



Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Eighteenth Meeting**

Geneva, 28 - 30 September 2016

Item 2 of the provisional agenda

Assessing costs and benefits of censuses**Using cost modelling in the 2021 England and Wales Census
to inform design decisions¹****Office for National Statistics, United Kingdom***Abstract*

Understanding the costs of the census is a crucial part of planning and managing a successful census. This paper will provide some background information on the use of the cost model in planning for the 2021 Census and then explore some specific aspects of the 2021 Census design that the cost model has been used to identify risk and refine the census design

I. Introduction

1. With an increasing demand on public sector resources there is a greater need to understand and demonstrate the costs and benefits required to undertake a census. A clear requirement and demonstration of benefits against a well presented and robust assessment of the costs will increase the chances of success of securing funding for the census, or particular aspects of the census.

2. A key element of census planning and management is the development and maintenance of a cost model – or a detailed estimate of costs (UN, 2016).

¹ Prepared by Cal Ghee.

3. For the purposes of this paper a cost model is a simplistic model that enables the planning of costs by spend item (e.g. field staff pay, printing, etc) to be estimated by year over the length of the census life-cycle. The model may also identify those costs that are fixed and variable and enable the parameters around the variable costs to be changed. Using different parameters in this way may help to understand the different impacts if assumptions in operational performance do not meet expectations – for example if the proportion of online responses is less than planned what is the additional cost of paper data capture?

4. For the 2021 Census in England and Wales, we have prepared a robust cost model which covering the entire programme and which provides the estimated yearly costs for planning, developing and undertaking a census operation. Some of the main reasons and uses of developing the cost model are outlined in this paper. However, the paper focuses on some specific examples where the cost model has been used to identify areas where modelling costs have been instrumental in informing decisions, directing further development and testing and assessing the cost impacts of different quality assessments.

II. Background – 2021 Census in England and Wales

5. The planning and conduct of the 2021 Census in England and Wales is the responsibility of the Office for National Statistics (ONS) in partnership with the Welsh Government (WG). A main goal for ONS however, is for the production of consistent UK-wide statistics. ONS is working closely with the National Registers of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA) who are responsible for the development and conduct of the censuses in Scotland and Northern Ireland, respectively.

A. Drivers for change and design principles for 2021 Census design

6. The initial high-level design for a 2021 Census is informed by:

- lessons learned and successes, where appropriate, from the 2011 Census
- developments and lessons learned in international census taking
- requirements from the user community about the types, quality, frequency and detail of outputs required
- changes in technology, in particular the opportunities offered by the internet, and changes in the propensity for the public to interact with government, enabling a move away from a traditional paper based census
- improvements in administrative data sources (such as the patient register and tax and benefits data) giving potential for increased use in the production of statistics
- a desire to maximise the benefit from the investment in the Census by developing systems and services so they can be used across the wider statistical system, and
- the continued and ongoing need to make the most effective use of public money

7. The design will evolve as the programme develops and the different design elements are tested; initially independently of each other, and then (at the appropriate time) together. We have set design principles that will guide the development and finalisation of the design for 2021. These include:

- utilise the elements of the 2011 Census that worked well and are still relevant
- embrace new technologies and methods wherever possible
- design the operational and statistical processes for online first
- make it as easy as possible for the public to respond independently, whilst ensuring that the design complies with recognised standards on equality and accessibility
- seek to minimise the respondent burden on the public – the length of time to complete the questionnaire (online or paper if applicable) should be about the same, or less, as that of the 2011 Census
- protect, and be seen to protect, confidential personal data
- testing the census design iteratively to assure us and stakeholders of the underlying system, processes, and security of the overall design
- attempt to get a response from every person and household in England and Wales
- the operation and statistical methods employed should be developed to deliver the highest quality for the population estimates by age and sex at the local authority level
- overall seek to achieve as good or better quality results than in the 2011 Census
- maximise the use of administrative data in all areas of the operation
- estimate and adjust the results to account for over and under-enumeration as in the 2001 and 2011 Censuses
- for every question (apart from any voluntary question, eg religion) where no response is provided, estimate a response to provide a complete set of outputs
- make the first results available more quickly than results from the 2011 Census and also seek to complete the full suite of outputs (still to be defined) more quickly than those of the 2011 Census.

