Jun 25th, 9:00 AM

Improving our understanding of user trial samples using survey data

Anya Petyaeva
University of Cambridge, United Kingdom

Joy Goodman-Deane
University of Cambridge, United Kingdom

Mike Bradley
University of Cambridge, United Kingdom

Sam Waller
University of Cambridge, United Kingdom

John Clarkson
University of Cambridge, United Kingdom

Citation

Petyaeva, A., Goodman-Deane, J., Bradley, M., Waller, S., and Clarkson, J. (2022) Improving our understanding of user trial samples using survey data, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), *DRS2022: Bilbao*, 25 June - 3 July, Bilbao, Spain. https://doi.org/10.21606/drs.2022.676

Improving our understanding of user trial samples using survey data

Anya Petyaeva*, Joy Goodman-Deane, Mike Bradley, Sam Waller, John Clarkson

University of Cambridge, UK

*corresponding e-mail: ap2160@cam.ac.uk

doi.org/10.21606/drs.2022.676

Abstract: User research such as user trials provides valuable information on users and how they respond to interfaces in practice. However, it can be hard to ensure a representative sample. We propose a methodology to improve the understanding of a sample's skew and to identify the characteristics of those who are missing by comparing the sample with survey data. This can improve the interpretation of results and inform further recruitment to improve the sample. We provide a case study of this methodology in practice. 30 participants were recruited using quota sampling with significant effort to obtain people with low technology experience. Nevertheless, comparison with UK survey data on technology experience, competence and attitudes identified four key groups of people not included in the sample, covering 29% of the population. We discuss how these missing people would likely respond on the tasks, based on the characteristics of similar people in the survey.

Keywords: User Trials, Survey Data, Sampling, Inclusive Design

1. Introduction

User research is vitally important for understanding the end users' needs and how a product or service will perform with a particular target population. In particular, user trials can provide valuable insights into people's interaction, experience and use of a particular design (Norman, 2002; Stanton et al., 2013). Potential problems and successful features can be identified, and they may provide insight into possible solutions to challenges (Green & Jordan 1999).

However, it can be hard to obtain a representative sample, particularly when designing for a diverse target population. For example, this is the case when designing products, services or systems that are intended to be used by the population as a whole, such as banking, healthcare, government and public transport services. Participant recruitment can be challenging due to limitations in time, cost, opportunity and resources. Samples can be skewed due to participant self-selection, and individual differences among volunteers can create bias (Berry, 1983: Sutton & Edlund, 2019).

Using a skewed sample misses the experiences of sections of the target population. This is important because different people respond to interfaces differently depending on a wide range of variables. These include their previous technology experience and consequent mental models, their competence at basic digital interface tasks, their willingness to explore an interface and their self-efficacy (Goodman-Deane et al., 2020). For example, previous work has shown that differences in technology experience affect how people use a new technology design (O'Brien et al., 2008).

A skewed sample in user research means that the results will not be representative of the actual target population. Conclusions drawn from such results will be representative of only part of the target population, and may result in interfaces that are difficult for other people to use or that do not meet their needs. This can lead to a need for modifications or add-ons or potentially even design failure.

In this paper, we propose a methodology to better understand the skew in user research samples, and identify the characteristics of people who are missing from a study. This can be used to better interpret user research results, and to inform further targeted recruitment to improve the representativeness of the sample. The methodology involves asking the participants a set of questions on technology experience, competence and attitudes, taken from the UK Digital Exclusion Survey (Goodman-Deane et al., 2021a). The data from the survey can provide insight into how those omitted from the sample might have performed in the user trials.

2. Related work

Inclusive design is one approach that considers the diversity in the target population in the design of mainstream products and services (Clarkson & Coleman, 2015). Inclusive design involves a range of methods to do this. In particular, there has been work on involving people with extremes of ability in design to inspire design and to ensure that the viewpoints of a diverse range of users are considered (e.g. Dong et al., 2015). However, in doing so, it can be difficult to ensure that the full range of users are considered. The work described in this paper can help inclusive designers consider who they have already involved and who they have not yet heard from.

