From automated to data-driven large-scale dietary assessment

Timur Osadchiy

Supervisor: Dr. Y. Guan
Dr. I. Poliakov
Prof. P. Olivier
Dr. P. Lord

Advisor: Dr. E. Foster

School of Computing
Newcastle University

This dissertation is submitted for the degree of
Doctor of Philosophy

January 2020

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 57,000 words including bibliography, footnotes, tables and equations and has 53 figures.

Timur Osadchiy
January 2020

Acknowledgements

I would like to thank the School of Computing and the Faculty of Science, Agriculture and Engineering of Newcastle University for funding my research.

I would like to thank my now external supervisor Patrick Olivier who encouraged me to pursue this degree and to become a proud member of the Open Lab. Thank you for many opportunities and challenges that helped me to build my research and software engineering experience.

Special thanks to my supervisor Ivan Poliakov for sharing his massive software engineering experience with me, introducing me to concepts of functional programming and helping me to believe in myself. You are a true friend. Most of my research would not happen without your support and help.

A huge thanks to my supervisor Yu Guan and Pete Wright for guiding and supporting me in my final but probably the most important year of my PhD journey.

I would also like to thank my former supervisor Emma Foster and Maisie Rowland for guiding my way in the fields of human health, nutrition and dietary assessment. I would like to thank all members of the Open Lab. It is an amazing and truly inspiring environment to work in.

I would like to thank all of my friends and especially Dinislam, Stas, Maksim and Alexey. You guys are awesome! I am so grateful for meeting my Australian friends Geremy and Ashley. Thank you for your support and inspiration!

I wish to thank my grandmother Flora who is always full of energy and teaches me to stay positive no matter happens. I would like to thank my aunt Dilyara and my cousin Nariman for their love and support. I would like to thank my uncle Adel for being my example of success. I would like to thank my mother in law Irina for giving the best life to our doggo Bush, who we could not bring to the UK. I would like to thank all my family. You all are my heroes. I miss you a lot.

So much thanks to my mum Gulshat for raising, protecting and motivating me. There is no way to describe how lucky I am that you made everything possible to give me the best in my life. I would not be where I am today without you.

I would absolutely not make it without the support from my loving wife and angel Ekaterina, who patiently walked this journey with me. You are my wings, my energy, my inspiration.

Abstract

Dietary assessment surveys are an important tool for measuring and/or monitoring the nutritional profile of a population. The analysis of data that is collected in these surveys helps to develop health care guidelines and policies that minimise the risk of diet related diseases on a national scale. For years these surveys had to be conducted in a form of an interview by trained researchers with a nutritional background. The emergence of systems that automate interviewer-led protocols and transform these interviews into online surveys has addressed financial limitations and brought scalability into dietary assessment studies. In the meantime, online dietary assessment surveys mostly copy the interviewer-led procedures and inherit some of their methodological issues that lead to misreporting of dietary intake and lower the accuracy of assessment. This thesis primarily focuses on the issues related to human-memory, motivation of respondents to take part in dietary assessment studies, and the usability of survey interfaces. This work pinpoints the elements of automated dietary assessment systems, where these issues affect the accuracy of results. This analysis is then translated into three research questions of this thesis. Challenges related to human-memory are then addressed by developing and evaluating a recommender system for prompting omitted foods in online dietary assessment surveys. This work also explores short retention intervals (i.e. time between an intake and recall) as another method for recall assistance. As a way to motivate respondents to take part in dietary assessment surveys this thesis explores tailored dietary feedback provided to respondents at the end of a survey. Usability and performance of new methods are analysed in real-life dietary assessment surveys using a usability framework developed for this research. Acceptance of the methods is analysed using thematic analysis of transcribed interviews with respondents. Research activities conducted during this work provide some support to hypotheses defined in the research questions.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Background. Why is large-scale dietary assessment important?

Diet is a major life-style factor in our health and longevity globally affecting health care systems and economic development [150]. Certain patterns of dietary habits at different stages of our life are related to a wide range of chronic diseases, including cancer and cardiovascular disease; and nutrient-dense diet is associated with lower risk of all-cause mortality [147]. Diet and nutrition are estimated to account for approximately 30% of all cancers in developed countries and 20% in developing countries [155] and according to the World Cancer Research Fund (WCRF), there is strong evidence of an association between specific diet-related factors and increased risk of cancer [105]. The Mediterranean diet, a dietary pattern considered to be healthy, categorized in particular by a high consumption of vegetables, fruit, nuts and unsaturated fats like olive oil has been associated with a reduced risk of non-communicable diseases [104]. More specifically, a diet high in dietary fibre (found in foods such as wholegrain cereals) has been found to reduce the risk of colorectal cancer [20], however the UK national dietary survey - The National Diet and Nutrition Survey (NDNS) Rolling Programme 2014/15-2015/16 found that only nine percent of adults aged 19-64 years met the recommended intake of dietary fibre [134]. Micronutrients are also key in reducing risk of disease; iron is a major component of haemoglobin and an essential mineral in the diet enabling the distribution of oxygen around the body among many other functions. Iron deficiency anaemia can lead to many health problems and during pregnancy, is associated with low birth weight of the child and increased risk pre-eclampsia and bleeding in the mother [10]. Women of reproductive age are particularly at risk of iron deficiency due to menstrual losses and the NDNS 2012/13 - 2013/14 found that 48% of girls aged from 11 to 18 years had iron intakes below the lower recommended nutrient intake [134].

In addition to nutritional deficiencies, over nutrition is a public health epidemic. Statistics from the UK National Health Service (NHS) show that there were 617,000 admissions with a primary or secondary diagnosis of obesity in 2016/17, which is an 18% increase from 2015/16 [7]. The latest data published by NHS digital in 2019 stated that 29% of adults were classified as obese in 2017, and one in five children in Year 6 were classified as obese in 2017/18 [8]. Obesity is one of the main determinants of an increased risk of CVD [163] and in 2012, CVD was the leading cause female deaths in the UK (28% of all female deaths) and was the second most common cause of male deaths (29% of all male deaths) [25]. Overweight and obesity related ill health was estimated to cost the NHS £6.1 billion in 2014-15 [2]. The ability to identify and examine correlations between dietary habits and health on a national scale is essential and therefore, adequate, reliable and economically effective tools to measure population dietary intake are important in identifying these relationships.

Adequate dietary intake assessment is not only important in identifying relationships between dietary habits and well-being. Where associations are already known healthcare systems need reliable and economically effective tools to monitor the intake of populations. An accurate and up-to-date picture of a nation's nutritional status is essential for implementing an effective intervention strategy and evaluating it. An example of such programme is the aforementioned UK National Diet and Nutrition Survey (NDNS) programme, first set up in 1992 and jointly funded by Public Health England and UK Food Standards Agency [169]. NDNS is a rolling survey designed to collect quantitative data on the food consumption, nutrient intake and nutritional status of the general population of the UK aged 1.5 years and over. The aim of the NDNS is to estimate the proportion of individuals that follow dietary recommendations and the proportion with potentially compromised nutritional status. Data collected is also used to track the progress towards existing dietary targets and to identify areas of concern. Collected data also serves as the basis for research and the development of intervention programmes [169]. The NDNS has used different methods to collect dietary data, previously 7-day weighed food diaries (WFD) were used, and in 2014/15-2015/16 the decision was made to change to age specific 4-day non-WFD [4].

1.2 How is diet assessed?

There are a number of different methods for assessing diet of an individual by either measuring markers of nutrient intake or by measuring the intake of foods and drinks. An accurate dietary assessment ideally represents the true habitual intake of the individual. The habitual intake of a person can be described as an average of dietary intake over a prolonged period of time (i.e. weeks or months rather than days) [94]. Habitual intake is expected to provide energy and other nutrients that maintain weight stability, a steady physiological state, nutritional status and health of the individual in both the short and long term. However, measuring habitual intake is a highly complex task due to natural variations in diets of subjects [94]. For that reason, in practice, an accurate dietary assessment method is expected to measure the true intake during the period of a study. On a large-scale a dietary assessment method needs to capture a representative (average) habitual intake for a population and for its various demographical groups (e.g. age, gender, socioeconomic status).

Dietary assessment methods can be broadly split into three categories, nutrient biomarkers, direct observations and subjective methods (i.e. self-reported intake by subjects) [176, 147]. Specific nutritional biomarkers can be used to measure intake of some nutrients (e.g. fat and fatty acid intake) and energy can be assessed through energy expenditure measured using doubly-labelled water (DLW), assuming the participants are weight stable [16, 27]. Previous studies have demonstrated a number of advantages of biomarkers including a relatively high accuracy for assessing intake of some nutrients (such as energy and protein), the lack of social desirability bias, independence of subject's memory and ability to self-report their intake [128, 55]. Nevertheless, this method imposes a number of practical and economical challenges including the need to collect, store and analyse blood, urine, or other biological specimens for most of the nutritional biomarkers [159]. Biomarkers are still highly useful in calibrating the measurement error in dietary reports and, for example, DLW is widely considered as a gold standard for estimating energy intake that other dietary assessment methods are compared with [27, 94, 34]. Another approach to collecting data for dietary assessment involves skilled researchers directly observing and recording subject's intake [147]. Records are often collected in subject's home environment and can include not only information about food consumption but also about preparation methods. This method is especially useful in developing countries with subjects with literacy difficulties or where food is prepared in large quantities for a group of people (e.g. family). However, direct observations are very costly and impractical on large scale.

A successful dietary assessment method that can be applied on a large-scale is expected not only to be cost-effective, scalable and to estimate dietary intake with acceptable accuracy, but also to impose a low subject burden to reduce the likelihood of misreporting, participant attrition and changes in subjects' diets [100]. In other words, the method, ideally, needs to ensure that the choice of reported foods and drinks has not been influenced by the act of recording or being observed and that subjects would consume exactly the same foods and drinks were they not involved in a study [94]. For those reasons, methods that are based on subjective self-reported intake that can be applied in a form of surveys are widely used for population dietary assessment.

Subjective dietary assessment methods include WFD, food frequency questionnaires (FFQ) and 24-hour recall [147]. The WFD method requires subjects to weight and record each item of food and drink consumed before and after every eating occasion. To collect accurate records this method assumes subjects have access to scales at the time of preparing their food and are able to use them competently [135, 161]. Thus, WFD require a high level of motivation and pose a relatively large burden on subjects [147]. Respondents surveyed with this method were found to reduce the number of consumed foods and snacks and to substitute their normally consumed foods by those that are simpler to record [132]. For these reasons, portion size estimation through validated photographs of foods with known serving sizes and through common household measures and dishes (e.g. cups, spoons, bowls, and glasses) is considered to cause less burden to respondents [118, 135, 161, 161]. This method of portion size estimation is for example commonly used in FFQ and 24-hour recalls. FFQ enable individual's intake to be captured over a long period of time (e.g. a month or a year) in a relatively simple, cost- and time- efficient manner [159, 147]. FFQ is an advanced form of a checklist that normally contains 100-150 foods and asks a subject to report, which foods and the amount of them they consumed over a specific period [147]. Answers for the questionnaire can be collected either through an interview or via a self-administered approach (e.g. postal or online survey). In contrast to the WFD, FFQ poses a low subject burden [159]. The accuracy of this method, however, is relatively low. FFQ is prone to misreporting as the method relies on the ability of subjects to remember their diet over a long period of time [85, 159]. One of the most widely adopted approaches is the multiple-pass 24-hour recall, which is considered to offer a favourable balance of a high accuracy and low subject burden [77]. This method was designed as an interviewer-led method to collect information about all foods and drinks individuals consumed for a previous day [69]. The multiple-pass 24-hour recall was developed by the US Department of Agriculture (USDA)-Human Nutrition Information Service (HNIS) to limit the extent of under-reporting that occurs with self-reported food intake [69, 78]. Capturing habitual intake using this method requires multiple non-consecutive interviews to be conducted over a long period of time (e.g. weeks, months). To cover a wider variation of foods in a subject's diet this method can also be combined with an FFQ [147]. The estimation of energy intake with the multiple-pass 24-hour recall is relatively accurate when validated with DLW [97]. However, as with FFQ's, 24-hour recalls are also prone to misreporting due its reliance on subjects self-reporting their dietary intake [159].

The interviewer-led nature of a 24-hour recall involves the need for skilled professionals to conduct interviews and analyse complex dietary data which poses economic and scalability implications. To address those issues a number of systems have been developed that replace an interviewer in the 24-hour recall method with an online survey [77, 154, 39, 30]. The process of collecting and processing dietary data using software is generally referred to as "automated dietary assessment" [159]. Automated dietary assessment allows scaling a survey to thousands of subjects and receiving immediate results in a short time frame at a relatively low cost [149, 26]. Another advantage is that the procedure of the 24-hour recall is strictly standardized. Moreover, a database of foods can be adapted for a specific country, population group or for the purpose of a particular study (e.g. medical intervention or a disease) [149, 26]. One such dietary assessment system, Intake24, has been developed by a multidisciplinary team from the fields of Nutrition, Human-computer interaction (HCI) and Medical Statistics at Newcastle University [30]. The accuracy of nutrient intake estimates by the system has demonstrated an agreement with the interviewer-led multiple-pass 24-h recall [30, 139, 56].

1.3 Why is measuring dietary intake still a problem?

There are a number of key elements of subjective dietary assessment methods where their accuracy is affected. Gathering information on habitual intake requires capturing every food and drink consumed with an accurate estimation of their portion sizes. Dietary assessment methods that rely on self-reported intake depend on an individual's motivation and willingness to report, their ability to remember and accurately estimate portion sizes [100, 29]. The accuracy of assessment also depends on all components of the method that are used for collecting dietary data and calculating the results (e.g. the complexity of a questionnaire, food composition database, portion size estimation methods).

Misreporting and its most common form under-reporting are among the fundamental issues with producing accurate dietary assessment results based on self-reported intake [100]. Macdiarmid and Blundell described two categories of under-reporting: intentional and unintentional [100]. Intentional misreporting may be caused by social desirability bias leading to some individuals knowingly under-reporting foods that may be seen as 'unhealthy' or choosing not to report them at all [147]. Dietary assessment methods that involve long questionnaires require a considerable amount of time and effort from respondents. Thus, intentional under-reporting may be also contributed by the lack of motivation of respondents to go through a survey [63, 101].

Unintentional under-reporting in dietary assessment is largely due to lack of attention and by errors of human memory that cause omissions of foods and inaccuracies in estimation of serving sizes [66, 116, 117]. Memories about consumed foods and drinks start fading an hour after consumption [24, 84]. For that reason, a dietary assessment method that requires subjects to remember the intake for the previous day or for longer periods of time is prone to memory errors [100]. Misreporting may also occur when subjects are genuinely not aware of some details about the food they ate. For example, when they are asked about recipes used in preparation of reported foods but they did not cook the meal themselves [66]. Another source of under-reporting is the lack of training for participants when taking part in dietary assessment studies [166]. For example, some individuals may not fully understand the level of detail needed for a study (e.g. some may think that coffee with milk and sugar is the same as without them). In addition to that, the use of technology in dietary assessment is characterised by a range of additional limitations in populations with a low level of technological illiteracy and/or with limited access to technology (e.g. computer, mobile device) [147, 129, 129]. Lastly, both types of misreporting (intentional and unintentional) can be confused with genuine low-energy intake reports caused by factors such as illness, weight loss diets or irregular eating patterns [127]. Genuine under-reporting may also occur when respondents start eating less during the course of a study as a result of completing a prospective method such as a food diary and becoming more aware of how much they were eating [147].

Some level of error in dietary assessment that is based on in self-reported intake is inevitable [147, 75, 64]. A reliable method has to measure the nutritional profile of a population or an individual accurately enough to be able to reveal related health implications [133, 70]. An acceptable level of accuracy will likely depend on the purpose of a study (e.g. monitoring of the nutrient status of a population, individual dietary intervention, disease prevention) [75].

1.4 Aim of this thesis

The use of online surveys has greatly simplified the logistics of dietary assessment and increased its scalability while bringing the costs down. However, despite their increasing use, the development of dietary assessment systems is still in its infancy. The designers and developers of the systems have gone a little beyond simple mimicking interviewer-led approaches and have not to date systematically addressed matters of either user experience or usability that impact on accuracy of results. The next step for automated dietary assessment is to embrace the benefits enabled by other fields of computing including data science and HCI. The goal of this thesis is to explore new digital methods for improving the accuracy of automated dietary assessment systems. The thesis achieves this aim by answering three research questions.

1.5 Research questions

1.5.1 Research question 1 (RQ1). How can recall assistance be improved and evaluated in automated dietary assessment?

Existing dietary assessment systems mainly automate long-established interviewer-led procedures. In the meantime, existing protocols for collecting dietary data from respondents are largely dictated by the logistics and complications involved in using those interviewer-led approaches. For example, one potential reason for the 24-hour time frame in 24-hour recall could be that it allows the interviewer to collect dietary information on a single occasion. At the same time, the 24-hour period is short enough for an individual to remember meals consumed by them. However, as it was mentioned above the accuracy of recall may reduce even an hour after an eating occasion [24, 84]. Replacing the interviewer with an online survey gives an opportunity to rethink prevailing methods of recall assistance and discover new methods. For example, Baxter et al. have demonstrated that shortening the time period between an intake and an interview (i.e. retention interval) may be beneficial for the accuracy of recall [22, 23]. In an online survey that approach can be taken to a next step and respondents can be asked to record multiple recalls per day. Recording meals shortly after the intake could potentially reduce stress on the memory of respondents. At the same time, changes to previously validated dietary assessment methods need to be examined for potential negative/positive effects on their accuracy. The known approaches to evaluation of dietary assessment methods include comparison of estimated energy to that measured by interviewer-led methods or by DLW [78, 34, 30]. However, these approaches pose considerable financial and practical implications. Most importantly, these cannot be used to analyse issues with usability of newly implemented techniques and of dietary assessment systems in general. The usability of dietary assessment system may affect the accuracy of results through reduced/increased complexity of the procedure for respondents [139]. For that reason, the focus of this research question includes the development of a framework that provides quantitative indicators for the evaluation of accuracy and usability of designed recall assistance methods.

1.5.2 Research question 2 (RQ2). Can data driven methods facilitate the accuracy of dietary assessment?

Interviewer-led protocols employ various methods to assist respondent's recall. For example, in a 24-hour recall, if there is a considerably long time period in the report during which the respondent has not reported any meals, the interviewer is expected to ask about potentially omitted eating occasions (e.g. snacks or drinks). Similarly, the interviewer asks questions about foods commonly eaten with foods reported by the respondent (associated foods). For instance, if respondent reports having eaten toast, the interviewer asks whether it was eaten with butter. Dietary assessment systems commonly mimic that behaviour of the interviewer and display such prompts and reminders to respondents. Food association rules that trigger prompting are currently either 'hard-coded' or entered manually by a professional with a nutritional background. As in interviewer-led dietary assessment those associations are based on experience from previous interviews with respondents [112, 30]. However, eating habits depend on region, culture, diet, and a number of other factors, thus generalized rules for prompting questions are not necessarily the most effective approach where systems are deployed beyond the narrow population for whom they are generally developed for. New foods and recipes emerge and dietary trends change over time, which requires food associations to be constantly revised. At the moment of writing this thesis no published study has evaluated the appropriateness of hand-coded food associations or explored alternative approaches to discovering them and to prompting for associated foods. It seems that technology assisted dietary assessment is missing the opportunity to apply data-mining and machine learning techniques to data collected from previous dietary surveys. Dietary assessment systems could then build dynamic knowledge-based models and adapt more easily to specific contexts. For example, a system could learn about foods that are statistically very likely to be consumed together (i.e. associated foods), or during a certain time of the day (e.g. fried eggs for breakfast), or by respondents of particular demographics. Such models can then be used to probe users about potentially forgotten foods after they reported part of a meal. Furthermore, these dynamic models could facilitate the usability of some other key elements of the system, for example, the elements responsible for searching for foods in the database by respondents. Dietary assessment systems store thousands of foods in their databases and a lot of the foods have similar names (e.g. different types of bread). The provision of adequate search results in response to free text queries of respondents is pivotal for the accuracy of assessment [58]. Respondents failing to quickly identify their foods may choose to report a food with a similar name but different nutritional content or to skip reporting that item at all. The knowledge of items that are more likely to be selected next based on the previous selection of foods and a given context may increase the relevance of presented search results.

1.5.3 Research question 3 (RQ3). Can tailored dietary feedback improve participatory engagement in online dietary assessment surveys?

Recall assistance methods alone cannot stop respondents who intentionally report lower or higher than actual intake for some foods as well as those who omit other foods. One reason for such behaviour is social desirability bias [147, 101]. That is a subject feeling uncomfortable (i.e. feeling guilty) to report a food or a portion size that might be outside of accepted societal norms (e.g. fast food). Similarly social desirability bias may cause respondents to exaggerate their intake of healthy foods. The complexity of a survey is another known factor affecting the accuracy of dietary assessment [147, 63, 101]. Some subjects may choose to drop out from a study completely which reduces the sample size of a study and may impact on the representativeness of the study population. Others may skip reporting some foods to complete the survey faster. The accuracy of self-reported intake of an individual can be analysed using direct observations or nutritional biomarkers [27, 94, 34]. However, there are no known methods to detect intentional or unintentional under-reporting on a large scale. Nevertheless, automated dietary assessment could take lessons from research in the area of participatory engagement in crowdsourcing tasks including online surveys. For example, according to Nakamura and Csikszentmihalyi respondents who see a personal gain in outcomes of a task are more likely to be motivated to complete it [115]. This is to some degree supported by another study by Powers et al. that demonstrates that providing feedback on questionnaire results at the end of a survey increases participation rates [130]. What form of personal value could be provided in a dietary assessment survey? There is a growing number of people interested in technology for aggregating information about their well-being (including dietary data) and analysing it (e.g. Fitbit, MyFitnessPal) [45, 62, 98]. Thus, in a dietary assessment survey, that personal value could be provided in a form of individual feedback about the nutrient intake of participants based on their responses. In other words, an opportunity to receive a higher quality of dietary feedback could motivate respondents to provide more accurate answers to a dietary survey. A higher quality of responses on an individual level could lead to a higher quality of dietary assessment on the level of a population. Thus, RQ3 focuses on understanding whether tailored dietary feedback can be used as a driver for participatory engagement in population dietary assessment surveys and identifying characteristics that facilitate acceptability and usability of dietary feedback. At the same time, the answer to the RQ3 has to fit into practices of the conventional approach to conducting dietary assessment studies, which suggests to refrain from either positive or negative feedback about dietary habits of respondents since this may change their diets and limit the reliability of the outcomes [157, 100].

