

## **“Alexa, I just ate a donut”: A pilot study collecting food and drink intake data with voice input**

Louise AC Millard<sup>\*1,2</sup>, Laura Johnson<sup>1,2,3,4</sup>, Samuel R Neaves<sup>1,2</sup>, Peter A Flach<sup>5</sup>, Kate Tilling<sup>1,2</sup>, Deborah A Lawlor<sup>1,2</sup>

\* Corresponding author: [louise.millard@bristol.ac.uk](mailto:louise.millard@bristol.ac.uk)

<sup>1</sup> MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, United Kingdom

<sup>2</sup> Population health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

<sup>3</sup> Centre for Exercise, Nutrition and Health Sciences, School for Policy Studies, University of Bristol, Bristol, United Kingdom

<sup>4</sup> Centre for Health, NatCen Social Research London, UK

<sup>5</sup> Faculty of Engineering, University of Bristol, Bristol, United Kingdom

WORD COUNT: 3915

## ABSTRACT

### **Background**

Voice-based systems such as Amazon Alexa may be useful to collect self-reported information in real-time from participants of epidemiology studies, using verbal input. We demonstrate the technical feasibility of using Alexa, investigate participant acceptability, and provide an initial evaluation of the validity of the collected data. We use food and drink information as an exemplar.

### **Methods**

We recruited 45 staff and students at the University of Bristol (UK). Participants were asked to tell Alexa what they ate or drank for 7 days, and also to submit this information using a web form. Questionnaires asked for basic demographic information and about their experience during the study and acceptability of using Alexa.

### **Results**

Of the 37 participants with valid data, most were 20-39 years old (N=30; 81%) and 23 (62%) were female. Across 29 participants with Alexa and web entries corresponding to the same intake event, 357 Alexa entries (61%) contained the same food/drink information as the corresponding web entry. Participants often reported that Alexa interjected, and this was worse when entering the food and drink information compared with the event date and time. The majority said they would be happy to use a voice-controlled system for future research.

### **Conclusions**

While usability of our skill was poor, largely due to the conversational nature and because Alexa interjected if there was a pause in speech, participants were mostly open to participating in future

research studies using Alexa. Many more studies are needed, in particular, to trial less conversational interfaces.

KEYWORDS: data collection, voice-based approaches, Amazon Alexa, self-reported data, food and drink.

#### KEY MESSAGES

- Over the last few years voice-controlled ‘smart’ systems have emerged giving the possibility of collecting self-reported data using a voice-based approach.
- We successfully collected epidemiology food and drink information in real-time, demonstrating that voice-based collection of self-reported data is technically feasible.
- The conversational design of our skill meant that usability was poor, for example, most participants (86%) reported that Alexa either occasionally, often or always interjected during use, and the majority of participants who had previously used a paper diary or my fitness pal did not find Alexa as efficient to use compared with these approaches.
- After participating in this study, the majority of participants would be happy to use Alexa again, either at home or on a wearable device.
- Our results highlight that further work is needed to evaluate use of voice-based systems, including comparing Amazon Alexa with the Google Assistant, and trialling less conversational interfaces.

## INTRODUCTION

Epidemiology cohorts typically collect data at widely spaced time points (e.g. every 1-5 years)<sup>1,2</sup>.

While some types of traits (e.g. weight or height) are fairly stable or change gradually across time, others such as activity levels, blood glucose levels, mental wellbeing, and dietary intake can vary more acutely, for example, within days, hours or even minutes. For these traits, prospectively capturing how they vary across time means we can assess how this variability relates to other traits and disease. Some acutely varying traits can be collected continuously and objectively using wearable digital devices, for example physical activity using accelerometers or blood glucose using continuous glucose monitors<sup>3</sup>. For others, such as mental health traits and dietary intake, no objective approach to measuring within-day variation in these traits exists, and they need to be collected via self-report.

One possible approach to providing real-time self-reported information is verbal input, which could enable participants to conveniently enter free-text rather than selecting from a set of prescribed options (in contrast to Ecological Momentary Assessment (EMA)<sup>4-6</sup>). Over the last few years voice-controlled 'smart' systems have been released by several technology companies. These systems, such as Amazon Alexa and the Google Assistant, allow users to talk to a device rather than typing or pressing a button. They each have core functionality available by default (e.g. saying the time when asked for this), but also have developer platforms that allow anyone to produce and publish a custom voice-based app. This means it is now technically possible to collect self-reported data continuously across a day or several days, using verbal input.

Voice-based data collection may be most useful (in contrast to EMA) for collecting self-reported data that is both complex and variable across a day. One possible example is the food and drink a person consumes, and when they consume it. Traditionally, cohorts have collected dietary intake

information using paper or online food frequency questionnaires or (less commonly) diaries.

Limitations of these include retrospective recording, needing converting to electronic form, potential for missing data because participants are not prompted for missing information, and the inconvenience of having to carry a diary. More recently, some other approaches have been developed; a web-based dietary recall tool<sup>7</sup> and an approach using photographs<sup>8</sup>. While these methods are able to collect detailed dietary information, they are burdensome so can only be used for short periods by highly motivated participants<sup>9</sup>.

In this pilot study we explore the potential of voice-based data collection in epidemiology research, using food and drink diaries as an exemplar. We have three key aims: 1) to demonstrate the technical feasibility of collecting data using Alexa, 2) to gain an initial insight into participant acceptability, and 3) to provide an initial evaluation of the validity of the collected data. In general, we view the capture and processing of the information as separate steps, and in this study focus on demonstrating and evaluating the former.

## METHODS

Ethical approval was given by the University of Bristol Faculty of Health Sciences Research Ethics Committee (approval number 63861).

### **Study participants**

Power calculations based on two measures suggested a sample size of at least 35 is needed (see details in Supplementary section S1). We recruited volunteers from University of Bristol staff and postgraduate student email lists. Participants were compensated with a £30 voucher.

### **Description of system architecture: a voice-based system using Amazon Alexa**

In this work we use the Amazon Alexa voice system. The Alexa system enables development of custom functionality, referred to as a custom skill. Alexa skills are comprised of intents that each define an interaction that a user can have with the skill. We developed a custom skill to collect food and drink intake events, with intents that allow participants to: 1) Add the date and time of an intake event, 2) add one or more items they ate or drank at this time, 3) cancel the event, 4) cancel the last item added to the event, and 5) submit the event. See example utterances in Supplementary table S1 and an example conversation in Figure 1. Supplementary section S2 provides further details.

### **Data collection protocol**

An overview of the data collection protocol is shown in Figure 2. Due to the COVID-19 pandemic, participants took part at home. An initial email with accompanying participation information sheet (Supplementary file 1) invited staff and students to take part in this study. On replying, participants were sent a pre-participation questionnaire asking for basic demographic information such as their age and sex (Q1, Supplementary file S2). On completion, participants were booked for a 7-day Alexa data collection period.