The HADRIAN tool and dataset (Porter et al., 2004) provides another way of considering extreme users. This tool provides data on 100 people with a variety of abilities and ages, and describes how they interact with systems and services and everyday products. However, the sample was not intended to represent the whole population, and the study did not examine technology experience or competence.

Another inclusive design approach is to use exclusion audits and expert appraisals in combination with user trials (Goodman-Deane et al., 2014). An exclusion audit involves identifying the demands that the product or service places on the user, comparing these with population data on users' capabilities to provide an estimate of how many people in the population would be unable to use the product successfully (Waller et al., 2010).

This is a useful complement to user trial results and helps to frame the insights in terms of the wider population. However, previous work on exclusion audits focused on non-digital products and did not examine technology experience or technology competence. In addition, the user trials and exclusion audits have generally been treated as separate methods, rather than directly using the population data to understand the user trial samples and results. In this paper we propose a method for doing this that takes digital aspects into consideration.

Oudshoorn et al. (2004) discuss how user trials, at times, have little importance in the overall design, and that designers claim user trial success even when limitations were apparent in sample groups. Their case studies highlight gender bias and limitations in design assumptions made by researchers who had male dominant user samples. They suggest that when user contact with groups of individuals is missing, designers can still explicitly consider their needs and how they may use the design. The insights about how future user groups may use the product or service, can then be used to improve the design.

3. Method

The method described in this paper elaborates on a method suggested by Goodman-Deane et al. (2020). The participants in a user trial or other user research are given a simplified version of the UK Digital Exclusion Survey questionnaire (Engineering Design Centre University of Cambridge, 2021). This is in addition to the standard user trial procedure, which involves completing tasks with the product, service or interface under consideration. The full version of the questionnaire was previously administered to a larger, more population-representative sample. The questionnaire responses can thus be used to identify which people in the wider survey are similar to the participants in the user trial in terms of their characteristics that affect digital exclusion. The survey data can also be used to identify which types of people in the general population are not covered in the sample. This method can be used in all kinds of user research with both small and large samples to improve the understanding of the sample.

3.1 Survey sample

Details on the UK Digital Exclusion Survey can be found in (Goodman-Deane et al., 2021a), and the questionnaire itself and dataset are available at (Engineering Design Centre University of Cambridge, 2021). The survey was developed by the authors and conducted by an independent market research company. It included 338 people from across England and Wales. Quota sampling was used with quotas on gender, age, social grade, technology experience and education. The sample achieved good matches on most of these, although the figures for education did not match the census data so closely. As a result, the data from the survey was weighted to better match the education levels in the population, and this weighted data is used throughout this paper.

The majority of participants were recruited on-street using a screening questionnaire at 9 locations across England and Wales. The selected participants then completed the questionnaire in test centres at each location. They received a £10 high street shopping voucher as thanks for their participation. An additional ten interviews were conducted with participants who left their houses once a week or less, to address the fact that on-street recruitment under samples those who do not leave the house frequently. These participants were recruited through a third-party recruitment agency and were then interviewed in their homes by interviewers from a market research company. They received £20 in cash for taking part. Ethical approval was given by the Engineering Department ethics committee at the University where the study was based.

3.2 Survey questions

The questionnaire and showcards used in the UK Digital Exclusion Survey can be found in full at (Engineering Design Centre University of Cambridge, 2021). The survey questions covered a range of characteristics that affect digital exclusion, including technology access and use; basic digital interface competence; attitudes towards technology; motor, sensory and cognitive capability; and demographic variables. Questions were mostly multiple-choice self-report, except for interface competence, which was measured using paper prototype performance tests as described below. The questionnaire took approximately 20 minutes to administer.

In the modules on technology access and use, participants answered questions about computer, internet, mobile phone, smartphone and tablet access and use. Participants were then asked whether they had completed various tasks on digital devices in the last few months. The questions on technology access and use were adapted from those in the Office for National Statistics (2017) to enable comparison with national figures over time.

