Assessing bias in online surveys using alternative survey modes

*Neil H. Spencer, Dag S. Syrdal, Matthew Coates
and Ursula Huws*

Neil H. Spencer is Professor of Applied Statistics at
Hertfordshire Business School, University of Hertfordshire,
Hatfield, UK
Dag S. Syrdal is Statistical Consulting Associate at
Hertfordshire Business School, University of Hertfordshire,
Hatfield, UK
Matthew Coates is Statistical Consulting Associate at
Hertfordshire Business School, University of Hertfordshire,
Hatfield, UK
Ursula Huws is Director of Analytica Social and Economic
Research, London, UK

ABSTRACT
**Due to concerns that respondents to online surveys are different from populations
of interest, parallel offline surveys can be undertaken and results compared. In this
article we create a set of principles to compare results from online surveys with
those from surveys using other survey modes. Rather than just comparing estimates
and confidence intervals from the different modes, these principles consider biases
that each survey mode introduces and whether the results obtained are compatible
with each other, given these different biases. Using the example of a survey of
platform work, we demonstrate that this approach can be used effectively and be
applied to a variety of social science studies that use online surveys.**

Introduction

This article is concerned with the practical issue of assessing the validity of the results
from an online survey and assessing the impact of the online mode by using other
survey modes. The motivation for this work comes from the authors' involvement in a

survey of platform work across 13 European countries between 2016 and 2019 (Huws et al., 2019). The term 'platform work' is not something which has a precise definition but, for the purposes of the study discussed here, it is sufficient to say that it is employment mediated through an internet platform. Thus, the surveys undertaken need to determine whether respondents have not just found work via the internet but have had that employment actively facilitated and managed online via the platform, be that via an app or a website. For further details, see Huws et al. (2019). However, despite the surveys discussed being rooted in one particular setting, the issues discussed in this article can be applied more generally to online surveys in other scenarios.

The surveys in the study were designed to collect data about platform work in an efficient and effective manner. In the second section, we give the background and rationale as to how these surveys were conceived and carried out. At all times, the prime motivation was, as it should have been, to collect valid, reliable and unbiased data from samples that reflected the patterns existing in the populations. The motivation to be able to assess the validity of the survey results and, in particular, the impact of the survey mode chosen, although important, was of course secondary. This means that the surveys were not specifically designed with the intent of being used in a multi-mode environment (i.e. online and also in-person or telephone modes). Thus, when trying to assess the effect of the main survey mode used, it is not necessarily a simple matter to convert the online survey used into other (offline) modes. As a result, one cannot expect the different modes to deliver identical results. The difficulties that this introduces make these surveys ideal for a paper such as this in which we seek to assert that not only is it desirable that the impact of the online survey mode be examined but that it is possible to do so using alternative survey modes, even when the nature of the surveys means that direct comparisons are not possible.

In the third section, we turn to the issue of how to assess the impact of the online survey mode. We discuss the need to utilise other survey modes and the difficulties of doing so when the survey mode may affect the results obtained. A rationale for the comparison of the results from the different survey modes is created and then we apply it to the surveys described here. To conclude, in the fourth section we argue that the rationale can be carried beyond the confines of the current study to any online survey and that we have demonstrated that any online survey can, and should, have the impact of its online survey mode assessed without having to compromise the integrity of the survey or the benefits of the online survey mode.

Creating a survey to measure platform work

In this section, we give a detailed outline of the rationale for the way we conducted the surveys in the study. We do not claim to have developed the perfect design for these complex surveys but, as discussed in the following sections, the design was created with the potential issues in mind and we believe that the decisions that were made were, at the very least, reasonable.

When designing the survey of platform work for the first survey, conducted in 2016, we were able to draw on the work of Williams & Schneider (2016) who discuss methods used to measure the size of the 'shadow economy', which they define as 'paid activities that are lawful as regards their nature but not declared to the public

authorities'. Platform work does not necessarily fall precisely within this definition but, nevertheless, there are parallels with, for example, both types of work involving non-standard methods of recruitment and work not being undertaken in traditional workplaces. As such, considering Williams & Schneider (2016) and associated literature, and the limited budget available, it was decided to carry out the surveys online, utilising omnibus surveys (where our questions would be added to those from other bodies undertaking research on a variety of other issues) or (for some countries where an omnibus survey was not available) as a stand-alone survey.

In recent years, there has been no consensus arrived at as to how best measure platform work. O'Farrell & Montagnier (2020) provide a review of methods for surveying platform workers and conclude that there is no one universal method that should be used. They also suggest that the method used in our work (Huws et al., 2019), asking a series of questions associated with different aspects of platform work and the researcher then deciding, from responses to these whether or not the respondent is a platform worker, has much to recommend it. Other authors have taken approaches such as using Facebook to recruit respondents (Felstead, 2021) or using figures released by platforms' website themselves (Tubaro, Le Ludec & Casilli, 2020). Galvin, Bierman & Schieman (2021) have used a large survey which included questions on platform work (the *Canadian Quality of Work and Economic Life Study*). This uses an online panel which aims to be representative of the working age population, in a similar manner to the surveys discussed in this article.

Principles for creating a survey instrument to measure platform work

In early 2016, the European Foundation for Progressive Studies (FEPS), in collaboration with UNI Europa, commissioned the University of Hertfordshire to undertake a series of surveys to explore the extent and characteristics of platform work across a number of European countries (with co-funding from other organisations within each country). Altogether 14 surveys were carried out in 13 countries, with the UK, Sweden, the Netherlands, Germany and Austria being surveyed in 2016, Switzerland and Italy in 2017, Estonia, Finland and Spain in 2018 and Slovenia, the Czech Republic, the UK (repeat survey) and France in 2019. The full report can be found at Huws et al. (2019).