The equipment was kept at the home of the principal investigator (LACM). On day 1 of a participants' Alexa data collection period the equipment was delivered to their home by courier, along with a participant guide (Supplementary file 3). The participant was asked to set up the equipment and start using it as soon as they were able. Participants were instructed "After you have had something to eat or drink, we would like you to submit your food and drink information to Alexa first, and then submit it on the web form." Entering the food and drink information using both Alexa and a web form (Q4, Supplementary file S2) allows us to compare the data entered using these approaches (i.e. relative validity<sup>10</sup>). On day 7 the equipment was returned to the principal investigator's home by courier.

Participants were then asked to complete a post-participation questionnaire on their experiences during the study and acceptability of using Alexa (Q2, Supplementary file S2).

To understand views on acceptability of using voice-based interfaces more widely (beyond our participant group) we also sent a further invitation (to the same email lists) asking for those who did not participate to complete a short questionnaire about their feelings on using voice-based devices and their reasons for not participating (Q3, Supplementary file S2).

Questionnaires were deployed via the University of Bristol REDCap secure web platform [<https://sscmredcap.bris.ac.uk/redcap/>]. Study emails are provided in Supplementary file S4.

### **Analytical sample**

A participant flow diagram is shown in Supplementary figure S1. Of the 45 participants who registered to participate, 1 withdrew, and 7 were excluded with incorrect equipment (see Supplementary section S3). The remaining 37 participants comprised our analytical sample. Of these, 1 participant did not attempt to use Alexa. In addition, 7 participants had Alexa entries, but no web diary entries completed within the 30 minutes directly following the Alexa submission. As Alexa and web entries must be within 30 minutes to be identified as corresponding to the same intake event in our data processing approach (see below), the Alexa and web entries from these participants could not be compared. The remaining 29 participants were used to compare the information entered via the web form versus Alexa ('comparison' sample).

### **Data preprocessing**

#### *Mapping web food form entries to Alexa intake events*

The web and Alexa entries both include the following information: a) intake timestamp: the date and time the participant (said they) ate or drank; b) submission timestamp: the date and time the

participant submitted the entry; and c) intake items: one or more intake food and drink items. To compare the content of the web and Alexa entries, we first undertook an automated process to identify the Alexa and web entry pairs that correspond to the same intake event, referred to as counterpart entries. This was non-trivial because a participant might not have entered each entry with the web form immediately after entering it via Alexa, or the intake timestamp entered by Alexa might have been recorded incorrectly (i.e., Alexa might have heard the day or time stated by the participant incorrectly).

We identified counterpart entries using the intake and submission timestamps. The process we used was as follows (illustrated in Supplementary figure S5):

- 1) Identify counterparts as the set of entries where the web and Alexa intake timestamps were within 5 minutes of each other, and the Alexa submission timestamp is up to 30 minutes before the web submission timestamp. The non-exact match of the intake time was because participants can tell Alexa this using a phrase such as ‘just now’ or ‘ten minutes ago’ so may not correspond exactly to the intake time entered using the web form.
- 2) Identify web counterpart entries of the Alexa submissions not matched in step (1), as the nearest subsequent web entry where one occurs within 30 minutes of the Alexa entry.

#### *Comparing counterpart web and Alexa entries food and drink description*

We compared counterpart entries using two approaches, an automated approach, and a systematic manual approach.

Automated approach: We compared the text content of the counterpart entries, by comparing the set of words contained in each. Entries were preprocessed to remove plurality of words (e.g., crisps



becomes crisp)<sup>11</sup>, and convert numbers to numeric values (e.g., “one” and “a” both become 1). For each counterpart pair we calculated the number of words in a) the web word set but not the Alexa word set, b) the Alexa word set but not the web word set, and c) both word sets.

Systematic manual approach: Our systematic manual approach was conducted by LACM. As this approach has some degree of judgement, we also asked 5 researchers independent to the project (within the same Unit but not involved in this study) to review 10 random entries (none repeated across researchers) so that we can evaluate inter-researcher variability of these manual evaluations.

To conduct this manual review we used a two-step process. First, the intake items of each counterpart pair were compared to determine if there was any similarity at all. If the set of items was completely different then these were marked as most likely corresponding to different intake events (i.e., the counterpart pairing did not work in this case – e.g. ‘a cup of coffee with milk’ versus ‘spaghetti bolognese’). All other entries are taken forward to step two.

Step two involved reviewing each counterpart entry and for each, recording the number of food or drink items in a counterpart pair in the following categories:

- a) Same item semantically (the two entries are equivalent with no additional or different information in each)
- b) Same item but with different detail (e.g., ‘cup of tea’ vs ‘mug of tea’)
- c) Same item, Alexa information has less detail (e.g., ‘cheese and salad sandwich’ vs ‘a sandwich’)
- d) Same item, Alexa item has more detail
- e) Same item, misspelling in Alexa input, but still understandable i.e., there is no loss of information (e.g. ‘to bagels’ versus ‘two bagels’)
- f) Same item, misspelling in web form input, but still understandable

g) Same item, with Alexa entry issue, where the consumed item is still identifiable (e.g., “ball of yoghurt” rather than “bowl of yoghurt”)

h) Item with major entry issue, such that it contains no food or drink information, or the main essence of the meal/drink is missing (e.g. a “cough with milk” rather than “coffee with milk”)

i) Extra Alexa item with major entry issue (which can happen if a participant makes a mistake or then stops talking, then tries again so there is an extra item, e.g., “two”)

j) Extra Alexa item that is recognizable as a food or drink (i.e., should not be assigned to (i))

k) Extra web item

Supplementary table S2 shows some example assignments using this approach.

The independent researchers who completed 10 entries were provided with an information sheet describing the task (Supplementary file S5). We evaluated agreement between the assignments of LACM and the independent researchers visually using a stacked bar chart.

The automated and systematic manual approaches are complementary as the former is objective but is likely to be a more pessimistic assessment of agreement. This is because, participant may not write an entry the same way as they would speak it. For example, a participant might write “1 x apple. 1 bar of chocolate” but say “one apple and a chocolate bar”, which have differences in the words used even though they are semantically the same.

## **Statistical analyses**

### *Usage summary*

We summarised the participants’ usage of the web and Alexa approaches using the median and interquartile range (IQR) of the number of submitted web and Alexa entries, respectively.

### *Comparison of counterpart diary entries*

We compared the intake timestamps in the counterpart pairs using a plot similar a Bland-Altman plot but where the x-axis is the intake time entered using the web form rather than the average.

Assuming the intake timestamp entered on the web form will be largely correct, this is to help show whether intake time submitted via Alexa may be less accurate for particular times of day. We summarised the automated and systematic manual comparisons using stacked bar charts.

### *Summarizing the number of incomplete attempts*

The Alexa skill saves partial entries (i.e., those that have not been submitted perhaps because the internet connection was interrupted) in addition to completed entries. We estimated the median (IQR) number of unsuccessful attempts across participants.