In the module on interface competence, participants were shown pictures of smartphone interfaces (such as the one in Figure 1) on a paper showcard. They were asked to indicate on the showcard what they would do to achieve a particular goal, such as finding an event in the calendar in Figure 1. Some goals would take a few steps to complete, but participants were asked to indicate just the first thing that they would do to achieve the goal. A set of pre-determined response options was used by interviewers to record responses. This paper prototyping method was chosen because it requires minimal training to administer, and fits within the standard protocols for Computer Assisted Personal Interviewing, accompanied by showcards.

The interfaces and questions in these tests were chosen to cover a range of common, basic interface patterns on a smartphone: search, changing settings, creating a new item, opening a menu with more options, going back to a previous screen, activating a drop-down menu, activating an on-screen keyboard, and setting favourites. They therefore examined a basic level of digital interface competence.

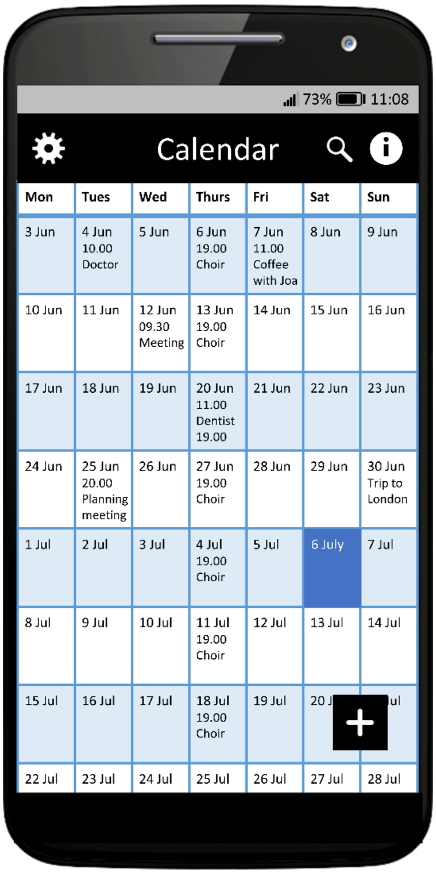

Figure 1. Example of one of the interfaces used in the digital interface competence tests.

The module on attitudes towards technology used the Affinity for Technology Interaction Scale (ATI) developed by Franke et al (2018). The response options for the ATI questions were additionally used in some extra questions developed by the authors, to examine the willingness to explore an unfamiliar interface.

3.3 Clustering survey participants

A cluster analysis for the survey participants was carried out using five clustering variables covering frequency of technology use, the number of technology activities conducted recently, ATI, willingness to explore an unfamiliar interface and basic digital interface competence (measured using the number of interface tests carried out correctly). This identified 12 distinct clusters, where the variation within each cluster was acceptably small. Full details can be found in (Goodman-Deane et al, 2021b).

Each cluster was then represented by a persona. A persona is a fictional profile of a user that can help designers to understand and consider the needs of target users during the design process (Cooper, 1999). This particular set of personas was created to capture the range of technology experience and competence in the population (Goodman-Deane et al, 2021b). It can help to indicate the characteristics and potential requirements of end-users who may be under-represented in a user research sample. The final set of 12 personas is available at: https://www.inclusivedesigntoolkit.com/digitalpersonas/.

3.4 Method within the user research

Recruitment for a piece of user research, such as a user trial, is carried out in the normal way. For example, quota sampling may be used to ensure adequate representation of a range of ages, genders and key user groups. It may be helpful to use the survey data to help set these quotas, but this is not essential.

Within the user research, each participant completes the survey questionnaire. If desired, a shortened version of this questionnaire could be used, focusing on the variables used in the cluster analysis. This shortened version would provide less insight into other variables but still allow an understanding of how the cluster variables are distributed across the user research sample. The participants then complete the standard operating procedure for the user research as normal. For example, in a user trial, they might be asked to try to achieve certain goals with the product, service or interface under examination. The survey performance tests involve extremely basic tasks on a smartphone so are unlikely to cause an order effect for most user trials. However, if the aim of the trial is to examine very basic operations then it could be suggested to counter-balance the trial tasks and the survey questions, or administer these on separate days.

