

NatCen

Social Research that works for society

Considering alternative fieldwork designs for a biennial face-to-face longitudinal survey

Authors: Joanna D'Ardenne, Martin Wood and Bruno Pacchiotti

Date: 26/06/2019

Prepared for: ESRC



**Economic
and Social
Research Council**

At NatCen Social Research we believe that social research has the power to make life better. By really understanding the complexity of people's lives and what they think about the issues that affect them, we give the public a powerful and influential role in shaping decisions and services that can make a difference to everyone. And as an independent, not for profit organisation we're able to put all our time and energy into delivering social research that works for society.

NatCen Social Research
35 Northampton Square
London EC1V 0AX
T 020 7250 1866
www.natcen.ac.uk

A Company Limited by Guarantee
Registered in England No.4392418.
A Charity registered in England and Wales (1091768) and Scotland (SC038454)
This project was carried out in compliance with ISO20252

Contents

Summary	1
1 Introduction	2
2 Methodology	3
3 Questionnaire review findings	4
4 Response considerations for alternative fieldwork designs	7
5 Summary reflections	10
References	12
Appendix A. Desk review code-frame	13
Appendix B. Questionnaire review detailed findings – all modules	17

Summary

This report provides some initial considerations for how online data collection might be used to supplement and replace elements of the current biennial face-to-face fieldwork design of the English Longitudinal Study of Ageing (ELSA). The approach used a desk-based review and stakeholder consultation to assess the potential risks to longitudinal data quality / comparability and an assessment of potential response rates based on those of comparable online panel studies.

- The review found that there would be gains and losses in data quality in the transition to an online methodology, with more complex/ burdensome modules being more difficult to transfer, but much that could move online with minimal re-design.
- The scale of the current ELSA interview and the opportunities for more frequent data collection suggested a modular approach that varies the rotation of topics and enables specific events to be identified and explored more accurately.
- Response analysis from the random probability panels suggests that it would be possible to achieve relatively high levels of response to more frequent online surveys, particularly where backed by a CATI mode, except in the case of the oldest age group.
- This would be unlikely to replace the need for CAPI, particularly given the range of biomeasures that ELSA routinely collects, but the role of nurses and interviewers may change over time to facilitate alternative designs.
- However, the challenges in terms of complexity and resourcing of implementation of more extensive options are considerable in terms of sample management, survey implementation, data archiving and analysis.
- The report provides an initial step, but a cautious development approach that provides evidence for the likely impact on quality, comparability and sustainability is important, and should include a realistic assessment of costs.

1 Introduction

Background

Longitudinal studies in the past have often comprised periodic large-scale fieldwork operations that target their entire sample with lengthy interviews at each fieldwork sweep. However, social phenomena change at different rates and there is variation between them in the level of detail analysts require participants to report. Individuals' occupations will be relatively memorable and will change relatively infrequently: we may be content with the level of detail that it is possible to collect in work histories every few years. In contrast, self-assessed wellbeing is not easily collected retrospectively and there may be interest in its variation over relatively short timeframes. There are other types of information, such as physical activity, that might ideally be collected continuously.

A further limitation for longitudinal research of periodic data collection is the omission or late identification of life events such as leaving the parental home, separation, retirement or the onset of a health condition that researchers would wish to identify close to the point of their occurrence to obtain crucial detail. Although there are benefits in terms of simplicity of analysis from a single-interview model in having all measures available for all participants from a single point in time, the relatively long periods between survey sweeps often just reflects the costs and complexity of setting up large-scale face-to-face surveys rather than reflecting analytical needs.

Web and CATI surveys that are cost-effective and more quickly administered change this calculation, potentially enabling more frequent collection of existing measures or the collection of new types of measure. In this context, the English Longitudinal Study of Ageing (ELSA)¹, in common with other longitudinal and cohort studies, is considering designs that would enable more efficient and frequent data collection in some areas, compared with its current cycle of biennial face-to-face interviews with nurse visits every four years.

Aims of the research

This report provides some early considerations of the basic proposition of trying to move some or all data collection for ELSA to more efficient and flexible modes, specifically web or sequential web-CATI. A key consideration is the risk to an invaluable, consistent data series and a committed cohort of future research participants. We address three questions related to the risks and opportunities:

- Looking across the latest Wave 9 data collection instrument, what topics and measures is it feasible to move online with minimal risks to consistent data?
- What might be the areas of substantive analysis that would benefit from more frequent data collection?
- What are the likely patterns of response for a web-only or a web-CATI design that included more frequent data collection than currently?

We then consider the implications of the findings for the development of new approaches and the steps required for a development process that will preserve the value of this key resource for science.

¹ <https://www.elsa-project.ac.uk/>

2 Methodology

Our review comprised of three elements:

1. A desk review of ELSA CAPI questions to ascertain which questionnaire modules would be more at risk of mode effects occurring if transitioned online, and what steps if any would need to be taken to minimise these risks. This included a review of the dress rehearsal timing estimates for each module to provide further evidence as to which modules would be suitable.
2. Consultation with the ELSA senior management team to explore the perceived utility to moving towards more frequent data collection and which survey topics this could apply to.
3. Review of the response patterns for two probability-based general population panels that use predominantly online data collection in the delivery of multiple short surveys over a one-year period to help guide understanding of the likely implications of more frequent data collection.

Desk review of questions

A desk review was conducted of the Wave 9 ELSA CAPI questionnaire to establish whether an alternative fieldwork design could work in practice for this survey.² The aim was to identify all questions with a 'risk factor' associated with mode effects. The coding allows us to comment on what proportion of ELSA items could, theoretically, work well online (i.e. with limited risks of mode effects) and what questions may have some level of risk attached to such a transition. The reason/type for each risk is also coded. This allows us to propose mitigating steps to address the risks identified based on feature of the question is flagged. While the focus is ELSA, the findings will be applicable to these topics and specific measures asked in face-to-face settings in other longitudinal studies.

The methodology used to classify 'at risk' ELSA questions is based on a code scheme developed by NatCen's Questionnaire Development and Testing (QDT) Hub. This code scheme was developed and used to review all items proposed for waves 7-10 of Understanding Society (d'Ardenne *et al* 2017). Some codes from the Understanding Society code-frame were dropped as the earlier review indicated they rarely applied – 'Fear of Disclosure' code relating to illegal or illicit behaviours was dropped on this basis. Similarly, the code 'Document consultation' was dropped as this was not recorded on an item-by-item basis. Other codes were dropped to streamline the review in cases where there was overlap between the codes used. Details of all the codes used for the ELSA desk review are provided in **Appendix A** and the populated spreadsheet used for in the review is available alongside this report. The risk coding related to three overall categories:

- **Interviewer effects** (including socially desirable reporting and positivity bias);
- **Difficult questions** (including long / complex wording, computations or ranking and open questions);
- **Presentation effects** (including the number of answer options and hidden codes for interviewers).

For the purposes of this review we are only looking at substantive sections of the ELSA Wave 9 CAPI questionnaire. For brevity the household grid and demographic questions were excluded from the coding. The self-completion items were also excluded from the coding exercise because the transition from a paper self-completion to a web-completion should be relatively straightforward, and it was felt that the self-completion module could work in an equivalent way online and on paper. The measurement and cognitive functioning modules were excluded on the basis that

² Wave 9 is currently in the field, but the Wave 8 questionnaire is largely similar in content and can be accessed at <https://www.elsa-project.ac.uk/>.

these measures are sufficiently different from standard survey questions that the code-frame developed was largely not applicable to the items.

To complement the coding exercise we looked at timing estimates for each of the modules. These were taken from the Wave 9 dress rehearsal data and only include timings for single person rather than multi-person/ concurrent interviews.

Consultation with stakeholders

In addition to the desk review, a short consultation was conducted with key ELSA stakeholders. A meeting was held (and followed up in correspondence) in which the ELSA scientific leadership team were asked to reflect on the following questions:

1. Cost aside, would there be any appetite for more frequent, online only, sweeps of ELSA? What would be the main advantages and disadvantages?
2. Setting aside any concerns regarding online response/ response bias, are there any ELSA measures that would be beneficial to collect on more regular basis (e.g. annually, monthly or even more regularly?) What types of measures?
3. Setting aside any concerns regarding online response/ response bias, would there be any interest in collecting specific measures that are triggered by a life event? What type of life event?