### *Evaluating questionnaire responses on acceptability*

We summarised responses of the participation questionnaire (Q2) and the non-participation questionnaire (Q3) by calculating the number of participants (and percentage) that responded to each questionnaire item option. Responses to free text items were read and reread in order to identify key themes.

The Alexa skill and web service code is available at <https://github.com/MRCIEU/food-diary-alexa-pilot-aws/>. Analysis code is available at <https://github.com/MRCIEU/food-diary-alexa-pilot-analysis/> (git tag v0.1 corresponds to the version of the analyses presented here).

## RESULTS

### **Participant characteristics**

Participant characteristics are summarized in Table 1. Most participants were in early adulthood (81% 20-39 years old). Our sample included more female participants than male participants (62% female). The majority (84%) reported that they did not believe they had a strong UK regional accent, with 68% reporting they did not have an accent due to English being a second language. 43% of our participants have an Alexa at home that they use.

### **Usage summary**

On average, participants completed more web diary entries compared to Alexa entries (median number of entries 17 [IQR: 13, 27] compared with 11 [IQR: 7, 21]; paired t-test P value <0.001) (histograms shown in Supplementary figure S6). The median number of partial Alexa attempts was across all participants was 6 [IQR: 1, 9].

### **Comparison of counterpart diary entries**

#### *Intake timestamp comparison of web form versus Alexa entry*

Across all participants in the counterpart subsample 72% of the completed counterpart entries had a matching timestamp (Figure 3). The median proportion of completed counterpart entries with a matching timestamp across participants was 0.67 [IQR: 0.5, 1].

#### *Food and drink description comparison of web form versus Alexa entry*

Across the 29 participants in the comparison subsample there were 310 counterpart entries. Results comparing the submitted food and drink information using the automated and manual comparison approaches are shown in Figure 4 and Figure 5, respectively. Of the 310 counterpart entries manually reviewed, 21 (7%) were classified as corresponding to different intake events. The remaining 289 counterpart entries included 612 web form items and 588 Alexa items, with 33 extra web items (not identified in the counterpart Alexa entry) compared to 9 extra Alexa items (not found in the counterpart web entry). The majority (N=357 [58% and 61% for the web and Alexa items,

respectively]) of the items entered via the web form and Alexa were the same, containing the same information. Of the 194 items that were identified as corresponding to the same intake item but containing different information, 64 (33%) had less detail from Alexa, 12 (6%) had more detail from Alexa, 15 (8%) had different detail, 3 (2%) items had a web entry issue, 36 (19%) had an Alexa entry issue, 4 items (2%) had spelling mistakes in the web version not the Alexa version, 59 (30%) had a misspelling in the Alexa only, and 1 (1%) had a misspelling in both the Alexa and web input. Of the 59 items with an Alexa misspelling, 40 of these (68%) were recording the word 'to' rather than 'two'. 28 (5%) of the 588 items entered via Alexa were classified as having a major entry issue.

We did not identify systematic differences in the assignments of LACM for the systematic manual approach compared to those of independent researchers (Supplementary figure S7).

#### **Evaluating participants questionnaire responses on usability and acceptability**

Summaries of post-participation questionnaire responses are provided in Table 2. Of the 35 participants who completed the post-participation questionnaire 26 (74%) said they would be happy to use voice-controlled system at home for future research and 28 (80%) said they would be happy to use one on a wearable device (e.g., a smart watch). Alexa sometimes interjected when participants were telling her *when* they ate or drank, with 7 participants (20%) saying this happened often or always, and 11 (31%) saying this happened occasionally. Alexa often interjected when participants were telling her *what* they ate or drank, with 18 participants (51%) saying this happened often or always and 12 (34%) saying this happened occasionally. In terms of convenience, enjoyment, and efficiency, 18 (51%), 21 (60%) and 15 (43%) participants, respectively, said they found using Alexa OK or better.

Of the 35 participants who completed the post-participation questionnaire, 25 (71%) had previously used another approach to record their food and drink intake (Table 2). Of the 13 participants who have used a traditional diary (on paper or a computer), 7 (54%) found using Alexa at least as

convenient, 5 (56%) at least as enjoyable, 2 (22%) at least as efficient. Of the 12 (34%) participants who have used MyFitnessPal, 9 (75%) found Alexa at least as convenient, 7 (64%) at least as enjoyable and 4 (36%) at least as efficient.

### **Evaluating non-participation questionnaire responses**

Of the 69 participants who responded, 11 (16%) did not take part due to privacy concerns (either around Amazon or researchers collecting their diet information or Alexa inadvertently listening to other conversations (Table 3). 61% (N=42) stated they would be happy to use Alexa at home for research in the future, while 57% (N=39) said they would be happy to use Alexa on a wearable device for research purposes.

## **DISCUSSION**

In this study we have demonstrated the technical feasibility of collecting data using Alexa for epidemiology research, by successfully developing an Alexa skill to collect food and drink information and using it to collect data in 37 participants across a period of 7 days (5 full days). Our results provide a useful initial insight into participant acceptability of using this approach and validity of the collected data. On average, more entries were submitted via the web form versus Alexa. Our results suggest that intake date and time was largely entered accurately via Alexa. The majority (61%) of Alexa entries contained the same food/drink information as the corresponding web entry, according to our systematic manual approach. The most common differences were the Alexa information having less detail or a homophone error (most often 'to' rather than 'two').

Overall, usability of our Alexa skill was fairly poor. The majority of participants reported that Alexa interjected while they were trying to enter food and drink information (34% sometimes, 51%

often/always), with better results for the date/time of the intake event (31% sometimes, 20% often/always). Several participants reported finding it difficult to avoid pausing while articulating what they ate or drank, which might cause Alexa to interject or cut-out. Some reported reducing the information they provided so that Alexa would be more likely to accept it. Participants also reported that Alexa would sometimes not understand or exit the skill during use.

The voice interface we have trialed comprises of our Alexa skill implementation and the Amazon back-end logic, and only the former is under our control. The implementation and deployment of the Alexa skill has several components, with many choices regarding the design of the voice interface, the technical infrastructure, and the study protocol (e.g., location of data collection which was home-based in our study). Each of these may have affected the usability of the skill to collect food and drink information. Most notably, we conclude that the conversational interface of our skill (where participants first tell Alexa the time, then each of the items consumed) was not successful, as when the skill inadvertently cut out (e.g., because of multiple failed attempts conversing with Alexa or a poor internet connection) the participant would have to start that entry from the beginning. A less conversational interface where the participant states the information without the separate prompts, would likely be more usable. While our results suggest Alexa may be more appropriate for entering short summaries of information, in the longer term, integration of this approach with other approaches such as a phone app, can be used to supplement the voice-collected data. For example, using Alexa to log events directly after eating or drinking, then entering more detail via a phone app when convenient to do so.