3.5 Comparison between user research sample and survey data

The sample of people involved in the user research study is then compared against the survey dataset. Each participant in the study is matched to the cluster from the dataset that is closest to them, in terms of their questionnaire responses. To do this, we calculate the root mean square (Gaudette & Japkowicz, 2009) of the difference between their responses and the centre of each of the clusters on the five variables used in the cluster analysis. This identifies which of the clusters are included in the study sample, and which ones are not. This can inform further recruitment to improve the representativeness of the sample.

In addition, it may be possible to estimate how survey participants from the missing clusters would respond to the questions or tasks within the study, based on the questionnaire data available from these survey participants. In particular, the survey data includes information on people’s performance on each of the basic digital interface competence tests, as well as their willingness to explore an unfamiliar interface, and whether they had performed various digital tasks in the last few months (e.g. e-mail or installing an app on a smartphone). The responses to these survey questions can help make a judgement regarding the direction in which the sample for the user research study was skewed. For example, if the clusters that are not represented in the sample had the lowest scores on the performance tests and attitude variables, then it is highly likely that the user research results are skewed towards over estimating levels of effectiveness, efficiency and satisfaction, and under estimating the significance of problems with an interface.

4. Case study

A user trial was conducted following the method described in this paper. The primary aim of the study was to validate the paper prototype testing method used in the survey questionnaire. However, this paper does not focus on this aim but rather on how the method proposed in this paper worked out in this case study in practice.

4.1 Methods

The study took place in a university department. Participants first gave informed consent, then completed the survey questionnaire with a researcher (see Section 3.1). They then attempted a variety of tasks on digital devices, as described in Section 4.3. The study took approximately 1 hour to conduct in total, with about 20 minutes for the questionnaire, an optional 5 minute break and 35 minutes for the digital device tasks. Participants had opportunity throughout the tasks to ask questions. Ethical approval was obtained from the University of Cambridge Engineering Department ethics committee.

4.2 Sample

The aim of the sampling was to obtain people with a range of ages, genders and technology experience including those with no smartphone experience or computer use, rather than to obtain a representative sample with specific percentages in different groups. This was to ensure that the issues for different types of people were represented. The different categories for technology use were as follows:

1. Daily or almost daily smartphone use
2. Less frequent technology use: Little or no use of a computer OR little or no use of smartphone. Little use was defined as less often than daily.

The resultant sample distribution is shown in Table 1. Initially we hoped to obtain people with no computer or smartphone use, but this proved challenging and only 3% of the sample had never used a computer.

Potential participants were asked about their age, gender and technology use for determining their match to the quotas. Some additional questions were asked to ensure that vulnerable adults were not included, for ethical reasons.

Participants were recruited by advertising within the university department and on notice boards and in community centres in the local area. Recruitment was also conducted in local cafes, where a researcher spent time liaising with residents and letting them know about the research study. Word of mouth was also employed with participants thinking of friends who might want to take part. Participants were given a £10 voucher from either Love to Shop or Amazon for their participation.

Table 1 Frequency table of achieved figures in the user trial (numbers are rounded so do not add up to 100%). See Section 4.5 for a comparison of the user trial sample and the sample from the full survey.

Variable	Value	% of user trial sample
Gender	Male	46.7%
	Female	53.3%
Age	16-34	26.7%
	35-64	26.6%
	35-74	26.7%
	75+	20.0%
Technology Use	Use a smartphone daily	73.0%
	Little smartphone use (less than daily)	00.0%
	Never used a smartphone	27.0%
	Use a computer daily	83.0%
	Little computer use (less than daily)	13.0%
	Never used a computer	03.0%

4.3 Digital device tests

Participants were asked to complete tasks on several digital devices: a smartphone, a touchscreen laptop and a simple bespoke touchscreen device. Five tasks were performed in total:

1. **GP surgery check-in simulation:** A simulation of a GP surgery check-in machine was created on a touchscreen laptop, to navigate through a fictitious GP surgery check-in machine. The simulation was created in PowerPoint. An example screen is shown in Figure 2. Participants were provided with details to use for the check-in to avoid them having to enter their own potentially sensitive information. They asked to check-in on the device as they would normally using the details provided.
2. **Health tracker:** This task used a bespoke handheld touchscreen device designed for monitoring possible urinary issues. Participants were given instructions to enter one entry using the fictional information provided.