1.6 Contribution to knowledge

This thesis explores new approaches for improving the accuracy of data collected in large-scale dietary assessment surveys conducted online. These approaches are implemented and evaluated in a system that collects data from respondents using a multiple pass 24-hour recall method [69, 78]. However, most answers to research questions defined in this thesis are intended to be applicable to other methods of gathering dietary data that are used in a form of online surveys. For example, the usability framework that is described in in chapter 3 and applies HCI research in the field of user experience and usability to the evaluation of recall assistance methods can potentially be relevant for online surveys that implement the FFQ method. Same applies to the recommender system for prompting omitted foods that is described in chapter 4 and in chapter 5. Similarly, other types of online dietary assessment surveys could benefit of the use of tailored dietary feedback as an incentive that is explored in chapter 7 and takes its inspiration from research in HCI for health and personal informatics. In the meantime, some of this research could find applications outside of the scope of automated dietary assessment. The recommender system developed for the purpose of prompting survey respondents about foods potentially omitted by them could be used in other contexts that do not collect or store an extensive history of user behaviour and/or where the range of indicators representing user interests is limited. For example, that could be systems characterised by high privacy concerns or irregular usage. Lastly, new components that were developed in Intake24 in the course of producing this research and that are described in chapter 3 are planned to be used in future dietary assessment studies including the next UK NDNS by the National Centre for Social Research (NatCen) and MRC Epidemiology Unit at the University of Cambridge [169]. Different parts of this thesis were rewritten by the author of this thesis into multiple scientific publications. Relevant publications are also acknowledged at the beginning of corresponding chapters.

[122] Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., and Foster, E. (2019a). Recommender system based on pairwise association rules. *Expert Systems with Applications*, 115:535–542

[123] Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., and Foster, E. (2019b). Validation of a recommender system for prompting omitted foods in online dietary assessment surveys. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 208–215. ACM

[pending publication] Osadchiy, T., Poliakov, I., Olivier, P., Foster, E., and Rowland, M. Progressive 24-hour recall: Feasibility and acceptability of short retention intervals in online dietary assessment surveys. *Journal of Medical Internet Research*

1.7 Thesis structure

1.7.1 Introduction

This chapter provides a background for this research, outlines existing challenges in large-scale dietary assessment, which are translated into research purpose and questions. Additionally, this chapter outlines contribution of this thesis and provides its structure.

1.7.2 Literature Review

This chapter provides a detailed review of appropriate literature pertaining to dietary assessment systems and the research questions defined in chapter 1. It also covers research that provides potential solutions for the challenges contained in the research questions.

1.7.3 Methodology

This chapter provides a detailed description of a methodology that was undertaken to inform the research questions defined in this thesis.

1.7.4 Recommender system based on pairwise association rules

This chapter informs the answers to RQ1 and RQ2. This chapter describes the development and an offline evaluation of a recommender system for prompting omitted foods in online dietary assessment surveys.

1.7.5 Validation of a recommender system for prompting omitted foods in online dietary assessment surveys

This chapter extends the answer to RQ2 and describes the deployment and the evaluation of the recommender system in real online dietary assessment surveys.

1.7.6 Progressive 24-hour recall: Feasibility and acceptability of short retention intervals in online dietary assessment surveys

This chapter extends the answer to RQ1 and describes the implementation of short retention intervals in an online dietary assessment survey as a method for reducing a burden on human memory. The chapter describes the observed effects of the method on data collected in the survey and the acceptability of this method by respondents.

1.7.7 Tailored dietary feedback as an incentive in large-scale dietary assessment surveys

This chapter informs the answer to RQ3 and extends the answer to RQ1. This chapter explores the use of tailored dietary feedback to incentivise respondents to provide more accurate answers in online dietary assessment surveys.

1.7.8 Discussion, Relevance, and Conclusion

This chapter reviews the major outcomes of this thesis, limitations of these outcomes, and describes the implications for research in the area of online dietary assessment surveys.

Chapter 2

Literature Review

2.1 Abstract

This chapter provides an overview of the multiple pass 24-hour recall method. The chapter also reviews the general architecture of dietary assessment systems that implement the 24-hour recall method. The description of the architecture includes the key components of such systems, survey interface for respondents and existing methods of recall assistance, researcher interface and ontology of foods. The description of the architecture is based on three existing dietary assessment systems ASA24, MyFood24 and Intake24. The chapter also examines challenges that affect the accuracy of automated dietary assessment systems. Finally, the chapter explores directions for addressing those challenges.

2.2 The multiple-pass 24-hour recall

The original 24-hour recall is an interviewer-led method for inquiring information about subject's food intake for the past 24 hours. Compared to other dietary assessment methods the 24-hour recall poses relatively low subject burden [77, 78] and is less likely to cause changes in a respondent's diet in the course of a study, misreporting and participants' dropouts [100]. Self-reported food intake is highly dependent on human memory, which makes it prone to omissions and misreporting [100, 159]. To limit the extent of under-reporting that occurs with self-reported food intake the US Department of Agriculture (USDA)-Human Nutrition Information Service (HNIS) developed the multiple-pass modification of the 24-hour recall method [69, 78]. The multiple-pass 24-hour recall interview generally consists of five distinct passes (i.e. steps). In the first, so called "quick-list" pass, the respondent is asked to recall all foods they ate the previous day. In the second pass, the interviewer probes about possible forgotten and associated foods (e.g. beverages, sweeteners, spreads, condiments). In the third pass, the respondent is asked about time the foods were eaten and eating occasions. In the fourth pass, the respondent is asked to give more details about those foods. The interviewer may ask, for example, to clarify ingredients of foods or specific brands, if that affects their nutrient content. As another example, if the respondent reports porridge, the interviewer asks whether it was cooked with water or milk, and if so the type of milk, and the amount added. In this pass the interviewer also asks the respondent to estimate portion sizes of foods. Portion sizes can be estimated using common household measures and dishes (e.g. cups, spoons, bowls, and glasses), geometric shapes (e.g. circles, rectangles, wedges) and food-specific gram weight descriptions. Portion sizes can also be estimated using validated photographs of foods with known serving sizes [161]. In addition to that the interviewer uses various prompts, such as questions about left over foods (e.g. Did you leave any of your porridge?). In the fifth pass the interviewer reviews the list of meals and foods, and probes for additional eating occasions. For example, if there is a long time gap between lunch and dinner, the interviewer asks whether any afternoon snacks were eaten in that period.

Such interviews are normally conducted by professionals with a nutritional background either face-to-face or over the phone. To estimate nutrient contents of the reported foods a composition table is used that contains most common foods and their validated nutrient values for the region where the study is conducted [54]. Having collected foods and portion sizes researchers can map them to records in a composition table to estimate the nutrient content for each individual dietary report. The 24-hour recall is commonly used to estimate nutrient intake for a population rather than on an individual level. For that reason, the interviews are typically conducted on three to four non-consecutive days to capture a wider variety of foods in the diet, including foods which are important contributors to nutrient intake but are less frequently consumed (e.g. oily fish). The time interval between the days of the survey depends on the purpose of the study. For example, to garner habitual intake over a year the survey has to be administered across all seasons within the year to capture seasonal variations [36, 28]. For the most accurate assessment respondents should be not aware about specific days of interviews, since they might change their eating habits on those days otherwise. Hence, this procedure involves considerably complex planning, logistics and costs due to intense labour and involvement of high skilled interviewers.

A number of computer-based systems that automate the multiple pass 24-hour recall method have been developed to address scalability and economic implications of the interviewer-led method including ASA24, Myfood24 and Intake24 [39, 154, 30]. Such systems provide an interface for respondents that standardizes the multiple-pass procedure and replaces an interviewer in the original method. Respondents submit their intake remotely without the additional burden of having to be present for a face-to-face or telephone interview. Population intake of nutrients can be automatically estimated from submitted recalls. Further the chapter describes the architecture of dietary assessment systems based on examples of three existing systems, ASA24, Myfood24 and Intake24. The elements of the systems covered in the description include survey interface for respondents and existing methods of recall assistance, researcher interface and ontology of foods. Intake24 was developed in Newcastle University, which allowed a wider access to the system in the course of writing this thesis. For that reason, the description of the survey interface is based on all three systems, whereas the description of the researcher interface and the ontology of foods is based only on Intake24.

2.3 Automated multiple-pass 24-hour recall

2.3.1 Overview

ASA24, MyFood24 and Intake24 are three systems that are based on the multiple-pass 24-hour recall method and are widely used for conducting online dietary assessment surveys [39, 154, 30, 37]. These and similar systems offer a multitude of benefits to researchers in the field of human health and nutrition. An online dietary assessment survey essentially replaces a trained interviewer while collecting the same information and probing for additional details as if it was an interviewer-led recall. Instead of scheduling and conducting face-to-face interviews researchers can disseminate a link to an online survey by email to respondents with instructions on completing it. Thus, dietary data can be simultaneously collected from large and geographically spread cohorts. In addition to that, the link to the survey can be sent right when the respondent is expected to complete their recall meaning that they are not aware about the specific date and time of the recall in advance and do not change their diet because of that. Online dietary assessment surveys enable instantaneous generation of nutritional output, ensuring consistency and eliminating the need for manual coding and data entry typically associated with traditional dietary assessment [148, 93]. Scalability and efficiency of online surveys lead to a significant reduction in the cost and effort involved in a dietary assessment study.

The development of the automated SelfAdministered 24-hour dietary recall (ASA24) has begun in 2006 in the US by the National Cancer Institute, in collaboration with a research firm Westat and with research groups at the National Institutes of Health [154]. Since then the system has been continuously refined based on the input from stakeholders, continuous usability testing and from analysed recalls collected from respondents of various demographic profiles [154]. ASA24 has also undergone validation in a feeding study with 83 adults aged 20-70 years and in observational feeding study with 10-13 years old children [82, 80].

MyFood24 was developed by the Nutritional Epidemiology Group in the School of Food Science & Nutrition at the University of Leeds [39]. The system's database of foods comprises more than 50,000 branded food items with six thousand of the most common foods linked to photographs of serving sizes to aid portion size estimation by respondents [39]. The system went though numerous studies including the usability and acceptability of the tool among adolescents (11-18 years old) [167, 15]. The accuracy of results provided by MyFood24 was compared to that of biomarkers and standard face-to-face interviews with adults (mean age for men and women is 43 and 44 respectively) and adolescents (11-18 years old) [167, 14].

Intake24 is an open source system developed at Newcastle University for conducting large-scale population dietary surveys online using a multiple pass 24-hour dietary recall method [30]. Development of the system started in 2009 when it was named 'SCRAN24' and has since undergone several cycles of user testing and development [30, 58, 138, 56, 148]. The original system was developed as software that had to be installed on a personal computer (PC). Respondents completed a dietary survey after installation and the results of the completed survey had to be collected from that computer. SCRAN24 served as a prototype to validate the concept of automated dietary surveys on a small scale and enabled the collection of user feedback from 38 participants 11-16 years old [58]. This study informed the development of a scalable cloud-based system - Intake24. A validation of Intake24 against interviewer-led recalls with 180 participants aged 11-24 years demonstrated that Intake24 under-estimated energy intake on average by only 1% compared to the interviewer-led method [30]. The interviewer-led method itself was found to underestimate energy intake on average by 33% compared to that measured with DLW [97]. The accuracy of energy intake estimated by Intake24 was also validated using DLW [60]. The user experience of Intake24 was evaluated with 80 participants aged 11-24 years old [148] and the system has also been field tested in those aged from 11 years to older adults, to examine the feasibility of using Intake24 with the Scottish population on a large scale, where it was found by participants to be user friendly and enjoyable to use [139]. Intake24 has been used to conduct at least 129 dietary assessment surveys with 13,633 respondents, who submitted more than 238,000 meals. The system is translated into Portuguese, Danish and Arabic and at the time of writing this paper is being used in six countries. It is also adapted for use in the UK coeliac population and has been adapted for use in India. Intake24 is planned to be used for the next UK NDNS by the NatCen and MRC Epidemiology Unit at the University of Cambridge [169] as well as being pilot tested in the Scottish Health Survey (SHeS) 2018/19 to assess the feasibility of integrating it into SHeS [6].

Intake24 is backed by a rich ontology of more than 4,800 foods tailored to each region and linked to local food composition tables. For example, for the UK version of the system the food database is linked to the Nutrient Databank from Public Health England [148]. In addition to nutritional information, for many items in the UK version of the food database, information about the carbon emissions of foods is included. One of the key features of Intake24 is the method used to assist subjects' portion size estimation using a collection of over 2,000 photographs of different portion sizes of a range of common foods, many of which have been extensively validated in a feeding study and against concurrent WFD [61, 57]. Intake24 based its selection of foods, and estimation methods, on the portion sizes of foods reported in NDNS and has been updated overtime [148]. The system consists of two main components, a survey interface for respondents, and an interface for researchers and administrators of the system. Further the chapter describes these two components in detail.

2.3.2 Survey interface

Survey interfaces of ASA24, MyFood24 and Intake24 generally follow the questionnaire of the multiple pass 24-hour recall method with some deviations. In all three systems respondents log into the survey interface using a unique and non-personally identifiable username and a password. In the system, the respondent is presented with a customizable welcome screen that provides instructions and a video tutorial . Intake24 and MyFood24 allow logging in with an authentication URL provided by the administrators of a study. In the system, the respondent is presented with a welcome screen that provides instructions and a tutorial. This page can also give an overview of a study (Fig. 2.1-2.3). After the welcome screen, there is a sequence of screens that represent the steps of the multiple pass 24-hour recall method.

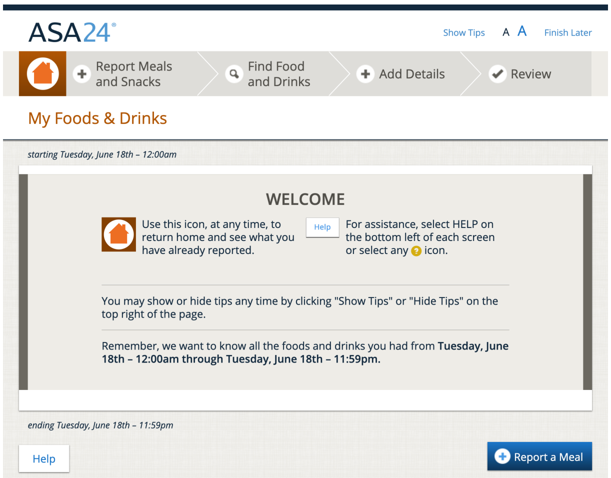


Fig. 2.1 Welcome screen in the ASA24 survey interface

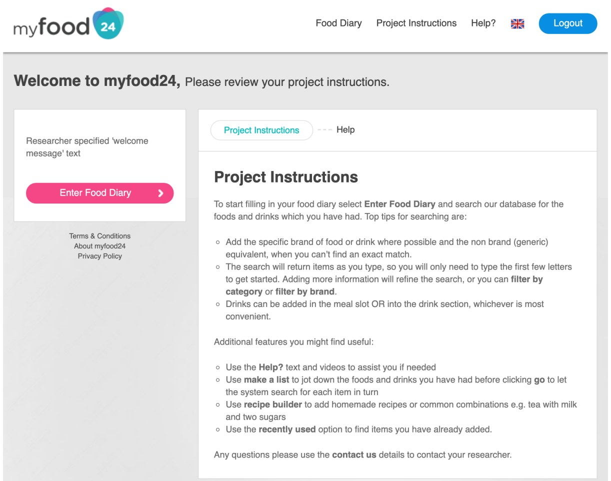


Fig. 2.2 Welcome screen in the MyFood24 survey interface

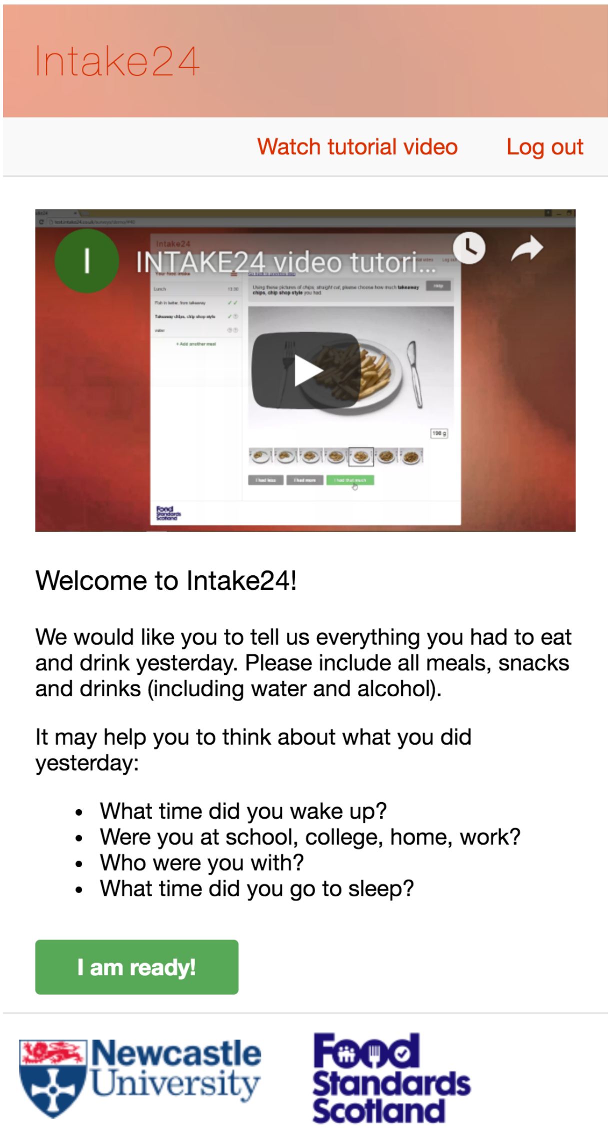


Fig. 2.3 Welcome screen in the Intake24 survey interface

The original multiple pass 24-hour recall procedure was modified in all three systems. In ASA24 and Intake24, the questionnaire consists of three passes (steps) instead of the five defined in the original method. In the first "quick-list" pass, the respondent is asked to recall all meals for a previous day along with the time they had them. ASA24 offers selecting meals from a standard list. Intake24 also allows the addition of custom meal names to fit with the respondents intake (Fig. 2.4). In ASA24, after selecting time of a meal the respondent searches and selects food items from a database of foods for each meal (Fig. 2.6). In Intake24, in the first pass, the respondent only types in names of all foods and drinks they had for that meal in a free text format (Fig. 2.5). ASA24 identifies long time gaps between meals on this step and asks the respondent about additional eating occasions.

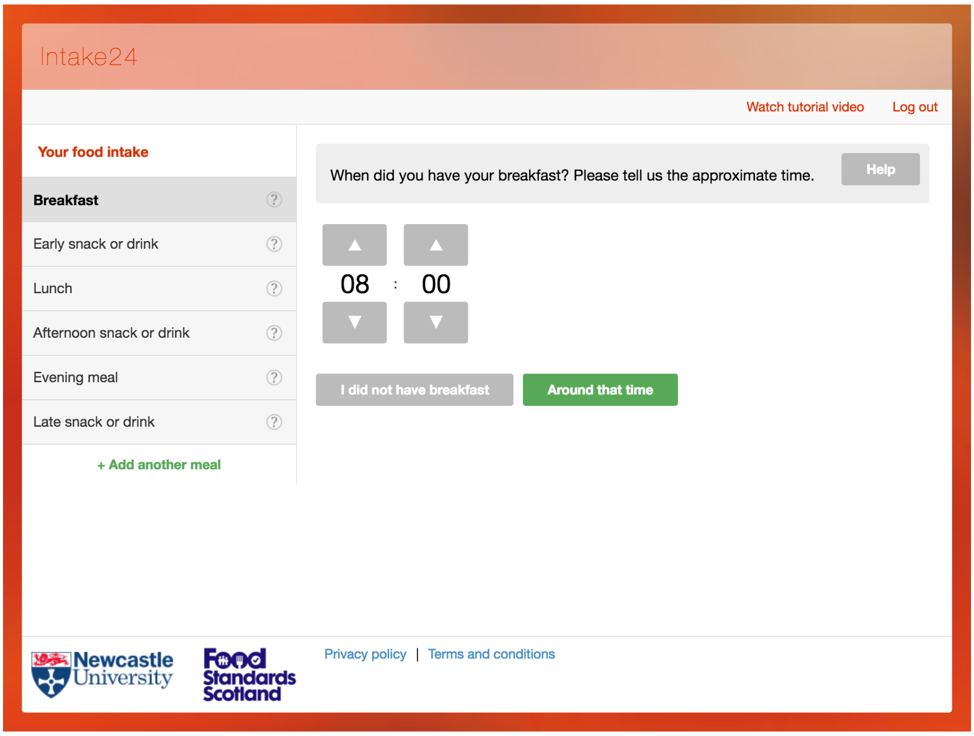


Fig. 2.4 Standard list of meals (left panel) and time question for a meal in the survey interface of Intake24

In the second pass, in Intake24, the respondent is asked to select specific items from the list of search results returned in response to the typed food names in pass 1 (Fig. 2.7). Respondents can browse foods in categories that matched their text query by clicking on them in the list of results. The system offers to browse the full ontology of foods in Intake24 by pressing "Browse all foods". Respondents can also refine search results by typing in a new free text query and pressing the "Search again" button. In ASA24 and Intake24, the respondent also estimates the amount of food/drink they had. For most foods, this is estimated by selecting the closest matching photograph from a range of predetermined serving or packaging sizes, e.g. a slice of bread, biscuit, soda can or a bowl of porridge (Fig. 2.9-2.11). For other foods, the system offers common household measures such as bowls, tablespoons or handfuls. To try to accurately capture portion size, for some foods in Intake24, participants are also asked if they had any food "left over" for example, "Did you leave some of your porridge, made with semi skimmed milk?".