Strengths of our study include the collection of pilot data 'in the wild' rather than in a controlled lab-based setting. We collected food and drink information via a web form in addition to Alexa, to allow comparison of the collected data via these approaches. Our study has several limitations. While asking participants to provide information via both Alexa and a web form was valuable, interactions

with one of these approaches may have impacted their interaction and perceived feelings towards the other. The Alexa and web entries in our data had no explicit link and identifying entries that corresponded to the same intake event was difficult. We could only assess relative validity of the Alexa entries (relative to the web form entries) i.e., we have no absolute ground truth. While the intake date and time could be easily compared between the Alexa and web form entries in an automated way, comparing the free-text food and drink information was non-trivial as differences in the way the participant conveyed this information would not necessarily amount to meaningful differences in the submitted information. Most of these limitations could be rectified by integrating this voice-based approach with a phone app where the participant can review each entry and either correct it or mark it as correct, instead of requiring a web diary, so that validity can be assessed in an automated way by evaluating the corrections the participant makes. Further strengths and limitations and details are provided in Supplementary section S5.

While other studies have used voice-based approaches in other health settings<sup>12,13</sup>, to our knowledge, this is the first study to assess collecting self-reported epidemiology data with a voice-based system (a previous grant that sought to create a voice-based interface did not to our knowledge achieve this objective<sup>14</sup>). Many more studies are needed to understand the strengths and limitations of different approaches to collect epidemiology data using voice, for example, with different voice-based systems (e.g. comparing Amazon Alexa versus the Google Assistant), different voice interface designs particularly those that are less conversational, and to further evaluate biases in the collected data<sup>15</sup>. Voice-based approaches may be particularly useful in populations that might not be able to write (or write with ease), for example those with learning difficulties such as dyslexia, or certain diseases such as motor neurone disease.



#### ETHICS APPROVAL

Ethical approval for this study was given by the University of Bristol Faculty of Health Sciences Research Ethics Committee (approval number 63861).

#### AUTHOR CONTRIBUTIONS

LACM and SRN conceptualized the use of Alexa for epidemiology data collection. LACM conceptualized the study, led design of the study, wrote the first version of the manuscript. All authors contributed to the design of the study. All authors critically reviewed and revised the manuscript. LACM acts as the guarantor for this paper.

#### DATA AVAILABILITY

Data cannot be shared for privacy reasons.

#### SUPPLEMENTARY DATA

Supplementary data are available at *IJE* online

#### FUNDING

This work was supported by the University of Bristol and UK Medical Research Council [grant numbers MC\_UU\_00011/3 and MC\_UU\_00011/6]. LACM was funded by a University of Bristol Vice-Chancellor's Fellowship. This study was also supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol (specifically via a BRC Directors Fund grant). DAL's contribution is supported by the British Heart Foundation (CH/F/20/90003 and AA/18/7/34219).

#### ACKNOWLEDGEMENTS

We are extremely grateful to the participants that took part in this study. We are grateful to Mrs Shirley Jenkins for providing administrative support for this study. Thank you to Professor Katrina Turner for helpful feedback on the qualitative review of questionnaire responses. Thank you to Mr Tom Clark, Dr Emily Kawabata, Dr Michael Lawton, Dr Nancy McBride, and Dr Tom Palmer who conducted independent manual assessments of subsamples of the food and drink entries.

#### CONFLICT OF INTEREST

DAL has received support from numerous national and international government and charitable funders, as well as Roche Diagnostics and Medtronic Ltd, for work unrelated to this paper. Other authors report no conflicts of interest.