In this section, we outline the principles that were considered in order to develop a survey instrument to accomplish the goal of investigating platform work. We separate this task from the decision as to what survey mode to use and the development of the actual questions. These activities, which were undertaken in parallel and informed by the principles developed here, will be considered later in this article along with issues around obtaining a sample representative of the population of interest.

The basic principles for the construction of a survey instrument for measuring platform work were the same as those for any survey. However, there are particular aspects of platform work which require special consideration to be given. The basic principles are threefold (e.g. Moser & Kalton, 1971): first, questions need to be correctly and fully understood by respondents; second, the data requested need to be fully and correctly recalled or identified by respondents; and third, the data need to be fully and correctly reported by respondents.

In the context of platform work, the first principle needs particular attention because there is no widespread agreement on the definition of platform work amongst researchers (Huws, 2016; Mandl, 2016; Jesnes & Braesemann, 2019; Vallas & Schor, 2020), let alone amongst members of the general population. Thus, one cannot simply ask respondents if they have undertaken 'platform work'. Furthermore, because most of the tasks undertaken by platform workers are also undertaken by workers who do not have their work mediated by an internet platform (Katz & Krueger, 2016; Armano, Bove & Murgia, 2017; Johnston & Land-Kazlauskas, 2018), one cannot ascertain whether or not someone is a platform worker by simply asking them about the work they have undertaken. Again, the meaning of work being 'mediated' through an internet platform does not have a simple definition (Eurofound, 2015; Johnston & Land-Kazlauskas, 2018; Vallas & Schor, 2020) and thus one cannot expect respondents to correctly and fully interpret the meaning of a question which asks about this directly. A further issue concerns the fact that platform work may be occasional in nature (Ross et al., 2010; Eurofound, 2015) and a respondent may fail to understand that researchers may want them to record details of their platform work if they also have other, perhaps better remunerated work that they undertake. As a result, it is necessary not only to ascertain whether any work of a platform nature has been undertaken but also the frequency with which it is performed. A relatively simple basic principle thus has several nuances in the context of this study.

In the context of the second principle, even if the questions asked are correctly and fully understood by respondents, there is still a concern that they may not correctly recall or identify the platform work that they have undertaken. As mentioned above, platform work may be occasional in nature (Green et al., 2014; Wood et al., 2019) and respondents may not readily recall that they have undertaken work of this type. Additionally, even if questions are correctly and fully understood, it may be that a respondent does not identify the website with which they have been engaging as being a platform (Zumbrun & Sussman, 2015; Katz & Krueger, 2016). It is thus important that the survey instrument takes account of the specific issues concerning platform work when considering this principle.

The third principle, concerning the full reporting of the information, is of particular relevance to platform work as it is often undertaken as part of the informal economy (Vallas & Schor, 2020). With platform workers frequently being treated as independent contractors or self-employed rather than employees (De Stefano, 2016; Prassl, 2018), the reporting of such work to relevant authorities is not subject to standard procedures and respondents may be concerned that there could be repercussions if they report that they have undertaken such work (Williams & Schneider, 2016). As a result, the survey instrument will need to be designed so as to draw out information from respondents without giving them cause for concern in this regard. The work of Williams and Schneider (2016) on collecting data on the shadow economy is relevant here. From another point of view, it may be that the reporting of platform work by respondents may be affected by the wish to give socially desirable answers (Dench, Iphofen & Huws, 2004) Some platform work may be regarded by some as being of low status (Deng, Galliers & Joshi, 2016; van Doorn, 2017) or respondents may be undertaking such work because they are unable to obtain more regular work (Green et al., 2014; Wood et al., 2019) The

survey instrument will thus need to be designed so as to ameliorate this issue. The desire for responses to be fully and correctly reported under the third principle also means that respondents should have response options available to them that will cover all eventualities. In the context of platform work, which can be very varied in nature (Rangaswamy, 2019; Vallas & Schor, 2020), this is also a particularly important issue.

As a result of the above issues relating to the three principles it was decided to ask respondents a series of questions about activities they had actually undertaken. It is well known that simple questions of this type, asking about facts personally experienced by respondents, typically yield accurate information (Moser & Kalton, 1971) and satisfy the first principle. Initially respondents were to be asked: 'How often, if at all, do you do each of the following online?', followed by a definition of what doing something 'online' meant and a number of activities such as selling possessions online, finding guests for accommodation, looking for a job on a job search website, looking on named websites (or similar) for work they could carry out from their own home, looking on named websites (or similar) for work they could carry out outside their home via a website, looking on named websites (or similar) for driving work.

By asking the question in this way, there was no need for respondents to understand what was meant by platform work and they were only being asked to report factual information. Next, respondents who had responded that they had looked on named websites (or similar) for work they could carry out in their own home, outside their home or that involved driving were asked: 'You indicated that you have used an online platform such as [*example websites*] to look for work . . . How often, if at all, do you personally tend to find new paid jobs, using any of these platforms?' Examples of such work were given to help the understanding of respondents. They were asked how often they found new jobs they could carry out in their own home, outside their home or that involved driving, giving options ranging from 'Every day' to 'Less often than once a year' as well as 'Never', 'Don't know' and 'Prefer not to say'.

Those respondents who replied that they had used websites like those specified to find new paid jobs were then asked about the nature of the jobs they undertook: 'How often, if at all, do you do each of the following types of work using any of these online platforms?' A list of various types of work were then presented and respondents invited to say how often they undertook each type.

By taking this multi-step approach, it was possible to filter out those respondents who had not undertaken work that would be regarded as platform work, without the respondents themselves having to understand whether or not they had done work that would be classed as such. At each stage, the fact that only basic information is being requested and a wide range of options is presented helps address the third principle. By prompting respondents to recall particular websites they may have used and particular jobs they have undertaken, the issues concerning the second principle are addressed.

Matching survey mode to principles for questionnaire

We concluded that, due to the principles involved and the specific issues concerning platform work, we needed to pose a series of questions to respondents. To actually design the wording of these questions, one needs to consider the survey mode to be used (de Leeuw, 2018).