3. **Calendar application on a smartphone:** Participants were asked to find a particular calendar application on the provided Android smartphone, successfully launch it and then complete two tasks:
    - Enter new event
    - Search for a specified event in the calendar
4. **Mapping application on a smartphone:** Participants were asked to find a specific mapping application on the same smartphone, successfully launch it and then bring up directions to specified place.
5. **Specific tasks on a smartphone:** Participants were asked to complete two specific tasks on Google maps on the smartphone above. These were chosen to match some of the interface competence tasks in the survey very closely. The tasks were:
    - Go back to the previous screen from a search
    - Bring up a menu with more options

Figure 2. A screen from the GP survey check-in simulation

The study was designed to validate the paper prototype testing method used in the survey questionnaire. It intentionally included some tasks that closely matched the tasks in the survey questionnaire and others that were different.

After each validation task, a few questions were asked about how easy or hard participants found each task. Data was collected about previous experience using similar technology. The interviewer recorded observations made during the task, such as unusual behaviour and alternative digital pathways. The interviewer also recorded comments that were made during the task and whether interviewer prompts were needed. The digital technology tasks lasted around 35 minutes in total including time for additional instructions and questions if required.

4.4 Results

The focus of this paper is not on the results from the study itself, but on its use as a case study of the method for understanding the sample better. Nevertheless, it is worth noting that some of the issues with the sample impacted the results. Although there were generally good correlations between the questionnaire results and the tasks on the real digital devices, there was not enough discrimination on some of the tasks to obtain a statistically significant result. For example, on some tasks, all of the validation study participants completed the task correctly, so we cannot be sure how well it would match for people who got it wrong. Examining the sample in more detail by comparing it with the survey results can help us to understand and address this.

4.5 Comparison between user trial sample and survey

Each participant in the user trial was matched to the cluster from the survey dataset that was closest to them on the five variables used in the cluster analysis. The results are shown in Table 2.

Cluster 12 represented 16% of the digital survey participants. However, 46% of the user trial participants matched this cluster as their closest match, resulting in a large over sampling for this cluster. As cluster 12 had the highest level of technology competence, it is likely to skew user trial results towards over-estimating the usability of the interface. However, this is not a big issue for this particular case study, because its aim was to understand the experiences of a wide range of users rather than evaluate the usability of the interfaces under examination.

Of greater concern, results show that no participants matched 4 of the clusters (cluster 1, 2, 3 and 6), so the issues that these groups of participants might experience were not picked up during the user trial. These included the 3 clusters with 'very low' technology competence and one with 'low' competence. Together these clusters made up 29% of the survey sample. It is notable that these very low technology clusters are missing even though efforts were made to recruit people with low technology use. This is discussed further in Section 5.

The comparison of the user trial sample with the survey data could inform further recruitment to improve the sample by focusing on people in the missing clusters. It is hard to screen people based on their technology competence directly as measuring this is more time consuming and hard to do over the phone. However, simple questions on technology use can be helpful.

Some estimates can also be made about how the missing clusters would have performed on the user trial tasks, based on their survey responses. For example, clusters 1,2 and 3 had very low technology competence scores. It is likely that they would have performed worse overall than the user trial sample on the technology based user trial tasks. This means that greater care and caution is needed in interpreting the trial results. In particular, the features that appeared easy-to-use in this user trial might not have been easy to use for the population as a whole.

Table 2 Numbers of participants matching each of the clusters in the survey dataset. 'Basic digital interface competence' is calculated by evaluation of scores achieved on paper-based tests.