Web-CATI panel response analysis

There are two random probability online / online-CATI panels in Britain that are comparable to ELSA (and broadly to other longitudinal surveys) in that they follow-up face-to-face surveys to form their samples. These panels deliver multiple short surveys over a relatively short period. Analysis of their response rates over these periods will guide consideration of whether event-triggered surveys or approaches that require the frequent collection of measures from the same sample are feasible.

Finally, based on practical experience of the delivery of ELSA and the online panels, we will reflect on the operational implications for moving to a more frequent survey design, and the implications of the findings from the desk review and consultation on the potential for an alternative fieldwork model for ELSA and more broadly.

3 Questionnaire review findings

This section presents illustrative findings from the desk review and stakeholder consultation that highlight issues relevant to the consideration of moving ELSA, or any other face-to-face longitudinal study, to a data collection design with an online component. We focus on ELSA's **health module** to do this and reference findings from other modules where they raise additional issues. We also reflect here on the appropriateness of a move to an online mode overall, taking account of the detailed review of each module. Findings from the review for the individual modules are provided at Appendix B in full.

The ELSA Health module includes a wide range of questions including those on eye-sight, hearing, diagnosis of any health conditions, date of diagnosis, medication, attendance at health screening, pain and the impact of health-condition on activities. Feed-forward data is used to ask about previously reported health conditions. The average length of the module is 12 minutes (and can be up to 25 minutes for some individuals who have multiple conditions), making it the longest module to administer.

Table 3.1 summarises findings for the Health module against the risk factors for mode effects. It lists the types of risk and the percentage of items (n=164) for which the risk was present and a summary of the nature of the issues found.

Table 3.1: Summary review findings for Wave 9 Health module (N=164 questions)

Type of effect	% Qs flagged	Summary of issues detected
Interviewer effects	27%	Some of the health items were flagged as at risk of interviewer effects (e.g. cancer diagnosis, mental health problems). A number of questions on 'self-rated' experience of health were flagged as at risk from positivity bias.
Difficult questions	35%	Several health items were flagged as 'difficult' questions. These tended to be recall questions i.e. age at first diagnosis or date of diagnosis if a condition first presented since the last wave of data collection. Other questions flagged were 'open' questions in which interviewers must record specific name of conditions or medication. There may be a reduction in data quality at such open questions if they are asked online.
Presentation effects	40%	Some items had 'presentation' issue including interviewer prompts not currently viewed by respondents (e.g. giving example medications); interviewer coded responses (e.g. coding reasons for feed-forward disparities); and long lists of activities.

Interviewer effects: socially desirable reporting

Over a quarter of the measures in the Health module were affected by some form of interviewer effect. There are both advantages and disadvantages to an interviewer-administered approach for data quality in this module. It is conceivable that participants may feel more comfortable disclosing certain health conditions online compared to face-to-face (e.g. a cancer diagnosis, mental health problems). In this case the move to online reporting could increase data quality, for example if participants feel comfortable disclosing conditions that have previously been under-reported. Sensitive questions that might benefit from a self-completion mode were also found in the Female Health and Income and Assets modules and the in relation to behaviours in the Smoking and Exercise module.

On the other hand, as the Health questions asked are quite detailed, and the administration time quite long, an interviewer may motivate participants to provide full information and prevent break-offs (this is a rare event in ELSA). An interviewer would also be expected to provide a suitable amount of information at open questions, including for ICD coding in the health module or for SOC coding in the Work and Pensions module or drug coding with the nurse visit. Online approaches to these questions would need to develop, perhaps in future including sophisticated look-ups that use AI, to ensure sufficiently accurate coding was maintained. A similar issue is apparent in other modules including the Work and Pensions module, where participants are encouraged to consult payslips.

Interviewer effects: positivity bias

Several items on self-rated health were considered at risk of positivity bias, where responses to an interviewer are more positive than they would be the case with self-completion (Ye *et al* 2011). Examples include questions rating general health, eye-sight, hearing, memory and pain. Positivity bias may be more prevalent in interviewer-administered modes, so self-reported health may be worse during online waves and improve in face-to-face waves, leading to problems with wave-on-wave comparability. One solution to this might be to include all the 'self-rated' health questions in the self-completion module. This way, the risk of mode induced fluctuations would be minimized.

Difficult questions

Overall, 35% of items were flagged as 'difficult' questions. These tended to be **recall** questions i.e. age at first diagnosis or date of diagnosis if a condition first presented since the last wave of data collection. Although recall difficulties were flagged we have no reason to believe an interviewer's presence would improve recall. In fact, an online approach that enables more frequent data collection would reduce the time between the event and the survey and thereby make recall easier.

A different issue is presented by the **looping structure** of the Social Care module in relation to participant burden. Its looping structure around the nature of help with each type of care need is potentially problematic online, particularly given that it will be those with the most complex care needs who will have the most loops. An interviewer may be better placed to encourage and enable completion and avoid short-cutting and satisficing behaviours that may be more common online.

A further issue found in that module is the reliance for **routing on earlier questions** from the household grid which means it could not be a standalone online module in its current form.

Within that same module, some questions were flagged as being relatively **long**. This raises the point that if a **mobile-first** approach is required, where surveys are encouraged to be completed on any convenient device, this may require considerable redesign of questions across the modules, with a consequent potential break in comparability.

Presentation effects

Two-fifths of the health items were flagged as having some form of 'presentation' issue which could impact how well they work online. These included interviewer **prompts not seen by respondents** (e.g. giving example medications) and interviewer coded responses (e.g. coding reasons for feed-forward disparities, don't know and refusals). Some of these issues could be resolved by including respondent-facing help-screens online and by dropping the interviewer checks. A survey model whereby a face-to-face mode is used in some waves means there would be an opportunity to check data gaps in self-reports during the face-to-face waves (e.g. checking the details of health conditions first reported online).

A further aspect of a move away from interviewer-administration relates to **control over who completes the survey**. There may be some ethical concerns regarding the use of feed-forward data online. Currently, the health module specifically prompts people about health conditions mentioned in previous waves e.g. 'Do you still have [condition].' It is unclear whether this type of questioning is appropriate online, as we will not know whether prior responses could be read by non-respondents. For example, one could imagine a scenario whereby an ELSA respondent is assisted to complete a web instrument by a child or grandchild, who then becomes aware of a diagnosis. One could also imagine a situation whereby ELSA respondents are completing the survey on mobile devices in public places. Therefore, if the health module is moved online we would recommend removing the feed-forward check questions.

Another dimension to this control relates to the **negotiation around whether proxy interviews** can and should be conducted. Proxy interviews for those who are unable to complete the interview are an important principle for ELSA and difficult to administer with an online approach where an interviewer is not present to ensure a proxy is appropriate and consented.

Finally, there is a related **verification** aspect to the interviewer role that is not easily replicable in a web context. We cannot easily control who completes the interview in a web context or enforce rules around things like household membership for the household grid.

Preferred frequency for the Health module

Health is a key topic area for ELSA analysts from a policy perspective. The ELSA scientific leadership team discussed how more frequent waves of administration may allow for new types of data to be collected. One example given was the occurrence of acute, time-limited conditions such as respiratory and urinary tract infections (UTIs). It was suggested that in an elderly/ frail population these conditions could lead to long-term more serious impacts, and that various factors could moderate these. However, it is not feasible to get highly accurate records on this topic with

the current biennial approach, as respondents will be unable to accurately recall details of all ‘minor’ ailments that occur between waves. It was discussed how a more frequent mode of data collection that identified health and other events would allow for this type of data to be collected. Participants could receive an online questionnaire that includes a number of ‘trigger’ items i.e. to establish whether they have had an infectious disease in the last month/ three months etc. Additional follow up questions could be asked for participants who confirm this is the case.