From a practical point of view, the choice of survey mode was between a telephone survey, a face-to-face survey or an online survey. Of course, each of these would introduce an element of bias into the study. For now, we concentrate on the choice of survey mode from the point of view of the best fit to the sorts of questions to be asked and the needs of the study as a whole, considering the issues of bias and obtaining a sample.

Examining the sort of questions that are to be asked, one can see that, whatever the final wording, there is a fair degree of complexity involved. The initial stage of asking about the search for work via websites involved giving examples and also making distinctions between different sorts of website. The second and third stages also contained complexities which needed to be communicated to respondents. The instructions to respondents are thus not going to be simple and a survey mode which enabled respondents to read instructions, either on a card presented by a face-to-face interviewer or on a web page, would have an advantage here (Fowler, 2014; de Leeuw & Berzelak, 2016). The questions were also seeking to find out how frequently respondents had searched for work and found new jobs and what sorts of jobs were undertaken, prompted by being given a range of options. Thus, respondents would have a good number of options from which to choose and, again, a survey mode which enabled respondents to read instructions would have an advantage (presuming that they were literate).

In the discussion of the third principle, the issue of respondents needing to be reassured of confidentiality and being inclined to give socially desirable answers is addressed. Although the interaction with a person available through a telephone survey or face-to-face survey may give greater opportunity for the reassurance of confidentiality to be given, it has nevertheless been found that the presence of a human being often gives rise to greater concerns over confidentiality than the interaction with an impersonal computer. Similarly, an online survey suffers less from the desire of respondents to give socially desirable answers (de Leeuw, 2018).

The above discussion points towards an online survey suffering from fewer problems in coping with the questions to be asked than a telephone survey or face-to-face survey. Williams & Schneider (2016) have shown that online surveys have been useful in obtaining data relating to the shadow economy. There are sufficient similarities between this area and platform work (Rangaswamy, 2019) to suggest that their advantages would transfer across to the subject of the study discussed here.

Of course, it is also the case that an online survey can be less demanding on budgets than a telephone survey or face-to-face survey, for a fixed sample size (Sarracino, Riillo & Mikucka, 2017). Alternatively, for a set budget, a larger sample size can be obtained and this may be of particular use in our study as platform work is carried out by a relatively small proportion of the population. Thus, in order to identify sufficient platform workers to explore their characteristics in a meaningful way, it is useful to work with as large a sample size as possible and the relatively low cost 'per respondent' of an online survey is useful.

Consideration of bias

In our previous discussions, we have already introduced the issue of bias, developing the rationale for the study design by considering the best means to obtain data which are as unbiased as possible. Taking a wider perspective, Total Survey Error encompasses both non-observation error and observation error (see e.g. Biemer et al., 2017; Tourangeau,

2018). Non-observation error occurs because a survey does not collect data from every member of the population and the sample obtained may not be representative of the whole population. Observation error occurs because methods of data collection can introduce non-random errors into the measurements/recordings made. Of particular relevance to this work is the fact that different survey modes may introduce different types, levels and directions of observation error (see e.g. de Leeuw, 2018).

Whichever survey mode is to be used, a careful approach to the recruitment of respondents is key to minimising non-observation error in the form of sampling bias and non-response bias. For online surveys, self-selection is a particular concern and thus we chose to enlist the expertise of a large experienced polling company, Ipsos MORI. Further details of how the sample were obtained and non-observation error minimised are discussed below.

The issue of observation error cannot be separated from the choice of survey mode (de Leeuw, 2018). Of particular relevance in the current study of platform work, and the requirement to ask a series of questions with multiple response options, are those potential errors introduced by the following four factors.

First, respondents may misunderstand the questions asked, particularly if they cannot see them written down when the telephone survey mode is used.

Second, respondents may be confused by being given a large number of response options, again particularly if they cannot see them written down when the telephone survey mode is used.

Third, respondents may exhibit satisficing behaviour in order to complete the survey quickly and easily. This is a problem for all survey modes, with respondents giving the simplest possible responses so as to avoid cognitive burden (Krosnick & Alwin, 1987). For telephone and face-to-face survey modes, respondents are less in control of the process than with an online survey because of the presence of an interviewer. As a result, they may put less cognitive effort into thinking about responses (Gooch, 2015). Thus, the issue of recall bias is of relevance, with telephone and face-to-face survey modes suffering from this to a greater degree than the online survey mode.

Fourth, respondents may wish to give socially desirable answers (de Leeuw, 2018). This is particularly an issue for those survey modes where an interviewer is involved but is less of an issue for the online survey mode.

In any survey, there are risks of sampling and non-sampling bias. In the specific instance of our study of platform work, the risk of sampling bias due to using an online survey mode was a concern. The decision to use Ipsos MORI to obtain the sample was important in ameliorating this issue. From the point of view of non-sampling errors, there are concerns that both the telephone and face-to-face survey modes might introduce biases which would be greater than those introduced by an online survey. The benefits/losses resulting from each survey mode need to be balanced and individual researchers may come to different conclusions as to the best way to proceed. We adopted a cautious approach, proceeding with the online survey mode but simultaneously commissioning additional surveys using alternative survey modes. These additional surveys are discussed later in the article along with an assessment of how the choice of an online survey mode impacted on the study of platform work. We

were fortunate to have the backing of FEPS for these plans and they are to be applauded for funding us to include these offline assessments in the study.

Obtaining a sample

As discussed above, a major issue when conducting an online survey is the difficulty of obtaining a sample which is representative of the population of interest. Associated with this is the known difficulty of low response rates in business and management research identified by Mellahi & Harris (2016). One of the major deficiencies of online surveys is that those who do not use the internet are unable to be part of the sample and this inevitably introduces biases. However, in recent years this gap in coverage has narrowed with the growth in internet use and mobile device usage, particularly across countries of Europe where the surveys in this study were to be conducted (Bahia & Suardi, 2019).