FIGURES

Figure 1: Example conversation of study participant with custom food diary skill

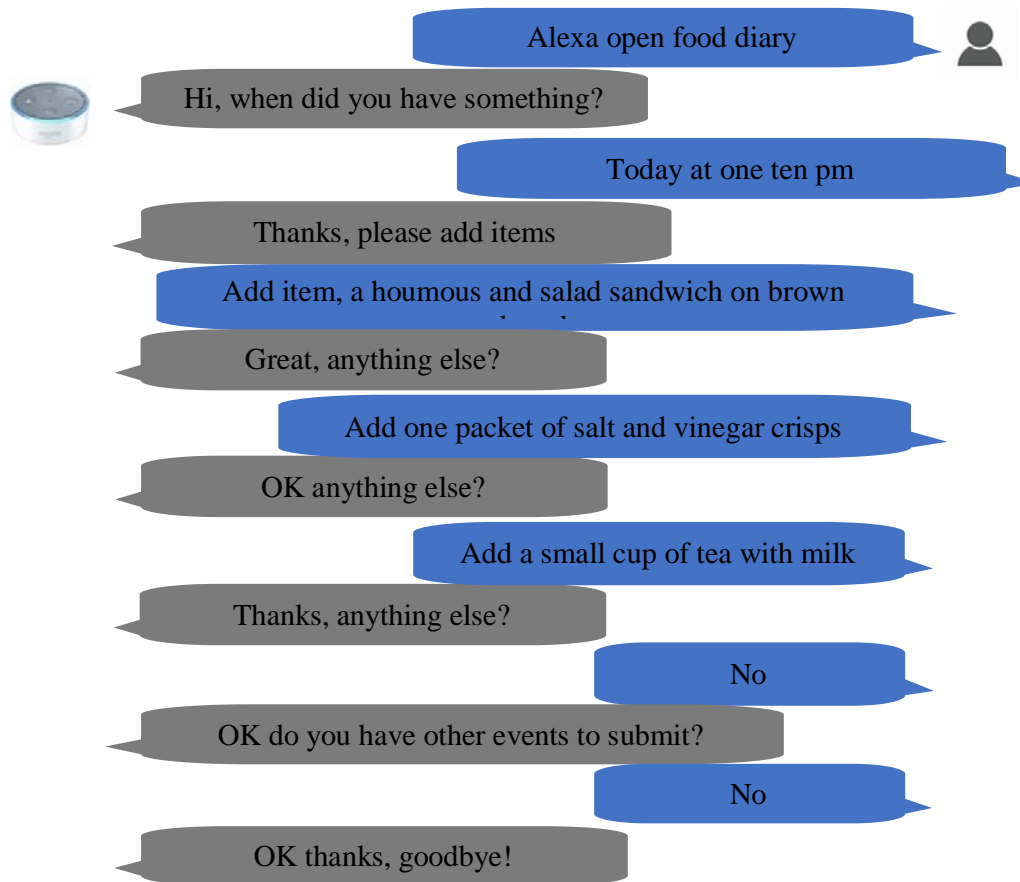


Figure 2: Overview of data collection protocol

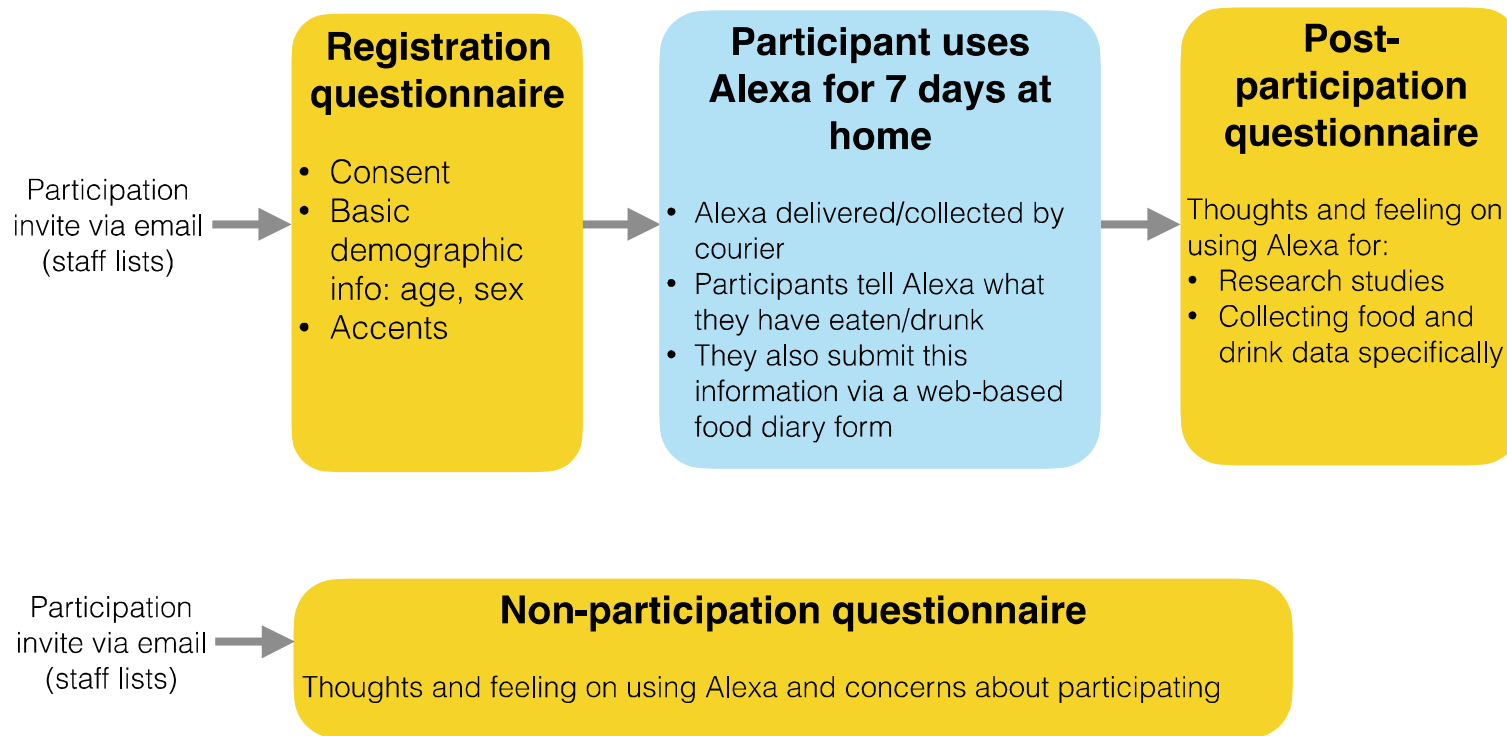
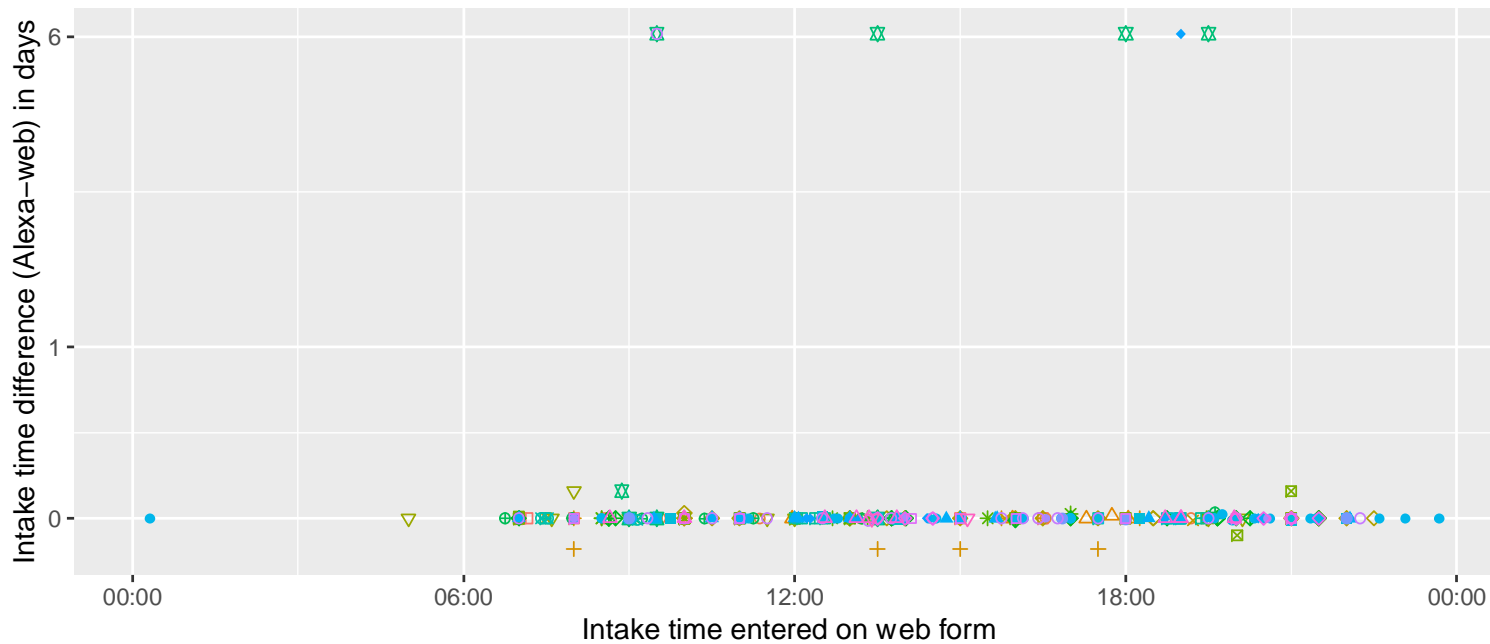
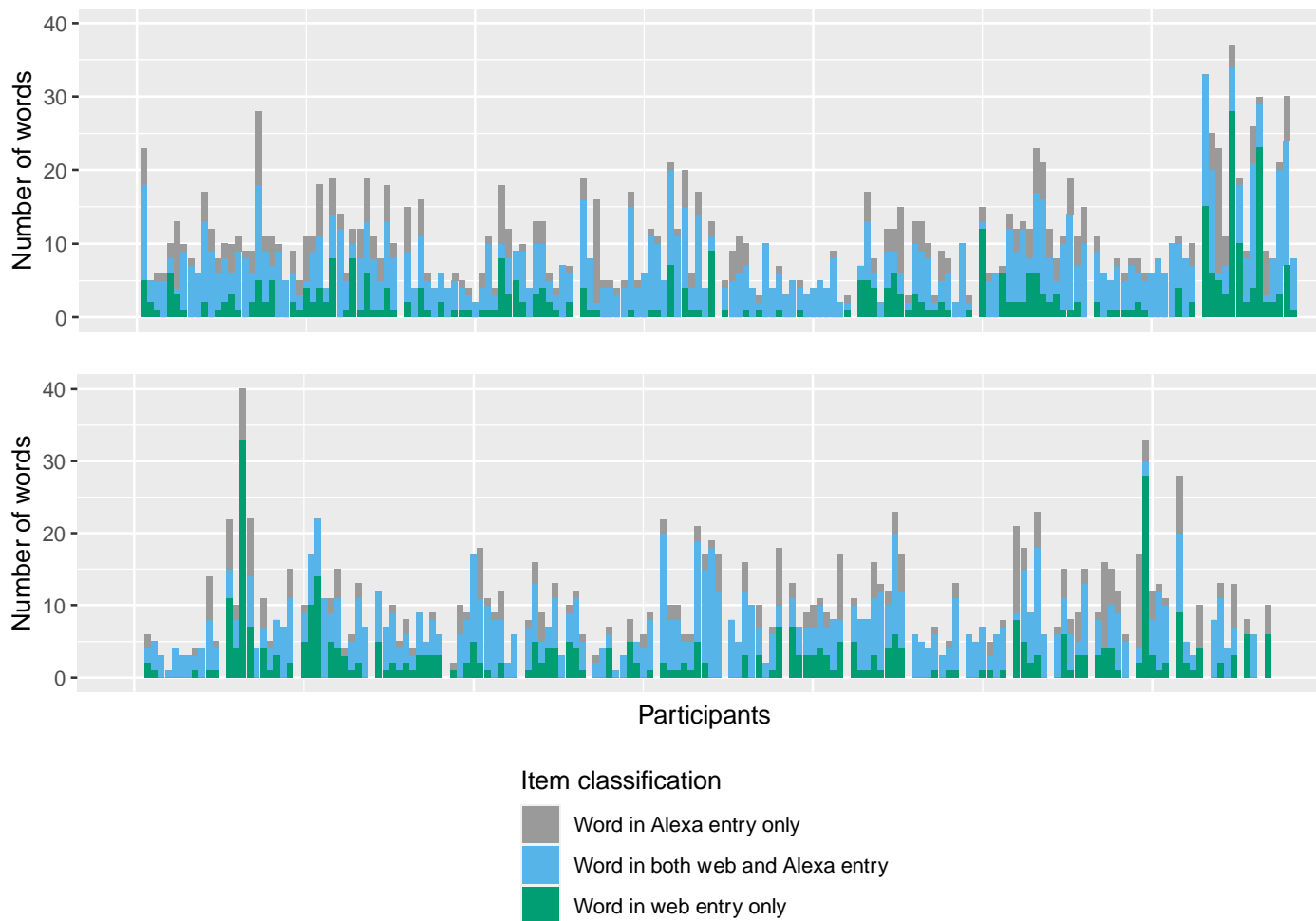