B. Main areas of change in the 2021 Census

8. ONS is aiming to maximise online responses in 2021 with a current planning assumption of 75% online. The move to a predominantly online census has many opportunities for all aspects of the 2021 Census, from the design of the questionnaire to the management of the field operation and the subsequent processing and production of outputs. This section provides an overview of the main design features and, where relevant, lists some of the opportunities for particular aspects that will be considered when finalising the design for the 2021 Census

1. Accessing census questionnaires

9. In previous censuses a paper questionnaire was delivered to every household. As a result of the 2021 Census being predominantly online, the way households are presented their census questionnaire will be different. The 2011 Census online questionnaire was accessed using a 20 character authentication code printed on the paper questionnaire that was posted to households, and provided a unique link between a response and the address. Work is underway to understand the most effective way of contacting respondents in a predominantly online solution.

10. To encourage online response, we need to ensure that services are in place that take account of respondents who would like to complete online but are unable to. Understanding our respondents and how they wish to interact with a census collection exercise, based on an understanding of interactions with other government services, is vital to achieving the required response rate. Current research takes account of assisted digital requirements to meet UK government guidelines and best practice.

2. Target population and digital inclusion

11. A primarily online census introduces new challenges. There are always likely to be hard-to-count populations who are at risk of low levels of engagement or response. These are by their nature specific and require special attention. These groups were identified and prioritised in the 2011 Census. We will build on this successful strategy for the 2021 Census, ensuring that any emerging groups are included and engaged with, at an early stage, to understand their needs and explore appropriate ways to achieve higher levels of response.

12. In addition new digital services need to meet the government's digital by default standard and provide services that are so straightforward and convenient to use that all those who can use digital services will choose to do so, while those who can't are not excluded. An estimated 10 per cent of the adult population (not households) may never be able to gain basic digital capabilities because of disabilities or basic literacy skills (HMCO, 2014). Digital exclusion typically affects some of the most vulnerable and disadvantaged groups in society. The design and operation of the census will therefore need to take particular account of the requirements of these individuals.

3. Follow-up of non-responding householders

13. As with previous censuses, we will use different strategies to contact non-responding households and encourage them to take part. This process will involve the use of reminder letters and household visits from census collectors. As with the 2011 Census, there will also be support available in local venues such as libraries, drop-in centres and religious centres.

14. Collectors will make multiple visits to non-responding households to persuade non-responders and offer support where it is needed. If an address has been incorrectly included, they will provide verification before removing from the system. As time progresses, the emphasis would shift from support and encouragement to the legal requirement for completing the census and the penalty for not doing so. Later on, a small, separate field force will manage more persistent non-compliance.

4. Operational management

15. Collectors and their managers will be provided with up to date information about which households have not completed a questionnaire. With the 2011 Census, this approach was used to enable managers to effectively refine and target the follow-up operation, flexibly deploying staff to the areas with the lower response rates. A greater online collection will also enable up to date information about the types of populations (eg students) that have not responded. This could be used to further target publicity or community engagement activities. The prioritisation of collector resource, publicity and community engagement will be important factors in minimising variability in response rates between areas and also between population sub-groups.

5. Data processing

16. A higher proportion of online completions will also provide opportunities to improve the quality of information collected and reduce the amount of processing to validate and correct incomplete or inaccurate responses. As shown in the 2011 Census, the quality of online completions was significantly better than that of paper questionnaires. There were substantially lower rates of question non-response for internet returns (in comparison with the Census Quality Survey), and internet responses were generally more accurately completed than paper responses.

17. A substantial online response also provides opportunities for the potential delivery of census outputs earlier through greater expediency and accuracy in the processing of the census information, compared to a predominantly paper operation. We aim to exploit the early availability of data and publish results earlier.

6. Census outputs

18. For 2021 we will continue to seek ways of making outputs available more quickly, while maintaining the high levels of quality and confidentiality of the information provided by the census. We will be reviewing the way that we process, protect and disseminate a range of census outputs to ensure that we continue to meet evolving user needs..

III. Developing a cost model

19. In order to plan for the effective management of the costs of the programme, ONS has developed a robust cost plan that covers all activities within the programme, and is being controlled and refined as the design of the 2021 Census service model is developed. The cost plan is underpinned by several working assumptions and dependencies, which are regularly reviewed and appraised to ensure the forecast out-turn costs of the programme are known and consistently available. This approach also ensures that ONS can assess the costs of changes (to the assumptions and dependencies) in the whole, with all implications assessed, making analysis of different options more straightforward.

20. A key starting point is building on the costs of the previous census. This provides a wealth of information from which to start to and compile costs estimates or at least the key components that need to be included; likely areas of change (for example the number of households and people who need to complete the questionnaire); and areas where the previous census may have been under (or over) funded.