The ELSA stakeholders also cited pain as an example of a phenomenon where more frequent data collected may be beneficial. Levels of pain can vary considerably over time and may be hard to report on retrospectively. It is possible that more regular data collection on specific topics such as this could be conducted online to capture ‘in the moment’ measures.

Summary of risks

There were advantages and disadvantages to the Health module moving online, but the majority of measures could be moved with relatively low risk of mode effects. Looking across the modules, some were more likely to be easily carried online than others. The detailed Social Care, Work and Pensions and Income and Assets modules appeared more challenging to transition than the Female Health, Smoking and Exercise, Social Participation, Housing and Consumption or Efforts and Rewards modules. All would require some adaptation, and there would be an expectation of some loss of comparability for questions in, for instance, the Expectations module, which was at risk of positivity bias.

4 Response considerations for alternative fieldwork designs

The alternative fieldwork designs considered in this report relate specifically to online data collection, which offers the most opportunity in terms of efficiency and flexibility. The precise approach and timing would be varied depending on the nature of the analytical requirement and type of instrument. ELSA includes an online dietary questionnaire in Wave 9 that requires two days to be selected at random and data collected on those days; time-use diaries might require logging in every day or more frequently. However, core measures might be rotated at more sustainable intervals (in terms of cohort member burden): for ELSA this might be an annual wave between the standard biennial face-to-face waves, perhaps including quarterly or even more frequent waves for some measures.

An alternative fieldwork design that includes relatively frequent online only or online-CATI surveys would expect to achieve a lower level of response compared to a design that included a full CAPI mode. An assessment of the analytical potential of such a survey should include both the degree of bias in the sample achieved and its size for analysis.

For ELSA, whose sample is aged 50 and over, a major question relates to the extent to which its older sample members would be able and willing to participate online. There are early suggestions from its online dietary questionnaire in Wave 9 that the older age groups in the sample are less likely to participate – sometimes for reasons related to their health but more often due to a lack of access / ability to use an internet-enabled device (among the general population, the BSA, for instance, finds 14% do not have regular internet access and are a distinctive demographic group). In this context, a CATI mode, as part of a sequential design with web first to maximise cost efficiency, would considerably increase the proportion of cohort members able to participate in principle.

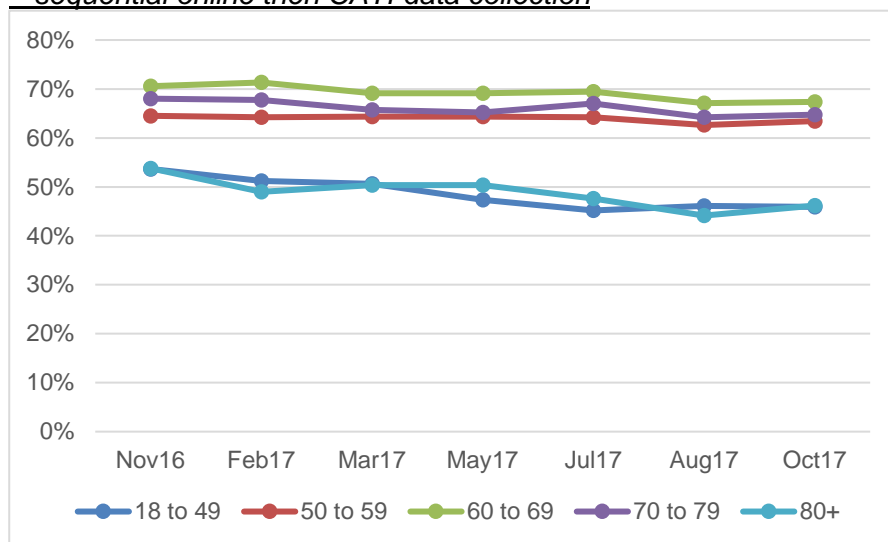
To provide some insight into the pattern of response that might be expected with a frequently-administered online (or online-CATI) approach, we analysed response for two random probability

panels (the only two of their kind in Great Britain): the NatCen online-CATI panel and the British arm of the CRONOS³ online panel run by ESS ERIC. These are useful for comparison with ELSA (and other longitudinal studies) in that in both cases the recruitment of panel members follows a face-to-face interview which is part of random probability survey (the British Social Attitudes survey and the European Social Survey respectively). This is comparable to ELSA, which follows up the Health Survey for England (HSE).

The two panels are recruited in similar ways and achieve a similar level of response when taking account of the initial BSA/ESS survey response rate and agreement to join the panel (together, estimated to be between 25% and 28% for the general population – a lower starting point compared with ELSA but not dramatically lower once the HSE response rate and the Wave 1 / refreshment ELSA response rates are taken into account. A difference to note for the two panels is that CRONOS is online-only but enables the participation of the ‘offline’ population by providing devices to those who report not having internet access. Despite this, analysis of response rates and response patterns for any given survey are very similar between the panels, as is participation in multiple surveys over time by a given cohort. Given this, the findings presented here focus on the NatCen Panel which has a larger sample size, enabling within more reliable age group breakdowns of response rates.

Our findings focus on a one-year period over which seven individual online-CATI surveys of around 15 minutes in length were administered to the same cohort of NatCen Panel members. In the November 2016 survey the response rate to the survey for the general population was 60%, rising to 67% among those 50 and over. Figure 4.1 shows response rates from November 2016 to October 2017 broken down by age groups. We see a relatively consistent response rate over time – the slight downward trend apparent for all age groups over the period was largely driven by attrition from the panel rather than declining response rates among those issued. This was usually triggered by the CATI contact. Analysis of the overall bias in the NatCen Panel suggests that there is little change in overall representativeness over time [reference to forthcoming R-Indicator ESRC publication from NatCen].

Figure 4.1 Cross-sectional response rates to successive surveys for a cohort on the NatCen Panel – sequential online then CATI data collection



Those aged 50 to 79 had a relatively high response rate throughout the period, suggesting cross-sectional web-CATI analysis would be feasible. Four-fifths (78%) of this group participated in at least one survey in this period, suggesting attempts to identify between-wave events would be effective. However, those aged 80 and over were substantially less likely to respond (54% compared to 71% among those aged 60 to 69 in

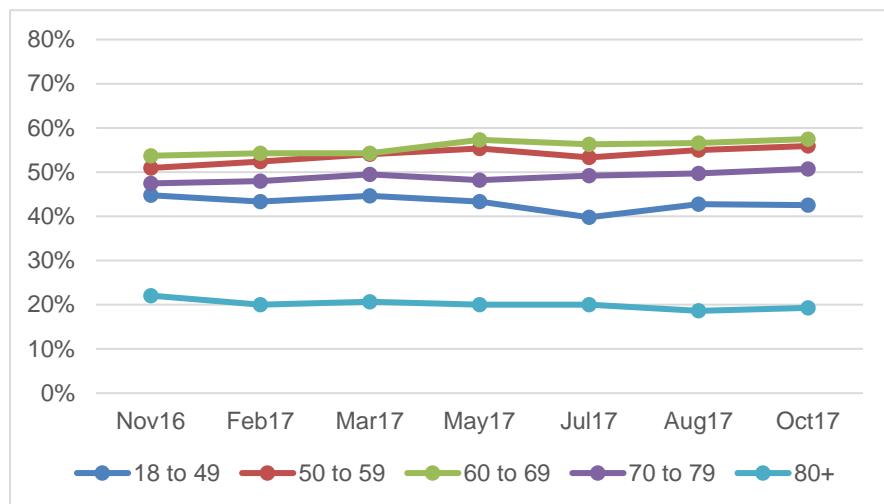
November 2016 and 59% participating in at least one survey in the period). This was the same level as for those aged under 50, although the reasons for non-response are likely to be quite different.

3

http://www.europeansocialsurvey.org/methodology/methodological_research/modes_of_data_collection/cronos.html

Figure 4.2 shows response rates for the same set of surveys but focused just on the online response (so excluding CATI interviews). Rates are stable over the period with a suggestion of increasing web response for those in the 50 to 79 age groups (54% for those aged 60 to 69 in November 2016, rising to 57% in October 2017). There may be scope for analysis that only focused on these age groups with online-only approaches.