To help address the issues inherent in obtaining a sample for an online survey, we partnered with Ipsos MORI to ensure that we would obtain samples of sufficient size that would be as representative as possible. As members of ESOMAR (formerly the European Society for Opinion and Marketing Research), they are committed to following high standards in their data collection activities.[1] We were also encouraged by the fact that, as a leading global survey company, they had a reputation to protect and thus an incentive not just to provide data but to provide data from representative samples. Self-selection issues are minimised by the fact that our questions were embedded within an omnibus survey in most countries, in which questions from other organisations were also present. Any self-selection into the survey that takes place is thus less likely to be associated with the questions being asked for our survey and thus the biases introduced are likely to be smaller. Samples of approximately 2,000 respondents of working age were provided by Ipsos MORI, stratified by demographic factors, and post-stratification weights were also provided for use in analyses to help balance relatively small deviations from population characteristics (see Huws et al., 2019).

Where possible, we included the survey questions within an omnibus survey completed by Ipsos MORI's online panel. The survey questions were thus included alongside questions asked on a variety of topics for a number of Ipsos MORI's clients. As a result, the risk of respondents being concerned that their responses might fall into official hands was reduced, helping to provide more complete responses. We acknowledge that using online panels has particular inherent issues (Callegaro, Manfreda & Vehovat, 2015) but believe that partnering with Ipsos MORI minimised these problems.

Assessing the effect of survey mode

In this section we address the issue of how potential biases in a survey carried out online can be assessed by additionally using other survey modes. We depart from traditional methods of assessing the validity of survey results and argue that, for the

---

1   See https://www.ipsos.com/en/ipsos-answers-esomar-28-questions-help-online-research-buyers

vast majority of studies, it is not reasonable to compare summary statistics and expect them to be nearly identical. Rather, we argue that the survey results must be compared holistically, giving due regard to the likely biases that are inherent in all the different survey modes used.

This desire to compare the effect of survey modes is of particular relevance for online surveys because they involve less direct contact between those running the survey and the respondents. It is sometimes suggested (e.g. Fowler, 2014) that this reduced contact leads to less control of the survey process and thus greater risks of bias.

Adapting the online survey for offline survey modes

As discussed above, the questions in the online survey were created as a series which ask about searching for work via websites, finding new jobs via certain types of website and what sorts of work have been undertaken, in each case asking how frequently these happened. We asserted that the online survey mode is best suited to a survey containing these questions, which implies that the offline survey modes of telephone and face-to-face are not best suited to these questions. Indeed, when moving beyond the outline of questions to the actual wording of the questions themselves, we designed these with particular attention to the online survey mode which was being used. If we were now to conduct the survey using two further survey modes, we needed to decide how to adapt the online survey to these new circumstances.

It is an accepted aspect of survey design that one needs to take account of the survey mode to be used, or how this should be adapted if a mixture of modes is to be used (see e.g. de Leeuw, 2018). If we had been obliged to conduct this survey into platform work using a telephone or face-to-face survey mode or using mixed modes, then we would, inevitably, have generated a series of questions which might be quite different from those used in the online survey mode. One option was, therefore, to run the offline surveys with new sets of questions and subsequently compare the results of the three survey modes. Another option would have been to use the existing questions from the online survey in the new survey modes. It is clear that neither option is ideal. The first option – rewriting the questions – has the additional difficulty that there is no single 'best' way of wording the questions for the new survey modes so the effect of the change of survey mode is heavily confounded with the change of questions (which also add their own biases) and, of course, the unavoidable sampling bias. We would also argue that (for reasons outlined in above) it is not possible to create a set of questions for the telephone and face-to-face survey modes which would be appropriate for the purposes of the study. The second option means that, when assessing the results from the different survey modes, one cannot simply compare summary statistics but the sub-optimal nature of the questions used also needs to be taken into account, as well as the change of mode and the sampling bias. However, there would be no need to generate revised questions and, as a result, have the choice of these questions add further biases. For this reason, we chose to carry out the offline surveys using the same core questions that were used in the online survey, although, for reasons of cost, we removed from the survey a number of supplementary questions that were not germane to the purpose of comparing survey modes.

Principles for comparing results from different survey modes

As already noted, we argue that it is not appropriate to simply calculate summary statistics and draw direct comparisons without taking into account the sub-optimal nature of the questions asked and the survey modes. Although the calculation of confidence intervals and the assessment of any overlap may be of interest, these should merely be regarded as some of the pieces of information that contribute towards an overall assessment of the performance of the survey modes.

We expect every survey to contain biases and the key to making relevant comparisons between the results from the different survey modes lies in understanding what these biases may be and the way in which they affect the results obtained. We then need to look at these biases together and assess whether the differences being observed are consistent with the effect we would expect these biases to have on the results.

We carefully separate the issue of sampling errors and non-sampling errors. Any biases introduced by sampling errors are assumed to be due to the survey mode used and we are thus implicitly assuming that the sampling is being carried out as appropriately as possible for the relevant survey mode. Non-sampling errors that are related to the activity of respondents engaging with the questions asked (perhaps via an interviewer) are considered to be the sources of biases that may differ between survey modes. Non-sampling errors due to other factors, such as the recording and processing of the data, are likely to be random in nature. Thus, although these are a source of variation, we do not expect these latter errors to be a source of bias.

In the context of the current study of platform work, we have outlined four key sources of non-sampling response bias. These are listed again below, along with the effect we would expect them to have on the survey results. Once we have taken account of these sources of bias, any additional differences in results from the different survey modes can be considered to be due to the survey mode used.