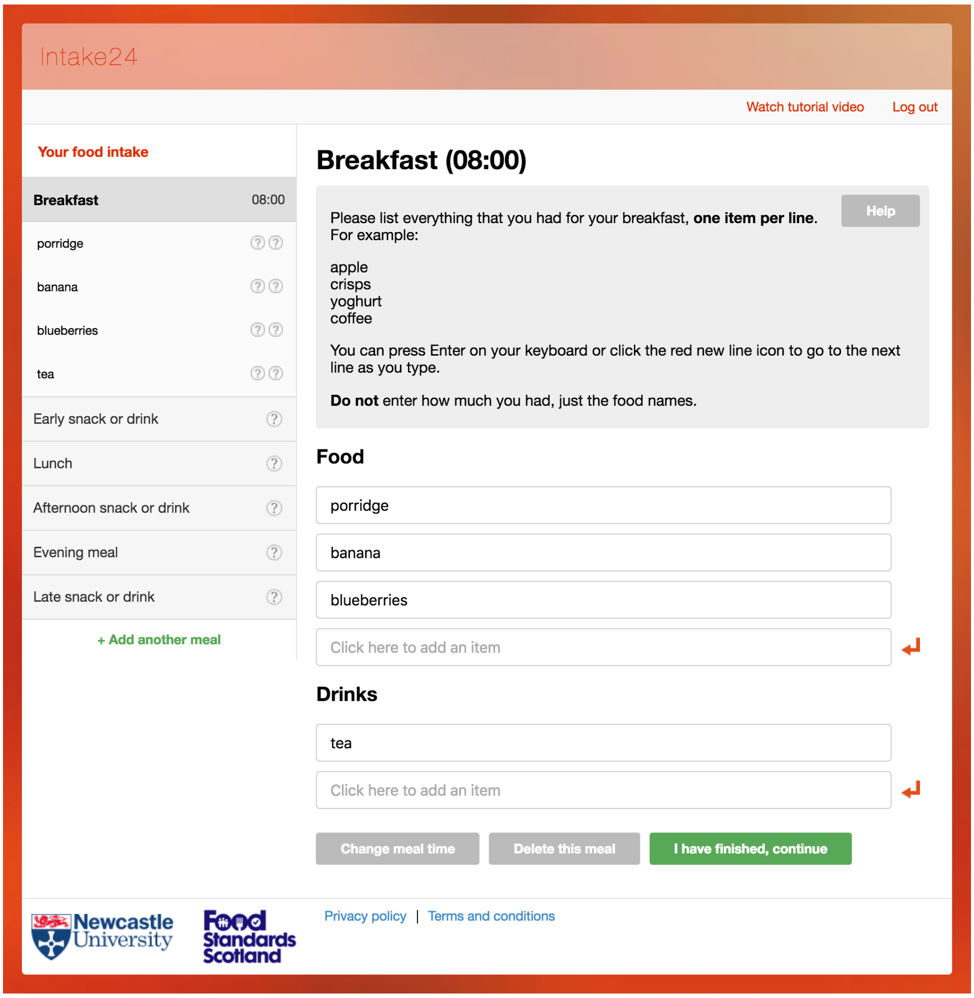


Fig. 2.5 Quick list of foods typed in a free text format for a meal in the survey interface of Intake24

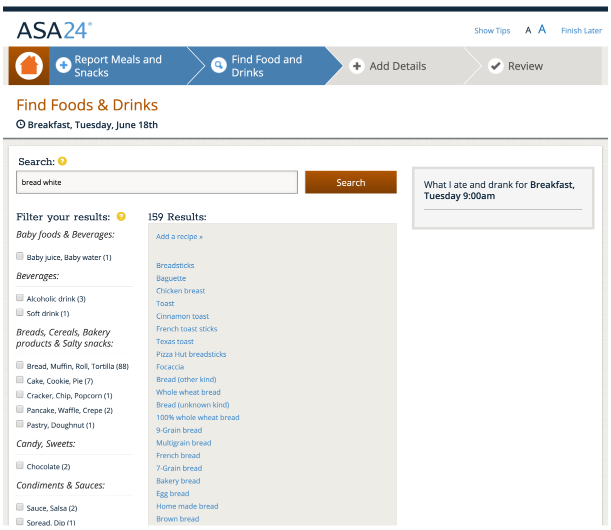


Fig. 2.6 Food search results in ASA24

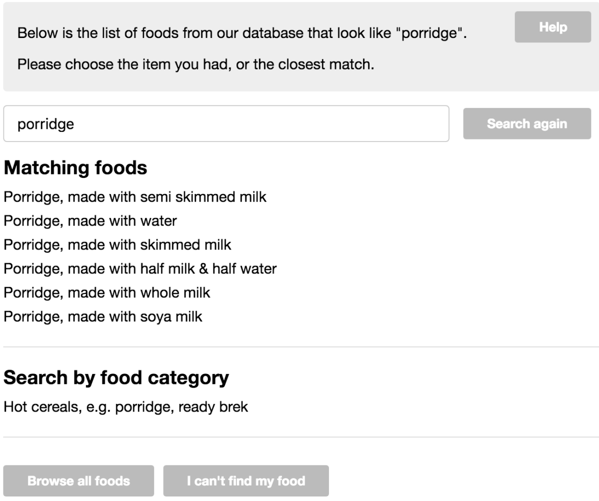


Fig. 2.7 Search results returned in response to a free text food name query in the survey interface of Intake24

If the participant chooses that they left some food, the system asks the respondent to estimate the amount of food left using the same interface for portion size estimation. After this step, both Intake24 and ASA24 asks a question about commonly associated foods, for example "Did you have any sugar or syrup on your porridge?" The respondent can accept or reject the prompt. If the prompt is accepted, the respondent selects the food item and portion size they had. If no drinks are reported in a meal time, Intake24 flags this and asks the participant if any drinks were consumed. In the third pass, both systems ask the respondent to review the list of reported meals and foods. Intake24 also probes for additional eating occasions at this step, for example, if there were any long time intervals between the meals or if the calorie intake is low, participants are asked to check they haven't missed anything from their recall. After these checks, respondents are asked to submit their recall.

MyFood24 combines the first and the second passes into a single one. The respondent is given a predefined list of meals on the left part of the screen. They select a meal that they want to report and search foods for that meal on the right part of the screen. Upon selecting the food from the search results the respondent is also asked to select the serving size of that food either using a common measure or a photograph (Fig. 2.8). As the other two systems, for some foods MyFood24 asks questions about common associated foods, for example, about milk in tea. In the process of analysing the survey interface of MyFood24 for this thesis the system did not request some information that the other systems did. For example, meal times, potentially missed eating occasions in case of abnormally low number of meals / foods in the report or long time gaps between the reported meals. The mean completion time of a single recall is 23 minutes in ASA24, 19 minutes in in MyFood24 and 14 minutes in Intake24 [152, 39, 139].

The three dietary assessment systems reviewed in this thesis aim to include all foods in the database that are available to a given population, which can be determined, for example, from previous national dietary surveys or from surveying foods that can be purchased in supermarkets of a given region [148, 154, 40]. A single food can be produced by multiple brands or may have minor recipe variations. Thus, the selection of foods in the database of a system may exceed

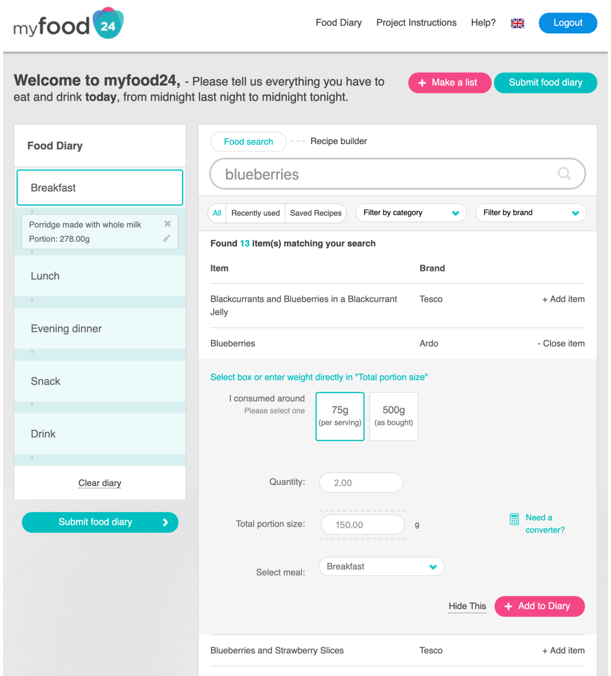


Fig. 2.8 Survey interface in Myfood24

the number of foods in a composition table that is used to estimate nutrient values of foods reported by respondents. Therefore, in many cases, a single generic food record in the regional composition table represents a range of foods in a dietary assessment system that have similar nutrient content.

Although the database is updated to be as comprehensive as possible, there are still cases when respondents might not be able to identify foods/drinks they had from the search results. There are several reasons why foods may be missing from the database. Firstly it is a challenging task to keep the food database up to date, especially when determining the nutrient content of newly developed foods, and in a multi-cultural population context, respondents may report home-cooked meals that are less common and therefore not in the database. Furthermore, there might be a literacy issues whereby the respondent does not know the exact name of the food they consumed. In these cases, Intake24 offers respondents to report a ‘missing food’ by pressing the "I can’t find my food" button which takes the respondent to a form for reporting missing foods. To allow administrators of the system to identify the missing food, the form asks various details including whether the food was homemade; if not, details such as the brand name, a description of the food/drink and details about portion size in a free text format are asked. Administrators can later calculate nutrient data which can be added to the respondent’s recall. The food/drink highlighted as missing can also either be added into the database if there is appropriate nutritional data available, or it can be added and linked to an existing nutrient code [148]. When the respondent knows the ingredients of the missing food, MyFood24 and Intake24 offer participants to list compound foods along with their amounts in a recipe builder. Foods available to choose in the recipe builder (e.g. vegetables and flour) have to be present in a system’s database. The process of reporting portion sizes of compound foods follows the same procedure as that for the normal foods with the additional option of reporting in grams (g).

In the process of the development of Intake24, the team of researchers has found that it is practically infeasible to keep record of all possible variations of sandwiches and salads in the database up to date [148]. At the same time, nutrient content of these foods is highly dependent on specific ingredients (e.g. bread, fillings, dressing). For this reason, the system offers custom recipe builders for salads and sandwiches (Fig. 2.12). In contrast to the missing food recipe builder, these two interfaces are triggered when the respondent types a food search query that contains the word "sandwich" or "salad" in a local language. Another distinct feature of these two interfaces is that the system asks about the ingredients in a guided order. For example, in the sandwich builder the respondent is asked to select bread first, then the type of spread, then any fillings and so on.

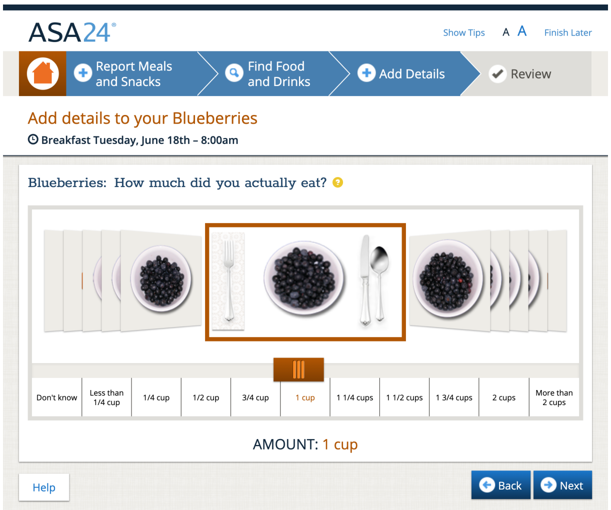


Fig. 2.9 Portion size estimation of blueberries in the survey interface of ASA24

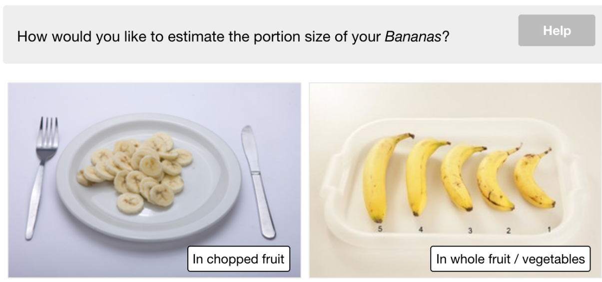


Fig. 2.10 Portion size estimation of a banana in the survey interface of Intake24

2.3.3 Researcher interface

Intake24 researcher interface was initially designed for the research team behind the project to maintain the ontology of foods. The ontology of foods in Intake24 consists of food categories and specific food items (Fig. 2.14). For each food there needs to be defined a name in English and in local languages, a unique identifier, a record in a composition table and portion size estimation methods (Fig. 2.14, 2.15). If the food is commonly consumed with other associated foods, those are linked in this interface as well (Fig. 2.16). This is to allow the system to prompt

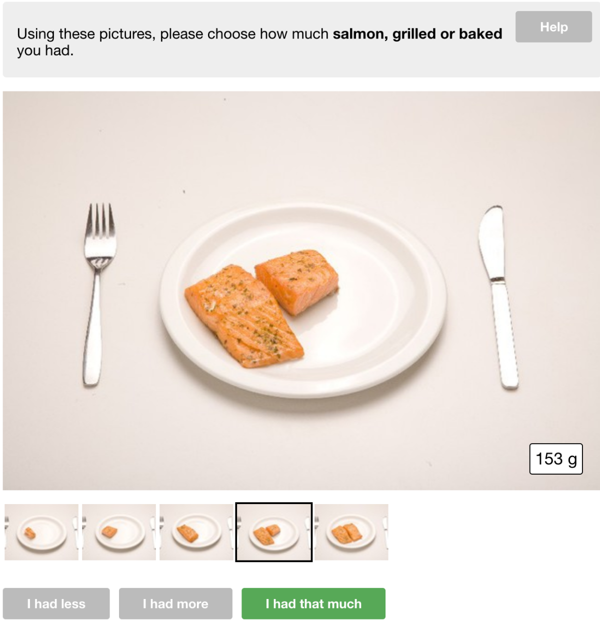


Fig. 2.11 Portion size estimation of salmon in the survey interface of Intake24

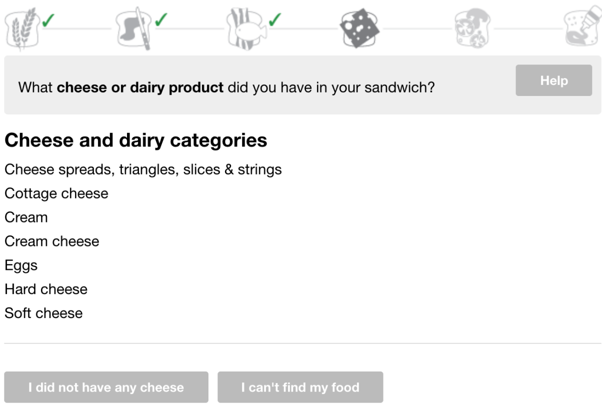


Fig. 2.12 Sandwich builder in the survey interface of Intake24

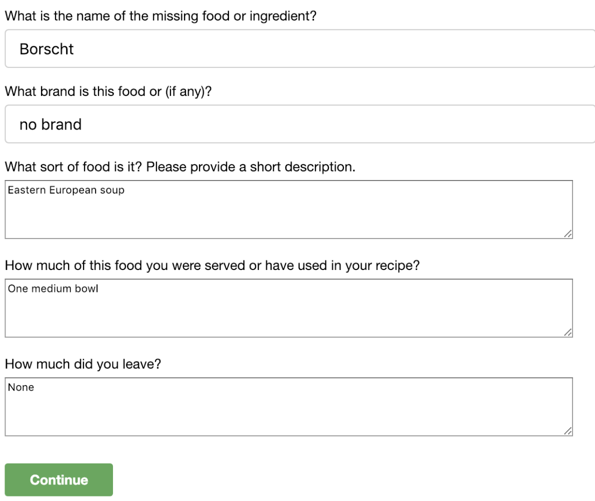


Fig. 2.13 Form for reporting a missing food in the survey interface of Intake24

the user about associated foods (e.g. butter, cheese) once the respondent reports the linked food (e.g. bread). For that an administrator needs to select associated foods and define questions that are needed to be asked about them during the survey. The remaining features of the interface were developed in the course of this thesis and are described in chapter 5.

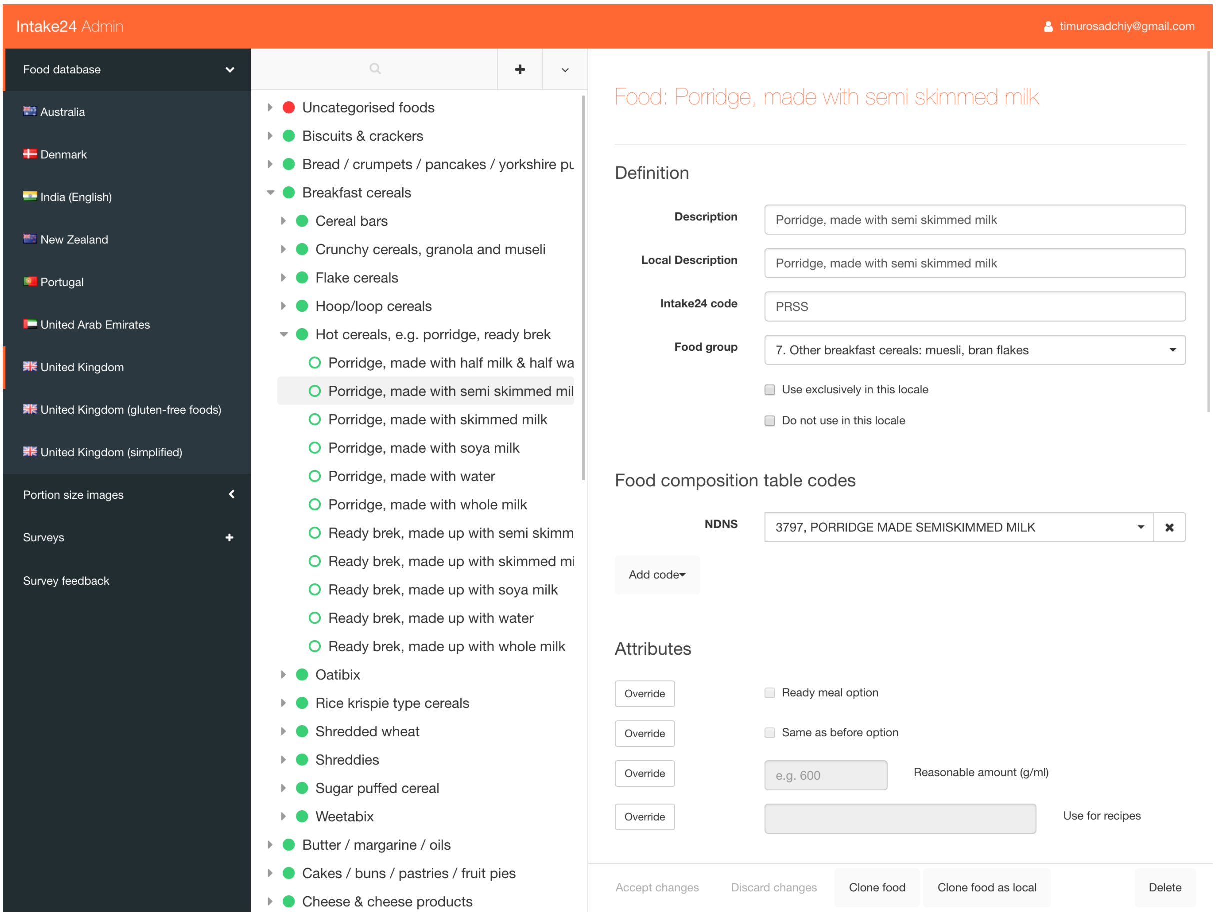


Fig. 2.14 Food ontology in the researcher interface of Intake24

2.4 Challenges in online dietary assessment surveys

2.4.1 Heredity

Since Intake24 and similar dietary assessment systems mostly mimic the interviewer-led multiple pass 24-hour recall procedure they do not seem to overcome methodological problems of a self-administered dietary recall. Thus, they inherit most issues of the original method that lead to errors in estimation of energy and nutrient intake [147]. For example, as with interviewer-led methods, online dietary surveys are prone to omissions. In a recent evaluation of Intake24, alcohol and spreads in particular were widely misreported [30]. Other 24-hour dietary recall systems, such as ASA24 and MyFood24 have demonstrated similar characteristics to Intake24 [77, 154]. In an interviewer-led procedure, these memory errors and omissions can be minimized by a trained interviewer who can probe the subject for additional details, where that is necessary.

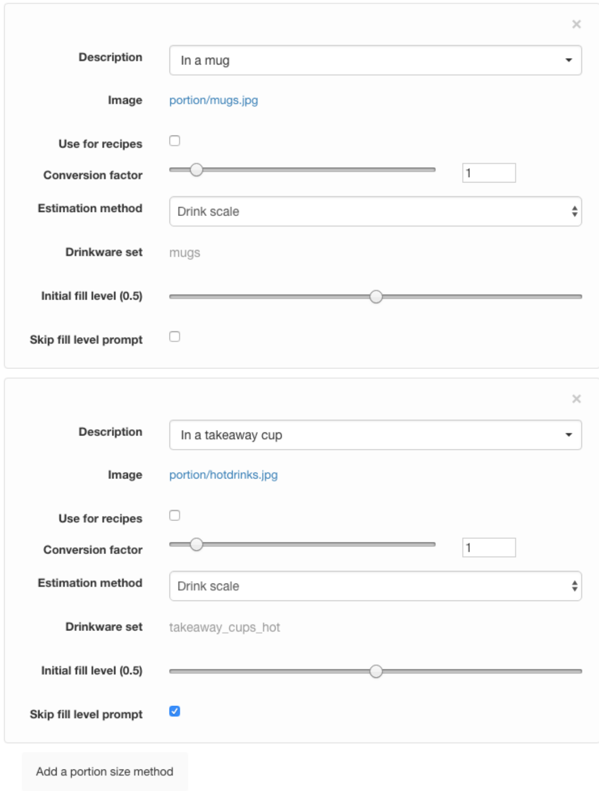


Fig. 2.15 Adding portion size estmination methods for coffee in the researcher interface of Intake24

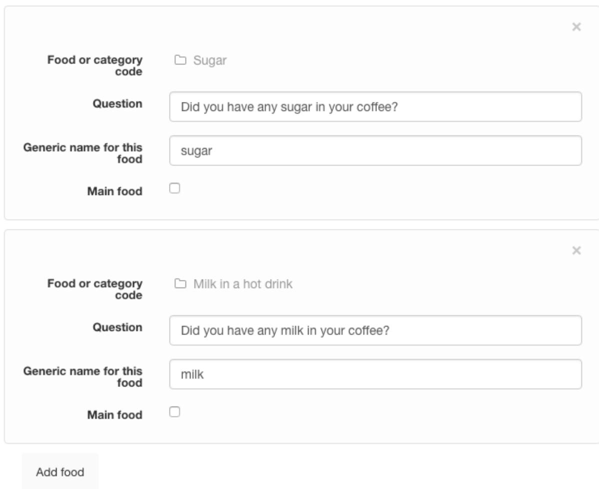


Fig. 2.16 Adding associated food questions for coffee in the researcher interface of Intake24

In an online survey, this interviewer behaviour has to be emulated in a form of digital prompts and reminders.