Figure 4.2 Response rates for multiple surveys for a cohort on the NatCen Panel surveys – online completions only

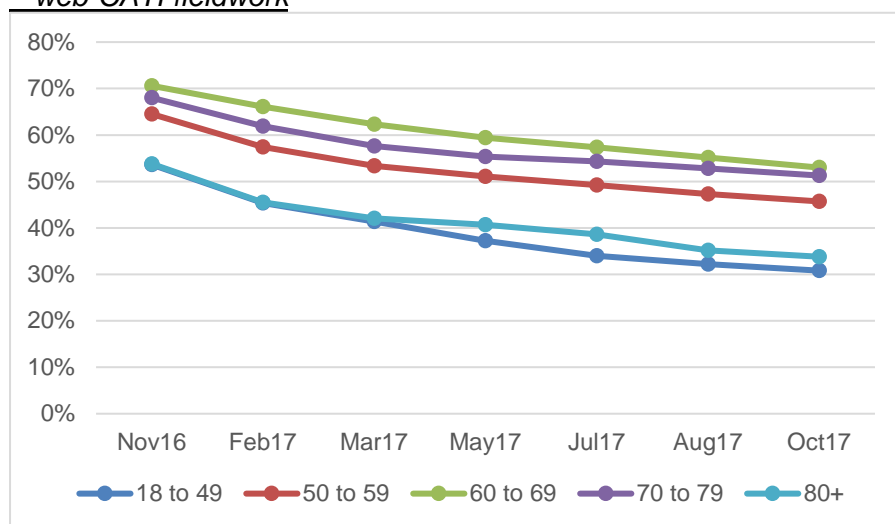


However, for the older population the level of response is considerably lower (22% in November 2017) and it seems clear that an online-only approach would not be viable. The CRONOS panel provided those who did not access the internet with devices, but consideration would need to be given to whether the oldest age group would be able and willing to take this

up without additional support.

Figure 4.3 considers the potential for *longitudinal* analysis (or analysis that requires regular collection of the same measures for an individual) over the same period. The chart shows the proportion of the cohort responding to all surveys to that point. At the higher end of participation, 53% of 60 to 69-year olds participated in all seven surveys in the series. This compared with 34% of those aged 80 and over.

Figure 4.3 Longitudinal response rates – percentage of cohort responding at all waves to that point – web-CATI fieldwork



In summary, with four-fifths of panel members aged 50 to 79 participating in at least one survey over an annual cycle, and half of this age group participating in all seven, both event-based analysis and that which requires regular measures to be taken from the same group of people would be feasible. However, those aged 80 and over were less likely to participate and would certainly require a CATI

option or extensive support to participate online.

5 Summary reflections

Gains and losses in a move online

Our review of the ELSA Wave 9 core interview material suggested that there would be gains and losses in data quality in the transition to an online methodology. Quality would be improved in instances where interviewers cause social desirability or positivity bias but reduced in relation to more complex questions or where records might be consulted. More burdensome modules including the Social Care, Work and Pensions and Income and Assets, were identified as more challenging to switch without significant changes. A predictable minority of measures were at risk from mode effects that could potentially harm comparability with data from previous sweeps, for instance with the module on future expectations which was at risk of positivity bias. However, there was much that could be transferred with minimal re-design that was mode agnostic.

Risks of web-first for the full core survey

A move to a full web-first approach to the core survey would require substantial redesigning of the structure of some modules and some questions, particularly if the goal was 'mobile first'. At 70 minutes on average – and considerably longer for those who have more health conditions and care needs – the ELSA interview is relatively long for an online survey, risking break-offs and reduced commitment to the study over time. This could be mitigated with an approach that targeted specific groups with web as the first mode in a sequential design, with those who would prefer/ need a CAPI interview allocated to that mode from the outset, as has been implemented on Understanding Society to date.

Modularisation and varying frequency of data collection

Another option to reduce burden and improve engagement is to modularise the core interview and deliver these modules (or sub-sections of modules) to sample members at different frequencies according to analytical needs. While in principle all areas of data would benefit from more frequent collection, a framework can be applied to decide what could be collected every four years and what would benefit from biennial, annual, quarterly or even monthly cycles. The highest frequency option may be aimed at:

- Measures such as wellbeing that may change frequently or attitudes / self-assessments that cannot be collected retrospectively
- Events where detailed recall of context is required (e.g. the context for a financial decision) and therefore need to be asked relatively close to the event
- Events that are important for analysis, but which may not be recalled in an interview some months / years later (e.g. minor infections / diseases)
- Frequent events where we otherwise need to talk about the 'most recent' or 'current situation', for instance in relation to episodic health conditions;

In addition to rotating modular surveys on specific topics, this might suggest short event-seeking surveys that then trigger longer online follow-ups where specific events are reported. Response analysis from the random probability panels suggests that it would be possible to achieve relatively high levels of response to more frequent surveys, particularly where backed by a CATI mode, except in the case of the oldest age group. Whether such a tempo of data collection could be sustained over multiple years without causing significant attrition is a question to be addressed.

Retaining a face-to-face component

A face-to-face mode at some point in the cycle would seem necessary to ensure full coverage of the sample, even if part of a multimode design. The review also noted parts of the interview process that would be difficult to manage if an online-only design were adopted for some individuals, notably that of establishing the need for / availability of proxy respondents.

In addition, there is the consideration of the collection of biomeasures. A highly valued aspect of ELSA's design is the consistent collection of biomeasures including hair samples, saliva samples,

anthropometry, balance tests, lung function and fasting whole blood samples, facilitated by a nurse visit every four years. Although Understanding Society is experimenting with self-collection of dried blood spots, there are currently no proven alternatives to the nurse visit for the full range of measures required. There are also barriers to the collection of cognitive function, the approach for which currently requires substantial interviewer interaction on ELSA.

Given the concerns around moving away from a face-to-face approach entirely in terms of the process and quality of data collection, the nurse visit may provide a potential solution (although it is currently only delivered every four years). However, this element would itself be under response rate pressure, given the current design is that nurses follow-up an interviewer visit – we would anticipate a decline in agreement to the follow-up if the first stage was completed online. In any event, it would be important to re-think of the role of interviewers (including telephone interviewers) and nurses to optimise their activity in the field. Their primary role should be to collect data that can only be collected in person but they can also help facilitate the high-quality implementation of a web-based operation.

Implementation considerations

The potential designs described above would entail significant sample and data management challenges, particularly where a CAPI mode is retained and used to ‘mop-up’ missing questionnaire areas and check quality. A modular approach with varying frequency of delivery would introduce multiple moving parts to be documented and managed. It would require sample management, including the production of clean feed-forward data, to operate at a high tempo, with updates from each round of data collection quality assured and built into the next round – and to deal with blurring fieldwork periods for overlapping surveys where multiple invitations and reminders may be going to individuals over short periods.

Software platforms are available that can manage frequent online surveys, but these would not lend themselves to multimode, and would be unlikely to cope with the inevitable complexity that longitudinal surveys introduce. Flexible, bespoke solutions are likely to be needed, and the manual effort to develop them over time and quality assure their operation should not be underestimated.

In addition to operational complexity, producing data for analysis would become considerably more challenging – and would in turn increase the complexity for analysts. There would need to be a strong focus on the strategy for archiving the data to avoid a reduction in its use by researchers due to the technical barriers being increased. However, with the anticipated rise in importance of linked administrative data (ELSA has been successfully linked to health and economic data), this is something that will need to be addressed in any case.

Finally, an important question that would need to be answered during a feasibility phase is whether the level of online completion achieved would reduce fieldwork costs sufficiently to offset the additional costs of delivering the more complex and resource-intensive survey management and data management task with a multimode approach. Significant up-front costs would be incurred and fixed costs would remain substantial to cope with the extra demands.

Development work

All of this suggests careful implementation and evaluation. This report represents a first step in a development process to assess a redesign of ELSA that maximises the potential benefits of online data collection while preserving data quality and comparability over time for analysts.