21. The main areas of spend from the 2011 Census are shown in Figure 1 with the spread over the census life-cycle shown in Figure 2 (from ONS, 2015).

Figure 1 - 2011 Census costs by component activity 2003/04 to 2015/16

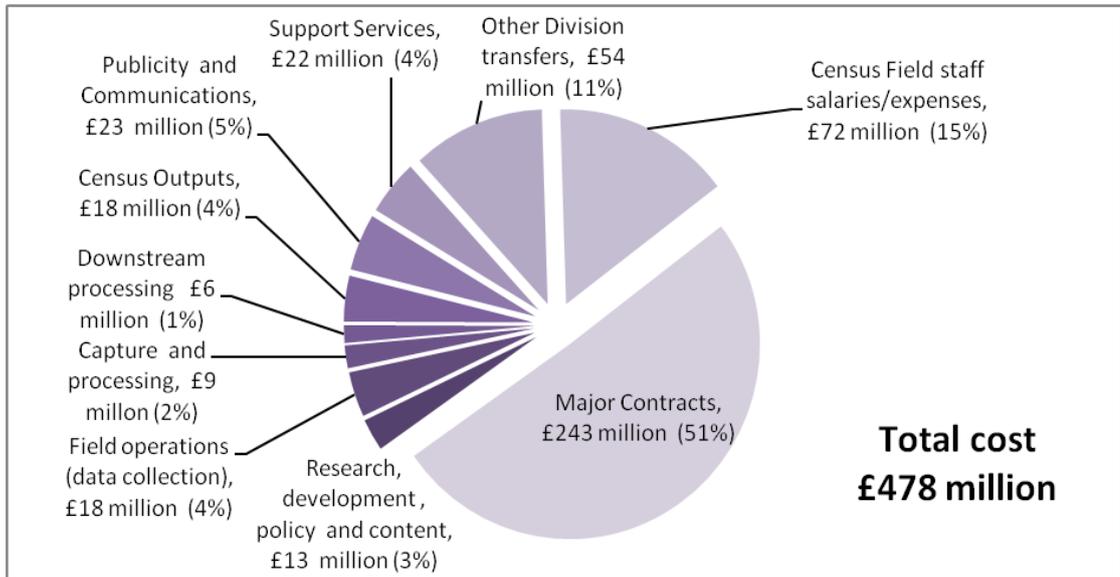
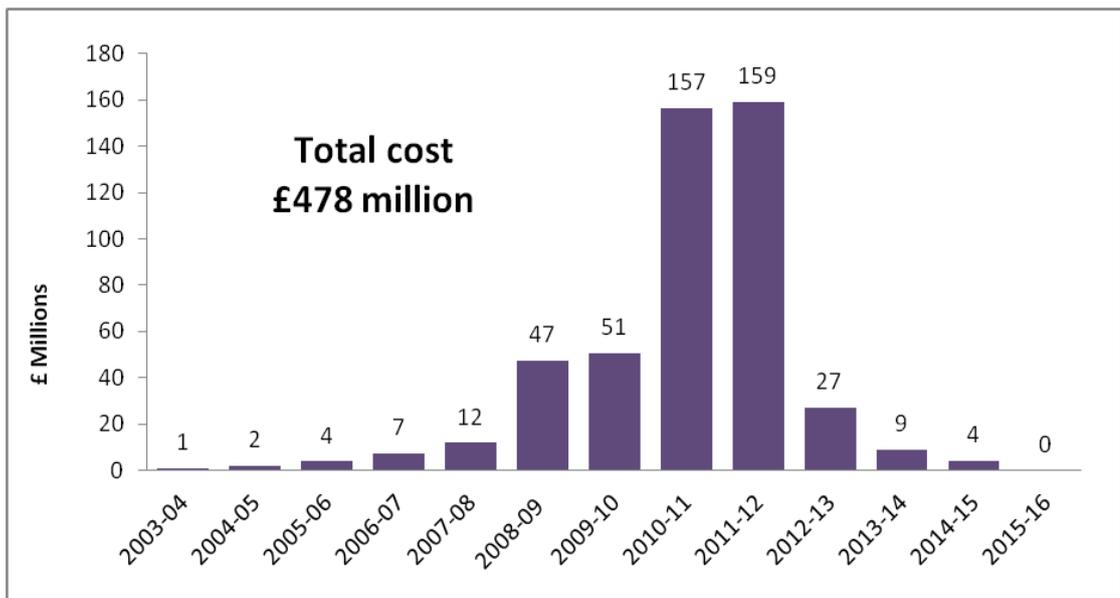