In their review of human errors in dietary intake recalls, Macdiarmid and Blundell have broadly divided misreporting into unintentional and intentional [100]. Under-reporting of energy intake is the most common form of misreporting and is one of the key factors affecting the accuracy of dietary assessment [100]. Among the most common forms of unintentional under-reporting are omissions due to poor human memory or lack of attention, which is especially common in population groups with limited cognitive and/or memory function [83, 100]. Other respondents may be not aware of the level of precision they need to describe their intake with due to not receiving appropriate training before taking part in a dietary survey (e.g. missing out sweeteners, sauces) [166]. This issue is particularly relevant for automated dietary assessment surveys that lack the interviewer who can inform the respondent about the missing details at any point of the survey. In should be noted that probing by an interviewer is still limited, for example, with respondents who report home made meals that were not cooked by them and they do not know the recipe details [159, 66]. Potential challenge for some individuals in recalling their intake comes from recent changes in their diet due to illness, pregnancy, economic uncertainties or due to intake of seasonal foods (i.e. episodical foods) [100, 159, 160]. Social desirability bias commonly causes respondents of dietary assessment studies to deliberately and knowingly alter their intake towards accepted societal norms [74, 101, 100, 94].

2.4.2 Usability

The use of online surveys for dietary assessment creates a range of challenges specific to technology that may lead to under-reporting. Other dietary assessment methods based on self-reported intake are known for causing under-reporting and participant dropouts due to the complexity of procedure [100]. For example, in a study by Macdiarmid et al. the second most common reason for under-reporting in a 7-day WFD pointed by respondents was the amount of effort they had to invest into weighing and recording all foods [101]. The interviewer-led procedure of the 24-hour recall is known to cause less burden compared to the WFD [77, 78, 100]. However, the use of online surveys makes the dietary recall self-administered and an overly complicated user experience is an additional difficulty in giving accurate answers [63]. The task of completing a survey requires a certain level of motivation from respondents [139]. The complexity of the procedure can be affected by every task that is needed to be performed during the survey including searching for specific food items [58]. A food name which the respondent types in to a free text format to search for the food they ate may differ from how this food is recorded in the system (e.g. typed ’pizza’ vs. selected ’hawaiian pizza’). If the respondent is not able to readily identify the food they had in the list of food items returned in response to the typed food name, they might select a different food or skip reporting the intake at all. Thus, even a long list of search results or results that are not prioritized appropriately for a particular context may affect the accuracy of a dietary assessment [58]. An extensive database that contains foods with similar names (e.g. same name, different brand name) in combination with respondents commonly mistyping food names may further complicate the issue. For example, at the moment of writing this thesis, search engines of MyFood24 and ASA24 demonstrated both examples of surplus of foods and irrelevant foods on top of the search results (Fig. 2.17, 2.6).

The search engine of Intake24 in response to an arbitrary text query prioritises foods based on two types of scores. The first score is the matching cost (i.e. similarity) of a food description against the text query. The engine takes into account that the respondent could have not just typed only a part of the food name but also misspelled it. For that reason, the matching cost itself consists of multiple metrics that include the edit distance between matched words (the approximate string matching is performed using Levenshtein automata [144]); phonetic similarity of words (using a pluggable phonetic encoding algorithm that depends on the localization language, e.g. Soundex or Metaphone for English [50]); the relative ordering of words; the number of words not matched. The lower the calculated matching cost for the food name, the better that food matches the search query. The second score that is used for sorting search results is the likelihood of the food being reported, which is inferred from the number of times it has been reported in the past. Thereby, the search engine first selects out food names that contain similar words to the words that were typed by the respondent from all foods registered in the system using the approximate string matching and phonetic similarity algorithm [144, 50]. On the next step, the matched words are assigned with the matching cost and the likelihood of being reported. Finally, the respondent receives search results of foods sorted by the decreasing likelihood of being reported and then by the increasing matching cost.

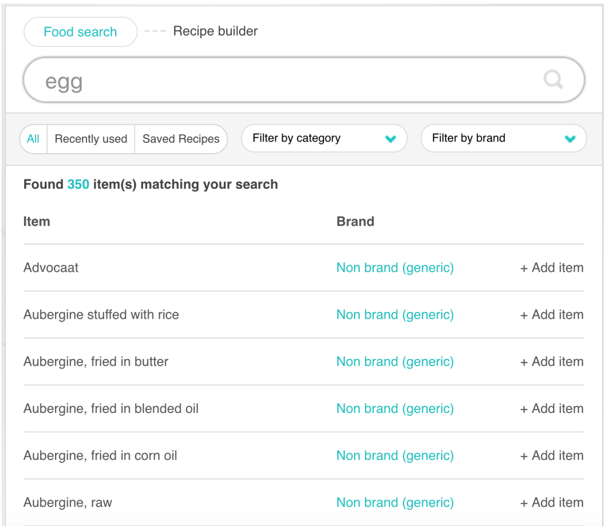


Fig. 2.17 Food search results in Myfood24

2.4.3 Cultural and demographic contexts

Diets in different regions and cultures may vary in the type common foods, ingredients, portion sizes and estimation methods (e.g. plates vs. handfuls) used. A dietary assessment survey could also target a specific health condition (e.g. diabetes) or a diet (e.g. vegans). For those reason, a scalable dietary assessment system should be able to adapt to those variations. The localization procedure of Intake24 for a targeted population consists of two key elements. First, text in the elements of the survey interface including reminding prompts are translated into a local language.

The second step involves producing a database of foods along with portion size estimation methods and nutrient contents adequate for a given context (e.g. regional, health specific). The majority of food items are shared between multiple localizations in Intake24 and the names of those foods are translated into languages of populations where the system is used. Food items that are specific for the targeted context are identified based on data collected in previous dietary assessment surveys and by surveying foods that are available in that context (e.g. local recipes, websites of local supermarkets) [148]. Each food is then linked to a record in a local food composition table that contains validated information about nutrient contents of common foods available in a given region [54]. Such tables are, in most cases, obtained from local official public authorities. For example, for the studies conducted in the UK Intake24 currently uses the Nutrient Databank provided by Public Health England [148]. An example of such dataset in the US are the Food composition databases (FCDB) maintained by the US Department of Agriculture's (USDA) [73]. Another element of the food database that can be context specific and may require localization are the portion size estimation methods. As with the food items in the database, there are a lot of estimation methods that are shared across different localizations (e.g. glass, plate), however food packaging sizes may differ depending on the region. Moreover, different cultures may have specific methods of serving foods. For example, in Tanzania food is often consumed communally using hands from a shared household dish [173]. For those reasons, Intake24 provides a protocol for surveying packaging and portion sizes for the targeted contexts, adding them into the systems and linking with the foods in the database. Since one of the methods of portion size estimation in Intake24 is based on photographs of predetermined serving and packaging sizes those also have to be prepared and uploaded into the system.

To build a clear picture of a population's diet, the analysis needs to cover a wide range of socio-economic statuses and demographic groups with varying education levels. Recording dietary intake requires some level of literacy and under-reporting has been linked with lower levels of education [100, 159]. The use of technology in dietary assessment is characterized by a range of additional limitations in populations with a low level of technological literacy and/or with limited access to technology (e.g. computer, mobile device) [147, 129]. To report their intake through the user interface subjects are required to have typing and spellings skills [26]. Respondents in this case need to get additional training on using a computer or a mobile device, accessing the Internet and addressing technical difficulties during a self-administered dietary recall [147, 26]. In an interviewer-led procedure, a trained interviewer aims to minimize errors caused by the lack of experience by carefully explaining the procedure to respondents, guiding them through each step and probing for additional details, where that is necessary. An online survey has to emulate that behavior. For that reason, Intake24 implements a range of user experience elements that aim to assist dietary recall of respondents. For example, the first screen that is displayed to respondents during the survey contains a video tutorial and instructions that explain the required level of detail of a dietary recall. The instructions can also be edited to fit with the purpose of a specific study. Another method for assisting the recall is the predefined list of meals on the left side of the screen that includes common eating occasions (e.g. breakfast, lunch, morning snacks). Respondents are not expected to report foods for all of those meals.

They can delete any of them, edit their names or add new meals. The list aims to provide examples and remind about common meals respondents can report for the previous day.

2.4.4 Evaluation of dietary assessment systems

The accuracy of dietary assessment systems is generally assessed by comparing it to that of other assessment methods with a validated accuracy, for example, of the interviewer-led 24-hour recall or of the gold standard method of assessing energy intake - DLW [78, 34, 30]. An interviewer-led recall conducted following the completion of the online recall, in addition to allowing estimation of the accuracy can help to identify food and drink items omitted during the online recall [148]. This, for example, informs the development of targeted food prompts and helps to reveal common foods missing in the database [148]. However, these methods only establish the accuracy of the system and might not reveal issues affecting it, for example, issues with the usability. Researchers behind Intake24, MyFood24 and ASA24 touched various aspects of user experience with their systems [139, 51, 15]. One of the key methods used for eliciting feedback were usability questionnaires. Such questionnaires normally give a list of statements on various aspects of usability of the system and the user needs to agree or disagree with those statements. An example statement can be "I would like to use this system frequently" and the user is expected to give their answer on a scale from 1 for 'strongly disagree' to 5 for 'strongly agree'. The usability evaluation of ASA24 has also addressed the efficiency of different types of photographs for the assessment of portion size (e.g. aerial photographs, angled photographs, images of mounds, and household measures for portion size estimation), of image sizes (large vs small), and of the number of images with different portion sizes [153]. The results of this research indicate that no single image type significantly affected the accuracy of portion size estimation but the use of 8 images resulted in a more accurate estimation than the use of 4 images [153]. The usability analysis of MyFood24 was performed in two stages. The usability of the beta version of the system was analysed in a study with 23 participants (various age groups) that measured the average completion time of a recall and asked participants to complete a (generic) System Usability Scale questionnaire [15, 32]. To evaluate the resulting improvements in the first release version of the system another user testing session with 94 participants of various age groups was conducted. Respondents completed the same usability questionnaire as before and the mean recall completion time was compared to that recorded with the previous version of the system. The evaluation resulted in the system usability score changing from moderate to good as reported by adolescent and adult respondents but staying poor for elder respondents using both versions. The research team behind Intake24 designed a feedback questionnaire to evaluate the level of satisfaction with the system which was completed by 182 respondents [139]. Most participants (80%) agreed that the system is user friendly and 84% disagreed that they would require any assistance in using Intake24. Around 88% of respondents agreed that they were able to complete the survey in a reasonable time. In the development of Intake24, in addition to usability questionnaires the researchers used other techniques to maximise feedback from respondents such as think aloud and eye tracking/screen grab software [148].

Surveying or even interviewing users gives an invaluable insight into their experience with a system. However, usability analysis that is solely based on user feedback risks being subjective. Intake24 in addition to eliciting user feedback collects quantitative usage metrics from system logs (e.g. number of times help button was pressed, number of reported missing foods, number of times tutorial video was watched). However, to date these have not been utilized in a systematic assessment of the usability of Intake24. None of the researchers behind other projects have presented a framework with quantitative metrics for evaluating and refining the effectiveness of key elements of their systems, for example, food search and associated food prompts. Such a framework could play an important role in the analysis of performance and accuracy of the system with newly developed and updated features (e.g. interface design, extended food database, new localisations).

2.5 Improving the accuracy of online dietary assessment surveys

By minimizing human involvement automated dietary assessment systems lack the intelligence of a trained interviewer who can tailor their questioning to the food knowledge of the respondent and help identify forgotten or misreported foods in a face-to-face interaction [30]. That is somewhat addressed in automated dietary surveys by associated food questions that are being entered into the system by researchers with nutritional background. However, manually going through thousands of foods to identify and record all associations is prone to errors even with highly trained and skilled dieticians.

Dietary assessment research has only explored a small number of technology-based methods to assist respondents' memory in self-reporting their intake. For example, various studies have demonstrated that images captured by respondents using handheld devices or wearable cameras can enhance the quality of self-reported dietary intake on individual level by revealing unreported foods and misreporting errors [65]. However, no study has discussed the feasibility of image assistance for large-scale dietary aseessment studies. No research was found while writing this thesis that identifies ways to minimise intentional under-reporting of intake and along with unintentional it is commonly detected using biomarkers and excluded. A less costly method is to filter results based on confidence levels calculated using the so-called Goldberg cut-off [127]. The benefits of this method are arguable though according to Wrieden et al. [170]. Detecting misreporting in already recorded recalls in general is a challenging task. For example, genuine low-energy intake report caused by pregnancy, illness, diet, economic uncertainties or irregular eating patterns or due to other reasons can be misclassified as a result of under-reporting [100, 159]. For those reasons, a better strategy for addressing misreporting could be its prevention.

Some data driven approaches could be borrowed from such online shopping and entertainment systems as Amazon or Netflix to address under-reporting due to issues with human memory and usability in dietary assessment systems [91, 67]. For example, an online retail website may recommend items that the current user purchased previously but did not add into their virtual basket in the current shopping session. An online shop may also recommend items that are often purchased by other consumers together with products that are already in the basket of the current user. Algorithms that are responsible for providing those recommendation are called recommender systems [91, 67]. An intake record in a dietary assessment system consisting of meals and foods can be considered similar to a basket with products in an online shop. Thus, recommender algorithms could be used to produce automated associated food prompts as well as to improve the quality of search results based on foods that the respondent has already reported. However, none of the existing dietary assessment systems found to be using such data-driven approaches for recall assistance.

Common techniques for implementing recommender systems are collaborative filtering and content-based filtering or a combination of these two methods. Collaborative filtering relies on models that are generated from user behaviour data that describes interests of each individual (e.g. giving ratings or purchasing items) [91, 67]. For example, items that were purchased by one user can be recommended to other users with similar interests. In this case, similarity of user interests can be determined from correlations in their purchasing history. The knowledge of demographic profiles (e.g. age, gender, occupation) can improve the quality of those recommendations. Content-based filtering provides recommendations based on similarities between items that are based on their descriptors (e.g. similar content, same manufacturer, author, artist) [125]. Items similar to previously purchased by the current user can be used as recommendations for them. In the meantime, the lack of information about interests of a new user or the lack of ratings for a new item limits the ability to generate recommendations with collaborative and content-based filtering, which is known as the cold-start problem [146]. This issue is highly relevant to dietary assessment systems since respondents use them for a short period of time to complete online surveys. Such systems generally do not store any personal information about respondents to address privacy concerns. These factors limit the application of conventional approaches to recommender systems. However, Shaw et al. demonstrated that the cold-start problem can be addressed using a data-mining technique called association rules mining. This method is applied to large transaction datasets for discovering items that frequently appear together in a single transaction [146, 11]. For example, it can be used for the so called market basket analysis task that aims to discover items that are frequently purchased together [131]. Similarly, association rules could be applied to dietary recalls to discover foods that are commonly reported together in a single meal and use those rules as associated food prompts in dietary assessment surveys.

Holding accurate memories of foods, drinks and serving sizes even for a day is a challenging task especially for people with limited cognitive and/or memory function [24, 84]. Shortening retention interval, the time between the intake and the interview (i.e. recall), has been demonstrated to improve the accuracy of intake recall among children [22, 23]. The study of the effects of the retention interval also revealed that the highest correspondence rate for energy and macronutrient intake was observed for the interviews conducted in the afternoon and in the evening for the immediate prior 24 hours, and the lowest for the previous day (midnight to midnight) recalls conducted in the afternoon and in the evening [22]. This work has examined the correlation between the retention interval and the accuracy of reporting energy and macronutrient intake only among children and is yet to be replicated for other age groups. Since automated dietary assessment is more flexible compared to conventional face-to-face interviews, it could allow fine tuning (including shortening) of the retention interval. For example, a respondent could be asked to report a single meal right after the intake rather than reporting all meals for the previous day in the morning. Nevertheless, Intake24, MyFood24 and ASA24 generally offer respondents to report their intake for a previous day at any time during the next day.

Research in participatory engagement in crowdsourcing tasks and online surveys indicates that participants that see a personal value in accurate survey results may feel more motivated to provide reliable responses [96, 115, 130]. Hence, intentional misreporting could be potentially addressed by creating survey conditions, where respondents of a dietary survey are personally interested in its accurate outcomes. In the context of dietary assessment such a condition could be the provision of dietary feedback about nutrient intake of respondents and the quality of their intake at the end of a survey. This assumption is based on a growing interest of people in technologies that facilitate tracking and understanding information related to their health and personal well-being generally referred to as personal informatics [45, 62]. The information collected by a dietary system about a respondent's intake is broken down into the values of consumed nutrients (e.g. free sugars, saturated fat). By applying certain calculations and recommendations from health care services (e.g. NHS) the system can tailor dietary feedback to advise the respondent to try to promote healthier eating and cooking practices or it could be used to assess specific clinical intervention targets with those with diseases which can be aided by diet (e.g. diabetes). Aside from using tailored dietary feedback as a driver for participatory engagement, it could be used for developing nutrition knowledge of a surveyed population [48, 171, 120]. Nutrition knowledge can be described as a theoretical or practical understanding of relationships between nutrition and health, this particularly applies to the understanding of relationships between diet and health, diet and disease, awareness of major sources of nutrients, dietary guidelines and recommendations [108]. Despite health care services making efforts in educating populations and adopting such measures as detailed food labelling and diet-related public health interventions the vast majority of people are not aware of their aggregated nutritional and energy intake [150].

Chapter 3

Methodology

3.1 Abstract

The Literature Review covered in chapter 2 examined the composition of existing systems for conducting online dietary assessment surveys and discussed the three major factors that negatively affect the accuracy of these systems. Those are issues with human memory, motivation of survey participants, and usability of the survey interface. The chapter discovered three research opportunities that were translated into three key research questions. The literature review also revealed a need for a usability framework for evaluating the efficiency of implemented methods.

This chapter reviews research activities that were undertaken to evaluate new methods for improving the accuracy of dietary assessment and inform research questions defined in this thesis. Each of these research activities compares two versions of Intake24. The first version is the system in its current state with validated accuracy [30]. The second version is the system that implements the new method. The accuracy of results produced by each version of the system individually is NOT the primary goal of this comparison. The research activities look for deviations in performance metrics of the updated system from its validated state. Should a deviation that supports a hypothesis be found, the new method can then be validated on a larger scale and/or with more precise methods (e.g. direct observations, nutrient biomarkers) [77]. This chapter also describes a usability framework that was developed to provide quantified metrics and tools for monitoring, analysing and refining the performance of implemented recall assistance methods. Lastly, the chapter provides an overview of software modules developed for Intake24 that supported research activities described in the course of this thesis.

3.2 Research activity 1: Development of the recommender system based on pairwise association rules (RA1)

This research activity informs the answers to RQ1 and RQ2. Two of the main contributors of under-reporting in automated dietary assessment are discussed in chapter 2. Those are omissions due to inability of respondents to retain all necessary details about their intake in their memory, and the complexity of a task of identifying specific food items in a large database of foods (i.e. searching foods). The chapter proposed addressing both issues by applying a recommender algorithm to past survey responses collected in the system to mine a model of foods commonly eaten together. Intake24 could then use that model to prompt respondents about possible forgotten foods and to improve the prioritization (i.e. sorting) of foods in search results based on foods already reported by the respondent. In the development of the recommender algorithm existing approaches designed for addressing similar tasks were reviewed. That is producing recommendations in the context of a system that lacks personal user profiles based on transactions from all users. Chapter 4 then proposed a new alternative method for mining pairwise association rules designed specifically for this task. The comparison of candidate algorithms is performed using data collected in past real-life dietary surveys conducted in Intake24. This data is essentially a list of meal records collected from respondents. Each meal record consists of the meal name (e.g. "breakfast"), time, a list of food identifiers (e.g. toast, boiled egg, orange juice) and search queries in a free text format that were entered to seek for those foods (e.g. egg). The performance of the algorithms was analysed through the metrics concerning associated food prompts described in section 3.6. These metrics initiated the developement of the usability framework for validating recall assistance methods that is covered in section 3.6. The development of the recommender system and the results of this research are described in chapter 4.

The participants and the researchers who took part in past Intake24 surveys have not given an explicit consent for the data collected in these surveys to be used for the current research activity. For that reason, this research activity considered the risks of respondents being identified and exposed in an unlikely event of an unauthorized access to the data used in this study. Intake24 generally does not collect and store personally identifiable data (e.g. name, email) to avoid concerns related to General Data Protection Regulation (GDPR) [165]. Thus, this research activity had no access to such data. Each respondent visits the system on average 3 times, meaning that user identification through eating patterns is unlikely. Thus, the data used in this research activity was considered to be non-personal. The use of non-personally identifiable data collected in past dietary surveys using Intake24 for the validation of the recommender system received ethical approval from the Newcastle University Research Office (Ref: 1377/2017). This research activity received an additional consultation and approval from the Faculty of Science, Agriculture & Engineering (SAgE) Ethics Committee.

3.3 Research activity 2: Deployment and evaluation of recall assistance methods (RA2)

This research activity informs the answers to RQ1, RQ2 and RQ3. Another method of recall assistance discussed in chapter 2 is reducing the burden on a person's memory by shortening the period of time between their intake and recall (i.e. retention interval). Based on that this thesis proposed changes to the 24-hour recall procedure that allow reporting meals for the day progressively. The progressive 24-hour recall procedure along with the associated foods recommender algorithm that was developed in chapter 4 were implemented in Intake24. Both methods of recall assistance were validated independently in real-life settings in a study, where the performance of the system with each of the implemented methods was compared to that with the current state of the system (24-hour recall with hand-coded food prompts). The performance was analysed through the metrics defined by the usability framework that was finalized in this research activity and that is described in section 3.6 of this chapter. To inform the answer to the RQ3 respondents were also presented with two versions of tailored dietary feedback.

To get more insights into the effects of the implemented methods on user experience with the system this research conducted interviews with participants of the study. Interviews were transcribed and thematically analysed [31]. During the thematic analysis sections of transcribed interviews were highlighted and coded with labels that described their content. Codes were then used to identify patterns in the interviews that were then used to identify broader themes related to the acceptability of the progressive 24-hour recall procedure and the usability of tailored dietary feedback. The emerged themes from the analysis informed answers to RQ1 and RQ3. The study received ethical approval from Newcastle University Research Office (Ref: 4971/2018). The outcomes and findings from this research activity are described in chapters 5 - 7.

To recruit participants for this study the procedure disseminated an advertisement describing the purpose of the study and a web form to register interest via the University internal email system. Individuals had to complete the web form to volunteer for the study and were asked to circulate the invitation to their relatives and friends for participation. The first page of the registration form presented more details about the study and an informed consent form, where all conditions had to be accepted to be able to register. As part of the registration process potential participants were asked to provide their contact details (e.g. name, email, phone number and an age range). This study aimed to follow the original procedure that was used in previous validations of Intake24 [30, 139] and asked participants to record their for the previous day. To help respondents to follow the procedure they were asked about their preferred times to receive reminders about completing a dietary recall. To be able to take part in this study individuals had to be over 18 years old, speak English, have a diet that is common for the UK. To minimize the likelihood of variations in diets of respondents affecting the results of the study they were informed that it is important to avoid changing their diet during the study. Participants were informed that they can withdraw from the study at any time. For completing the survey participants were offered a £30 Amazon voucher. Recruitment resulted in 50 participants, of whom $n = 26$ were men and $n = 24$ were women. One female participant chose to withdraw during the study. The participant's ages ranged between 18 and 64 years.