Ideally, we would have an experimental approach to understanding mode effects and the impact on response rates of different modes and additional data collection. Evidence from Understanding Society’s transition will be important and some sort of equivalent to the Innovation Panel that allows us to assess the impact of changes for ELSA’s specific context would be preferable. Costs will remain a significant question but given the potential of alternative fieldwork designs for analysis and long-term sustainability of the study, opportunities should be taken wherever possible to help move strategy forward.

References

Campanelli, P., Nicolaas, G., Jäckle, A., Lynn, P., Hope, S., Blake, M. and Gray, M. (2013) 'A Classification of Question Characteristics Relevant to Measurement Error and Consequently Important for Mixed Mode Questionnaire Design'. First presented at the Royal Statistical Society. London, 11 October 2011.

d'Ardenne J., Collins, D., Gray, M., Jessop, C., & Pilley, S. (2017) Assessing the risk of mode effects: review of proposed survey questions for waves 7-10 of Understanding Society, Understanding Society Working Paper Series, No.2017– 04:
<https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2017-04.pdf>

Dillman, D. A., Smyth, J. and Christian, L. (2009) Internet, mail, and mixed-mode surveys: The tailored design method. Hoboken, NJ: Wiley

Jessop, C. (2018) 'The NatCen Panel: developing an open probability-based mixed-mode panel in Great Britain' Issue 6 Summer 2018 Social Research Practice. Social Research Association.
<http://the-sra.org.uk/wp-content/uploads/social-research-practice-journal-issue-06-summer-2018.pdf>

Kaminska, O. and Foulsham, T. (2013) Real-world eye-tracking in face-to-face, web and SAQ modes. ISER Working Paper Series. 2013–7.

Krosnick, J. (1991) 'Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys', Applied Cognitive Psychology, 5, pp. 213–236.

Ye, C., Fulton, J. and Tourangeau, R. (2011) 'More Positive or More Extreme? A Meta-Analysis of Mode Differences in Response Choice', Public Opinion Quarterly, 75(2), pp. 349–365.

Appendix A. Desk review code-frame

Risk of interviewer effects

The content (or topic) of a question can impact on whether interviewer effects occur. For example, when answering sensitive questions people may respond differently in interviewer administered (IA) modes compared to self-completion (SC) modes. Socially desirable reporting is more likely to occur in interviewer administered modes. Some authors argue that interviewer presence is related to *positivity bias* in rating scales (where people are more likely to give 'positive' or 'agree' answers if a scale is interviewer-administered). Our final code-frame included two types of risk based on question content. The codes and coder instructions are shown in Table A.1 below.

Table A.1. Risk factors for interviewer effects

	Rationale	Coder instructions
1. Socially desirable reporting	Socially desirably reporting is more prevalent in interviewer administered modes compared to self-completion modes (e.g. Tourangeau, Rips and Rasinski: 2000). It could lead to a systematic bias in time-series data if there is a move from face-to-face to web data collection.	<p>Could participants edit their answers to 'look good' in front of an interviewer? Include anything related to socially desirable responses (e.g. drinking, voting, housework). Also include questions that there may be a cultural sensitivity or embarrassment when reporting (e.g. financial problems, mental health issues)</p> <p>Please try and think about how different groups of people could react to these questions. If any groups may edit their answers code '1. Yes.' If people may adjust their answer based on interviewer characteristics (e.g. sex, ethnicity etc) code '1. Yes'</p>
2. Positivity bias / Rating scales	Face-to-face responses are more likely to have extreme 'positive' ratings compared to the same questions asked as a self-completion (e.g. Ye et al 2011).	Is the question asking the participant to rate something on a scale e.g. their satisfaction with something or their levels of agreement with something? Include agree/disagree scales, fully labelled verbal scales (e.g. excellent-poor) or end point labelled scales (e.g. 0-10). Include three point scales

Difficult questions

It is generally assumed that the risk of respondent satisficing is greater for difficult questions than easy questions (Krosnick, 1991). In interviewer administered questionnaires, interviewers can:

- Ensure all pertinent information is read out;
- Motivate participants to perform more complex tasks in a way that maximises data quality (e.g. assist in computations and prompting them to provide full and adequate information);
- Explain more unusual question formats.

Therefore, interviewer presence could decrease satisficing behaviours by decreasing question difficulty, increasing respondent motivation and increasing optimal answering strategies. Our code-frame includes four codes on question difficulty. These are shown in Table A.2 below.

Table A.2: Risk of satisficing

	Rationale	Coder instructions
1. Complex question stem	Interviewers are trained to read out the entire question, including all instructions. In self-completion modes participants may not read the entire question, especially if it is long-winded or complex.	Does the question stem include lengthy instructions, introductions, or explanations? This code was automatically generated on whether the question stem (i.e. the question excluding the answer categories contained more than 240 characters).
2. Computation	In self-completion modes participants may be less motivated to carry out calculations and could adopt 'short-cutting' strategies. In interviewer administered modes assistance can be provided with calculations.	Does the question involve any mental calculations e.g. adding, subtracting or averaging? Only include questions where you suspect that respondents may have trouble adding up an answer or that might be susceptible to 'rounding.' E.g. age of child / number of people you look after are unlikely to require computation. Include recall of dates and numeric questions if appropriate but exclude verbal frequency scales (Always-Never).
3. Open questions	Less information is given in open questions in self-completions compared to interviewer administered questions. This can lead to differences in how open responses are coded.	Is a completely open <i>textual</i> answer required?
4. Ranking tasks	Ranking tasks are not recommended for self-completion modes (Campanelli <i>et al</i> , 2011).	Is the question a ranking task (example shown below)? <i>Please rank the following 3 items from 1 meaning most important to 3 meaning least important.</i> <ul style="list-style-type: none"> • <i>Less traffic [2]</i> • <i>Less crime [1]</i> • <i>More / better shops [3]</i>

Presentation effects

Some mode effects are associated with how questions and the answer categories are visually presented. Increasingly, survey participants may opt to complete online surveys on a mobile

device (such as a smart phone or tablet) rather than on a laptop or PC. This poses challenges for the questionnaire designer in terms of both making questions ‘user friendly’ but also maintaining comparability with other modes. For example, when viewing questions on a mobile device, participants may not be able to see all answer options on a screen without scrolling. This could increase the likelihood of primacy effects occurring in web (for long list questions) compared to a CAPI interview with a showcard.

In addition, some questions make use of substantive ‘interviewer coded’ or ‘hidden codes’ e.g. where interviewers are asked to record information about the respondent or the survey situation rather than asking the respondent to provide information directly. In some cases interviewers may also be provided with additional help screens, definitions and probes to assist the participant as necessary. These additional prompts can be included in a self-completion format but, depending on the level and complexity of the information, may need to be revised to make them more usable for members of the public.

Our final code-frame includes three codes on question and answer presentation that could result in mode effects in a CAPI/web survey. These are shown in table A.3 below.

Table A.3 Question and answer presentation issues

	Rationale	Coder instructions
1. Number of answer options	<p>In self-completion modes participants may be less likely to read all options compared to in a CAPI mode with a showcard. (Kaminska & Foulsham 2013).</p> <p>Higher numbers of answer options may be particularly problematic in terms of ‘mobile first’ design as not all options will be visible without the use of scrolling. This could lead to options lower down the list being less likely to be selected online rather than in face-to-face modes.</p>	<p>Are five or more answer options provided to the respondent. This code was automatically generated.</p>
2. Hidden codes and interviewer coded items	<p>There is limited evidence on how best to translate interviewer coded items into questions for participants and what impact this has on the data collected.</p> <p>Hidden codes may not be used in the same way in self-completion modes. Hidden modes can either be shown to participants (in which case they might be selected more in web) or remain hidden (in which case they may be selected less in web).</p>	<p>Does the item include an interviewer observation (that is not read out to respondents)?</p> <p>Are there any spontaneous or ‘hidden’ answer codes which are not shown to participants? Exclude hidden don’t know/ no opinion answer options.</p> <p>Are there any optional interviewer prompts or read-outs?</p>

3. Other 'non-standard' question format	During the course of the ELSA review the coding team made queries regarding a small number of questions that had 'non-standard' formats and how well these would translate to a mobile-first design. This final code was designed as a 'catch-all' to allow coders to report issues not captured elsewhere in the code-frame.	Are there any other questions that you think could be at risk of mode effects for issues not recorded above?
--	---	--

Overlap in conceptual framework

We acknowledge that there is a conceptual overlap on how the three areas described above interact to produce mode effects. For example, the positivity bias (described in Table A1) has been attributed by some authors to be due to acquiescent reporting in the presence of interviewers (see Ye *et al* 2011). However other authors ascribe the same effect as resulting from aural vs. visual pathways of question processing (Dillman *et al* 2009). For this project we are not aiming to definitively attribute causal mechanisms. The headings are being used for pragmatic reasons to help ensure all risk factors are fully mapped.