Figure 3: Plot comparing intake timestamps recorded using Alexa versus the web form



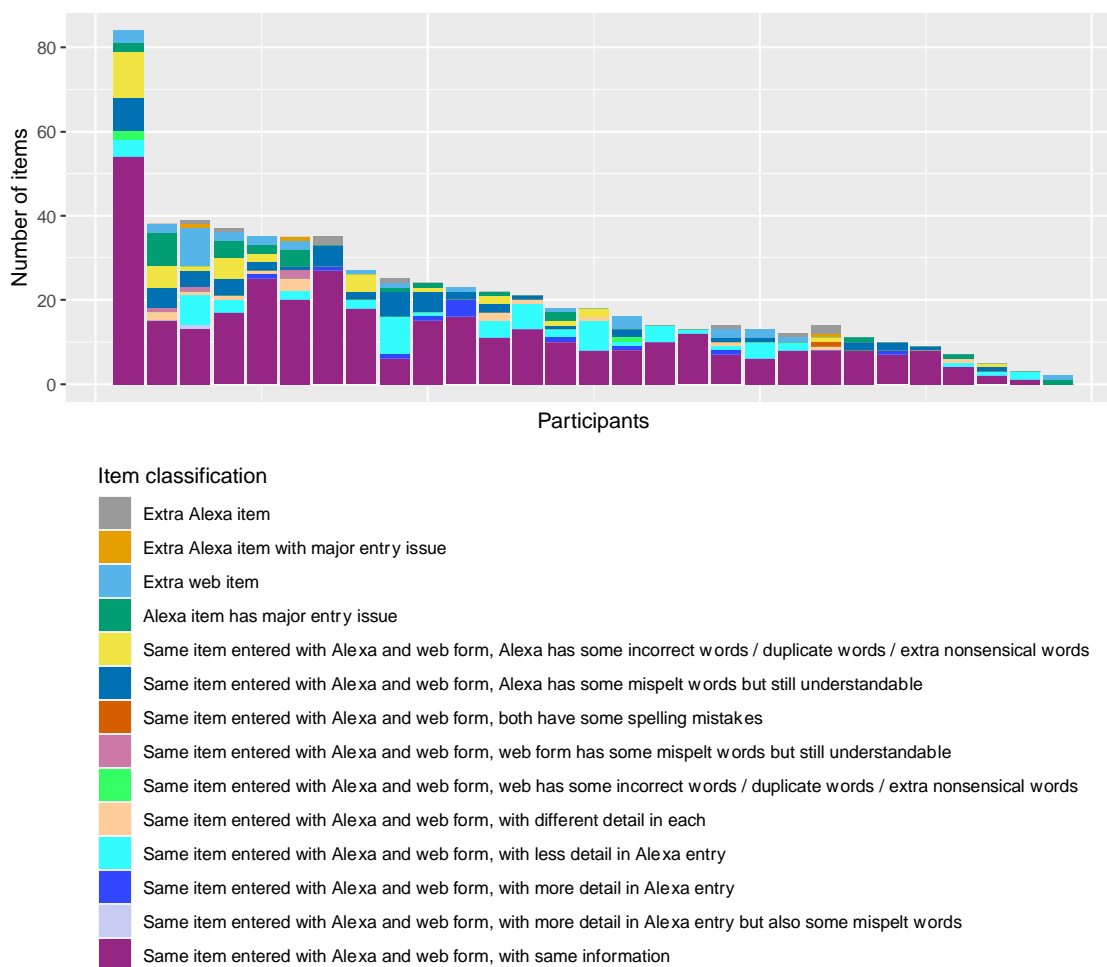
Each colour/symbol combination denotes a particular participant. Each point refers to an intake event for a participant, and shows the intake event time entered using the web form (x-axis) and the difference between the intake event time entered using Alexa versus the web form (y-axis).