Cluster number	**Cluster size (% of survey sample)**	**Variables in the cluster analysis**					**Number of participants in user trial sample**
		Basic digital interface competence	**Frequency of technology use**	**Number of technology activities**	**ATI (Affinity for Technology Interaction)**	**Willingness to explore an interface**	
1	12%	Very low (0.5)	Very low	Very low	Very low	Very low	0 (0%)
2	3%	Very low (2.0)	Very high	Very high	Moderate	Low	0 (0%)
3	8%	Very low (2.5)	Very low	Very low	Low	Moderate	0 (0%)
4	9%	Low (4.0)	Low	Low	Low	Low	2 (7.1%)
5	6%	Low (4.0)	High	High	Low	High	4 (14.3%)
6	6%	Low (4.5)	Very high	Very high	High	High	0 (0%)
7	7%	Moderate (6.0)	High	High	Low	Low	1 (3.6%)
8	14%	Moderate (6.5)	High	High	Moderate	High	2 (7.1%)
9	2%	High (7.0)	Very low	Very low	Low	Moderate	1 (3.6%)
10	8%	High (7.0)	High	High	Low	High	1 (3.6%)
11	9%	High (7.5)	Very high	Very high	High	Moderate	4 (14.3%)
12	16%	High (7.5)	Very high	Very high	High	Very high	13 (46.4%)

The authors are currently developing a method for doing a digital exclusion audit that compares the characteristics of the tasks within the user journey against the characteristics of each person in the survey dataset (e.g. their performance on the interface tests in the questionnaire, attitudes, etc). This can provide a more detailed understanding of how the people in the missing clusters might respond to the user trial tasks. Initial work on the digital exclusion audit is reported in (Bradley et al, 2021).

5. Discussion

Comparing the user trial sample with the survey dataset provided useful insight into the limitations of the sample and how it could be improved. It enabled a better understanding of those who were missing from the user trial sample and improved the interpretation of the results.

It was notable that the clusters with 'very low' technology performance were missing from the user trial sample, even though a quota was set on technology use, and efforts were made to recruit people with low technology use. It seems likely that people who are nervous about technology do not want to sign up for a study in which they will have to use technology. Furthermore, people with very low technology competence may be hesitant about taking part in a study which will demonstrate their low competence to others, as this could be potentially embarrassing for them. It may be possible to mitigate this to some extent by changing the way a study is designed and advertised to reduce the emphasis on technology, and to highlight the value of people's input from across the range of technology experience. Holding the study in a neutral location rather than a university department may also help, particularly if this is a location where the people already go frequently, such as a community centre or café.

In general, integrating a survey dataset that provides data on a wider population enables researchers to place user research results in the larger population context. However, the value of this is limited by the quality of the survey dataset and ideally a population representative survey would be used. The survey presented in this paper is limited by its size (338 participants) which is not typically large enough for a survey to be considered population-representative, although the use of quotas on a range of variables ensured that a good spread from across the population was obtained.

The authors have adapted the survey in this paper and are currently running it in five different European countries to obtain larger, more population-representative data. Initial population-representative results from 1010 people in Germany are now available (Goodman-Deane et al, 2021c). Once processed further, these results could be used to provide a more robust dataset for understanding user trial samples. This dataset and similar datasets from Italy, the Barcelona region in Spain, the Netherlands and the Flanders region in Belgium will be made available open access on the UPCommons repository (https://upcommons.upc.edu/). These datasets contain data on the same technology areas as the UK survey as well as key aspects related to travel.

In this paper, we have presented a cluster analysis covering technology experience, competence and attitudes. If these are relevant for a study, then comparing the user research sample to the personas from this previous cluster analysis is relatively quick and straightforward for someone with some basic statistical experience. In further work, we plan to automate this process further through publicly available software, thus reducing the need for statistical experience. We recommend using this previously completed cluster analysis for user research projects that focus on the digital aspects of a product or service.