Figure 2 - 2011 Census costs by year 2003/04-2015/16 (rounded to nearest million)



22. In developing the 2021 Census estimate of cost the following were key areas of change or uncertainty that needed to be addressed:

- Likely spend profile by year over the life-cycle of the census – and how did this fit with the Government’s commitment to review spending after each election, scheduled in May 2015 and May 2020;
- The level of assumed online take-up; with the move to an online census ONS has a working assumption to encourage 75% of households to respond online. However, that is a significant shift from the 16% achieved in 2011 and is higher than yet achieved in other countries. The assumed proportion here was critical as it influenced some other key cost drivers, such as printing and postage volumes and the amount and methods of paper data capture

- The number and nature of the questions: some questions can be automatically captured and coded, others require manual intervention.
- What levels of quality, primarily response rates, do we aim to achieve? Setting these early is important for a number of reasons (UNECE, 2015), but understanding the levels of quality/response rates is critical as it drives so much of the operation and in particular one of the most expensive aspects of the census operation the field operation.
- The level of self-response we expect to achieve and so the level of response that will require more expensive assistance and persuasion from field staff
- The impact of greater in-house development of IT systems rather than outsourcing – this will bring costs forward as they are incurred in system development not when final products are delivered
- Impact of different procurement approaches (a government requirement to disaggregate contracts to increase opportunities for smaller businesses) – this may bring cost savings and innovation in systems but additional integration costs

IV. Applying the cost model in developing the 2021 Census design

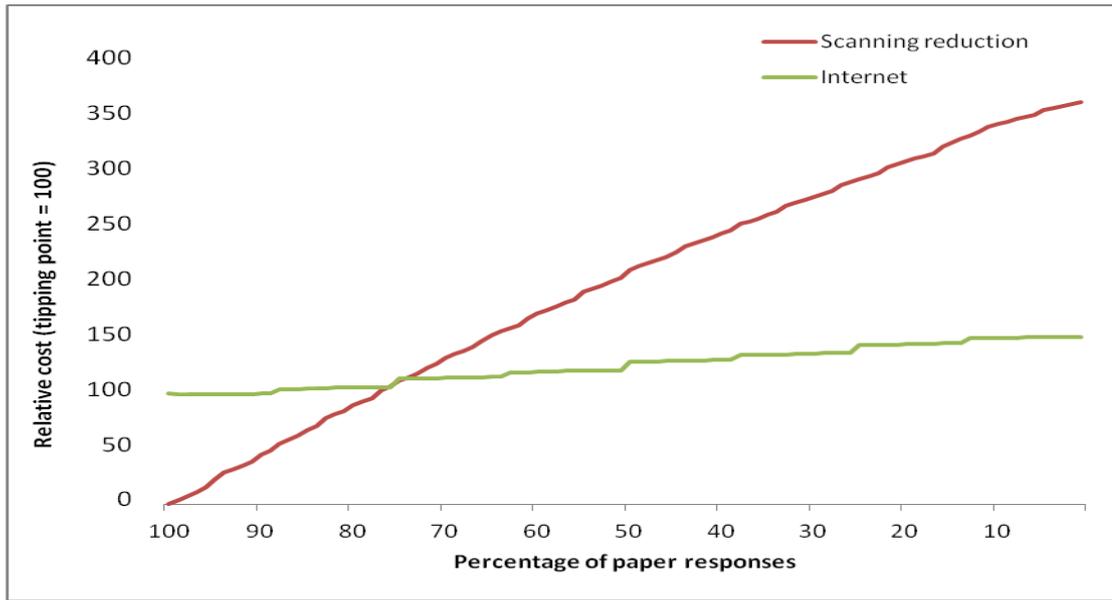
23. A good understanding of cost is therefore seen as being essential to conducting a successful census; it enables areas of significant cost to be identified early and plans put in place to continually review costs for the most effective and efficient use of the resources. Even at this stage of the programme, the cost model is being continually used to assist with making design decisions; identify development areas and where technology could reduce cost; and, to assess the impact on different quality targets. An example of each of these applications follows.

A. The 2011 model and the online provision tipping point

24. To demonstrate the use of a cost model, in the 2011 Census internet completion was offered as an alternative to paper completion on the Census. Work was commissioned to determine the point at which offering this option became cost efficient (ONS, 2013). Figure 3 below shows this tipping point.