First, respondents may misunderstand the questions asked, particularly if they cannot see them written down. Of particular concern for our survey of platform work is the fact that respondents cannot be expected to know that activities they may have undertaken may be regarded as platform work. It is possible that if a question is misunderstood, a respondent may report activity that has not actually taken place. However, for this to happen, the respondent must have undertaken some activity which is similar enough to platform work to fall within the scope of the misunderstood question without the activity actually being platform work. We consider this scenario to be less likely than the situation where activity fails to be reported because the respondent misunderstands the question and is thus not prompted to report the activity.

All three of the survey modes, telephone, face-to-face and online, may suffer from respondents misunderstanding questions. Of these, the telephone survey mode may suffer most because not only can respondents not reread the question themselves but they are also more distanced, both physically and psychologically, from the interviewer than is the case for the face-to-face survey mode. They are thus less likely to seek clarification than those with a face-to-face interviewer (de Leeuw, 2008). Although the respondents for the online survey are able to reread the question themselves, research suggests that these respondents are less motivated to put effort into understanding a

question than is the case with a face-to-face interview (de Leeuw, 2008). On the other hand, a respondent (even though they were able to read the interviewer's screen) may be reluctant to admit to a face-to-face interviewer that they have not understood a question and may not want to waste the interviewer's time by rereading the question or asking for it to be clarified. The impact of this potential source of bias thus appears to be that the telephone survey mode may under-report the level of platform work. While both the face-to-face and online survey modes may also suffer from some under-reporting, the effect is likely to be smaller than for the telephone survey mode.

Second, respondents may be confused by being given a large number of response options. The survey questions are of the form 'How often had you done X' with nine frequency options ranging from 'Never' to 'Every day' (along with 'Don't Know' and 'Prefer not to say'), appropriately giving respondents a means to fully report the extent of their engagement with 'X'. For both the face-to-face and online survey modes, the respondents can see a list of the options but for the telephone survey mode, this is not possible. As with the first issue, respondents in the telephone survey may be reluctant to waste the interviewer's time by asking for the options to be repeated and thus a default statement that they have 'Never' done 'X' is more likely. The face-to-face and online survey modes may suffer to some extent from having so many response options but it is likely to be to a lesser extent than for the telephone survey.

As a result, the impact of this potential source of bias appears to be that the telephone survey mode may under-report the level of platform work. While both the face-to-face and online survey modes may also suffer from some under-reporting, it is unlikely to be to as great an extent as for the telephone survey mode.

Third, respondents may exhibit satisficing behaviour in order to complete the survey quickly and easily. While this may be an issue for all the survey modes, it may exhibit itself differently depending on the mode. Where questions are not very straightforward, Gooch (2015) shows that survey modes where respondents are self-completing the survey give more accurate answers than those respondents who interact with interviewers. Thus, for the telephone and face-to-face survey modes in the study discussed here, the quickest and easiest thing is for respondents to tell the interviewer that they have never engaged in various activities prompted by the questions or, similarly, that they 'don't know'. However, with the online survey mode, respondents are more likely to correctly recall any instances of undertaking platform work.

Additionally, the issue of recall bias is relevant. Any respondent may be reluctant to spend time recalling past events to answer a survey question, but this may be even more the case when an interviewer is waiting for an answer to be given, either in a face-to-face situation or via the telephone (de Leeuw, 2008).

Considering these issues, it is thus reasonable to suppose that this form of bias may exhibit itself as the telephone and face-to-face survey modes showing lower levels of platform activity than the online survey mode.

Fourth, respondents may wish to give socially desirable answers (de Leeuw, 2018). As already mentioned, this may be a particular issue for a survey of platform work since some of the activities that are carried out in this way are regarded as being of low status (Flanagan, 2019). Furthermore, if the respondent is concerned about the confidentiality of the survey and information about their platform work finding its way to official

bodies, they may be reluctant to report it (Williams & Schneider, 2016). It is well-known that these issues are of greater importance when an interviewer is involved, be that via a telephone or face-to-face (Callegaro, Manfreda & Vehovar, 2015). On the other hand, an online survey is less likely to suffer from this (Callegaro, Manfreda & Vehovar, 2015) although the issue of trust in the organisation/person conducting the survey is also important (Dodou & de Winter, 2014).

As a result, it is likely that this form of bias may result in both the telephone and face-to-face surveys producing results that show lower rates of platform activity than is the case for an online survey.

To summarise, from the above we are led to hypothesise that we will see differences between the levels of platform work shown by the data from the three survey modes. Looking purely at the biases due to this non-sampling response error, it is likely that the telephone survey mode will show the most downwards bias in levels of platform work, with the face-to-face survey mode also showing a downwards bias but to a lesser extent. Lastly, the online survey mode is likely to have the lowest levels of downwards bias.

In addition to the biases due to the non-sampling response error discussed above, we need to consider the sampling biases introduced by the survey mode. We recall that the prime motivation for carrying out the surveys using the telephone and face-to-face survey modes was a concern that the online survey mode may be biased towards being completed by those who are more likely to undertake platform work. We thus need to interpret the results of the surveys in the context of both the expected biases due to non-sampling response error and also the potential sampling bias.

Comparison of online and offline platform work survey results

In this section, we give results from the online surveys carried out in the UK and Switzerland, the telephone survey conducted in Switzerland and the face-to-face survey conducted in the UK. The aim is to apply the principles outlined above and thus assess the impact of the online survey mode on the results obtained.

The UK online survey was conducting by Ipsos MORI using their online omnibus survey (where survey questions from a number of researchers are combined), with data collected between 22 and 26 January 2016. The sample was of UK residents aged 16–75, stratified by age, gender, region, social grade and working status. A sample of 2,238 respondents was obtained. This was followed up by a face-to-face (CAPI – Computer Aided Personal Interviewing) survey by Ipsos MORI, also as part of an omnibus, between 24 March and 7 April 2017 which was stratified by age, region, working status, social grade within gender, household tenure and ethnicity. The sample was of UK residents aged 16–75 and 1,794 respondents were obtained.