Appendix B. Questionnaire review detailed findings – all modules

The following sections look at each of the ELSA questionnaire modules reviewed. We describe findings from the desk review and stakeholder consultation and our reflections on whether a move online would be feasible and whether there would be any benefit of collecting data on a more frequent basis online.

Findings on Health module review

The ELSA Health module includes a wide range of questions including those on eye-sight, hearing, diagnosis of any health conditions, date of diagnosis, medication, attendance at health screening, pain and the impact of health-condition on activities. Feed-forward data is used to ask about previously reported health conditions. The average length of the module is 12 minutes (and can be up to 25 minutes for some individuals who have multiple conditions), making it the longest module to administer.

Table B.1. summarises the main issues identified when coding the Health module questions.

Table B.1: Findings on Wave 9 Health module (N=164 questions)

Type of effect	% Qs flagged	Discussion of issues detected
<i>Interviewer effects</i>	27%	Some of the health items were flagged as at risk of interviewer effects (e.g. cancer diagnosis, mental health problems). Several questions on 'self-rated' experience of health were flagged as at risk from positivity bias.
<i>Difficult questions</i>	35%	Several health items were flagged as 'difficult' questions. These tended to be recall questions i.e. age at first diagnosis or date of diagnosis if a condition first presented since the last wave of data collection. Other questions flagged were 'open' questions in which interviewers must record specific name of conditions or medication. There may be a reduction in data quality at such open questions if they are asked online.
<i>Presentation effects</i>	40%	Some items had 'presentation' issue including interviewer prompts not currently viewed by respondents (e.g. giving example medications); interviewer coded responses (e.g. coding reasons for feed-forward disparities); and long lists of activities.

See Chapter 3 of the main report for detailed findings for the Health Module.

Findings on the Female Health module

The Female Health module is a relatively short section of questions on menarche, menopause and use of HRT.

Table B.2: Findings on Wave 9 Female Health module (N=24)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	46%	Some items were flagged as potentially sensitive and at risk of interviewer effects. In this case the move to self-reporting could decrease this sensitivity.
<i>Difficult questions</i>	67%	Over half of the female health items were flagged as 'difficult' questions. These were recall based questions asking for details of age and dates. Although question difficulty was flagged we have no reason to believe an interviewer's presence would increase or decrease this difficulty, thus the risk of mode effects in this case would be minimal.
<i>Presentation effects</i>	4%	No major risks were identified in relation to presentation effects (one item was coded as having five or more response options).

Based on the type of risks code we suggest that the Female Health module would be suitable for online administration. In fact, having it as a self-completion may be preferential due to the subject matter under investigation. However, given the age of the cohort, we assume all data collected will rely on retrospective recall (and thus as already been collected or would be collected on the first interview for refreshment sample). Therefore, more frequent data collection would have no impact on recall, and it is unclear what the benefit would be of collecting this data at more frequent intervals.

Findings on Social Care module

The Social Care module includes detailed questions on what tasks, if any, the respondent has received help with, who has provided this help, duration and frequency of help received from each carer, funding of care and so on. It should be noted that in the Wave 9 questionnaire the smoking and exercise items were included under the Social Care sub-heading. They are therefore included in the below discussion.

Dress rehearsal data indicates the average (median) length of the Social Care module as four minutes. The maximum length recorded during the dress rehearsal was nine minutes.

Table B.3: Findings on Wave 9 Social Care Module (N= 94)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	41%	Multiple items were flagged as being potentially sensitive. Items flagged included those on type and extent of care needed and carer pay.
<i>Difficult questions</i>	47%	Multiple were flagged as being at risk as they were too long. Multiple items were coded as difficult in terms of recall/ calculation e.g. the questions determining hours and frequency of care provided by different individuals.
<i>Presentation effects</i>	42%	Multiple items were flagged as having some form of 'presentation' issue. Some items included interviewer prompts not currently viewed by respondents. Some questions were considered at risk due to number of response options. A small number of interviewer coded items are also present in the module i.e. on who is completing this section.

The Social Care module was considered to be a higher risk module for a transition online. A general structural issue exists in that the module consists of multiple repeated loops of questions

e.g. who provides help with each type of care need, hours and frequency of visit for each carer, pay for each carer etc. The people who would face the most loops would be the group least suited to deal with a longer web instrument i.e. those with complex care needs. It is unclear to what extent respondents would remain motivated to complete multiple loops accurately without the assistance of an interviewer. Respondent short-cutting behaviours or satisficing online could result in under-estimations in the amount of care being received, and inaccuracies regarding whom is providing each sort of care

Some of the questions within the module rely on data entry from either the Household Grid or the Health module. This means that the Social Care module would not, in its current format, work as a 'stand-alone' online module without the inclusion of relevant items from other questionnaire sections.

Therefore, if this module were to be asked online we would recommend that a truncated version of the care questions should be produced (i.e. with less reliance on loops and/ or less granularity on breakdowns of care provided). This could be supplemented by a more detailed set of questions during face-to-face waves.

Findings on Smoking and Exercise

The Smoking and Exercise questions were included in this Social Care section of the questionnaire. Both items were flagged as being at risk items in terms of socially desirable reporting. In this case the move to online reporting could increase data quality via a reduced risk of socially desirable reporting. If these questions are included in an online module we would also recommend that they be moved to the self-completion element during face-to-face waves.

There was some interest by stakeholders in asking the exercise questions more frequently, as exercise is a behaviour that is known to vary over time and that could be impacted on by both life events (e.g. injury, work transition) and season. Therefore, there would be an analytical benefit to collecting exercise data more frequently. The stakeholder consulted did not feel that smoking data needed to be collected more frequently as this behaviour tends to be more stable over time.

Findings on Social Participation module

The Wave 9 Social Participation Module in CAPI is a short set of questions that includes items on transportation e.g. access to private transport, lifts from friends and different means of public transportation, including community and hospital transportation. Questions on social contact are also included in ELSA W9 but are already part of the self-completion module (and thus automatically classified as low risk and not included in the below figures).

Dress rehearsal data indicates the average (median) length of the Social Participation module is very short- around one minute.

Table B.4: Findings on Wave 9 Social Participation module (N=12)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	0%	No major issues were detected with the module in terms of interviewer effects.
<i>Difficult questions</i>	17%	No major issues were detected in relation to difficult questions. One item was flagged as difficult for being long and one was flagged as difficult due to recall.
<i>Presentation effects</i>	75%	A high proportion of items included interviewer prompts and long lists of response options.

A relatively high proportion of the social participation (transport) questions were flagged as having presentation issues. These were due to interviewer prompts not currently viewed by respondents (e.g. definitions of different types of transport) and long lists of response options (e.g. reasons for not using transport). However, the presentation issues identified could be addressed by introducing respondent facing help screens and a Yes/No format on longer categorical response lists. Therefore, the questions could work in an online module with minimal risk if there is an interest in measuring transport issues more frequently.

It is possible that there would be an interest in collecting more frequent information on social contact if we consider the self-completion items in addition to the CAPI items. For example, it may be interesting to see if levels of social contact add any protection against negative outcomes during or directly after transitions in and out of work or adverse life events (as assessed via a trigger questionnaire). It would be feasible to use the current self-completion items on social contact in this way.