25. In Figure 3 the scanning reduction line shows the reduced costs associated with developing and running a scanning system to support the percentage of forms indicated by the x-axis. The internet line shows the costs associated with supporting internet responses. The point at which developing an internet solution becomes cost effective is the point at which the two lines cross. The net saving for the combined solution is the difference between comparable points on the two lines. In the scenario above the internet solution becomes cost effective at around 25%.

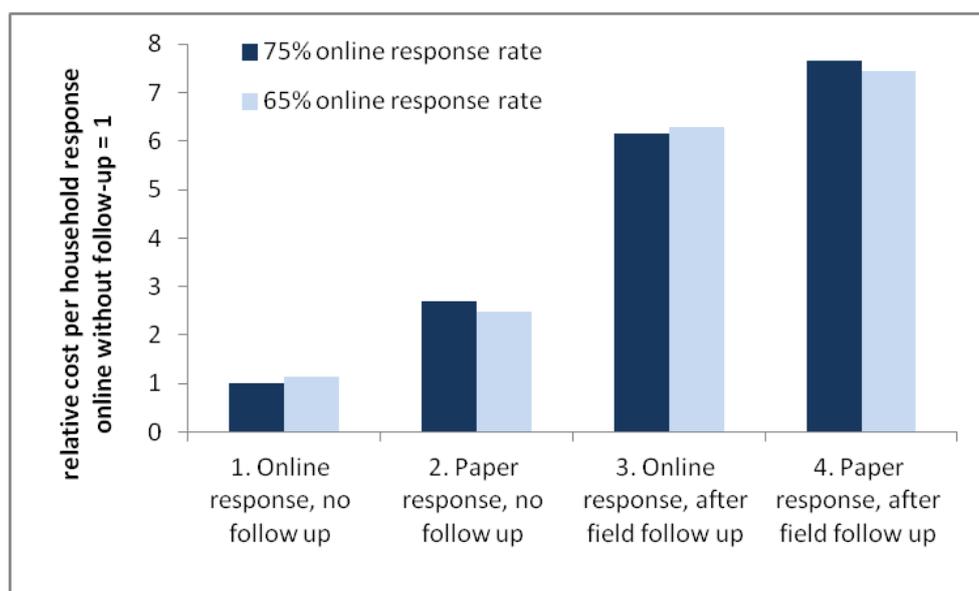
Figure 3 – The tipping point where online provision becomes cost-effective



26. This tipping point (25%) along with other supporting information helped to set our targets for online completion in the 2011 Census. Additionally this information was used in our contingency planning, guiding how much extra resource and data capture capacity might be needed should the online target not be met.

B. Online-first Census - demonstrating the need to target some paper-first

27. Given the opportunities afforded by an online census, we want to maximise online response rates. However, ensuring the overall response (whether online or paper) meets quality targets and is cost efficient is still a primary consideration for the 2021 Census. Figure 4 shows the relative cost per household response online versus paper, and also self-response versus needing field follow-up. It is clear that a paper response with no follow up is cheaper than any non-response follow up visit. The figure also shows the variability in the unit costs as the channel mix changes, still showing this clear cost difference.

Figure 4 – Relative costs per response per household

28. Understanding these costs has helped focus our design and development work and in particular the importance of maximising self-response and the cost savings incurred without having to employ field staff to undertake follow-up visits. In addition, this also shows a paper self-response (response before any follow-up activity) is 60 percent cheaper than an online response with field follow-up, highlighting the importance of maximising self-response. Therefore, one area of the design that we are focusing on is having enough information that will enable us to predict the types of people (or areas) who are more likely to self-respond if given a paper questionnaire from the start. This is one of the key aspects that our 2017 Test will inform.

C. Field follow-up - assessing the cost of different response rate targets

29. In order to inform the decision on the number of collectors that may be needed to meet a targeted response rate, a simulation model being developed as an input into the wider 2021 field operations cost model currently includes the ability to, for example, specify different census officer start/end dates, working hours or calling patterns (visiting at different times of day/week). It can also change aspects of follow-up, such as stopping all visits when a threshold return level is reached or varying the maximum number of visits per non-responding address.

30. The model can also be used to consider the cost effectiveness of different quality strategies. The accuracy of the Census estimates relies on consistent, relatively high response rates everywhere. This requires putting more resource into challenging areas, which costs more. There will be trade-offs that can be modelled between seeking to maximise the national response rate and getting every area above a set minimum. In 2011, we set – and met – targets for both but the local targets were more important to the accuracy of the outputs.