The online survey in Switzerland was again conducted by Ipsos MORI as part of an omnibus survey with data collected between 3 and 14 April 2017. The sample was of Swiss residents aged 16 to 70, stratified by age, gender, region and working status. A sample of 2,001 respondents was obtained. At about the same time (27 March to 7 April 2017), a telephone (CATI – Computer Aided Telephone Interviewing) survey was undertaken by Ipsos MORI, again as part of an omnibus survey. This was of Swiss residents aged 15 to 79 and was stratified by age, gender,

region and working status. A sample of 1,205 respondents was obtained which was reduced to 1,060 once the age range was restricted to be the same 16 to 70 age range as the online survey.

In the next section, we discuss the results obtained from the surveys. Although the surveys themselves cover a wide range of issues associated with platform work, we focus upon two measures relevant to the purpose of this article: assessing bias in online surveys using alternative survey modes. These are the proportion who undertake commercial activity online at three varying levels (ever having done so, doing so at least monthly and doing so at least weekly) and the proportion who undertake platform work at the same frequency levels. Other analyses which involve other data collected and relationships between these can be seen in Huws et al. (2019).

Results

In order to examine the effects of the different survey modes, we studied rates of undertaking online commercial activity (selling possessions/products via websites and/ or finding paying guests via websites) and rates of platform work (defined as having found and undertaken one or more relevant activities via an online platform). We considered three periods of time: weekly, monthly and ever. The results are shown in Figure 1 and Figure 2.
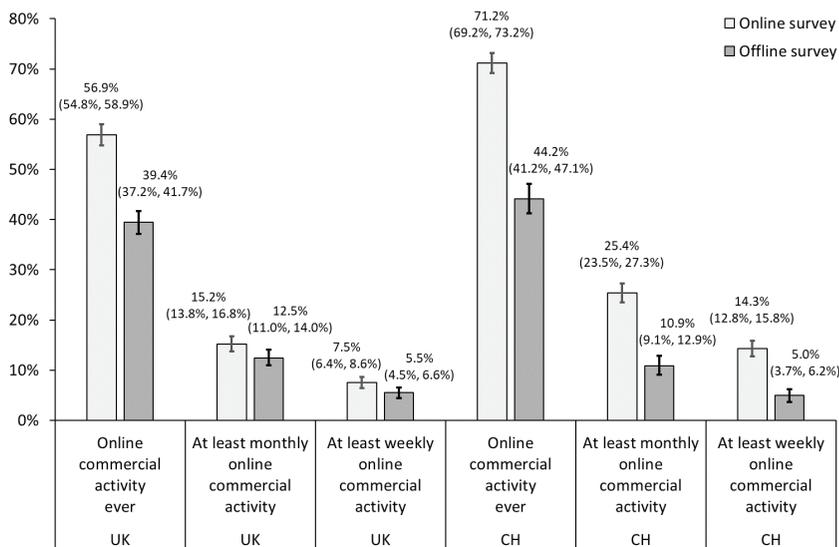
Figure 1:   Online commercial activity rates in UK/Swiss online and offline surveys (with 95% confidence intervals)

Base: 2,238 respondents in UK online survey, 1,794 in UK offline survey, 2,001 in Swiss online survey, 1,060 in Swiss offline survey (weights used; missing and don't knows excluded; percentages rounded to nearest whole number).
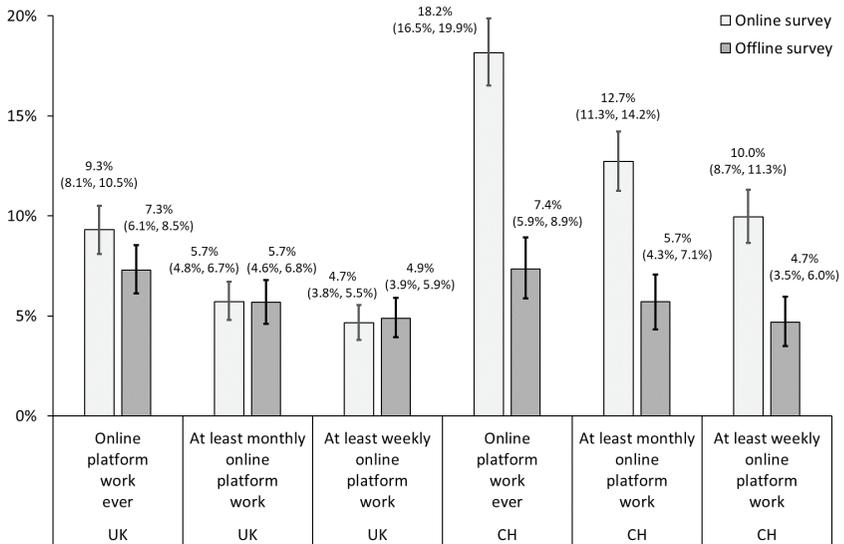
Figure 2: Platform work **rates in UK/Swiss online and offline surveys (with 95% confidence intervals)**

Base: 2,238 respondents in UK online survey, 1,794 in UK offline survey, 2,001 in Swiss online survey, 1,060 in Swiss offline survey (weights used; missing and don't knows excluded; percentages rounded to nearest whole number).

The rationale of this article is that at this point we do not immediately compare the confidence intervals of Figure 1 and Figure 2, but rather consider the principles discussed in the previous section.

From the predictions formed in that section, we note that, even with perfectly unbiased samples, we would expect the online survey mode to be showing higher rates of activity than the surveys which used the face-to-face or telephone modes. We would also expect the survey which used the face-to-face survey mode to be showing higher rates than the one which used the telephone mode. If we now consider the figures in Figure 1 and Figure 2, we note that this is indeed the pattern we observe.