Figure 4: Summary of automated comparison of the web and Alexa submitted food and drink information



Each stacked bar shows the number of unique words in a) both the web and Alexa entries, b) the Alexa entry only, and c) the web entry only. Each block of stacked bars shows the set of entries for a given participant.

Figure 5: Summary of manual evaluation of the submitted food and drink information



Each stacked bar shows the number of items in each category, for each participant. Manual evaluation conducted by LACM. Supplementary figure S7 shows comparison of assignments with those by independent assessors for random subsamples.

## TABLES

Table 1: Participant demographics

<b>Participant characteristic</b>	<b>N (%)</b>
Age band – <20	4 (11%)
– 20-29	16 (43%)
– 30-39	14 (38%)
– 40-49	2 (5%)
– 50-59	1 (3%)
Sex – N female	23 (62%)
Regional accent - No	31 (84%)
- A little	4 (11%)
- Yes	2 (5%)
Non-English accent - No	25 (68%)
- A little	9 (24%)
- Yes	3 (8%)
Has a voice-controlled device - No	16 (43%)
- Yes but it's used by others, not me	5 (14%)
- Yes and I use it	16 (43%)

These summaries use the usage sample, N = 37.

**Table 2: Post-participation questionnaire summary**

<b>Question</b>	<b>Total N</b>	<b>N (%)</b>
Reasons did take part – a) Interested in nutritional research	35	18 (51%)
– b) Wanted to support research.	35	27 (77%)
– c) Wanted to support others in the same research unit	35	12 (34%)
– d) Felt should be a research participant when asked, as they benefit from other participants.	35	11 (31%)
– e) Wanted to try a voice-controlled system	35	7 (20%)
– f) Wanted to try Alexa	35	12 (34%)
– g) Wanted to see how well Alexa could interact with them.	35	12 (34%)
– h) Wanted to see how well Alexa could record their dietary information.	35	17 (49%)
– i) Other reason *	35	2 (6%)
Participant was able to accurately tell Alexa what they ate and drank – Completely agree	35	3 (9%)
– Somewhat agree		16 (46%)
– Neither agree not disagree		2 (6%)
– Somewhat disagree		12 (34%)
– Completely disagree		2 (6%)
Participant was able to estimate accurate quantities describing how much they ate – Completely agree	35	3 (9%)
– Somewhat agree		11 (32%)
– Neither agree not disagree		10 (29%)
– Somewhat disagree		9 (26%)
– Completely disagree		1 (3%)
Participant chose not to record particular snacks/meals (e.g. because it was unhealthy) – Completely agree	35	1 (3%)
– Somewhat agree		5 (14%)
– Neither agree not disagree		0 (0%)
– Somewhat disagree		7 (20%)
– Completely disagree		22 (63%)
Participant sometimes chose to be selective with the truth – Completely agree	35	3 (9%)
– Somewhat agree		5 (14%)
– Neither agree not disagree		3 (9%)
– Somewhat disagree		6 (17%)
– Completely disagree		18 (51%)
Participant felt they remembered to submit food and drink information – Completely agree	35	9 (26%)
– Somewhat agree		19 (54%)



		1 (3%)
– Somewhat disagree		6 (17%)
– Completely disagree		0 (0%)
Other comments about your experience supplying dietary information via Alexa *	35	29 (83%)
Alexa interjected when I hadn't finished telling her when I ate or drank – Never	35	4 (11%)
– Rarely		13 (37%)
– Occasionally		11 (31%)
– Often		6 (17%)
– Always		1 (3%)
Alexa interjected when I hadn't finished telling her what I ate or drank – Never	35	2 (6%)
– Rarely		3 (9%)
– Occasionally		12 (34%)
– Often		17 (49%)
– Always		1 (3%)
Other things that happened when using Alexa *	35	30 (86%)
How convenient or inconvenient did you find providing information using Alexa? – Very inconvenient	35	4 (11%)
– Somewhat inconvenient		13 (37%)
– OK		10 (29%)
– Somewhat convenient		7 (20%)
– Very convenient		1 (3%)
How enjoyable or unenjoyable did you find providing information using Alexa? – Very unenjoyable	35	3 (9%)
– Somewhat unenjoyable		11 (31%)
– OK		15 (43%)
– Somewhat enjoyable		6 (17%)
– Very enjoyable		0 (0%)
How efficient or inefficient did you find providing information using Alexa? – Very inefficient	35	5 (14%)
– Somewhat inefficient		15 (43%)
– OK		6 (17%)
– Somewhat efficient		8 (23%)
– Very efficient		1 (3%)
How easy or hard did you find providing information using Alexa? – Could not use at all	35	0 (0%)
– Very hard		3 (9%)
– Somewhat hard		18 (51%)
– OK		7 (10%)
– Somewhat easy		7 (20%)

– Very easy		0 (0%)
Happy to use a voice-controlled system (e.g. Alexa) at home for research in the future - Yes	35	26 (74%)
- No		2 (6%)
- Not sure		7 (20%)
Not happy reasons: – Data privacy concerns around researchers collecting information about me	35	0 (0%)
– Data privacy concerns around Amazon collecting information about me.	35	2 (6%)
– Concerns that Alexa will inadvertently listen to other conversations, which may then be accessed by researchers	35	1 (3%)
– Concerns that the system will inadvertently listen to other conversations, which may then be used by the company that owns it (e.g. Amazon)	35	3 (9%)
– Other reason *	35	5 (14%)
Happy to use a voice-controlled system (e.g. Alexa) on a wearable device such as a smart watch, for research - Yes	35	28 (80%)
- No		1 (3%)
- Not sure		6 (17%)
Not happy reasons: – Data privacy concerns around researchers collecting information about me	35	1 (3%)
– Data privacy concerns around Amazon collecting information about me.	35	1 (3%)
– Concerns that Alexa will inadvertently listen to other conversations, which may then be accessed by researchers	35	1 (3%)
– Concerns that the system will inadvertently listen to other conversations, which may then be used by the company that owns it (e.g. Amazon)	35	3 (9%)
– I wouldn't like the idea of wearing a device all the time because of how it looks and / or feels.	35	2 (6%)
– Other reason *	35	1 (3%)
<b>Other approaches to collect dietary data</b>		
Participant has used another approach to record food and drink diary	35	25 (71%)
Participant has used a traditional diary	35	13 (37%)
Alexa Convenience compared with traditional diary – Much less convenient	13	3 (23%)
– A little less convenient		3 (23%)
– About the same		1 (8%)
– A little more convenient		5 (38%)
– Much more convenient		1 (8%)
Alexa how enjoyable compared with traditional diary – Much less enjoyable	9	2 (22%)
– A little less enjoyable		2 (22%)
– About the same		1 (11%)
– A little more enjoyable		3 (33%)
– Much more enjoyable		1 (11%)
Alexa efficiency compared with traditional diary – Much less efficient	9	4 (44%)