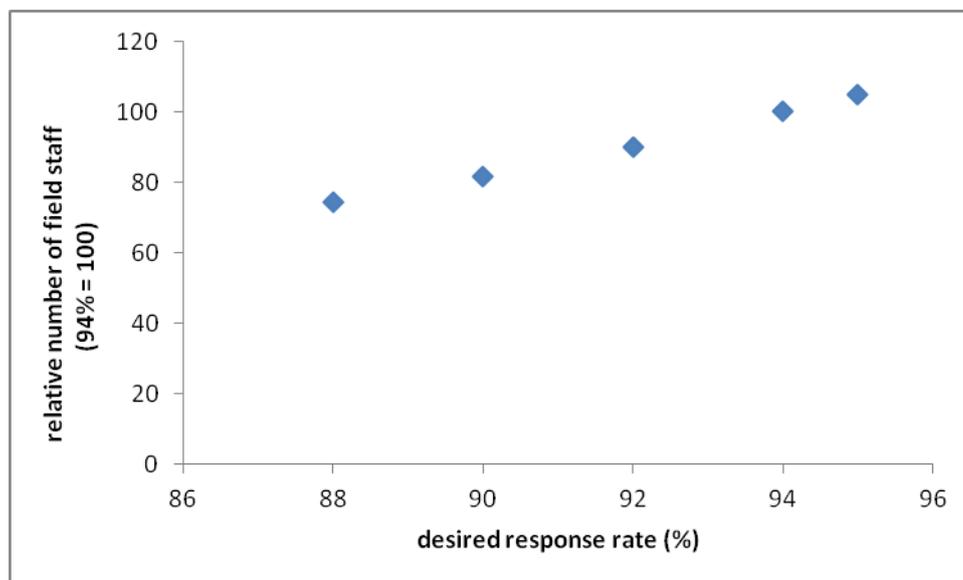
31. The estimated numbers of collectors are then fed into the cost model to see the effect on total costs. The models are based on a number of assumptions. These assumptions are being updated as we do test and get up-to-date information from other sources.

32. Use of the models is firstly to aid decision making, for example to demonstrate diminishing returns and optimum design decisions. But they can also be used ‘live’ to

demonstrate the impact on contingency plans if early response levels indicate a scenario we had predicted so that we can instigate a suitable cost-effective intervention.

33. One example of this demonstrates the number of collectors that would be needed to get an expected overall response rate. This uses assumptions including number of visits per hour, effectiveness of follow-up visits, length of follow-up period and variability of success depending on the type of area. Figure 5 below demonstrates this for a range of return rates.

Figure 5 – estimated relative number of field staff to achieve a given response rate



34. Note that the top return rate assess was 95%, as given rules we set in model, starting with our expected level of self-response and with a maximum of 10 visits to every non-responding address, we'd only get to 95.4% response.

D. Assessing the likely costs and benefits of different field follow-up start dates, and the likely benefits of reminder letters

35. We are conducting a number of test in 2017 which will enable us to assess whether the additional cost of starting field follow-up earlier gives value for money – i.e. do we wastes visits by calling on people who would have self-responded during that time anyway or reduce the overall amount of work we undertake during the whole operation. In 2011, we started non-response follow-up ten days after Census Day as evidence from 2001 showed this was the point when unprompted responses tailed off. It could be that with a predominantly online Census, this point is earlier (as there is not a lag between the householder posting their return and ONS receiving it).

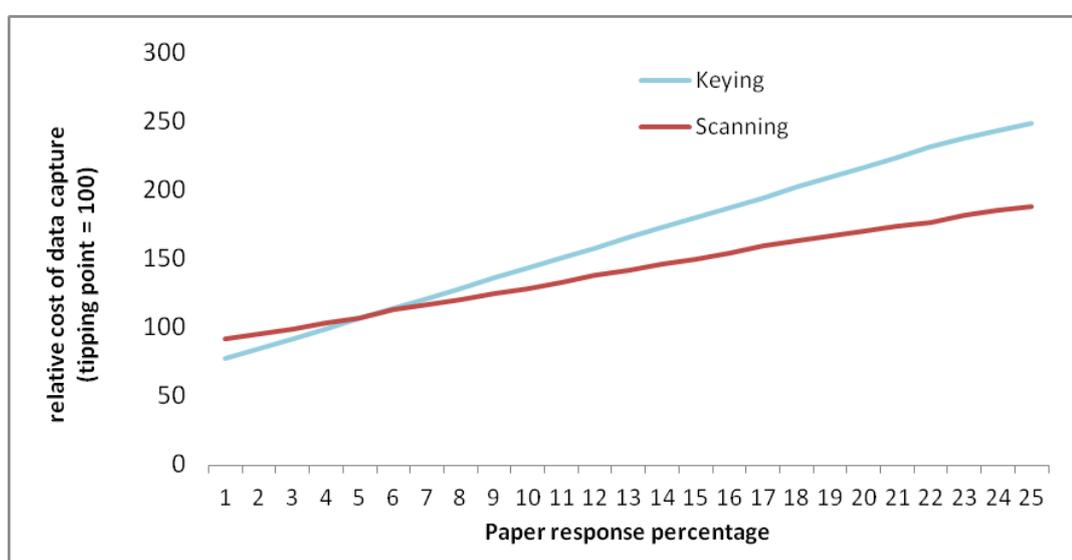
36. We are also looking at costs per visit, and costs per response conversion, across different types of areas, and the effectiveness of reminder letters across different types of areas. From these we are looking to see if we can effectively predict where more visits are effective, and where in some areas we don't need to visit but (much cheaper) reminder letters are effective.