The study was conducted over a six-week period with a separate group of participants in every two weeks. Each group of participants was asked to complete their recalls on two consecutive weeks. On one week participants were asked to complete five recalls (Monday - Friday) following the existing 24-hour recall procedure. During the first three days (Monday - Wednesday) the system presented associated food prompts of one type and in the remaining two days (Thursday - Friday) prompts of the other type. The study resulted in the system displaying food prompts hand-coded by researchers in the first three days of this week to $n = 19$ participants. For $n = 30$ participants the system displayed hand-coded food prompts in the last two days. On the other week respondents were asked to complete three consecutive recalls following the new progressive 24-hour recall method. For $n = 34$ participants the original 24-hour recall was used on the first week and the progressive recall was used on the second week. With the remaining $n = 15$ participants the two methods were applied in the reverse order. The first day of each week was used to minimize the learning effect by familiarizing participants with the interface of the system and the procedure, and was excluded from analysis. The prevalent approach to the multiple-pass 24-hour recall method suggests that respondents should complete their recalls on at least three non-consecutive weekdays and one weekend day to capture their habitual intake [77]. However, since the primary goal of this research is the comparison of the new recall assistance methods, this was not considered necessary for this study. For that reason, no recalls of weekend intake were used during the analysis. Respondents were informed about the schedule of their recalls two days before the first one, which could affect their diets. However, the procedure assumes that this factor affects the accuracy for all types of recall. Thus, if there is a difference in the accuracy of the methods, it still can be observed.

When participants were surveyed using the original 24-hour recall method, they received automated reminders by email and SMS in the morning to submit their intake for the previous day. When the progressive 24-hour recall method was used, participants received three reminders to add meals into the system - in the morning, in the afternoon and in the evening. The next morning, they received the last reminder to record late snacks and drinks, if they had any, and to complete their submission. Reminders were set through a researcher interface of Intake24 and circulated at the preferred times specified by respondents during registration for the study. Participants were asked not to record their meals elsewhere (e.g. notepads) to aid their recalls.

3.4 Research activity 3: Deployment of a recommender system for prompting omitted foods in a large-scale dietary assessment survey (RA3)

This research activity provides additional data for answering RQ2. To further examine the performance of the associated foods recommender system as a method of recall assistance it was deployed in a large-scale dietary survey. This dietary survey was conducted as part of a weight loss campaign Newcastle Can [5] that aims to cut down obesity levels in the North East of England. The campaign resulted in a series on BBC One "Britain's Fat Fight with Hugh Fearnley-Whittingstall" [3]. The campaign has its own website that describes information about the campaign, gives tips on healthy eating and provides a weight loss diary to track progress of participants of the campaign [5]. The website also offers participants to get feedback about their diet using Intake24. Participants in this survey were not paid. To access Intake24, participants had to first click a corresponding button in their Newcastle Can personal profile page. Before they were redirected to Intake24 a modal window was displayed informing participants that they are transferred to an external website and by proceeding they accept the privacy policy of Intake24. The privacy policy informs a reader that anonymized and aggregated information may be subject to processing for scientific purposes and used as a basis for publications (Fig. 3.6). All user data that is received from Newcastle Can and that is stored in Intake24 does not contain any personally identifiable information (e.g. name, date of birth). The performance of the recommender system was analysed through usability metrics described in section 3.6 of this chapter. Integration of Intake24 with the Newcastle Can website is described in section 3.8.1 of this chapter. During the dietary survey created for the campaign recalls were assisted by the associated foods recommender algorithm and without it. The evaluation of the associated foods recommender algorithm in the dietary survey created for the campaign was conducted under ethical approval from Newcastle University Research Office (Ref: 9232/2018). The results of this work are described in chapter 5.

3.5 Research activity 4: Deployment of a tailored dietary feedback system in large-scale dietary assessment surveys (RA4)

To inform the answer to RQ3 of this thesis a tailored dietary feedback system was deployed in Scottish Health Survey 2018 that was conducted by the Food Standards Scotland. Participants received £20 for completing two recalls. To avoid changes in diets of respondents the feedback was displayed only after completing the expected number of recalls. This research was performed under ethical approval from Newcastle University Research Office (Ref: 1377/2017) and with the permission of Food Standards Scotland. The results of this research activity are covered in chapter 7.

3.6 Usability framework

The usability framework that is used for the evaluation of recall assistance methods implemented in this thesis is based upon five quality components characterising usability of a system interface defined by Jakob Nielsen [119]. These are learnability, efficiency, memorability, errors and satisfaction. The section covers these components in the context of dietary assessment systems:

- Learnability: ease of navigating through and understanding the features of the system interface necessary to complete basic tasks that in this case are primarily completing a dietary survey and understanding dietary feedback by a respondent.
- Efficiency: accuracy and speed of a survey completed by the respondents that are familiar with the system; broadened nutrition knowledge from dietary feedback and more informed attitude to their eating habits.
- Memorability: changes in efficiency of tasks completed by the respondent after a period of non-use.
- Errors: number of errors respondents make (e.g. misreporting or forgetting foods, misunderstanding of the dietary feedback), severity of the errors and the ease of recovering from errors.

- Satisfaction: respondents' subjective feeling of appreciation of the system interface and positive emotion induced by the interaction with it.

The research interest of this thesis primarily concerns the development of methods for assisting dietary recall of respondents and for their participatory engagement through the provision of tailored dietary feedback. This thesis assumes that those methods mainly improve efficiency, reduce the error rate and improve satisfaction of respondents in dietary assessment systems. For that reason, the main goal of the usability framework developed for this thesis was to evaluate those three characteristics. However, the framework can be extended to analyse changes in learnability and memorability as well.

This research uses explicit and implicit metrics for analysing efficiency, error rate and satisfaction. Explicit metrics directly measure the performance of an implemented method. For example, the work described in chapters 4 and 5 as the measure of the accuracy of associated food prompts generated with a recommender algorithm uses the number of accepted/rejected prompts by respondents. The research measured the mean time it took respondents to complete a single recall to analyse how different types of food prompts affected the speed of survey completion. Implicit metrics are indirect indicators of the usability performance of the system based on a set of assumptions. These metrics are used where explicit metrics are not available. For example, a study by Lopes et al. with 83 adults aged between 20 and 60 years demonstrates that the interviewer-led multiple-pass 24-hour recall underestimates energy intake on average by 33% compared to that measured with DLW [97]. Validation of Intake24 demonstrates underestimation of energy intake by 1% compared to the interviewer-led method [30]. For those reasons, in analysing the benefits of a short retention interval for automated dietary assessment chapter 6 measures the mean number of reported foods and estimates the mean energy per recall assuming that an increase in those characteristics within 34% may indicate of an improved accuracy. Another example of an implicit indicator is using the average number submitted recalls per respondent to analyse the level of engagement of respondents with a dietary survey that presents dietary feedback in chapter 7. This metric assumes that respondents who submit a larger number of recalls than that which was required are more engaged with the survey. Table 3.1 lists all metrics that were used in the course of research for this thesis

Metric	**Type**	**Indicates**	**Description**
Food prompts acceptance rate	Explicit	Performance of food prompts	The mean number foods accepted from associated food prompts by respondents in a single recall
Coverage of food prompts	Explicit	Performance of food prompts	The number of unique foods accepted by respondents
Precision of food prompts	Explicit	Performance of food prompts	The number of accepted foods divided by the number of foods returned in prompts
Normalized discounted cumulative gain	Explicit	Performance of automated food prompts / food search	The quality of ranking of foods in automated food prompts / food search results

Mean energy reported per recall / meal	Implicit	Accuracy of dietary assessment	The mean energy value per recall / meal estimated by the system
Mean number of foods reported per recall / meal	Implicit	Accuracy of dietary assessment	The mean number of foods reported by respondents per recall / meal
Survey completion time	Explicit	Speed of survey completion	The mean time it takes for a respondent to complete a single recall
Mean number of recalls	Implicit	Engagement with a survey	The mean number of recalls submitted by a single user in a survey

Table 3.1 Usability metrics used for evaluating methods of memory assistance and participatory engagement

The selected metrics are measured in a series of offline and online experiments. Offline evaluation is performed on data collected from previous surveys. For example, in the development of automated associated food prompts chapter 4 evaluates candidate algorithms on a dataset of 20,000 meals reported in Intake24 that was used to simulate respondents' omitted foods. The performance of candidate algorithms is analysed by plotting their precision-recall (PR) curves [47]. In case of this work, precision is the number of foods correctly predicted by the algorithm (foods were reported by the respondent during their recall) divided by the number of all recommendations from the algorithm (includes foods that were reported by the respondent). Recall is the number of foods predicted correctly by the algorithm divided by the number of all foods reported by the respondent (includes foods that were not predicted by the algorithm). By changing the number of recommendations returned by the algorithm one can change its precision and recall. For example, the algorithm can return all possible foods, which will likely predict all foods selected by the respondent and will result in high recall. In the meantime, if such prediction mostly contains food items that were not reported by the user, it is characterised by low precision. PR-curves are built by gradually increasing the number of recommendations produced by each algorithm, calculating precision and recall at each threshold and using them as coordinates on a graph. The algorithm is then selected by the largest area under its PR-curve, which demonstrates a combination of high precision and recall across all thresholds. As another performance metric for the recommender algorithm this work analyses the ranking quality of recommendations, which is represented by Normalized Discounted Cumulative Gain (NDCG) [33]. The value of NDCG depends on the position of correctly predicted recommendations in the list of all recommendations. The higher the food items that were reported by the respondent are positioned in the list of all recommendations produced by the algorithm, the higher the NDCG is.

In online evaluations the implemented methods are deployed in real-life dietary surveys. The performance of those surveys is compared to a survey without any modifications (i.e. A/B testing). For example, after performing an offline evaluation of automated associated food prompts they were compared to food prompts hand-coded by nutritionists and dieticians in a study that involved real-life dietary surveys and is described in chapter 4. To measure the significance of deviations between the metrics measured for surveys with and without modifications this research uses Mann-Whitney U test. This non-parametric test of the null hypothesis is selected due to its lack of requirement of data being normally distributed [103, 114]. Thus, it allows this research to avoid testing data for normality. To visualize differences in quantitative metrics this work uses histograms and probability density plots.

In addition to quantitative metrics this research gathers qualitative data around user experience mainly from two sources. The first source are user interviews with participants of dietary surveys, where the new methods are applied. With permission from respondents interviews are audio recorded. The second source is feedback that respondents can leave in the system after completing a survey. Reflections from respondents are transcribed and thematically analysed to harvest topics that concern usability and acceptability of implemented methods [31].

3.7 Development of the survey interface

To conduct research studies and collect the data for the analysis that answers research questions of this thesis a set of essential but missing modules in Intake24 were designed and implemented. In addition to that some of the existing modules were revised. The key modules and features that were developed in the course of producing this thesis are:

1. Behaviour data collection
2. Survey participants notification
3. Survey state synchronisation
4. Survey administration
5. Portion size images administration
6. Interface for linking food records in the ontology of Intake24 to records in regional composition tables
7. Interface for managing portion size estimation methods of foods in the ontology of Intake24
8. Interface for managing food prompts in surveys

3.7.1 Behaviour data collection module

This section covers the behaviour data collection module that captures events related to user experience with the survey interface. Information contained in those events describes the state of the survey interface and is used to produce metrics defined in section 3.6 of this chapter. The module consists of two parts, the front-end and the back-end.

The front-end component of the module is embedded into the survey interface of Intake24 and provides an API for software developers for sending events to the back-end. Events are triggered by the respondent interacting with the survey interface. For example, accepting /

rejecting food prompts, searching foods, etc. The back-end component defines a list of accepted events and a web API for collecting and storing them in the system database. Since sending events takes considerably large amount of HTTP requests and space in the database the collection of events is activated only for specific surveys. For the same reason the range of collected events is currently limited to those concerning the usability of recall assistance elements of the survey interface. The list of events can be extended depending on the purpose of a study. Currently the system accepts the following events:

`SearchResultsReceived`. Respondent received a list of foods and categories in response to a search query in a free text format.

`BrowseCategoryResult`. Respondent received a list of foods and categories in response to selecting a food category in the list of search results, or in response to selecting "Browse all foods".

`SearchButtonClicked`. Respondent pressed "Search again" after receiving the list of foods and categories.

`CantFindButtonClicked`. Respondent pressed "I can't find my food" after receiving the list of foods and categories.

`BrowseAllFoodsButtonClicked`. Respondent pressed "Browse all foods" after receiving the list of foods and categories.

`BrowseBackButtonClicked`. Respondent pressed "Go back to search results" while browsing a food category, or the ontology of foods.

`SearchResultSelected`. Respondent selected a food or a category in the list of foods and categories.

`ManualAssociatedFoodReceived`. Respondent received a hand-coded associated food prompt.

`ManualAssociatedFoodConfirmed`. Respondent accepted a hand-coded associated food prompt.

`ManualAssociatedFoodRejected`. Respondent rejected a hand-coded associated food prompt.

`ManualAssociatedFoodAlreadyReported`. Respondent pressed "Yes, I have already entered it" in an associated food prompt.

`AutomaticAssociatedFoodsReceived`. Respondent received an automated associated food prompt.

`AutomaticAssociatedFoodsResponse`. Respondent submitted a response for an automated associated food prompt.

Events are accepted by the web API in a JSON format. The description of every event has a header. The header consists of the type of event; date and time on the device of the respondent, when the event was dispatched; a unique identifier of a recall session `sessionId`. The description of the event may optionally carry data relevant for that particular type of event. The full list of properties of events that are stored in the database of Intake24 by the behaviour data collection module is described in Tables 3.3-3.25. To describe some data types in tables (e.g. `Option[T]`, `Either[T1, T2]`) this chapter uses data types from the programming language Scala version 2.12.4.

A sequence of events with the same `sessionId` can be used to extract various performance metrics about individual elements of the survey interface. For example, from the `SearchResultsReceived` event and `SearchResultSelected` we can see that a respondent has selected a food in the list of search results and we can measure how long it took. If we observe the `SearchButtonClicked` between those events, that means that the respondent had complications with finding their food and had to refine their search query.

Property	Data format	Description
`eventType`	`String`	Unique identifier of the event type
`eventCategories`	`List[String]`	List of event categories / tags (e.g. search, associated food prompts). Used for filtering events in the system database
`created`	`ZonedDateTime`	Local date time on the Intake24 server, when the event was created
`sessionId`	`UUID`	Unique identifier of a recall session in the survey interface. Events created during a recall of the same respondent have the same `sessionId`
`userId`	`Long`	Unique identifier of a respondent who created the event.
`localTimestamp`	`Long`	Timestamp representing local date and time on the device of a respondent used for completing the survey, when the event was sent to the server.

Table 3.3 Header of a usability event in Intake24

Property	Data format	Description
`query`	`String`	Text query entered in the search field at the time when the event had been created.
`algorithmId`	`String`	Unique identifier of an algorithm used for sorting list of foods and categories.
`existing`	`List[String]`	List of unique identifiers of foods reported in the current meal, when received the list of foods and categories.
`result`	`LookupResult`	List of foods and categories received in response to search query or selecting a food category.

Table 3.5 Properties of SearchResultsReceived usability event

3.7.2 Implementation of the associated foods recommender algorithm

The associated foods recommender algorithm was developed as a recall assistance method in large scale population dietary surveys. To evaluate the algorithm in real-life settings it had to be integrated into Intake24, deployed in a dietary survey and compared to the performance of the system without the algorithm. The algorithm was implemented as an additional feature that can be activated for some surveys. When this setting is active the system replaces hand-coded associated food prompts with those generated by the algorithm. Chapter 5 provides more details about the study evaluating the algorithm including user interface design developed for the associated food prompts generated by the algorithm. One of the key challenges in the integration

Property	Data format	Description
categoryCode	String	Unique identifier of category selected in the list of foods and categories.
algorithmId	String	Unique identifier of an algorithm used for sorting list of foods and categories.
existing	List[String]	List of unique identifiers of foods reported in the current meal, when received the list of foods and categories.
result	LookupResult	List of foods and categories received in response to search query or selecting a food category.

Table 3.7 Properties of BrowseCategoryResult usability event

Property	Data format	Description
query	String	Text query entered in the search field at the time when the event had been created.

Table 3.9 Properties of SearchButtonClicked usability event

Property	Data format	Description
selectedFood	Option[FoodHeader]	Description of a selected food in the list of foods and categories.
selectedCategory	Option[CategoryHeader]	Description of a selected category in the list of foods and categories.
selectedIndex	Int	Index of a selected food or category in the list of foods and categories.

Table 3.11 Properties of SearchResultSelected usability event

Property	Data format	Description
food	FoodHeader	Description of a food that triggered a hand-coded associated food prompt.
prompt	AssociatedFood	Description of a hand-coded associated food prompt.

Table 3.13 Properties of ManualAssociatedFoodReceived, ManualAssociatedFoodRejected and ManualAssociatedFoodAlreadyReported usability events in Intake24

Property	Data format	Description
food	FoodHeader	Description of a food that triggered a hand-coded associated food prompt.
prompt	AssociatedFood	Description of a hand-coded associated food prompt.
givenFoods	List[FoodHeader]	List of food descriptions that user has already reported in the moment of accepting a hand-coded associated food prompt.
selectedFood	FoodHeader	Description of a food that was selected after accepting a hand-coded associated food prompt.

Table 3.15 Properties of ManualAssociatedFoodConfirmed usability event

foods	List[FoodHeader]	List of food descriptions that user has already reported in the moment of receiving a generated associated food prompt.
suggestedCategories	List[CategoryHeader]	List of food category descriptions returned by the system in a generated associated food prompt.

Table 3.17 Properties of AutomaticAssociatedFoodsReceived usability event

Property	**Data format**	**Description**
foods	List[FoodHeader]	List of food descriptions that user has already reported in the moment of receiving a generated associated food prompt.
suggestedCategories	List[CategoryHeader]	List of food category descriptions returned by the system in a generated associated food prompt.
selectedCategories	List[CategoryHeader]	List of food category descriptions selected by the respondent after receiving a generated associated food prompt.

Table 3.19 Properties of AutomaticAssociatedFoodsResponse usability event

Property	**Data format**	**Description**
foods	List[FoodHeader]	List of food descriptions in response to user query or selecting a food category.
categories	List[CategoryHeader]	List of category descriptions in response to user query or selecting a food category.

Table 3.21 Properties of LookupResult data format

Property	**Data format**	**Description**
code	String	Unique identifier of a food or a category.
localDescription	String	Description of a food or a category.

Table 3.23 Properties of FoodHeader and CategoryHeader data formats

Property	Data format	Description
`foodOrCategoryCode`	`Either[String, String]`	Unique identifier of a food or a category in a hand-coded associated food prompts. Left of Either data type is a food identifier, Right - category identifier
`promptText`	`String`	Text of a hand-coded associated food prompt (e.g. "Did you have sugar in your coffee?").
`linkAsMain`	`Boolean`	`foodOrCategoryCode` is the identifier of the triggering food of the prompt.
`genericName`	`String`	Generic name of food specified in `foodOrCategoryCode`.

Table 3.25 Properties of AssociatedFood data format

of the algorithm into the system was avoiding negative impact on the performance of the web server of the system. The mining process for analysing all recalls in the system and building association rules takes a considerable amount of time. For that reason, the mining was scheduled to run only at night, when the system is not actively used. Recalls reported during the next day were progressively added to the recommender model. For the study conducted in chapter 5 the recommender model was trained only on recalls reported in the UK. In the process of building the model, the system used only recalls from surveys that were considered to be representative of eating behaviour of the population based on their size and purpose.

3.7.3 Implementation of the progressive 24-hour recall

To evaluate short retention intervals as the method of recall assistance in automated dietary surveys this thesis developed the progressive version of the 24-hour recall method in Intake24. The new procedure that is described in chapter 6 allows respondents recording their meals progressively during the day as opposed to recalling all meals for the previous day the following morning. The new method is implemented as an additional feature that can be activated for selected surveys. Thus, the work that is described in chapter 6 compares the accuracy of dietary recalls reported with the original 24-hour recall method to those produced with the progressive 24-hour recall.

3.7.4 Dietary feedback module

Dietary feedback presented at the end of a dietary recall was a long-requested element by respondents of Intake24. As discussed in chapter 2 offering dietary feedback could potentially motivate participation of respondents and improve the accuracy of dietary assessment. Provision of dietary feedback in Intake24 also allowed it to be used in the Newcastle Can campaign. Nutrient rules for Intake24 Nutrient Feedback system were developed in collaboration with the Newcastle University Human Nutrition Research Centre. The rules are implemented as ranges of low, slightly low, good, slightly high and high intakes for various nutrients based on the Eatwell Guide, UK dietary reference values and the individual's physical characteristics that are voluntarily provided upon them requesting the feedback [35, 1]. These include gender, age, height, weight, weight target (maintain, lose or gain weight) and self-assessed physical activity level. Feedback is given for macro- and micronutrients including energy, fat, saturated fat, non-milk extrinsic sugars, protein, Englyst fiber, carbohydrate, calcium, vitamin C and A. Intake24 has the ability for feedback on other nutrients to be added to the system, if required. The interface also provides feedback on the respondent's consumption of red meat, as well as feedback on the number of portions of fruit and vegetables in relation to the five a day guidelines and details on the carbon footprint of the respondent's diet. The design of the presentation of the nutrient feedback assumed no specific diet related health conditions.

The thesis developed two styles of dietary feedback. The first is designed as a set of characters representing the selected nutrients, for example, a battery for the energy intake or a burger for the saturated fat intake. The characters show different sentiments (for example, sad or happy) depending on the calculated ranges (for example, low or good, respectively). In the second design of the feedback each nutrient is represented with a photograph of a food that is a source of the corresponding nutrient. In both designs of feedback users are given extracts of information related to the listed nutrients with links leading them to reliable sources containing more details on the topic, such as the NHS website. The feedback also lists foods reported by the individual that are highest in calories, saturated fat, added sugar and sums up their total intake to assist people in weight loss and the consumption of these nutrients. To get an insight into user experience the dietary feedback web page includes a widget that allows users to express their opinion by pressing 'like' or 'dislike' button and adding a comment about their experience. The development of the feedback is described in chapter 7.