However, if other variables are particularly relevant for a study (such as sensory capability or cultural aspects), then a new cluster analysis and possibly a different dataset may be necessary. Finding a suitable dataset and conducting a new cluster analysis requires more skill in statistics. The time taken depends on the skill of the analyst, the degree of automation that is supported by software and the amount of iteration required in order to get a useful result.

Another issue is that, in a standard cluster analysis, every data point belongs to a cluster. However, some data points may actually be quite a distance from their cluster centre. For the purpose of examining user research samples, it may be better to set a threshold for the distance from a cluster centre which is acceptable. If data points are further from their cluster centres than this threshold, it may not be useful to consider them as really belonging to that cluster.

The primary benefit of the approach presented in this paper is to identify the kinds of people who are under-represented within the user research sample. However, further time is then required to act on this information. This might involve targeted attempts to recruit the kinds of people who are under-represented and/or using the personas from the cluster analysis to try and infer how the under-represented people might have responded, if they had been able to take part in the user research.

6. Conclusions and further work

In conclusion, this paper presents a methodology for improving the understanding of the skew in a user research study sample by comparing it with survey data on technology experience, competence and attitudes. This can help to inform researchers and designers, providing a more accurate understanding of different groups within the target population and greater insight into those who have not been included in the user trial sample. This can help to inform further recruitment to improve the sample, as well as aiding in the interpretation of user research results. This is increasingly important, especially during the Covid-19 pandemic, when there will be larger groups of the population who may not be able to take part in user research. Further work is required to develop the methodology further to make it more readily applicable to a wider range of user research studies, reduce the need for researchers to have expertise in statistical analysis and make the methodology more accessible to the community of practice.

The survey that was described within this paper has been extended across multiple countries, and with larger samples (see, for example, Goodman-Deane et al, 2021c). The data from these surveys could be used to further develop the methodology. Methods could be developed to help cluster the survey participants into groups, and to take into consideration the sizes of these groups when considering the distribution of a user research sample . Future work will also develop an exclusion audit method to provide insights about how those underrepresented in user trials would interact with designs in practice, based on their responses on the survey.

Acknowledgements: The main body of this work was part of a project to improve the inclusivity of railway journeys and information systems. It was funded by DFT and delivered through RSSB's TOC'16 project: Towards the Inclusive Railway. Collaborative partners were Siemens Mobility, Keolis Amey Docklands and Astutim. Some of the work was also funded by the Dignity project which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 875542. We would also like to thank the University of Cambridge Statistics Clinic for statistics advice on cluster analysis.

7. References

Bradley, M., Goodman-Deane, J., Kluge, J., & Clarkson, P. J. (2021). Proceedings of MHCI 2021: Predicting population exclusion for services dependent on mobile digital interfaces, September 24 – October 1, 2021, held Toulouse, France. ACM

Berry, L. (1983). Residential Conservation Program Impacts: Methods of Reducing Self-Selection Bias. *Evaluation Review, 7* (6), 753-775. https://doi.org/10.1177/0193841X8300700603

Clarkson, P., Coleman, R. (2015). History of Inclusive Design in the UK. *Applied Ergonomics, 46 (B),* 235-247. https://doi.org/10.1016/j.apergo.2013.03.002

Cooper, A. (1999). *The Inmates Are Running the Asylum, SAMS.* Indianapolis, Indiana.

Dong, H., Clarkson, P. J., Cassim, J., & Keates, S. (2015). Critical user forums – an effective user research for inclusive design. *The Design Journal, 8 (2),* 49 – 59. https://doi.org/10.2752/146069205789331628

Engineering Design Centre University of Cambridge. (2021). UK Digital Exclusion Survey Dataset (Data Source for Inclusive Design – Towards the Inclusive Railway). SPARK: The Rail Knowledge Hub. https://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=27147

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction,* 35(6), 456-467.

Gaudette L., Japkowicz N. (2009) Evaluation Methods for Ordinal Classification. In: Gao Y., Japkowicz N. (eds) Advances in Artificial Intelligence. Canadian AI 2009. Lecture Notes in Computer Science, vol 5549. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01818-3_25

Goodman-Deane, J., Ward. J., Hosking, I., & Clarkson, P. J. (2014). A comparison of methods currently used in inclusive design. *Applied Ergonomics, 45* (4), 886 – 894.