We also note from above that recall bias may result in the face-to-face and telephone survey modes showing lower rates of activity than the online survey mode. Considering the patterns observed over time in Figure 1 and Figure 2, we see that they are consistent with this. Disparities between the results of the online survey mode and the face-to-face mode are greatest when considering the longest time period and at their least when considering the shorter time periods. We note that the disparities are greater for the comparisons between the online survey and telephone survey (the Swiss surveys) and less for the comparisons between the online survey and face-to-face survey (the UK surveys). This corresponds with our predictions.

We also seek to bring other sources of information to the investigation of the effect of the online survey mode. We explained earlier how efforts were made to minimise issues of self-selection and we thus do not expect non-response bias to have a great

effect on the results of the survey. Here we have further evidence against self-selection being an issue. In Figure 1 and Figure 2 we see that the greatest disparities between results from different survey modes exist when considering lower levels of online activity. However, if self-selection were an issue for the online survey mode, we might expect an over-abundance of internet-active respondents, leading to disparities between survey modes being greatest for those reporting more frequent online activity.

Clearly it is not possible to conclude definitively that any sample obtained from any of the survey modes is unbiased (and, indeed, none of them can be completely unbiased). However, as a result of the above considerations, the patterns observed (face-to-face results in Figure 1 and Figure 2 being somewhat smaller than those for the online survey mode and telephone results being smaller still) are what we expect to see, given the predicted effects on the survey results of the different survey modes. As a result, we are able to conclude that the patterns observed are consistent with the online surveys producing unbiased results.

Only at this point do we now move to include consideration of the confidence intervals in Figure 1 and Figure 2. We see that in the estimates from the online survey mode and the telephone mode (the Swiss surveys), the differences are such that none of the confidence intervals overlap. However, this greater difference between the online and telephone surveys (when compared with the difference between the online and face-to-face surveys) is to be expected, given the arguments made above. When we look at the results from the online and face-to-face surveys (in the UK), we note that for those estimates of commercial online activity and platform work at least once a week or at least monthly, the estimates are quite similar, with confidence intervals which overlap. The fact that it is these estimates which are most similar is consistent with the expectations given above for differences between the survey mode results.

Discussion

When any survey is undertaken, it is important that the effect of the survey mode chosen is examined. However, the methods to be used to carry out this examination should be secondary to the primary aim of collecting valid, reliable and timely data on whatever topic is being researched. This is of particular relevance to online surveys because of the combined effects of not having an interviewer/researcher present, the cheapness (meaning larger samples can be obtained) and the speed with which the data collection exercise can take place. The advantages of the latter two effects are likely to outweigh the difficulties introduced into the comparison of survey modes by the first mentioned effect which render the data collection instrument used to be less than ideal for other survey modes.

It could be argued that when new or poorly researched topics are being investigated, more priority should be given to facilitating an examination of the effect of survey mode. Indeed, this could be the case for the example given in this article of research into platform work. However, despite this, as discussed above, it was still decided that priority should be given to using methods that would give valid, reliable and timely data in preference to methods that facilitated a comparison of survey modes. It is probable that this would also be the case when many other new or poorly researched topics are being studied.

Despite the difficulties inherent in undertaking an examination of the effect of the online survey mode with that of other survey modes, we have argued, and demonstrated, that it is still possible to carry out such a comparison. Crucially, we have shown that the examination needs to go beyond a mere comparison of estimates and confidence intervals. It needs to pay due attention to the fact that all survey modes have their intrinsic problems and biases and that no set of results can be considered to be completely unbiased.

Some may be tempted to believe that the arguments we have developed in this article have simply been devised to overcome difficulties we found when considering the results of our surveys of platform work. However, if we were merely searching for evidence to justify the results we obtained, we could have simply relied on the comparison of figures in Figure 2 for 'at least weekly' and 'at least monthly' platform work between the UK surveys using online and face-to-face modes. We do not claim that the study design used was 'perfect' (a claim that no competent researcher would ever make) but we do claim to have constructed a 'reasonable' study design for the circumstances. We also claim to have taken the issue of survey mode seriously. This is why we undertook the comparison with face-to-face and telephone surveys and this is why we have gone beyond the simple comparison of estimates and confidence intervals to produce this article.

What emerges in this article is a message for quantitative researchers that it is possible to undertake a meaningful examination of the effect of the online survey mode, even when it is difficult to replicate the data collection using alternative modes. Further, because it is possible to undertake such an examination, it is an exercise which should form part of a good study rather than be abandoned because it is difficult.

© Neil H. Spencer, Dag S. Syrdal, Matthew Coates and Ursula Huws, 2022

REFERENCES

Armano E., A. Bove & A. Murgia (eds) (2017) *Mapping Precariousness, Labour Insecurity and Uncertain Livelihoods: Subjectivities and Resistance,* London: Routledge.

Bahia, K. & S. Suardi (2019) *Connected Society: The State of Mobile Internet Connectivity 2019,* London: GSMA.

Biemer, P. P., E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker & B. T. West (eds) (2017) *Total Survey Error in Practice*, New York: John Wiley & Sons.

Callegaro, M., K. L. Manfreda & V. Vehovar (2015) *Web Survey Methodology,* London: SAGE.

de Leeuw, E. D. (2008) 'Choosing the method of data collection' in E. D. de Leeuw, J. J. Hox & D. A. Dillman (eds) *International Handbook of Survey Methodology,* Hove: Psychology Press:113–135.

de Leeuw, E. D. (2018) 'Mixed-mode: past, present and future', *Survey Research Methods*, 12 (2):75–89.

de Leeuw, E. & N. Berzelak (2016) 'Survey mode or survey modes?' in C. Wolf, D. Joye, T. W. Smith & Y.-C. Fu (eds) *The Sage Handbook of Survey Methodology*. London: Sage Publications.

Dench, S., R. Iphofen & U. Huws (2004) *An EU Code of Ethics for Socio-Economic Research,* Brighton: Institute for Employment Studies.