3.7.5 Respondents notifications module

The aim of the notifications module is to maximize participation during the studies conducted in the course of this thesis and encourage respondents to report their intake at the times determined by the procedure (e.g. 24-hour recall, progressive recall) and enable the following use case. Respondents register for a study through a web form. As part of the registration, they are asked about their preferred times of going through a dietary recall within the limits determined by a procedure of a study. For example, for the study examining the progressive manner of the 24-hour recall described in chapter 6 participants are asked about times for morning, afternoon and evening notifications. The registration form also asks them to provide their mobile phone numbers and emails. The notifications module provides a web API for uploading collected responses to create a schedule of notifications for each individual respondent. On days of a study at the times specified by respondents the system sends notifications to respondents inviting them to report their meals. Notifications are sent to respondents by SMS and email. The message in such a notification contains a URL unique for the respondent receiving it. Once the respondent clicks on that URL they are transferred to the survey and automatically login to the system. If the respondent does not login to the system within a configurable period of time, the system sends another notification.

3.7.6 Survey state synchronisation for respondents

The existing architecture of Intake24 assumes respondents completing a 24-hour recall in the morning for a previous day using a single device. For that reason, to minimise the number of HTTP requests, the usage of storage and other server resources the state of respondent's recall in the survey interface is first saved in the respondent's browser. Once the respondent completes the survey the full set of their answers is sent to the server. However, while planning the study for the evaluation of the progressive recall described in chapter 6, respondents were anticipated to start their recall on one device but complete it on another. For example, the respondent might start their recall in the morning at their work place on a computer and finish it at home on their mobile device. For that reason, a survey state synchronisation module was developed that saves the state of the survey interface on the Intake24 server and propagates that state to all devices of the respondent.

3.8 Development of the researcher interface

A number of activities related to administration of Intake24 as well as to the administration of independent studies conducted within the system had to be performed through various command line interfaces by a software engineer. To make the system more accessible for researchers, clinicians, or educators for registering, configuring, and monitoring studies without a technical background the most commonly used command line tools were replaced with graphical user interfaces. This section gives an overview of the developed modules.

3.8.1 Survey administration module

Most administration tasks for dietary assessment surveys in Intake24 had been done through a command line interface. For example, to create a survey and user accounts for respondents researchers had to produce a spreadsheet in a specific CSV format and pass it to a software engineer. The software engineer then created the survey and uploaded user accounts to the server. After the survey was finished the results had to be downloaded by the software engineer via a command line interface and passed back to researchers. To make survey management possible for people without a technical background this thesis designed and developed a survey administration tool in the researcher interface of Intake24.

With the new module a user with administrator rights in Intake24 can now create a survey and give administrator access to an external researcher limited only to that survey (i.e. survey administrator). The survey administrator can then define the duration of the survey, start / pause it, upload a spreadsheet with respondent accounts and manage those accounts (e.g. change passwords, add contact details) (Fig. 3.1, 3.3). Responses from participants can be downloaded by the survey administrator at any point during the survey (Fig. 3.3). Survey administrators can also create new users with administrator rights limited to their survey.

Another key feature enabled by the module is the survey welcome page manager tool that allows editing the contents of the first page in the survey interface. This tool was, for example, used as one of the channels to communicate changes from 24-hour to a progressive recall procedure in chapter 6 (Fig. 3.4) or to add branding and description for the survey that was used in the weight loss campaign Newcastle Can (Fig. 3.5) [5]. The campaign also motivated the development of an integration API for authenticating users from the campaign website. The organisers of the campaign wanted to use Intake24 as a tool for giving dietary feedback to their participants. For this thesis it was an opportunity to evaluate the dietary feedback module described in chapter 7 and the associated foods recommender algorithm described in chapter 4. The integration API developed for the Newcastle Can campaign allowed authenticating users from their website in Intake24 and passing physical information necessary to calculate dietary feedback for them (i.e. weight, height, gender, age). To use the API the Newcastle Can website authenticates with a username and a password and is given administrator rights limited to a survey that was created specifically for the campaign.

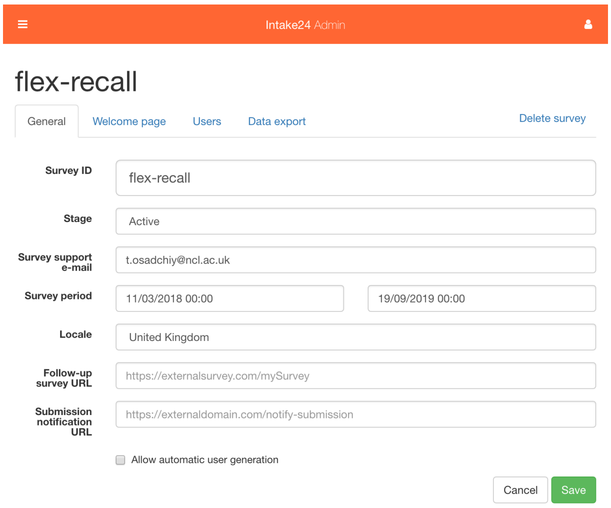


Fig. 3.1 Survey parameters in the survey administrator interface of Intake24

3.8.2 Portion size images administration module

One of the methods of food serving size estimation offered to respondents of dietary assessment surveys in Intake24 are validated images of portion sizes. Respondents are asked to select the closest image of a portion size to the portion size of a food they ate. These images are mapped to specific records of foods and their weights in the system database. Based on the estimated portion size and a food composition table Intake24 can estimate energy intake and intake of various macronutrients (e.g. carbohydrates, saturated fats). The administration module for portion size images is used for uploading images of serving sizes into the system and defining weights of foods that are represented on those images. The images are mapped to food records in the portion size estimation methods manager described in section 3.8.4 of this chapter. Before this module

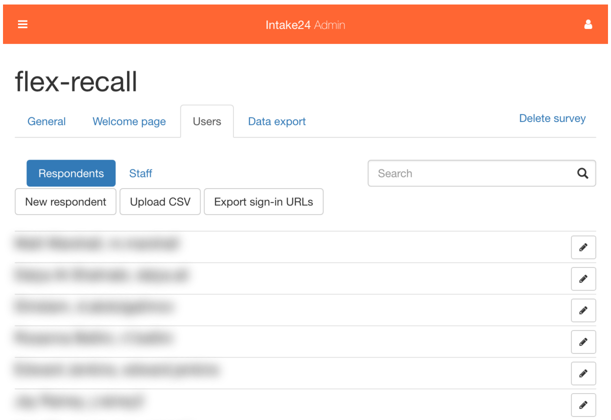


Fig. 3.2 Respondent accounts manager in the survey administrator interface of Intake24. Names are blurred for anonimisation purposes.

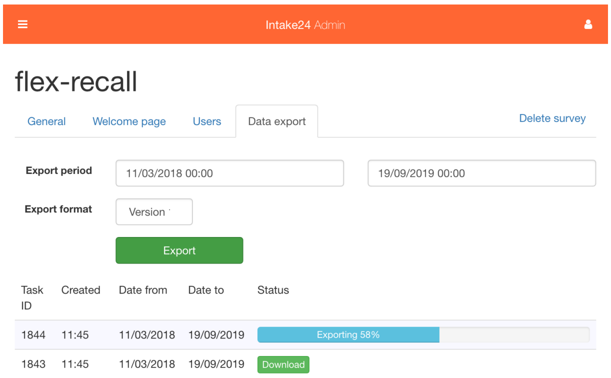


Fig. 3.3 Survey responses data export in the survey administrator interface of Intake24

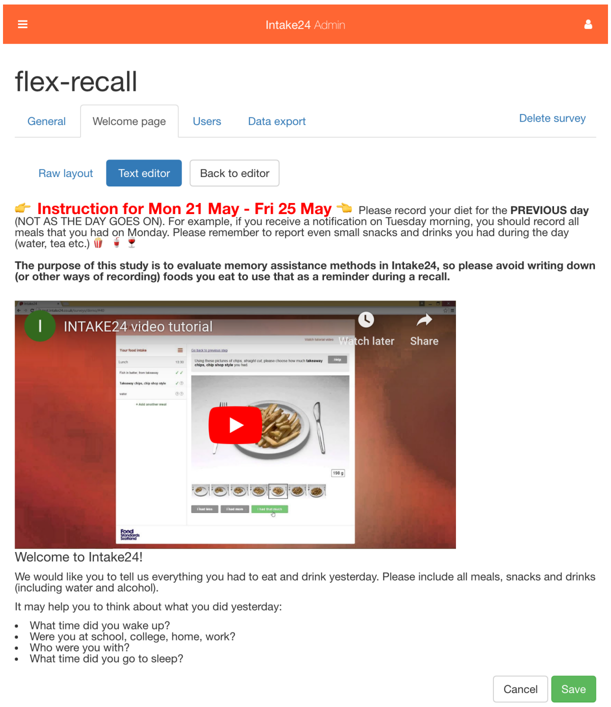


Fig. 3.4 Welcome page manager in the survey administrator interface of Intake24

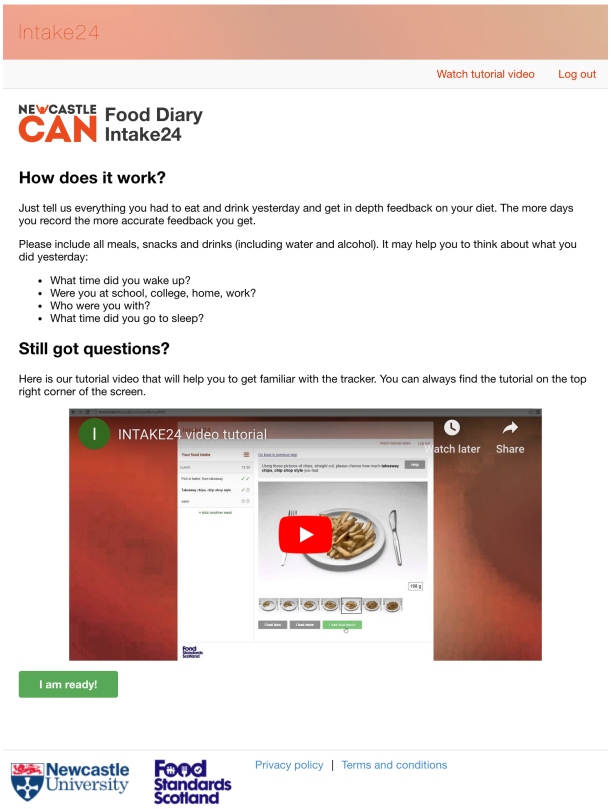


Fig. 3.5 Newcaslte Can Welcome page in the survey interface interface of Intake24

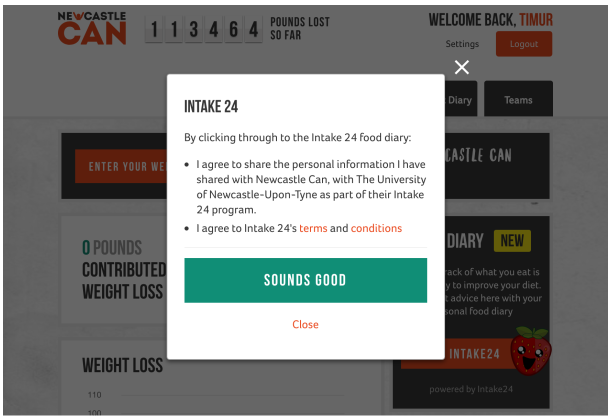


Fig. 3.6 Newcaslte Can privacy alert

was developed, all images had to be uploaded and mapped to food weights using a command line interface by a software engineer.

There are currently two types of estimation methods that can be managed in the portion size images administration module, 'as served' and 'guided'. 'As served' method offers respondents a set of separate photographs of portion sizes of the same food, where they need to select one photograph. This method is normally used for foods that are served on plates (e.g. sides, steakes). The administration module allows uploading multiple images into the system, grouping them into a set for a single food and defining weights that are represented in every photograph. Each set also gets a unique name and a description to be later identified in the database. For example, on Fig. 3.7 you can see photographs with salmon steaks of different sizes with weights typed in under each image. The corresponding component of the survey interface for selecting a portion size image of a salmon steak is presented in Fig. 2.11. A 'guide' image contains all portion sizes of a single food. A respondent selects a portion size by clicking on the corresponding area on the photograph. This method is generally used for fruits and packaged foods (e.g. different sizes of apple, different cans and bottles of soda). To create a 'guide' image an administrator uploads a photograph with multiple portion sizes on it and draws outlines of areas on the image, where respondents can click. The administrator then defines weights that are represented in the outlined areas. For example, Fig. 3.8 demonstrates the outlines drawn in the administrator interface of Intake24 on a photograph of a banana and food weights with descriptions for those outlined areas on the right of the interface.

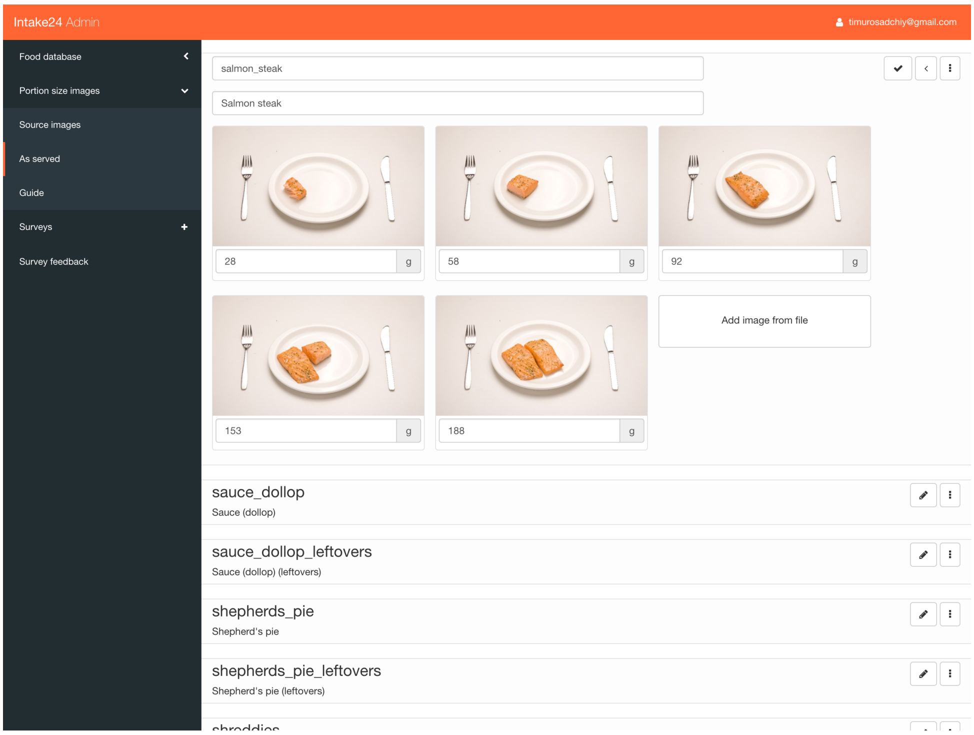


Fig. 3.7 As served images manager in the administrator interface of Intake24

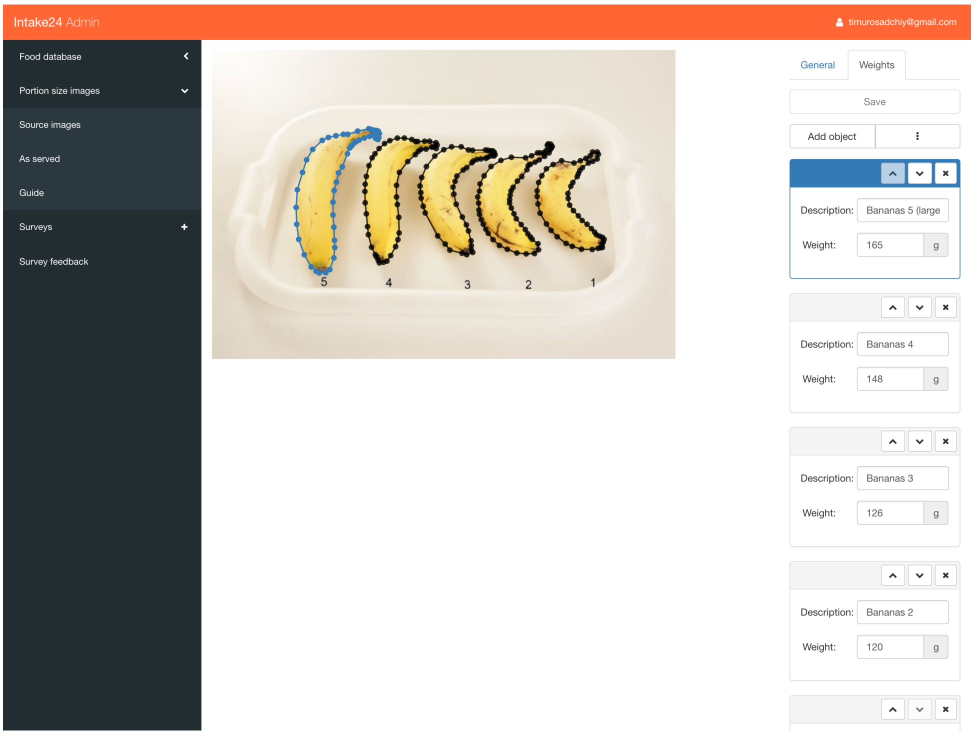


Fig. 3.8 Guided images manager in the administrator interface of Intake24

3.8.3 Mapping food ontology of Intake24 to regional composition tables

Food nutrient values that Intake24 uses for the assessment of dietary intake of a population are defined in national food composition tables. For populations from different regions Intake24 uses different composition tables. Composition tables are uploaded into Intake24 in a CSV format using a command line interface. Previously food records in the ontology of Intake24 could be mapped to the records in the composition tables also only using CSV spreadsheets that were then uploaded into the system via a command line tool. To make the process of adding new foods and locating them in composition tables more accessible a corresponding interface was developed in the Intake24 ontology of foods. To link a food in the ontology to a record in a composition table the system administrator needs first to select that food in the ontology (Fig. 3.9). Then in the right part of the interface with food details, they need to select that composition table in a dropdown that opens on pressing 'Add code'. Finally, they can search for and select a corresponding name of food as defined in the selected composition table (Fig. 3.9).

3.8.4 Portion size estimation methods manager

Portion size estimation methods manager is used to edit the list of portion size estimation methods for a food offered to respondents during a survey, when they report to have eaten that food. The tool is available to the administrator of the system in the food details window that opens on the right of the screen, when they select a food in the ontology. Portion size estimation methods are selected from a global dictionary of methods along with their human readable descriptions stored in the systems database (Fig. 3.10). Different estimation methods also require defining additional parameters relevant to that method. For example, for the 'as served' method the

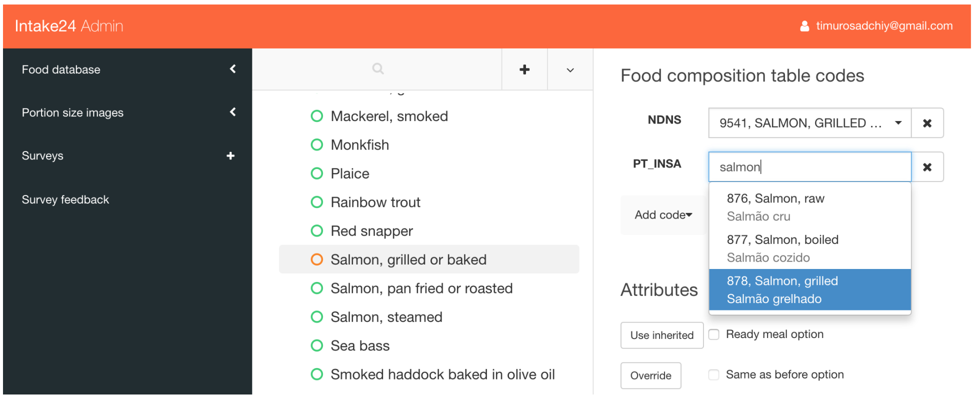


Fig. 3.9 Mapping food to a record in a national composition table in the food ontology of Intake24

administrator needs to select a corresponding set of images from the administration module described in section 3.8.2 of this chapter. Another example is the 'standard portion' method, where the administrator needs to define a list of mappings of human readable descriptions of standard portions (e.g. fillets) to different weights of the food (Fig. 3.11). During a survey, the system displays widgets to respondents that correspond to portion size estimation methods that are defined by the administrator. For the 'as served' method, for example, the system presents a set of images with different portion sizes of a food (Fig. 2.11), from which the respondent needs to select a single image.

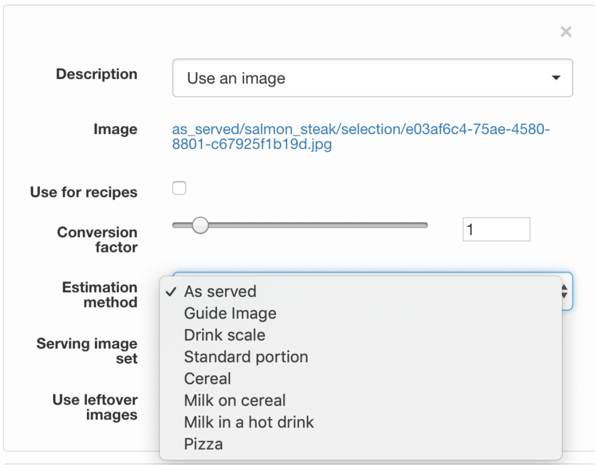


Fig. 3.10 Selecting portion size estimation method for a food in the ontology of Intake24

3.8.5 Associated food prompts manager

The associated food prompts manager is used to edit the list of associated food prompts for a food. These prompts are displayed to respondents during a survey, when they report to have eaten that food. The tool is available to the administrator of the system in the food details window that opens on the right of the screen, when they select a food in the ontology. For example, the Fig. 3.12 demonstrates associated food prompts defined in the manager for when a respondent reports having drunk a coffee.