E. Deciding on the method of paper data capture

37. For 2021 the internet will be offered as the first response mode. What then becomes of interest is how best to deal with the paper responses. The cost model assumes that systems will be built from scratch. Having a pre-existing scanning capability with no set-up costs would obviously make any comparisons meaningless, since scanning would always be the most cost-effective option. As this is not the case for the 2021 Census, we wanted to explore the costs of setting up a traditional paper data capture operation using OCR/OMR, and auto/manual coding compared to keying in responses to the paper questionnaire into the online questionnaire.

38. Figure 6 shows the tipping point at which keying becomes less cost effective compared with scanning in a joint online/paper solution. The chart below indicates that a 5% paper response is the point at which setting up a scanning solution becomes cost effective compared to keying responses (ONS, 2013).

Figure 6 – The tipping point where scanning becomes more cost effective



39. It is very difficult to accurately predict the proportion of online and paper responses. But what this modelling shows is that the savings of going with a keying model are very small compared with significantly additional cost involved in keying if our estimates are more than about 4 percentage points higher than expected. Given that we have decided that a traditional data capture operation would be more cost effective.

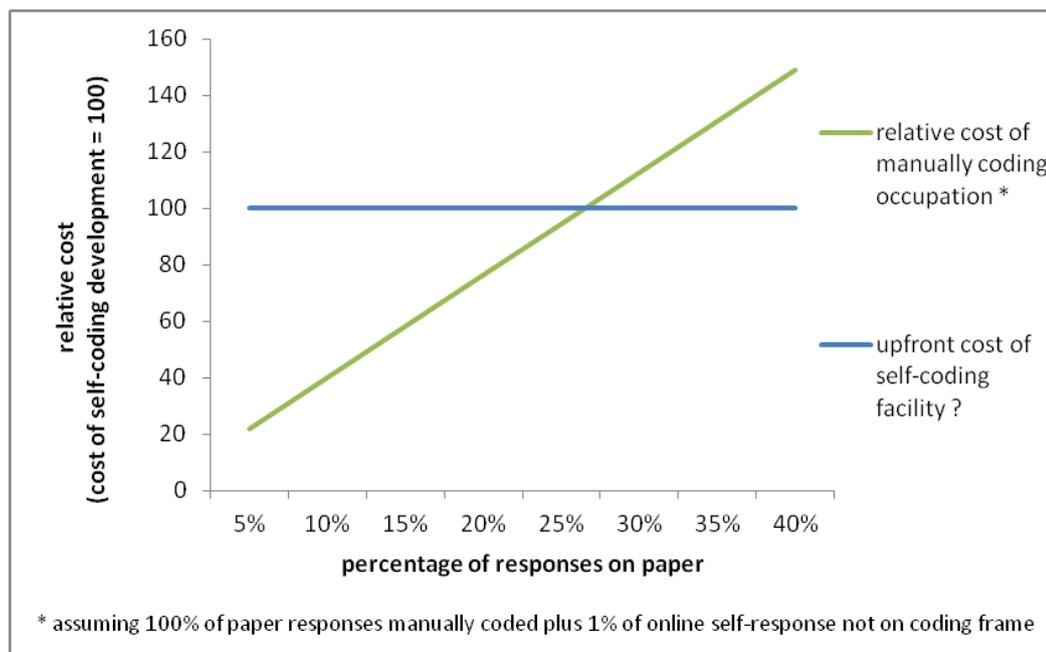
F. Developing the opportunity to reduce coding costs

40. As we move to primarily online collection, we have the opportunity to reduce the coding costs we had in 2011 (ONS, 2016). Although small compared to the overall budget for the 2011 census, the costs relating to all the different coding processes were significant, at £9 million (2% of the total). Coding of questions with write-in responses accounted for much of this cost. Occupation and industry combined accounted for 92% of the overall coding costs. Significant cost savings that can be made if the amount of manual coding could be reduced. The development of improvements will have cost implications, but this up-front investment will then be available across all ONS data collection activities.

