DLHE and Graduate Outcomes are another reason to avoid direct comparisons between the two surveys. DLHE was a survey of graduates conducted by

---

<sup>172</sup> Department for Education. 2016. *Employment and earnings outcomes for higher education graduates: experimental statistics using the Longitudinal Educational Outcomes (LEO) data*. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/543794/SFR36-2016\\_main\\_text\\_LEO.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543794/SFR36-2016_main_text_LEO.pdf)

<sup>173</sup> HESA. *Key principles of Graduate Outcomes*. <https://www.hesa.ac.uk/innovation/outcomes/about/principles>

<sup>174</sup> Graduate Outcomes methodology statement, part one: [https://www.hesa.ac.uk/files/Graduate\\_Outcomes\\_History\\_and\\_Background\\_20200330.pdf](https://www.hesa.ac.uk/files/Graduate_Outcomes_History_and_Background_20200330.pdf)

providers, whereas Graduate Outcomes is a survey of graduates directly; thus where DLHE was administered by providers who then returned data to HESA for processing and analysis, Graduate Outcomes is administered centrally. For DLHE, SOC coding was done by providers, whereas SIC and SOC coding for Graduate Outcomes is outsourced to the business data services company Oblong.<sup>175</sup> The central administration of both the Graduate Outcomes survey itself and its SIC and SOC coding ensures a greater degree of consistency than was possible with the DLHE survey.

Finally, although the two surveys cover similar ground, the specific questions asked by the two surveys are different. The list of activities which can be selected by respondents to Graduate Outcomes is different from the list available to DLHE respondents; the Graduate Outcomes survey gives respondents more options and, in particular, allows graduates who are in work to be more specific about the type of work they are doing. In addition to asking for more detail about areas which received less emphasis in the DLHE survey, Graduate Outcomes also includes new questions, such as the graduate voice questions, which reflect the new survey's emphasis on providing metrics for graduate success beyond employment and salary; similarly, the SWB questions, which were previously used in the final iteration of LDLHE, have been made part of the core Graduate Outcomes survey. Given these differences in survey design, much of the Graduate Outcomes data will have no direct equivalent in DLHE.

Having decided to replace DLHE with a new and fundamentally different survey, HESA has taken the decision not to undertake, publish, or otherwise disseminate any comparisons of data between the Graduate Outcomes survey and the DLHE survey. We likewise advise all users of the two surveys to avoid making any direct comparisons between the two datasets. The two surveys are not directly comparable and any attempts to make direct comparisons are likely to lead to questionable results which are open to misinterpretation.<sup>176</sup>

---

<sup>175</sup> For more detail on SIC and SOC coding methodology, see <https://www.hesa.ac.uk/definitions/operational-survey-information#data-classification-sicsoc>

<sup>176</sup> Graduate Outcomes dissemination policy: <https://www.hesa.ac.uk/files/Graduate-Outcomes-dissemination-policy-v1-20200529.pdf>

## 4. CONCLUSION

The Code of Practice for Statistics is based on three pillars: trustworthiness, quality, and value.<sup>177</sup> In order to comply fully with the Code of Practice, producers of statistics must therefore ensure that the statistics they produce reflect these three attributes. While this report has been primarily concerned with assessing the Graduate Outcomes survey in terms of quality, the mutually supportive nature of the three pillars means that any assessment of statistical quality will also, of necessity, have implications for the trustworthiness and value of the statistics in question.

Statistical trustworthiness depends on the conditions of statistical production. If statistics are to be trustworthy, there must be a high degree of public confidence in the people and organisations responsible for producing them. This confidence must extend to the honesty and integrity of statistical producers, to their independence, to their commitment to the orderly release of statistics, to the transparency of their operating processes, to their professional capability, and to their standards of data governance.

In producing this report on the Graduate Outcomes survey and the statistical outputs derived from it, we hope to have shed some additional light on the processes underlying the design and implementation of the survey, the processing of survey data, and the production of statistical outputs. In so doing, we have contributed to the transparency of HESA's operations, as required in section T4 of the Code of Practice; we hope that increased transparency will give users the information they need to have confidence also in the other elements which contribute to statistical trustworthiness. By explaining the processes by which we assess the accuracy and reliability of our data, for example, we hope to give users of the Graduate Outcomes survey confidence in the professional capability of HESA and the partner organisations involved in survey administration and data processing; similarly, by discussing the efforts we have taken to protect personal information, we hope to give users confidence in our data governance practices.