Findings on Work and Pensions module

The Work and Pensions module includes items on employment status, details of current job (or last job if retired), transitions between jobs/ retirement and motivations behind transitions. The module also includes questions current pension payments and future pensions plans (including predictions on whether pensions will be deferred or claimed and likely value of the pensions). This section of the questionnaire is relatively long (dress rehearsal data indicates the media length of the module as 9.5 minutes).

Table B.5: Findings on Wave 9 Work and Pensions Module (N= 141)

	%Qs flagged at risk	Main issues
<i>Interviewer effects</i>	23%	<p>A number of questions were flagged as sensitive e.g. salary, profit/loss</p> <p>A number of questions also ask respondents to give the percentage likelihood that they will receive more than a certain amount in pensions. It was felt that these questions may be sensitive but also at risk of 'positivity bias'.</p>
<i>Difficult questions</i>	43%	<p>A number of questions were flagged as difficult as they involve recall of financial details and calculations</p> <p>All questions on the percentage likelihood of receiving pensions of different values were flagged by coders as being 'difficult'.</p> <p>Open textual questions regarding current occupation are required for SOC coding.</p>
<i>Presentation effects</i>	23%	<p>There are a number of questions with five or more response options, for example questions on reasons for changing job/ retiring, types of pension, reasons for changing pension provider etc.</p> <p>There are some interviewer coded items (for example on whether respondents consult documentation) and interviewer facing prompts.</p>

There are both advantages and disadvantages to an online approach for this module. On the one hand respondents may be more comfortable providing details on salary and the profit/ loss of their business online. On the other hand, as the questions asked are quite detailed and quite long an interviewer may be beneficial to motivate respondents and prevent break-off. Respondents are encouraged to consult payslips when answering some questions and it is possible that interviewer presence encourages this behaviour.

Some specific questions may be problematic to administer online i.e. questions on percentage likelihood of receiving pensions of different sizes may be more problematic to transfer to an online format. It was felt that these questions may be sensitive but also at risk of ‘positivity bias’ whereby participants give higher likelihood of positive events in the presence of an interviewer compared to in a self-administration. It was queried whether this task would be understood by all participants without an interviewer providing help. It is also unclear what the analytical benefit would be of asking these types of questions on a more frequent basis.

Some of the questions in relation to work transitions may benefit from being asked from frequently i.e. using a modular online approach. For example, participants may be able to provide more accurate reasons for leaving work/ changing jobs if details are collected as soon as possible after the event, rather than being collected retrospectively in the face-to-face sweeps. This is one example where a ‘triggered’ online questionnaire may be useful, for example it would be possible to send an inter-wave online questionnaire out asking if the respondent has experienced any ‘transition’ in the last 6 months and then ask follow-up questions about this transition with groups who answer yes.

Therefore, an optimal approach may be to produce a short online questionnaire about work and pensions, focusing on the participant’s current situation and transitions, that can be administered online and keep some of the more detailed questions for the CAPI waves.

Findings on Income and Assets module

ELSA collects detailed information on income and assets. The questions include details of income from all sources (employment, pensions, rental income, benefits etc) and details of all assets held (different types of savings and investments). These questions are looped for each household member. Dress rehearsal data indicates the average (median) length of the module as 7 minutes. The maximum length recorded during the dress rehearsal was 16 minutes.

Table B.6: Findings on Wave 9 Income and Assets module (N=148)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	46%	Reporting of financial information is known to be sensitive. High value income and assets tends to be under-reported. Therefore, several items were classified as at risk of interviewer effects.
<i>Difficult questions</i>	67%	A high proportion of questions were flagged as difficult as they involve recall of financial details and mental calculations.
<i>Presentation effects</i>	32%	Includes some long showcards e.g. benefit types/ types of saving account. Includes some interviewer coded items including on interviewer estimates of whether data collected are broadly accurate or not.

The Income and Assets module is another example of module where there are both advantages and concerns regarding an online approach. On the one hand there are known sensitivities involved in

collecting data on income. Respondents may be more comfortable providing details on income, benefits and savings in a self-completion mode. On the other hand, as the questions asked are long and detailed, an interviewer may be beneficial to motivate respondents and prevent break-off. Respondents are encouraged to consult bank statements when answering some questions and it is possible that interviewer presence encourages this behaviour. Interviewers may also be beneficial in explaining the different types of benefit and savings account being described.

The income module relies on a number of looped questions based on respondents' prior answers. It is possible that short-cutting behaviours may be more prevalent online (i.e. where respondents reduce the number of responses they select in order to be routed to fewer questions). It is felt that this issue is most likely to happen with respondents who have multiple sources of income and who have multiple assets. This could lead to an under-reporting of income and value of assets in online administration in this group.

It could be possible to produce a truncated version of these questions for online use (e.g. just asking about income from work, or income from pensions) to supplement more detailed information collected during face-to-face waves.

Findings on Housing and Consumption module

The housing and Consumption module contains detailed questions on spending. It includes questions spending on housing, spending on all utilities, insurance, groceries, clothing, durables and so on. It also includes questions on home adaptations and how these are paid for.

The dress rehearsal data indicates the average (median) length of the module as 11.5 minutes (making it the second longest module, after the health module). The maximum length recorded during the dress rehearsal was 24 minutes. Table B.7 below summarises the main issues identified when coding these questions.

Table B.7: Findings on Wave 9 Housing and Consumption Module (N=198)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	6%	A small number of questions were coded as sensitive and therefore at risk of interviewer effects (e.g. mortgage and rent).
<i>Difficult questions</i>	35%	A relatively high proportion of questions were coded as being difficult e.g. items that required respondents to recall details of how much they spent on various items.
<i>Presentation effects</i>	34%	Some items were flagged as having some form of 'presentation' issue which could impact how they work online e.g. interviewer prompts and definitions not currently viewed by respondents This issue could addressed by introducing respondent facing help screens.

It was felt that the consumption questions could be asked online with minor adaptations with limited risk of mode effects occurring. A small proportion of questions were coded as sensitive and therefore at risk of interviewer effects. For example, it was felt that open items on mortgage and rent payment may be sensitive (in same way that asking for income data may be sensitive). Likewise, items on not being able to keep up with payments were classed as potentially sensitive

A relatively high proportion of questions were coded as being difficult e.g. items that required respondents to recall details of how much they spent on various items. Respondents are also

expected to recall differences in summer spending versus winter spending on all forms of fuel. Although multiple consumption questions were flagged as being difficult, we have little evidence as to whether interviewer presence would increase accuracy of reporting. Furthermore, some of the consumption questions may be made *less difficult* by more frequent administration. For example, instead of asking respondents to provide retrospective details on summer and winter spending, an approach could be devised where respondents are asked about ‘last months’ spending at various intervals throughout the year. This would allow for consumption questions to capture seasonal variations in spending more accurately rather than rely on retrospective reports based on season’s or average spending.

Although there is limited concern regarding the individual questions there are some concerns regarding the length of the module and whether participants would be motivated to spend time online providing these types of details. Therefore, if these questions are asked as more frequent online module it may be beneficial to ask them as a ‘standalone’ block (i.e. not to combine them with other modules) to online survey as short as possible.

Findings on the Expectations module

The Expectations module is a relatively short module in which respondents are asked to report their expectations for the future. Dress rehearsal data indicates the average (median) length of the module as 5 minutes.

In this module respondents are asked to give a percentage chance of certain events happening e.g. the likelihood of their having a life-expectancy above a set age or the likelihood of their needing residential care. There are a small number of questions on participants future expectations regarding their care needs e.g. who will provide this for them and how it will be paid for.

Table B.8: Findings on Wave 9 Expectations module (N=29)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	86%	The majority of the expectation questions were flagged as at risk from positivity bias.)
<i>Difficult questions</i>	79%	All likelihood rating scales in the module were flagged by coders as being ‘difficult’ questions as they require participants to be able to report perceived likelihood using a percentage.
<i>Presentation effect</i>	79%	The likelihood rating scales were also flagged as an unusual question format. A number of questions included interviewer prompts not currently viewed by respondents (e.g. residential care as opposed nursing care).