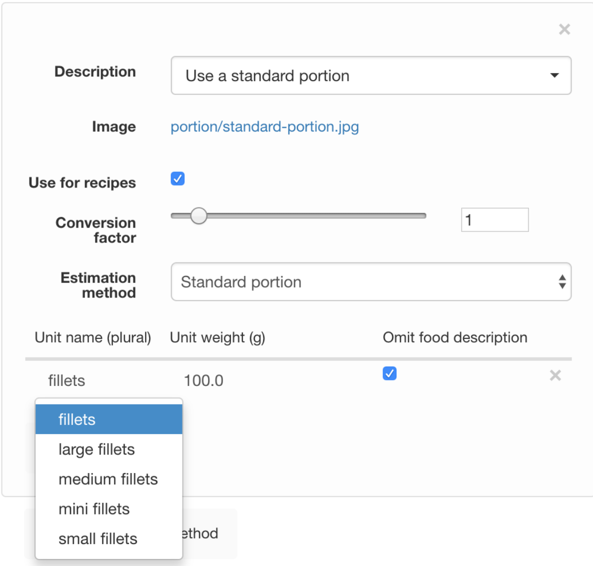


Fig. 3.11 Defining standard portions as a portion size estimation method for a food in the ontology of Intake24

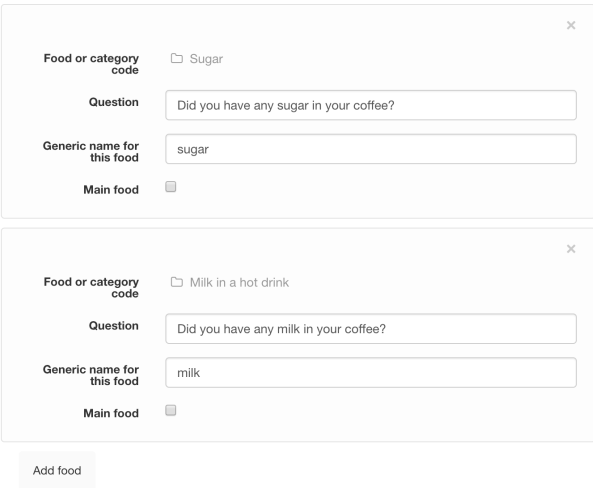


Fig. 3.12 Defining associated food prompts for coffee in the ontology of Intake24

Chapter 4

Recommender system based on pairwise association rules

4.1 Abstract

Recommender systems based on methods such as collaborative and content-based filtering rely on extensive user profiles and item descriptors as well as on an extensive history of user preferences. Such methods face a number of challenges; including the cold-start problem in systems characterized by irregular usage, privacy concerns, and contexts where the range of indicators representing user interests is limited. This chapter presents a recommender algorithm that builds a model of collective preferences independently of personal user interests and does not require a complex system of ratings. The performance of the algorithm is analysed on a large transactional data set generated by a real-world dietary intake recall system. The work described in this chapter has been undertaken during RA1 and informs the answers to RQ1 and RQ2. The content of this chapter has been rewritten as an (accepted) publication in a peer-reviewed journal Expert Systems with Applications: Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., and Foster, E. (2019a). Recommender system based on pairwise association rules. *Expert Systems with Applications*, 115:535–542 [122].

4.2 Introduction

Recommender systems aim to identify consumer preferences and accurately suggest relevant items (e.g. products, services, content). They are used in various application domains, including online retail, tourism and entertainment [91, 46]. Widely adopted recommendation techniques often utilize collaborative filtering or content-based recommendation methods [125].

Collaborative filtering produces recommendations based on user preference models that are generated from explicit and/or implicit characteristics and metrics corresponding to user interests. Explicit indicators normally imply users assigning ratings to items; for example, to products viewed in, or purchased from, an online store. Examples of implicit indicators include the amount of time users spend interacting with content (e.g. watching a video) or levels of interaction (e.g. scroll offset of a web page containing an article they are reading). Items that are positively rated or purchased by consumers with similar preference models are used as recommendations for target users. User similarity can be expressed through correlations in purchasing history or ratings given to the same products, which can be further amplified with demographics (e.g. age, gender, occupation). Content-based filtering identifies similarities between items based on a set of their descriptors (e.g. purpose of an item, author, artist, keywords). Items similar to those positively rated or purchased by the target user, are used as recommendations.

Collaborative and content-based filtering recommender systems are therefore heavily dependent on extensive user- or item- profile information and are most effective when there is a rich history of user preferences or behaviour. Sparse data sets and lean user profiles typically result in low-quality recommendations or an inability to produce recommendations at all. This is referred to as the cold-start problem, where new users are added into the system with empty behaviour profiles or new items are added that have not been reviewed or rated by anyone [146]. Many solutions to the cold-start problem have been considered, including hybrid methods that combine collaborative and content-based filtering [143], and methods that aim to predict user preferences from demographics [89], or knowledge of social relationships [38].

There are, however, a number of application contexts where users interact anonymously; for example, online shops where an unregistered user browses and adds products to their basket to check out later. In other contexts, such as email recipient recommendation [137], applications requiring high levels of privacy, or those where individual interactions with a system are necessarily infrequent (e.g. online dietary assessment surveys [30]) there are no features and ratings to exploit, and the construction of personal behaviour and preference models is not possible. This chapter presents a recommender algorithm that is designed to address these challenges. The algorithm is independent of any personal user model and does not require a complex system of ratings. Based on a set of observed items selected by a user, the algorithm produces a set of items ranked by confidence of their being observed next. In designing the underlying algorithm, this chapter reviews existing methods that aim to address similar tasks, adapts them to meet the constraints of the application context that is primary concern of the current research (online dietary assessment surveys), and proposes a novel alternative. The performance of three methods is compared through the task of recommending omitted foods in a real world dietary recall system.

4.3 Related work

While various approaches have been proposed to address the cold-start problem in recommender systems, the majority of these rely on knowledge of content [126] and users [89], including social relationships between users [38], whereas concern of this research is with contexts where such information is not available. Shaw et al. addressed the cold-start problem in recommender systems by using a data analysis technique, which is applied to large transaction data sets for discovering items that frequently appear together in a single transaction. This technique is known as association rules [146, 11]. Each association rule normally consists of a set of antecedent items that lead to a consequent item with a certain confidence. Pazzani and Billsus in developing a recommender system for books addressed the cold-start problem by considering the list of topics of books users voted for as transactions. This allowed them to extract association rules for topics that frequently appeared together as part of a user's interests [125]. To 'expand' preferences for each user, the algorithm then generates all possible combinations of topics for every book the user voted for and filters association rules, where the antecedent part of the rule matches one of the combinations. The consequent list of topics is added to the preferences of that user.

In combination with a domain ontology association rules can be effectively employed for extracting, understanding and formalizing new knowledge [141, 145]. However, association rules have to be adapted for recommendation tasks since they are primarily designed to be used as exploratory tools [140] to discover previously unknown relations that need to be analysed for their interestingness [19]. As we will probably want to provide more specificity, and recommend the exact titles of the books instead of generic categories, this potentially leads to a vast number of mined association rules and matching all possible combinations of the observed items may not result in rules being found. Furthermore, a consequent item may appear in multiple matching rules, meaning that a function must be introduced that aggregates the confidences of found rules into a single score for the consequent item. Finally, only the associations with a support (i.e. how often a rule holds as true across the data set) higher than a defined threshold are normally extracted [76]. The produced list of rules is supposed to be of a reasonable size, to allow manual examination. In a recommender system, even associations with low frequencies could still be relevant, if other relevant rules with higher confidence are not found. This requires the extraction of as many rules as possible, making the mining process a computationally expensive task [174].

Roth et al. [137] introduced a method for building implicit social graphs based on histories of interaction between users and estimations of their affinity and applied it to the problem of email recipient recommendation. Based on a set of email addresses selected by a user (the seed group) the algorithm extracts all groups of contacts with whom the user has ever exchanged emails. Here a group of contacts is a set of email contacts that were observed together in a recipient list. For each of the contacts in each group, excluding members of the seed group, an interaction score is calculated based on the volume of messages exchanged with that group, the recency of those messages, and the number of intersections of the seed group with the group of contacts that is being considered. Interaction scores of contacts that are present in multiple groups are aggregated into a single interaction score, which is then used to rank the set of email recommendations.

The implicit social graph is a promising alternative to mining association rules. It instead measures the confidence of recommended items based solely on observed transactions that are pre-filtered by intersections with given items. However, the method also estimates the relevance of a group of recommended emails based on the strength of social interaction of the target user with that group, which is not a meaningful metric for applications that do not assume communication (or other interaction) between users.

Association rules and implicit social graphs are data entities that represent item-to-group relationships. However, DuMouchel et al. [49] suggested that a more efficient approach to discovering "interesting" associations is to first find pairs of items that frequently appear together and then analyse larger sized item sets that contain those pairs. For example, if ABC appears in a data set with a certain frequency, then pairs AB, BC and AC would be at least as frequent as the triplet. Raeder et al. [131] effectively analysed associations through a graph of individual items connected to each other with edges that are weighed by the frequency of two items appearing together. Items that have stronger relationships with each other are compared to other items form clusters, which are then targeted for further analysis. Similar to the implicit social graph this method avoids the need to mine all possible association rules, but without requiring any additional indicators of relevance except for the item pairs frequencies. The relevance of produced recommendations is effectively inferred from the likelihood of their appearing with the observed items.

4.4 Associated food recommender algorithm

4.4.1 Intake24

This section introduces a new recommendation algorithm that was developed for Intake24, a system for conducting 24-hour multiple-pass dietary recall surveys [30]. Within a survey, a respondent typically records their dietary intake for the previous day on three separate occasions. A single day normally consists of four to seven meals (e.g. breakfast, morning snack, lunch etc.) which include a selection of foods, drinks, desserts, condiments, and such (referred to generically as foods). During the first step of a recall session, a respondent reports a list of names of foods consumed during each of the previous day's meals in a free-text format. For each text entry, the system returns a list of relevant foods selected from a food ontology, organized in a tree-based, multi-level structure. Specific foods are terminal nodes of this ontology and are linked to their nutrient values and portion size estimation methods. Respondents select one food from the returned list to add to their meal; for example: *Coca-Cola (not diet); Beef, stir fried (meat only); Tomatoes, tinned; Basmati rice; Onions, fried; Chilli powder; Kidney beans*.

The main application for the recommender algorithm is to automate the extraction of questions about foods that are commonly consumed together (associated food prompts). In Intake24, this feature is implemented as a link between an antecedent food (e.g. *toast, white bread*) and the consequent associated food category (e.g. *butter / margarine / oils*) along with a question that is asked if a respondent selects the antecedent food (e.g. *Did you have any butter or margarine on your toast?*). Such food associations prompts are currently hand-crafted by trained nutritional experts, which for thousands of foods is inevitably a time consuming process that is prone to omissions. At the same time, hand-coded associated food prompts in Intake24 represent a baseline for the recommender system developed in this chapter, meaning that a successful recommender system is expected to capture at least as many omitted foods as the hand-coded prompts.

Completing an accurate recall requires respondents to be able to identify foods they ate from a database that covers more than 4,800 foods; for example, there are more than 30 types of bread alone. Thus, one of the key features in determining the usability of a dietary recall system is its presentation of food search results. If respondents are not able to readily identify items from a list returned in response to their textual description of the food, they are more likely to select foods perceived as the closest match or even skip reporting the intake of that food [58, 30]. In other words, the relevance of search results, in terms of prioritizing them appropriately, may directly affect the the level of effort and time required to select the correct foods and report intake. For that reason, the second application for the recommender algorithm that is explored in this research is the task of ranking food search results.

4.4.2 Generic procedure

Our approach assumes that the patterns in eating behaviours of an observed population; that is, the respondents who took part in surveys conducted in a given country, has some relevance to those of an individual in that population. The recommender algorithm assumes no prior knowledge about an individual except their currently selected food items. The algorithm is trained on a large set of observed meals and produces a model of the eating behaviour of a given population, where a meal is a group of uniquely identifiable foods (e.g. vanilla ice cream, pear juice) reported to be eaten on a single occasion. Each individual food can be recorded as being eaten only once during a meal. During the recommendation step, the resulting model accepts a set of foods, which is referred to as input foods IF, and returns a set of recommended foods RF mapped to likelihoods of being reported along with IF (recommendation scores). IF are excluded from recommendations. The following section discusses three possible implementations that were considered for the recommender algorithm. The description of the methods provides examples of generated models and recommendations for a sample transaction data set.

4.4.3 Association rules

This section introduces a recommender algorithm based on association rules (AR) that generates a model of eating behaviour from a data set of meals (in the training step) in the form of association rules. Each rule consists of a set of antecedent foods and a single consequent food, together with the confidence that the consequent food will be present in a meal given the antecedent foods that were observed. The procedure for retrieving association rules is described in [11].

The AR algorithm makes predictions from stored association rules with antecedent part $antc$ similar to IF, and produces recommendations from the consequent parts of the rules. To do so, AR takes association rules that have a consequent food that is different from any of IF and the antecedent foods $antc$ that include at least one of IF (Alg. 1). The algorithm calculates the likelihood of a recommended food f to be selected next as the confidence of the rule c multiplied by the similarity between $antc$ and IF (i.e. match score ms). The match score ms is calculated as the number of foods that appear both in IF and $antc$ (i.e. intersections) raised to the power of two and divided by the size of IF and the size of $antc$. The match score allows the recommendations from the rules with $antc$ that are more similar to IF produce recommendations that will appear higher. The algorithm then sums the scores for every f as its single recommendation score $RF[f]$.

Algorithm 1: Recommendations based on association rules

function *Recommend*
  input : AM, association rules based model
  IF, foods selected by a respondent
  returns : RF, list of food recommendations
  $RF \leftarrow \varnothing$
  foreach $rule\ rl \in AM\ \&\ rl.consequent \notin IF$ **:**
    $f \leftarrow rl.consequent$
    if $\exists\ af : af \in rl.antecedent\ \&\ af \in IF$ **:**
      if $f \notin RF$ **:**
        $RF[f] \leftarrow 0$
      $antc \leftarrow rl.antecedent$
      $c \leftarrow rl.confidence$
      $intr \leftarrow size(\{af : af \in antc\ \&\ af \in IF\})$
      $ms \leftarrow intr^2/(size(antc) * size(IF))$
      $RF[f] \leftarrow RF[f] + c * ms$
  return RF

Recommendations produced by AR applied to the example transaction data set $\{abcd, ade, de, ab\}$ and given items $\{ab\}$ are provided below (Table 4.1).

Model based on AR	Filtered rules	Recommendations
1. a ⇒ b 0.67, d 0.67, c 0.33, e 0.33	1. **a** ⇒ d 0.67, c 0.33, e 0.33	d: 2.50
2. b ⇒ a 1.00, c 0.50, d 0.50	2. **b** ⇒ c 0.50, d 0.50	c: 1.96
3. c ⇒ a 1.00, b 1.00, d 1.00	6. **a**, **b** ⇒ c 0.50, d 0.50	e: 0.29
4. d ⇒ a 0.67, e 0.67, b 0.33, c 0.33	7. **a**, c ⇒ d 1.00	
5. e ⇒ d 1.00, a 0.50	8. **a**, d ⇒ c 0.50, e 0.50	
6. a, b ⇒ c 0.50, d 0.50	9. **a**, e ⇒ d 1.00	
7. a, c ⇒ b 1.00, d 1.00	10. **b**, c ⇒ d 1.00	
8. a, d ⇒ b 0.50, c 0.50, e 0.50	11. **b**, d ⇒ c 1.00	
9. a, e ⇒ d 1.00	14. **a**, **b**, c ⇒ d 1.00	
10. b, c ⇒ a 1.00, d 1.00	15. **a**, **b**, d ⇒ c 1.00	
11. b, d ⇒ a 1.00, c 1.00		
12. c, d ⇒ a 1.00, b 1.00		
13. d, e ⇒ a 0.50		
14. a, b, c ⇒ d 1.00		
15. a, b, d ⇒ c 1.00		
16. a, c, d ⇒ b 1.00		
17. b, c, d ⇒ a 1.00		

Table 4.1 Recommender algorithm based on association rules applied to the example data set

4.4.4 Transactional item confidence

This section adapts the implicit social graph method described in [137] for the food recommendation task, which resulted in a recommender algorithm based on transactional item confidence (TIC). One key difference to the food recommendation problem is that the original email recipient recommendation task for which the implicit social graph was developed assumed two types of relationships between items in a data set (outgoing and incoming emails). The data set used in this research assumes only one type of relationship, which is the co-occurrence of foods in a meal. For that reason, TIC produces recommendations based on similarity of historically observed transactions to IF and the frequency of foods appearing in those transactions.

During the training step, TIC converts all reported meals to a map of unique meals (or transactions) TM, so that there are no two transactions of the same length containing the same foods (Alg. 2). For every food f in a transaction m, the confidence (conditional probability) is calculated as $TM[m, f]$ of f being present in m given that the rest of the foods from m were observed. To do so, the algorithm counts the number cf of reported meals that contain all the foods from m, excluding f, and divides it by the number cm of reported meals containing all of the foods from m. This is similar to the confidence measured in AR, but in this case the algorithm calculates it only for the full-sized meals that were observed in the data set M, and not for all possible combinations of foods within those meals.

At the recommend step, the algorithm retrieves all transactions containing any of the input foods IF (Alg. 3). Within each of the retrieved transactions m, foods f that are not included in IF are mapped to a score that is calculated as the number of intersections of m with IF (i.e. similarity) multiplied by the food's confidence $TM[m, f]$. Multiple scores for f measured from different transactions are summed into a final recommendation score $RF[f]$.

Algorithm 2: Training the model based on transactional item confidence

function *Train*
    input : M, data set of all meals
    returns : TM, map of unique meals with confidence for every food
    $TM \leftarrow \varnothing$
    foreach *meal* $m \in M$ **:**
        if $m \notin TM$ **:**
            $TM[m] \leftarrow \varnothing$
        $cm \leftarrow size(\{m1 : m1 \in M \;\&\; m \in m1\})$
        foreach *food* $f \in m$ **:**
            $m2 \leftarrow \{f1 : f1 \in m \;\&\; f1 \neq f\}$
            $cf \leftarrow size(\{m3 : m3 \in M \;\&\; m2 \in m3\})$
            $TM[m, f] \leftarrow cf/cm$
    return *TM*

Algorithm 3: Recommendations based on transactional item confidence

function *Recommend*
  input : TM, map of unique meals with confidence for every food
     IF, foods selected by a respondent
  returns : RF, list of food recommendations
  $RF \leftarrow \varnothing$
  foreach $meal\ m \in TM$ **:**
    if $\exists\ f1 : f1 \in m\ \&\ f1 \in IF$ **:**
      foreach $food\ f \in m\ \&\ f \notin IF$ **:**
        if $f \notin RF$ **:**
          $RF[f] \leftarrow 0$
        $conf \leftarrow m[f]$
        $inter \leftarrow size(\{f2 : f2 \in m\ \&\ f2 \in IF\})$
        $RF[f] \leftarrow RF[f] + inter * conf$
  return RF

Recommendations produced by the TIC applied to an example transaction data set $\{abcd, ade, de, ab\}$ given items $\{ab\}$ are provided below (Table 4.2).

Model based on TIC	Filtered rules	Recommendations
1. a 1.00, b 1.00, c 1.00, d 1.00	1. **a** 1.0, **b** 1.00, c 1.00, d 1.00	d: 3.00
2. d 1.00, a 0.50, e 0.50	2. d 1.00, **a** 0.50, e 0.50	c: 2.00
3. d 1.00, e 0.67		e: 0.50
4. a 1.00, b 0.67		

Table 4.2 Recommender algorithm based on transactional confidence applied to the example data set

4.4.5 Pairwise association rules

Unlike the previous two algorithms, which produce recommendations from association rules and transactions similar to currently observed IF, the recommender algorithm based on pairwise association rules (PAR) recommends foods that are likely to be observed with any of IF in pairs. During the training stage (Alg. 4), PAR for every observed food f counts the number $OD[f]$ of meals that contain that food. For every observed pair of foods $\{f, f1\}$, it also counts the number $CD[f, f1]$ of reported meals that contain that pair. At the recommend step (Alg. 5), PAR retrieves pairs $CD[inf]$, where one food inf is observed in IF. For every pair $\{inf, f\}$, the algorithm calculates the conditional probability p, of f being in a meal, given that inf was observed as the number of meals that contain that pair $CD[inf, f]$, divided by the number of meals $OD[inf]$ that contain only inf. For example, if item A appeared 10 times in the data set and co-occurred with item B only 2 times, then the conditional probability that item B will occur the next time the A is present is 0.2. Multiple probabilities retrieved for f from different associations are summed into its single recommendation score $P[f]$.

As demonstrated in [137] the number of times items are observed together is an important relevance metric. Indeed, if the algorithm simply aggregates the probabilities derived from multiple associations, it loses information as to whether a recommended food has ever been observed with all IF. For example, given two input items C and D, the aggregation may produce two scores $R_{CD}(A) = 0.5$, where item A appeared with both C and D, and $R_C(B) = 0.5$, where item B appeared only with item C. Therefore, A should receive a higher score. Likewise, the algorithm takes into account the frequency of an input food inf that matched a retrieved pair. For example, we may have two equal scores, $R_C(A) = 0.5$ and $R_D(B) = 0.5$, where A and B historically appeared only with items C and D respectively; but C appeared 10 times and item D appeared 100 times, which implies that the recommendation produced by C should have a higher score. For these reasons, the algorithm weights aggregated probabilities $P[f]$ by multiplying them by the summed frequency of inf.

Algorithm 4: Training the model based on pairwise association rules

```
function Train
    input   : M, data set of all meals
    returns : PM, pairwise association rules
    OD ← ∅, food occurrences
    CD ← ∅, food co-occurrences
    foreach meal m ∈ M :
        foreach food f ∈ m :
            if f ∉ OD & f ∉ CD :
                OD[f] ← 0
                CD[f] ← ∅
            OD[f] ← OD[f] + 1
            foreach food f1 ∈ m & f1 ≠ f :
                if f1 ∉ CD[f] :
                    CD[f, f1] ← 0
                CD[f, f1] ← CD[f, f1] + 1
    PM ← [OD, CD]
    return PM
```

Algorithm 5: Recommendations based on pairwise association rules

function *Recommend*
  input : PM, pairwise association rules
         IF, foods selected by a respondent
  returns : RF, list of food recommendations
  $RF \leftarrow \varnothing$
  $P \leftarrow \varnothing$, conditional probabilities of foods
  $W \leftarrow \varnothing$, conditional probability weights
  $OD \leftarrow PM[OD]$, food occurrences
  $CD \leftarrow PM[CD]$, food co-occurrences
  foreach *input food* $inf \in IF$ **:**
    foreach *food* $f \in CD[inf]$ & $f \notin IF$ **:**
      if $f \notin P$ & $f \notin W$ **:**
        $P[f] \leftarrow \varnothing$
        $W[f] \leftarrow \varnothing$
      $p \leftarrow CD[inf, f] \ / \ OD[inf]$
      $P[f] \leftarrow P[f] + \{p\}$
      $W[f] \leftarrow W[f] + \{OD[inf]\}$
  foreach *food* $f \in P$ **:**
    $RF[f] \leftarrow sum(P[f]) * sum(W[f])$
  return RF

Recommendations produced by PAR applied to the example transaction data set $\{abcd, ade, de, ab\}$ given items $\{ab\}$ are provided below (Table 4.3).