Statistical quality is a characteristic of the statistical products themselves. It is not sufficient for statistical products to be produced in a trustworthy fashion; instead, the Code of Practice for Statistics stipulates that 'the statistics must be the best available estimate of what they aim to measure'.<sup>178</sup> Producing high quality statistical outputs depends on collecting data from suitable sources, on employing sound methodology in the collection, processing, and analysis of data, and on being able to provide users with clear information about how the quality of data and statistics has been assured.

Over the course of this quality report, we have guided users of the Graduate Outcomes survey through the processes used by HESA to assess the quality of the survey and the resulting statistical outputs. At each stage in the development and implementation of Graduate Outcomes, from the inception of the NewDLHE review in July 2015 to the Graduate Outcomes data releases in June 2020, HESA has considered how best to ensure that Graduate Outcomes would be a high quality survey, leading to high quality official statistics outputs. The survey was designed both to capture relevant data about the experiences of graduates after course completion and to reach as many members of our target population as possible. Rigorous quality assurance processes were built into our data collection and processing systems, and we have continued to take user feedback onboard and refine our methodology at each stage of the process. In cases where we are aware

---

<sup>177</sup> Code of Practice for Statistics. Introduction, Section xii. <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf>

<sup>178</sup> Code of Practice for Statistics. Introduction, Section xiv. <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf>

that improvements are still to be made, we are working to develop solutions which will enable us to improve the quality of our outputs further.

The final pillar of the Code of Practice for Statistics is value. While trustworthiness and quality refer to the how statistics are produced and the nature of the statistics themselves, statistical value depends on whether statistical products are fit for purpose. As is stated in the opening sentence of the introduction to the Code of Practice, 'official statistics are an essential public asset.'<sup>179</sup> Official statistics thus exist for the benefit of their users, and neither the quality of outputs nor the trustworthiness of their production can make up for a failure to consider user needs for statistics that contribute usefully to issues of public concern.

HE providers have collected information on the destinations of their graduates since at least the 19<sup>th</sup> century; as participation in higher education has expanded and debates about the value of higher education have grown increasingly prominent, the appetite for data on graduates has increased. In designing and implementing the new Graduate Outcomes survey, HESA has worked to iterate from and improve upon DLHE; we have retained questions from DLHE which were deemed to have value, but we have also refined old questions and added new questions to provide additional insight. Having worked with key users to design a survey that collects data on the most relevant questions about the outcomes of graduates, HESA has aimed to produce statistical outputs which present that data as clearly and accessibly as possible. The Graduate Outcomes data releases affirm our commitment to the principles of open data, and, even more importantly, they also ensure that all of our users have access to statistical outputs designed to meet their needs.

---

<sup>179</sup> Code of Practice for Statistics, Introduction. <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf>