Goodman-Deane, J., Bradley, M., Waller, S., & Clarkson, P. J. (2020). Proceeding of CWUAAT 2020: Quantifying exclusion for digital products and interfaces, March 23 – 25, 2020. Cambridge, UK.

Goodman-Deane, J., Bradley, M., & Clarkson, P.J. (2021a). Relating age, digital interface competence and exclusion*. Gerontechnology* 20(2).

Goodman-Deane, J., Bradley, M., Waller, S., & Clarkson, P. J. (2021b). *Proceedings of International Conference on Engineering Design 2021*: *Developing personas to help designers to understand digital exclusion,* August 16-20 2021, held in Gothenburg, Sweden.

Goodman-Deane, J., Kluge, J., Roca Bosch, E., Nesterova, Bradley, M., Clarkson, P.J. (2021c). *Proceedings of British HCI 2021: Digital mobility services: A population perspective,* July 19-21, London, UK.

Green, W., & Jordan, P. W. (1999). Human Factors in Product Design: Current Practice and Future Trends. Taylor & Francis.

Norman, D.A. (2002). The Design of Everyday Things. Basic Books.

Porter, J. M., Case, K., Marshall, R., Gyi, D., & Sims neé Oliver, R. (2004). Beyond Jack and Jill: designing for individuals using HADRIAN. *International Journal of Industrial Ergonomics, 33* (3), 249-264. https://doi.org/10.1016/j.ergon.2003.08.002

O'Brien, M. A., Olson, K. E., Charness, N., Czaja, A. D., Fisk, W., Rogers, A., & Sharit, J. (2008). Understanding technology usage in older adults. Proceedings of the 6th International Society for Gerontechnology (ISG08) Pisa; 2008.

Office for National Statistics (2017) Internet Access – households and individuals, Great Britain: 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017 (retrieved Nov 2021).

Oudshroon, N., Rommes, E., & Stienstra, M. (2004). Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. *Science Technology Human Values, 29*(1), 30-65. https://doi.org/10.1177/0162243903259190

Stanton, N. M., Salmon, P. M., Refferty, L. A., Walker, G. H. Baber, C., & Jenkins, D. P. (2013). *Human factors methods for Engineering Design.* Taylor & Francis Group.

Sutton, T. M., & Edlund, J. E. (2019). Assessing Self-Selection Bias as a Function of Experiment Title and Description: The Effect of Emotion and Personality. *North American Journal of Psychology*, *21*(2), 407. https://link.gale.com/apps/doc/A587973378/HRCA?u=anon~86d55e2&sid=googleScholar&xid=d3a08e29

Waller, S. D., Langdon, P.M., & Clarkson, P.J. (2010). Using disability data to estimate design exclusion. Universal Access, *International Journal of the Universal Access Information Society, 9* (3), 195 – 207.

About the Authors:

Anya Petyaeva is a PhD student in the Inclusive Design Group in the Engineering Design Centre (EDC), University of Cambridge. Research interests include developing and using inclusive design methods to improve design and help designers understand different stakeholders.

Dr Joy Goodman-Deane is a Senior Research Associate in the Inclusive Design Group in the EDC, University of Cambridge. Her research interests focus on developing tools and methods to support designers in putting inclusive design into practice.

Mike Bradley is a Senior Research Associate in the Inclusive Design Group in the EDC, University of Cambridge. He is interested

in interaction design for those who have little digital experience or who are reluctant to learn new interactions.

Dr Sam Waller is a Senior Research Associate in the Inclusive Design Group in the EDC, University of Cambridge. His research interests include analysing population statistics and creating tools that enable practicing designers to make more inclusive products.

Professor P John Clarkson is Professor of Engineering Design (University of Cambridge) and Healthcare Systems (TU Delft) and Director of the Cambridge Engineering Design Centre (EDC). Research interests are in the general area of engineering design, including inclusive design.