– A little less efficient		3 (33%)
– About the same		0 (0%)
– A little more efficient		1 (11%)
– Much more efficient		1 (11%)
Participant has used my fitness pal app on phone or tablet	35	12 (34%)
Alexa convenience compared with myfitnesspal app on phone or tablet – Much less convenient	12	3 (25%)
– A little less convenient		0 (0%)
– About the same		3 (25%)
– A little more convenient		6 (50%)
– Much more convenient		0 (0%)
Alexa how enjoyable compared with myfitnesspal app on phone or tablet – Much less enjoyable	11	2 (18%)
– A little less enjoyable		2 (18%)
– About the same		2 (18%)
– A little more enjoyable		5 (45%)
– Much more enjoyable		0 (0%)
Alexa efficiency compared with myfitnesspal app on phone or tablet – Much less efficient	11	3 (27%)
– A little less efficient		4 (36%)
– About the same		1 (9%)
– A little more efficient		0 (0%)
– Much more efficient		3 (27%)
Participant has used another method, or did not specify the method	35	5 (14%)

\* Participants who stated 'other' were able to then complete a free-text response, and these are summarized in Supplementary Section S4.

**Table 3: Non-participation questionnaire summary**

<b>Participant characteristic</b>	<b>N (%)</b>
Age band – <20	13 (19%)
– 20-29	36 (52%)
– 30-39	11 (16%)
– 40-49	5 (7%)
– 50-59	2 (3%)
– Prefer not to answer	2 (3%)
Sex – N female	51 (74%)
Reasons did not take part – Not available during the study session times	30 (43%)
– Data privacy concerns around Amazon collecting information on my diet	9 (13%)
– Data privacy concerns around researchers collecting information on my diet	3 (4%)
– Concerns that Alexa will inadvertently listen to other conversations	9 (13%)
– I don't eat or drink during my working hours	3 (4%)
– Picking up and returning the device was inconvenient	5 (7%)
– Other reason *	21 (30%)
Has a voice-controlled device - No	35 (51%)
- Yes but it's used by others, not me	10 (14%)
- Yes and I use it	24 (35%)
Happy to use a voice-controlled system (e.g. Alexa) at home for research in the future - Yes	42 (61%)
- No	13 (19%)
- Not sure	14 (20%)
Not happy reasons: – Data privacy concerns around researchers collecting information about me	10 (14%)
– Data privacy concerns around Amazon collecting information about me.	21 (30%)
– Concerns that Alexa will inadvertently listen to other conversations, which may then be accessed by researchers	16 (23%)
– Concerns that the system will inadvertently listen to other conversations, which may then be used by the company that owns it (e.g. Amazon)	20 (29%)
– Other reason *	3 (4%)
Happy to use a voice-controlled system (e.g. Alexa) on a wearable device such as a smart watch, for research - Yes	39 (57%)
- No	18 (26%)
- Not sure	12 (17%)
Not happy reasons: – Data privacy concerns around researchers collecting information about me	8 (12%)
– Data privacy concerns around Amazon collecting information about me.	19 (28%)
– Concerns that Alexa will inadvertently listen to other conversations, which may then be accessed by researchers	15 (22%)

– Concerns that the system will inadvertently listen to other conversations, which may then be used by the company that owns it (e.g. Amazon)	19 (28%)
– I wouldn't like the idea of wearing a device all the time because of how it looks and / or feels.	16 (23%)
– Other reason	0 (0%)

N = 69.

\* Participants who stated 'other' were able to then complete a free-text response, and these are summarized in Supplementary Section S4.

## REFERENCES

1. Boyd A, Golding J, Macleod J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013;**42**(1):111–127.
2. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med*. 2015;**12**(3):e1001779.
3. Lawlor DA, Lewcock M, Rena-Jones L, et al. The second generation of The Avon Longitudinal Study of Parents and Children (ALSPAC-G2): a cohort profile [version 2; peer review: 2 approved]. *Wellcome Open Res*. 2019;**4**(36).
4. Mason TB, Do B, Wang S, Dunton GF. Ecological momentary assessment of eating and dietary intake behaviors in children and adolescents: A systematic review of the literature. *Appetite*. 2020;**144**:104465.
5. Maugeri A, Barchitta M. A Systematic Review of Ecological Momentary Assessment of Diet: Implications and Perspectives for Nutritional Epidemiology. *Nutr*. 2019.
6. Hall M, Scherner P V, Kreidel Y, Rubel JA. A Systematic Review of Momentary Assessment Designs for Mood and Anxiety Symptoms. *Front Psychol*. 2021;**12**.
7. Simpson E, Bradley J, Poliakov I, et al. Iterative Development of an Online Dietary Recall Tool: INTAKE24. *Nutrients*. 2017;**9**(2).
8. Martin CK, Nicklas T, Gunturk B, Correa JB, Allen HR, Champagne C. Measuring food intake with digital photography. *J Hum Nutr Diet*. 2014 Jan 1;**27**(s1):72–81.
9. Lawlor DA, Lewcock M, Rena-Jones L, et al. The second generation of the avon longitudinal study of parents and children (ALSPAC-G2): A cohort profile [version 2; peer review: 2 approved]. *Wellcome Open Res*. 2019;**4**.
10. Cade JE, Warthon-Medina M, Albar S, et al. DIET@NET: Best Practice Guidelines for dietary assessment in health research. *BMC Med*. 2017;**15**(1):202.
11. Rudis B, Embrey B. R Package 'pluralize' [Internet]. 2020. Available from: <https://cran.r-project.org/package=pluralize>
12. Jian W-S, Wang J-Y, Rahmanti AR, et al. Voice-based control system for smart hospital wards: a pilot study of patient acceptance. *BMC Health Serv Res*. 2022;**22**(1):287.
13. Sharma A, Oulousian E, Ni J, et al. Voice-based screening for SARS-CoV-2 exposure in cardiovascular clinics. *Eur Hear J - Digit Heal*. 2021 Sep 1;**2**(3):521–527.
14. Weiss R. Mobile Food Intake Visualization and Voice Recognize (FIVR) [Internet]. Available from: <https://grantome.com/grant/NIH/U01-HL091738-04S2>
15. CLOSER. Preparing for the future: Tackling the key challenges facing the UK's longitudinal population studies [Internet]. 2019. Available from: <https://www.closer.ac.uk/wp-content/uploads/20190401-Preparing-for-the-future-Conference-report.pdf>