## 5. REFERENCES

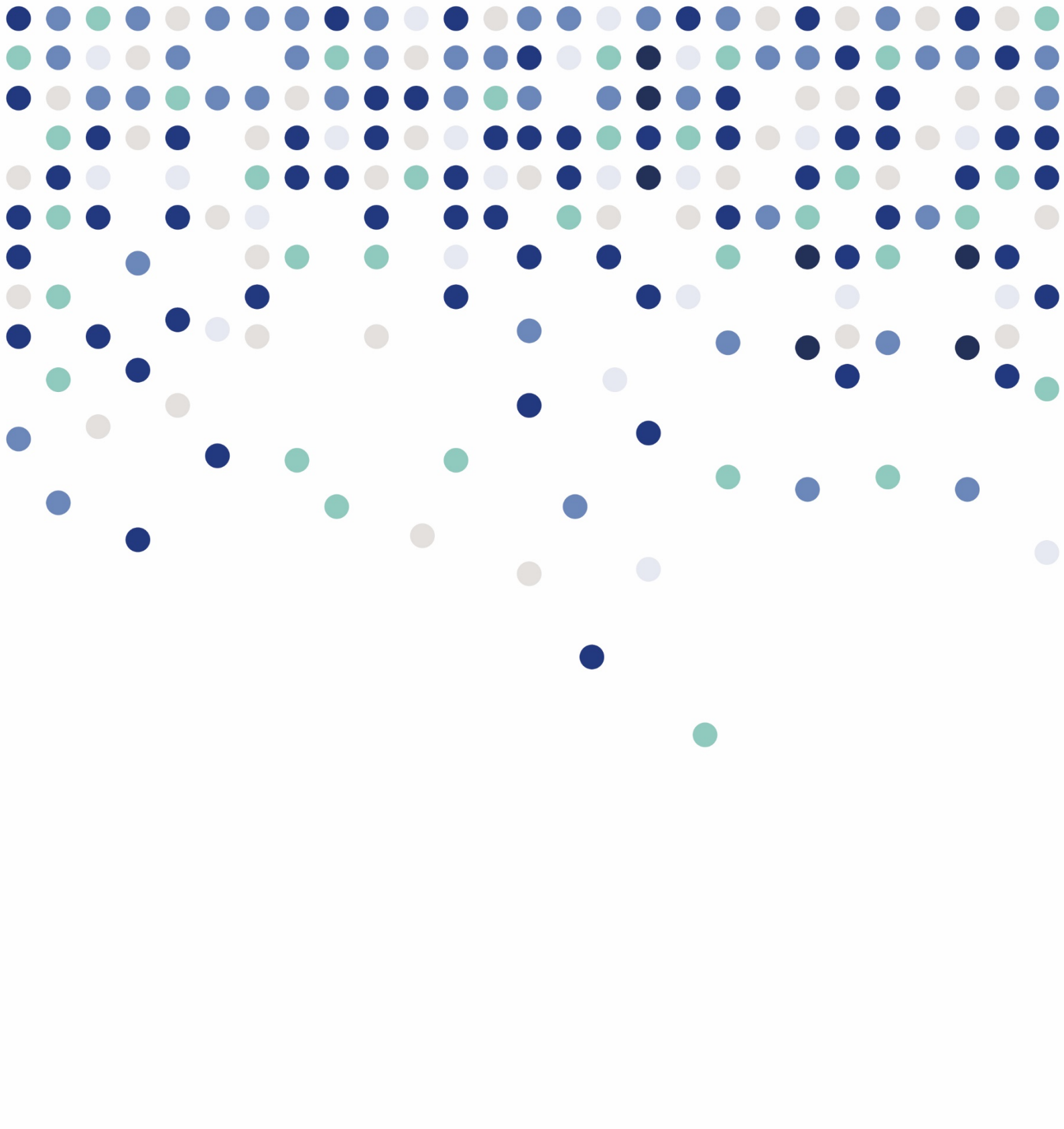
- Arab, M., Rafiei, H., Safarizadeh, M., Ahmadi, J. and Safarizadeh, M. (2016). 'Stress, anxiety and depression among medical university students and its relationship with their level of happiness'. *IOSR Journal of Nursing and Health Science (IOSR-JNHS)*, 5(1), 44-47.
- Curtin, R., Presser, S., & Singer, E. (2000). 'The effects of response rate changes on the index of consumer sentiment'. *Public Opinion Quarterly*, 64(4), 413–428.  
<https://doi.org/10.1086/318638>
- DAMA UK Working Group on 'Data Quality Dimensions'. (2013). The six Primary dimensions for Data Quality assessment. 17. [https://www.whitepapers.em360tech.com/wp-content/files\\_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf](https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf)
- Data Management Association. (2017). *DAMA-DMBOK: data management body of knowledge* (S. Earley & D. Henderson (eds.); 2nd edition). Technics Publications.
- Department for Business, Innovation and Skills. (2013). Privately funded providers of higher education in the UK (BIS Research Paper No. 111). Department for Business, Innovation and Skills. <https://www.gov.uk/government/publications/privately-funded-providers-of-higher-education-in-the-uk>
- Department for Education. (2020). *Graduate Outcomes (LEO): employment and earnings outcomes for higher education graduates by subject studied and graduate characteristics in 2017/18*.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/874410/2020\\_03\\_HE\\_LEO\\_main\\_text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/874410/2020_03_HE_LEO_main_text.pdf)
- Department for Education. (2020b). *Statistics: higher education graduate employment and earnings*. <https://www.gov.uk/government/collections/statistics-higher-education-graduate-employment-and-earnings>
- Department for Education. (2017). *Employment and earnings outcomes for higher education graduates by subject and institution: experimental statistics using the Longitudinal Educational Outcomes (LEO) data*.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/718225/SFR\\_18\\_2017\\_LEO\\_mainText.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/718225/SFR_18_2017_LEO_mainText.pdf)
- Department for Education. (2016). *Employment and earnings outcomes for higher education graduates: experimental statistics using the Longitudinal Educational Outcomes (LEO) data*.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/543794/SFR36-2016\\_main\\_text\\_LEO.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543794/SFR36-2016_main_text_LEO.pdf)
- Dolan, P., & Kavetsos, G. (2016) Happy talk: mode of administration effects on subjective well-being. *Journal of Happiness Studies*, 17 (3). pp. 1273-1291. [10.1007/s10902-015-9642-8](https://doi.org/10.1007/s10902-015-9642-8)
- Duffy, B., Smith, K., Terhanian, G. and Bremer, J. (2005). 'Comparing data from online and face-to-face surveys', *International Journal of Market Research*, 47(6), 615-639.
- Elias, P. and R. Ellison. (2012). 'Standard Occupational Classification (2010) for the Destinations of Leavers from Higher Education: SOC 2010 (DLHE)'. Warwick Institute for Employment Research. <https://www.hesa.ac.uk/collection/c14018/download/soc2010dlhe.pdf>
- European Statistical System Committee. (2017). *European Statistics Code of Practice*.  
<https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7>
- Groves, R. M. (2004). *Survey Errors and Survey Costs*. John Wiley & Sons.
- Higher Education and Research Act 2017*. Available at:  
<http://www.legislation.gov.uk/ukpga/2017/29/contents/enacted>