There is currently a query as to how well the expectations questions would work in an online module. Most of the expectation questions were flagged as at risk from positivity bias. As previously discussed, positivity bias is the phenomenon in which respondents are more inclined to give a more ‘positive’ response in the presence of an interviewer compared to in self-completion (Ye *et al* 2011). Positivity bias may be more prevalent in interviewer administered modes, so scores given could drop during online waves and peak in face-to-face waves. However, most of the literature on positivity bias discusses the issue in relation to agree-disagree scales and satisfaction type measures. There is room for further methodology work to see if the same issue would apply to rating/ expectation questions of this format. It may be useful to explore this issue further, potentially by including questions of this nature on Understanding Society’s mixed mode Innovation Panel.

From an analytical perspective it is not clear that there would be benefit of asking these questions at more frequent intervals. Coupled with the fact that an interviewer can provide assistance/ explain the question format, it may be that expectation questions are best retained in the face-to-face waves. If these questions are transitioned online consideration will need to be given to how the task is explained and how the questions are visually presented.

Findings on Efforts and Rewards module

The W9 CAPI questionnaire contains a short block of questions on efforts and rewards. These covered questions on volunteering, care giving and looking after grandchildren. Dress rehearsal data indicates the average (median) length of the module as 2.5 minutes. The maximum length recorded during the dress rehearsal was 8 minutes.

Table B.9: Findings on Wave 9 Efforts and Rewards module (N= 40)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	10%	The module includes a small number of agree-disagree scales that could be prone to positivity bias (e.g. satisfaction with volunteering and care-giving).
<i>Difficult questions</i>	25%	The module contains a small number of questions that were coded as difficult as they required computation. These included questions on average hours of care provided for grandchildren during different periods e.g. weekdays during term time, weekends and school holidays.
<i>Presentation effects</i>	20%	Some questions included longer lists of answer options. Some included spontaneous interviewer codes e.g. a code for 'one-off activities' that was not visible to respondents. These issues could potentially be resolved by reversing the order of answer options and by making all answer options visible to respondents in future survey administrations.

It would be feasible to adapt the Efforts and Rewards questions into an online ready module with minimal effort. There may also be some benefit of asking these types of questions at more frequent intervals as these behaviours are likely to fluctuate over time and they may be linked with other outcomes of interest e.g. health and wellbeing and work transitions. These questions could be asked after a life event 'trigger' for example retirement, taking on a caring role, after the diagnosis of a health condition and so on.

Findings on Psychosocial module

ELSA contains standardised measures on psychological wellbeing including scales on Quality of Life (CASP-19) and the ONS4 measures on subjective wellbeing. These items are included in the paper self-completion element of the survey and so were not included in the desk review. We would not expect measurement related mode-effects to occur when transferring questions from a paper self-completion to an online mode.

In addition to the self-completion psychosocial measures the CAPI contains a small number of additional questions on psychosocial wellbeing. These include binary (Yes/No) questions on mood and depressive symptoms over the past week. Dress rehearsal data indicates the average (median) length of the module as 2 minutes. The maximum length recorded during the dress rehearsal was 5 minutes.

Table B.10: Findings on Wave 9 Psychosocial module (N=11)

	% Qs flagged at risk	Main issues
<i>Interviewer effects</i>	64%	Questions on negative affect/ mood could be prone to social desirability bias.
<i>Difficult questions</i>	0%	No questions were coded as being difficult.
<i>Presentation effects</i>	0%	All questions were coded as being at risk for presentation effects/ All questions were short and simple. The binary response options (Yes/No) would be well suited for online administration.

It was felt that the CAPI psychosocial wellbeing questions could be prone to social desirability bias. For example, it was felt that respondents may be less inclined to report negative emotions and depressive symptoms to an interviewer. We believe a move to online reporting could increase data quality, for example if respondents disclose negative affect that has previously been under-reported.

Therefore, it is felt that psychosocial questions included in ELSA could work well as part of an online modular questionnaire. We would recommend that all psychosocial measures be administered in a self-completion format (including those currently in CAPI) to minimise the risk of fluctuations in reported wellbeing being confounded by different survey modes between waves.

In general, the psychosocial questions could benefit from more frequent administration as mood and wellbeing are likely to fluctuate over time. Likewise, variations in wellbeing may be of interest to assess after a trigger questionnaire (so the questions are asked as soon as possible after a transition or a key-life event).

Table B.11 Summary of desk-review of ELSA Wave 9 questionnaire - all modules

Module	No. of questions coded	Median length (mins)	Interviewer effects (%)	Difficult questions (%)	Presentation effects (%)	Benefits and risks of online/ more frequent data collection
Health	167	12	27	35	40	<ul style="list-style-type: none"> Improved reporting of sensitive conditions e.g. mental health Opportunity to collect information on conditions that are difficult to recall retrospectively e.g. respiratory infections and UTIs Opportunity to capture more granular data on pain BUT issues with module length and use of feed-forward data. Some items (e.g. rating scales) of higher risk of interviewer effects. This risk could be eliminated if they were asked as self-completion as standard.
Female Health	24	No data	46	67	4	<ul style="list-style-type: none"> No clear benefits of online/ more frequent administration
Social Care	94	4	41	47	42	<ul style="list-style-type: none"> Exercise questions (currently at the end of the social care module) would be of interest in terms of more frequent administration as this behaviour varies over time. BUT Social care questions may be more problematic to transfer online due to repeated loops and reliance on health module and household grid data for routing and text-fills.
Social Participation	12	1	0	17	75	<ul style="list-style-type: none"> Possible interest in more frequent collection e.g. as may change over time and be linked to transitions/ health outcomes. Some issues with question presentation noted but these could be addressed with minor changes in question design (respondent facing help screens and Yes/No formats).
Work and Pensions	141	9.5	23	43	23	<ul style="list-style-type: none"> Questions on work transitions may benefit from being asked from frequently e.g. improved recall of dates and reasons for transitions BUT issues with module length and loops and whether there would be equivalent motivation online to answer questions accurately. Risks of mode effects caused by positivity bias for certain questions (e.g. questions on likely size of future pension). Optimal solution may be a short 'triggered' version of the module that is given if respondents report a transition, e.g. changing job or retiring.

Module	No. of questions coded	Median length (mins)	Interviewer effects (%)	Difficult questions (%)	Presentation effects (%)	Benefits and risks of online/ more frequent data collection
Income and Assets	148	7	46	67	32	<ul style="list-style-type: none"> • Questions income may benefit from being asked from frequently e.g. if respondent suffers a financial shock or a change of income. BUT issues with looped questions and whether there would be equivalent motivation online to answer questions accurately. • Again, an optimal solution may be a short 'triggered' version of the module that is given if respondents report a transition, e.g. changing job or retiring with more detailed questions being asked in CAPI waves
Housing and Consumption	198	11.5	6	35	34	<ul style="list-style-type: none"> • Some of the consumption questions may be made less difficult by more frequent administration e.g. questions on seasonal spending. • BUT issues with overall module length and looped questions. accurately.
Expectations	29	5	86	79	79	<ul style="list-style-type: none"> • No clear benefits of online/ more frequent administration • Questions at risk of positivity bias so scores could peak in CAPI waves and drop in online waves. • An interviewer would be beneficial to explain the question format
Efforts and Rewards	40	2.5	10	25	20	<ul style="list-style-type: none"> • Some benefit of asking at more frequent intervals as behaviours mentioned are likely to fluctuate over time and they may be linked with other outcomes of interest e.g. health and wellbeing and work transitions. • Lower risk module in terms of mode effects.
Psychosocial	11	2	64	0	0	<ul style="list-style-type: none"> • Psychosocial questions could benefit from more frequent administration as mood/wellbeing are likely to fluctuate over time. • Questions are at risk of interviewer effects- this risk could be ameliorated if psychosocial questions are included in self-completion module as standard (including in CAPI waves).