Deng, X., R. D. Galliers & K. D. Joshi (2016) 'Crowdworking – a new digital divide? IS design and research implications', *Association for Information Systems Research Papers,* 148. https://aisel. aisnet.org/ecis2016_rp/148

De Stefano, V. (2016) *The Rise of the 'Just-in-Time Workforce': On-Demand Work, Crowdwork and Labour Protection in the 'Gig-Economy'*, Geneva: International Labour Office.

Dodou, D. & J. C. F. de Winter (2014) 'Social desirability is the same in offline, online, and paper surveys: A meta-analysis'. *Computers in Human Behavior*, 36:487–495.

Eurofound (2015) *New Forms of Employment,* Luxembourg: Publications Office of the European Union.

Felstead, A. (2021) 'Are online job quality quizzes of any value? Selecting questions, maximising quiz completions and estimating biases', *Employee Relations: The International Journal*, 43 (3):724–741.

Flanagan, F. (2019) 'Theorising the gig economy and home-based service work', *Journal of Industrial Relations*, 61 (1):57–78.

Fowler, F. J., Jr (2014) *Survey Research Methods*, 5th ed., Thousand Oaks, CA: SAGE.

Galvin, P., A. Bierman & S. Schieman (2021) 'Über-alienated: Powerless and alone in the gig economy', *Work and Occupations*, published online 22 June 2021.

Gooch, A. (2015) 'Measurements of cognitive skill by survey mode: Marginal differences and scaling similarities', *Research and Politics*, July–September:1–11.

Green, A., M. de Hoyos, S.-A. Barnes, B. Baldauf & H. Behle (2014) *Exploratory Research on Internet-Enabled Work Exchanges and Employability: Analysis and Synthesis of Qualitative Evidence on Crowdsourcing for Work, Funding and Volunteers*, European Commission Joint Research Centre Institute for Prospective Technological Studies, JRC Scientific and Policy Reports, Luxembourg: Publications Office of the European Union.

Huws, U. (2016) *A Review on the Future of Work: Online Labour Exchanges, or 'Crowdsourcing': Implications for Occupational Safety and Health,* Bilbao: European Agency for Safety and Health at Work.

Huws, U., N. H. Spencer, M. Coates & K. Holts (2019) *The Platformisation of Work in Europe: Results from Research in 13 European Countries,* FEPS – Foundation for European Progressive Studies, UNI Europa, Hatfield: Hertfordshire Business School, University of Hertfordshire.

Jesnes, K. & F. Braesemann (2019) *Measuring Online Labour: A Subcategory of Platform Work*, Nordic Future of Work, Brief 2, Oslo: Fafo Institute for Labour and Social Research.

Johnston, H. & C. Land-Kazlauskas (2018). *Organizing On-Demand: Representation, Voice, and Collective Bargaining in the Gig Economy,* Geneva: International Labour Office.

Katz, L. F. & A. B. Krueger (2016) 'The rise and nature of alternative work arrangements in the United States, 1995–2015', *NBER Working Paper* No. 22667, Cambridge, MA: National Bureau of Economic Research.

Krosnick, J. A. & D. F. Alwin (1987) 'An evaluation of a cognitive theory of response-order effects in survey measurement', *Public Opinion Quarterly*, 51 (2) 201–219.

Mandl, I. (2016) 'Working conditions in crowd employment' in W. Wobbe, E. Bova and C. Dragomirescu-Gaina (eds) *The Digital Economy and the Single Market,* Brussels: Foundation for European Progressive Studies:111–140.

Mellahi, K. & L. Harris (2016) 'Response rates in business and management research: An overview of current practice and suggestions for future direction', *British Journal of Management*, 27 (2):426–437.

Moser, C. A. & G. Kalton (1971) *Survey Methods in Social Investigation*, 2nd ed., London: Heinemann Educational Books.

O'Farrell, R. & P. Montagnier (2020) 'Measuring digital platform-mediated workers', *New Technology, Work and Employment*, 35 (1):130–144.

Prassl, J. (2018) *Humans as a Service: The Promise and Perils of Work in the Gig Economy,* Oxford: Oxford University Press.

Rangaswamy, N. (2019) 'A note on informal economy and ICT', *Electronic Journal of Information Systems in Developing Countries*, 85 (3).

Ross, J., L. Irani, M. S. Silberman, A. Zaldivar & B. Tomlinson (2010) 'Who are the crowdworkers? Shifting demographics in Mechanical Turk', *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Extended Abstracts Volume*: 2863–2872.

Sarracino, F., C. A. F. Riillo & M. Mikucka (2017) 'Comparability of web and telephone surveys for the measurement of subjective well-being', *Survey Research Methods*, 11 (2):141–169.

Tourangeau, R. (2018) 'Data collection mode' in D. L. Vannette & J. A. Krosnick (eds) *The Palgrave Handbook of Survey Research,* Cham: Palgrave Macmillan.

Tubaro, P., C. Le Ludec & A. A. Casilli (2020) 'Counting "micro-workers": Societal and methodological challenges around new forms of labour', *Work Organisation, Labour & Globalisation*, 14 (1):67–82.

Vallas, S. & J. B. Schor (2020) 'What do platforms do? Understanding the gig economy', *Annual Review of Sociology*, 46:273–294.

van Doorn, N. (2017) 'Platform labor: On the gendered and racialized exploitation of low-income service work in the "on-demand" economy', *Information, Communication & Society*, 20 (6):898–914.

Williams, C. C. & F. Schneider (2016) *Measuring the Global Shadow Economy: The Prevalence of Informal Work and Labour,* Cheltenham: Edward Elgar.

Wood, A. J., M. Graham, V. Lehdonvirta & I. Hjorth (2019) 'Good gig, bad gig: Autonomy and algorithmic control in the global economy', *Work, Employment & Society*, 33 (1):56–75.

Zumbrun, J. & A. L. Sussman (2015), 'Proof of a "gig economy" revolution is hard to find', *Wall Street Journal*, 26 July.