- Hunt, S., & Boliver, V. (2019). Private providers of higher education in the UK: mapping the terrain. Working paper no. 47. <https://www.researchcghe.org/perch/resources/publications/to-publishwp47.pdf>
- Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). 'Consequences of reducing nonresponse in a national telephone survey'. *Public Opinion Quarterly*, 64(2), 125–148. <https://doi.org/10.1086/317759>
- Koch, A., & Blohm, M. (2016). *Nonresponse Bias*. GESIS Survey Guidelines. [https://doi.org/10.15465/GESIS-SG\\_EN\\_004](https://doi.org/10.15465/GESIS-SG_EN_004)
- Office for National Statistics. (2020). *Labour Force Survey User Guide, Volume 2: User guide to the LFS questionnaire*. <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance>
- Office for National Statistics (2019). *A guide to labour market statistics*. <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/aquidetolabourmarketstatistics#introduction>
- Office for National Statistics. (2018). *Personal well-being user guidance*. <https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide>
- Office for National Statistics. (2018b). *Labour Force Survey User Guide, Volume 1: LFS Background and Methodology*. <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance>
- Office for National Statistics. (2016). *SOC 2010*. <https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2010>
- Office for National Statistics. (2016b). *UK SIC 2007*. <https://www.ons.gov.uk/methodology/classificationsandstandards/ukstandardindustrialclassificationofeconomicactivities/uksic2007>
- Office for National Statistics. (2013). *Guidelines for Measuring Statistical Output Quality*. Version 4.1. <https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/01/Guidelines-for-Measuring-Statistical-Outputs-Quality.pdf>
- Office for National Statistics. (2009). *UK Standard Industrial Classification of Economic Activities 2007 (SIC 2007): Structure and explanatory notes*. Available from <https://www.ons.gov.uk/methodology/classificationsandstandards/ukstandardindustrialclassificationofeconomicactivities/uksic2007>
- Office for Statistical Regulation. (2019). *Exploring the public value of statistics about post-16 education and skills in England*. Office for Statistical Regulation Systematic Review Programme. <https://www.statisticsauthority.gov.uk/publication/exploring-the-public-value-of-statistics-about-post-16-education-and-skills-in-england/>
- Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010). 'Reduction of Nonresponse Bias through Case Prioritization'. *Survey Research Methods*, 4(1), 21–29. <https://doi.org/10.18148/srm/2010.v4i1.3037>
- Rosen, J. A., Murphy, J., Peytchev, A., Holder, T., Dever, J. A., Herget, D. R., & Pratt, D. J. (2014). 'Prioritizing Low Propensity Sample Members in a Survey: Implications for Nonresponse Bias'. *Survey Practice*, 7(1), 1–8. <https://doi.org/10.29115/SP-2014-0001>
- Schouten, B., & Shlomo, N. (2017). 'Selecting Adaptive Survey Design Strata with Partial R-indicators'. *International Statistical Review*, 85(1), 143–163. <https://doi.org/10.1111/insr.12159>
- Sunak, R. 11 March 2020. *Budget Speech 2020*. <https://www.gov.uk/government/speeches/budget-speech-2020>

UK Statistics Authority. (2018). Code of Practice for Statistics. Edition 2.0.

<https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf>

Wagner, J. (2013). 'Adaptive Contact Strategies in Telephone and Face-to-Face Surveys'. *Survey Research Methods*, 7(1), 45–55.



**Official Statistics**

HESA  
95 Promenade  
Cheltenham  
GL50 1HZ  
E [official.statistics@hesa.ac.uk](mailto:official.statistics@hesa.ac.uk)  
T +44 (0) 1242 388 513  
W [www.hesa.ac.uk](http://www.hesa.ac.uk)