41. Developments could include self-coding by respondents², automatically deriving distinct industries from occupations where appropriate, or putting in clarification questions only where needed³. Expert coders may still be needed to code the residuals, for non-codable online responses and any paper responses, but the number of these residuals would be far fewer.

42. Figure 7 shows a potential⁴ tipping point at which investment in a respondent self-coding system would pay for itself by saving the cost of manual coding of the occupation question. Once developed, the vast majority of the online response occupation coding would be done by the respondent.

Figure 7 – A potential tipping point where self-coding justifies the development cost



43. It is worth noting that there was a large amount of auto-coding in 2011 (75% of occupations, 60% of industries). It was only the non auto-coded responses that needed manual coding and incurred most of the costs. If these potential approaches simply cover what was auto-coded, the benefit won't be as large. Consequently, other methods of reducing the costs of coding include improving auto-coding, but as rare and new occupations emerge some extent of manual coding is always likely to be required.

44. However, to offset against this, it should also be noted that such an investment in a self-coding facility (and the creation of the vast look-up table that it utilises) would also be of benefit across the wider statistical system.

² For example, respondents could write in their occupation or industry to trigger the instant display of a predictive list of codable options (also referred to 'semantic matching' and similar to a Google search). However, there is a risk of a reduction in quality compared with expert coding, and there is also an additional respondent burden must be taken into account in an assessment of suitability.

³ For example if someone has a clearly defined occupation no further clarification would be needed, but if not codable from the first question, subsequent questions could appear to the respondent.

⁴ Note that these are theoretical figures used to demonstrate the concept.

G. Assessing how much we use paper responses in processing

45. We are looking at what balance of online v paper responses would we need before we could reduce how much we use paper responses, ranging from not keying in some variables but imputing instead, or even not using paper responses at all. There are considerable gains to be had around faster outputs and cheaper processing, but testing needs to be done regarding quality concerns, especially to what extent there would be any bias in excluding those paper responses.

46. Part of this approach can be to hold off coding of some paper questions until later, which may have potential cost-savings in having a more balanced resource level, and keeping services and costs in-house. Other options include imputing write-in responses from paper, and only manually code a proportion of responses to validate imputation outcomes. There are obviously substantial timing/cost/quality trade-off questions to be addressed with these approaches.

V. Next Steps

47. The model will continue to evolve as we move closer to the 2021 Census. Testing and development will help to refine our census design and will provide greater certainty on costs in some aspects of the design; some systems and services will be procured which again will provide greater cost certainty. The examples above on the follow-up wave model, improvements to coding and to what extent we use paper responses in 2021 are examples of developments we are working on now.

References

(HMCO, 2014) Government Digital Inclusion Strategy policy paper. Cabinet Office, 2014.

<https://www.gov.uk/government/publications/government-digital-inclusion-strategy/government-digital-inclusion-strategy>

(ONS, 2013) ODP 4.4 Report on research into the tipping point at which keying becomes more financially viable than scanning. Ede, M. Office for National Statistics, 2013. (Unpublished report)

(ONS, 2015) 2011 Census General Report. ONS, 2015.

<https://www.ons.gov.uk/census/2011census/howourcensusworks/howdidwedoin2011/2011censusgeneralreport>

(ONS, 2016) Working towards online self-identification of occupation for the 2021 Census for England and Wales. Waruszynski, P, Office for National Statistics.

In: Are We There Yet? Where Technological Innovation is Leading Research Proceedings of the Association for Survey Computing, Volume 7. Edited by T. Macer et al. Association for Survey Computing. 2016.

<http://asc.org.uk/publications/asc2016/16%20Paul%20Wanuszynski.pdf>

(UN, 2016) Handbook on the management of population and housing censuses. UN, 2016.

(UNECE, 2015) Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing, UNECE, 2015. Chapter IV, Quality Management.

http://www.unece.org/fileadmin/DAM/stats/publications/2015/ECE_CES_41_WEB.pdf