Model based on PAR	**Filtered rules**	**Recommendations**
1. a 3.0 ⇒ b 2.0, d 2.0, c 1.0, e 1.0	1. **a** 3.0 ⇒ d 2.0, c 1.0, e 1.0	d: 5.8
2. b 2.0 ⇒ a 2.0, c 1.0, d 1.0	2. **b** 2.0 ⇒ c 1.0, d 1.0	c: 4.2
3. c 1.0 ⇒ a 1.0, b 1.0, d 1.0		e: 1
4. d 3.0 ⇒ a 2.0, e 2.0, b 1.0, c 1.0		
5. e 2.0 ⇒ d 2.0, a 1.0		

Table 4.3 Recommender algorithm based on pairwise association rules applied to the example data set

4.5 Methodology

The three algorithms are compared for 20,000 randomly sampled meals, each containing no fewer than two foods, reported by participants of various ages in the UK between 2014 and 2018. The order of foods in each meal is randomized. K-fold ($k = 10$) cross validation was used for segmenting the data set into training and testing sets [142]. On each step nine subsets for training a model are used, leaving out one subset for testing. The testing procedure is similar to the procedure described in [137]: a few foods are sampled from each meal (input foods), leaving the rest (at least one food) to simulate respondents' omitted foods to be guessed by the algorithm.

Every trained model makes predictions, starting from an input size of one food and gradually incrementing it to five.

In the course of the evaluation, the precision-recall (PR) curves for every algorithm on every increment are plotted. For the purposes of the evaluation, the recall is measured as the percentage of correct predictions out of the total number of foods selected by the respondent, and the precision as the percentage of correct predictions out of the total number of predictions made by the algorithm. Predictions that were present in the set of foods actually entered by the respondent (excluding input foods) are counted as correct (true positives). The quality of the top 15 recommendations is analysed, which is a slightly larger size than viewed by most users [162, 33]. As the measure of algorithm ranking quality for every size of input foods the mean value of Normalized Discounted Cumulative Gain (NDCG) at rank 15 is calculated [33] as $NDCG_{15} = DCG_{15}/IDCG_{15}$. Discounted cumulative gain is measured as $DCG_{15} = \sum_{i=1}^{15}(2^{r(i)} - 1)/log(i+1)$, where $r(i)$ is the relevance score of the ith food. As the relevance score, the evaluation uses 0 for a wrong prediction and 1 for a correct prediction. Thus, the Ideal Discounted Cumulative Gain (IDCG) in this case is always 1, which is a single correct prediction as the first result. The procedure then selects the implementation that demonstrates the highest performance and applies it to the task of recommending foods omitted by respondents with a lower level of specificity, and for ranking search results returned in response to their text queries.

In this research, the implementation of AR uses the FP-growth algorithm (frequent patterns algorithm) [76]. FP-growth is an efficient and scalable association rules mining algorithm that is based on building frequent-pattern tree structure. In contrast to Apriori-like algorithms that serve the same purpose [12], the FP-growth compresses a large database into a much smaller data structure avoiding costly repeated database scans and generation of a large number of candidate sets. This research uses a parallel version of FP-growth implemented in the Apache Spark framework [88, 106]. As a parameter, this implementation accepts the minimum support for an item set to be identified as frequent and the minimum confidence for the generated association rules. To gather as many association rules as possible both the minimum support and the minimum confidence are set to the lowest value (3×10^{-4}) that allows the completion of the mining process of the data set on a machine available at the time of conducting this evaluation (Mac Pro, 2.9 GHz Intel Core i5, 16 GB) within a time limit of 5 minutes.

4.6 Results

4.6.1 General performance

As can be observed from the PR curves (Fig. 4.1, 4.2), PAR produces the largest area under the curve, which increases with the size of input foods. PAR also demonstrates higher NDCG than TIC and AR for all input sizes (Fig. 4.3). PAR is the second fastest algorithm to produce a model (after AR) but the fastest to produce a single set of recommendations (Table 4.4). Based on this comparison, PAR is selected to be used for the implementation of the associated foods recommender algorithm. At the same time, these results demonstrate that the quality of predictions produced by PAR is still relatively low. Experiments in the following sections aim to improve the efficiency by exploiting the context of the task it is used for.

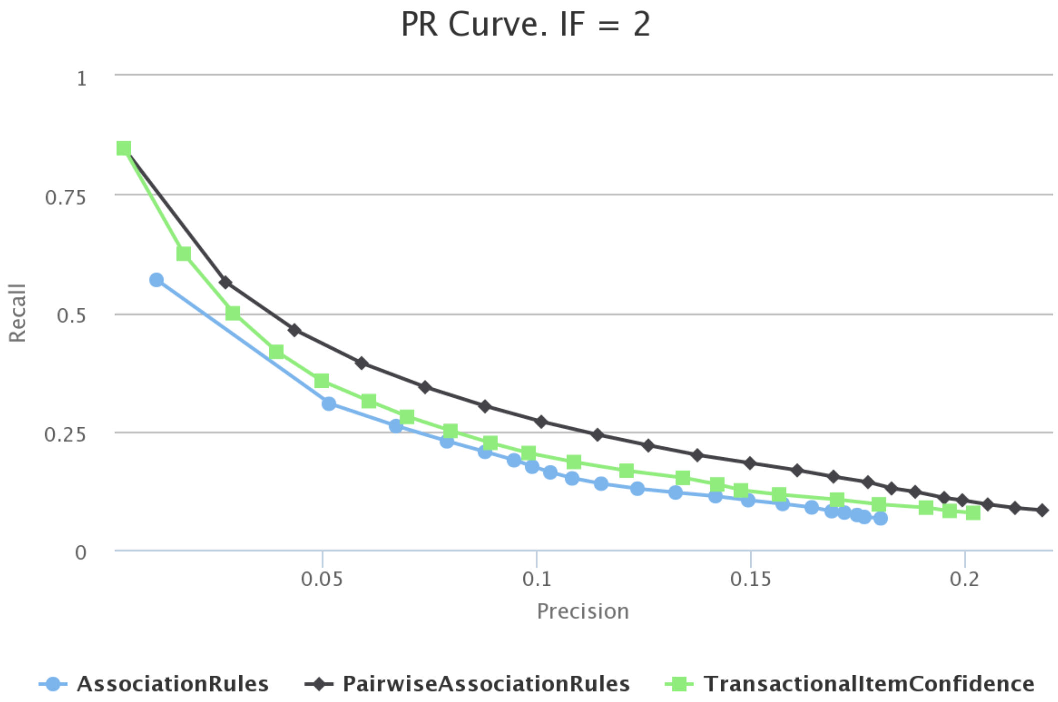


Fig. 4.1 Precision-Recall curves for an input size of 2 foods

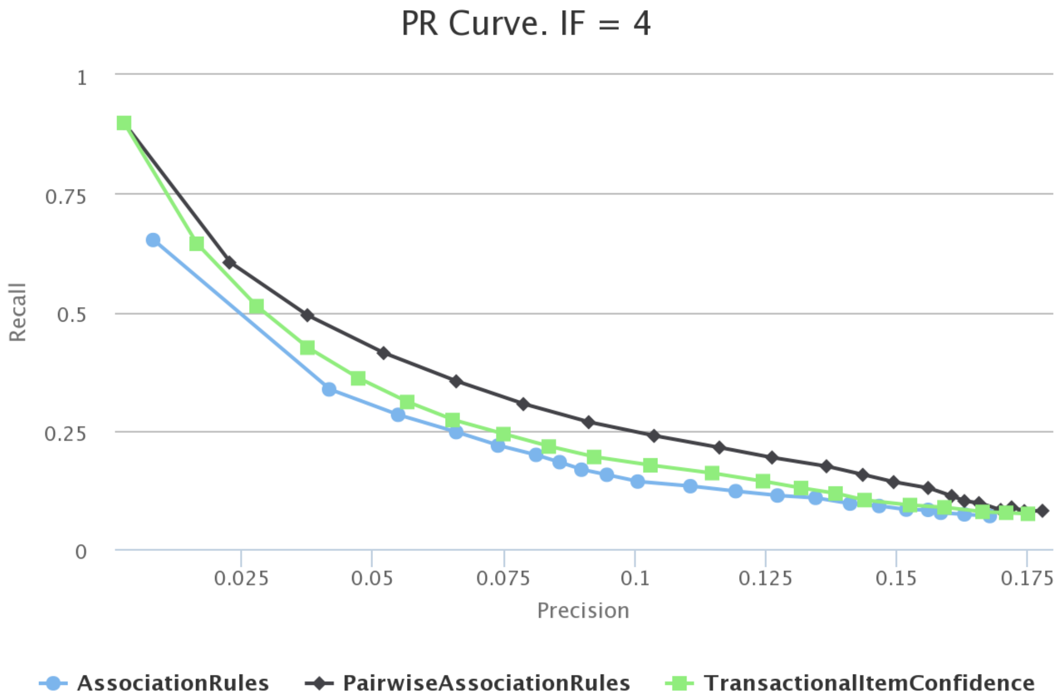


Fig. 4.2 Precision-Recall curves for an input size of 4 foods

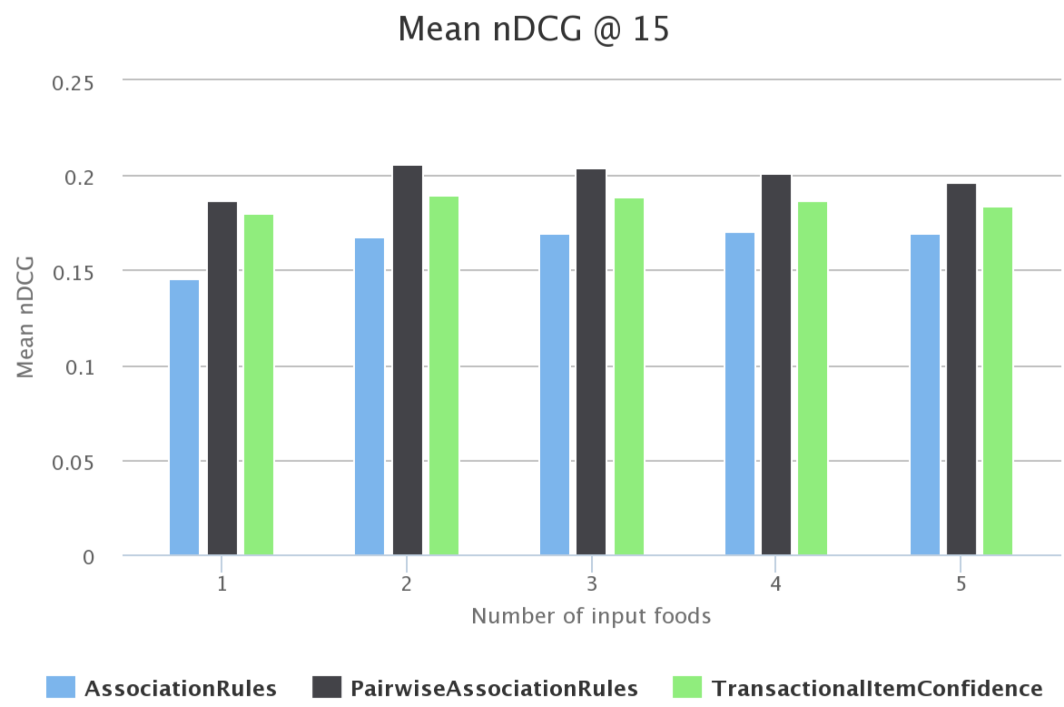


Fig. 4.3 The ratio of mean NDCG for the top 15 results to the number of input foods

Model	Training	Mean recommendation
AR	3905.1	39.5
PAR	6904.9	2.5
TIC	93710.2	32.0

Table 4.4 Mean training and recommendation times in milliseconds

4.6.2 Associated food questions

To compare the efficacy of recommendations produced by the recommender algorithm to the existing hand-coded associated food questions the evaluation procedure follows the same protocol as above, except that on the recommend step a trained model returns food categories instead of exact foods. In this case, true positives are considered to be foods selected by the respondent (excluding input foods) that belong to one of the food categories predicted by the recommender algorithm. The ontology of foods implemented in Intake24 allows control of the specificity of the returned categories. So, this experiment demonstrates the performance of the algorithm in returning the direct parent category of a food (first level, e.g. *Flake cereals* is the parent category for *Choco flakes*) and a more generic category (second level, e.g. *Breakfast cereals*) that is a parent of the category with the first-level specificity (Fig. 4.4). Since the existing associated food questions do not store any relevance scores plotting their PR curves or assessing their NDCG is impossible. For that reason this section compares the recall of the top 15 recommendations produced by the algorithm to the recall of all hand-coded associated foods rules extracted for given input foods.

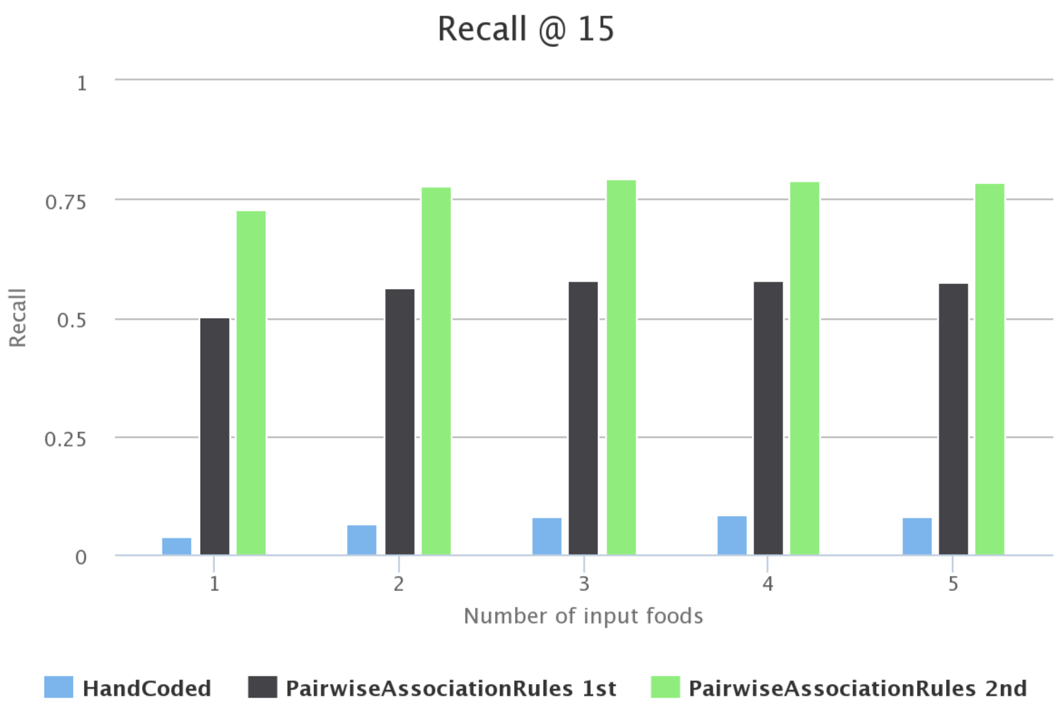


Fig. 4.4 The ratio of recall for the 15 results to the number of input foods for pairwise association rules with the first and the second levels of specificities and manually entered associated food prompts

In the simulation of respondents omitting foods hand-coded association food rules recognize 8.3% of omitted foods at most, whereas the recommender algorithm's peak recall is at 58.0% and 79.1% for the first and the second levels of specificity respectively.

Table 4.5 includes examples of commonly forgotten foods established in the validation of Intake24 [30] but correctly predicted by the recommender algorithm with two levels of specificity. At the time of writing this chapter, none of these associations were covered by hand-coded associated foods rules in Intake24. In addition to that, controlling the specificity of the returned recommendations allows us to address the cold-start problem, so that new foods that have not been reported by any respondents can still be captured by their categories. However, the names of some food categories predicted with the second-level specificity could be perceived as too generic (e.g. "Pickles, olives, dips and dressings") and may require being assigned names that would be easier to understand by respondents when displayed in associated food prompts.

Input foods	First-level specificity	Second-level specificity
Chicken breast; Fanta; Instant potato	Gravy	Sauces, condiments, gravy and stock
Bananas; Fruit and yoghurt smoothie; Semi skimmed milk	Sugar	Sugar, jams, marmalades, spreads and pates
Blackcurrant squash (juice), e.g ribena; Heinz beans and sausages	Brown bread toasted	Brown, wholemeal and 50:50 bread
Porridge, made with skimmed milk; Tea; White sugar	Butter	Butter / margarine / oils
Tuna mayo sandwich; Volvic mineral water, still or fizzy	Chocolate covered biscuits	Sweet biscuits
Bread sticks; Coffee	Dips	Pickles, olives, dips and dressings
Cheese and tomato pizza (includes homemade); Raspberries	Ice cream	Ice cream & ice lollies
Cheese sandwich; Tea	Crisps and snacks	Crisps, snacks and nuts
Green Olives; Water	Wine	Alcohol
Bottled mineral water; Chicken breast fillet; Chips, fried; Hot sauce	Fizzy drinks	Drinks
Still energy drink, eg Lucozade Hydroactive, Gatorade, Powerade; Tuna in brine, tinned; White bread sliced	Mayonnaise	Sauces / condiments / gravy / stock

Table 4.5 Omitted foods captured with pairwise association rules but not with manually entered associated food prompts

4.6.3 Search ranking

In response to a respondent's text query, the existing Intake24 search algorithm ranks foods based on two types of scores. The first is the matching cost of the known food description against the query. The matching cost is based on several metrics, including the edit distance between matched words (the approximate string matching is performed using Levenshtein automata [144]); phonetic similarity of words (using a pluggable phonetic encoding algorithm that depends on the localization language, e.g. Soundex or Metaphone for English [50]); the relative ordering of words; the number of words not matched; and so forth. The lower the matching cost, the better the food name matches the query. The second score is the likelihood of the food being selected, which is measured by the number of times the food was previously reported. The results are then sorted, first by decreasing food report count (FRC) and then by increasing matching cost.

The evaluation of the associated foods recommender algorithm applied to the task of ranking search results, follows the same evaluation procedure, with some variations. In response to each text query that was recorded into the Intake24 database for each reported food (excluding input foods), the evaluation retrieves a list of foods using the existing search algorithm. Foods selected by a respondent are counted as true positives and the rest of the results as false negatives. The mean NDCG produced by the existing search algorithm is compared to that of the new search algorithm, where FRC is replaced with PAR. As it can be seen from the figure below (Fig. 4.5), PAR slightly outperforms FRC starting from an input size of two foods, with the gap gradually widening as the number of input foods increases.

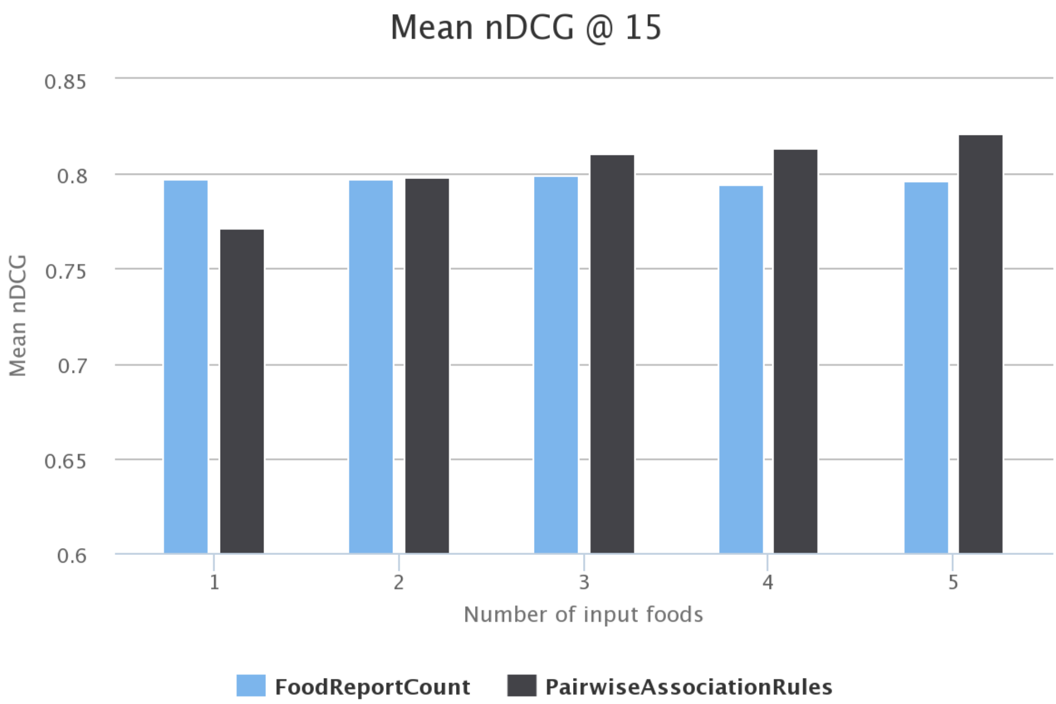


Fig. 4.5 The ratio of mean NDCG for the top 15 results to the number of input foods for the search results ranked based on pairwise association rules and FRC

4.7 Conclusions

This chapter aims to address one of the key issues in automated dietary assessment, which is unintentional under-reporting. To do so, this research developed an associated foods recommender algorithm to remind respondents of omitted foods and improve the ranking quality of search results returned in response to respondents' free-text food name queries. The algorithm, in contrast with collaborative and content-based filtering approaches, is independent of personal user profiles and does not require an extensive history of users' preferences or a multitude of item descriptors. Instead, the algorithm uses transactions performed by respondents from a given population to build a collective model of preferences.

This work considered three algorithms for the implementation of the recommender system, based on an implicit social graph [137], association rules [11], and analysing pairwise association rules [49]. The evaluation, performed on a large data set of real dietary recalls, has demonstrated that the implementation based on pairwise association rules performs better for the defined task. By controlling the specificity of the produced recommendations within a reasonable level the recommender system achieved a recall of 79.1%. That is significantly higher than food associations hand-coded by trained nutritionists, the recall for which reached only 8.3%. Where a respondent filled in at least one food, the recommender algorithm improves the ranking of search results. At the same time, these results do not necessarily mean that the hand-coded prompts are less accurate than those by the recommender system in a real survey.

The algorithm was evaluated on dietary recalls of respondents from the UK. The future extension of this research can analyse how dietary specificities of different regions affect the accuracy of the recommender algorithm. Although the evaluation results described in this chapter were produced by analysing food contents of meals reported by respondents in Intake24, the described methods apply to any recommender tasks where selection of items by the target user can be observed (e.g. email recipients or tags recommendations on community platforms).

Chapter 5

Validation of a recommender system for prompting omitted foods in online dietary assessment surveys

5.1 Abstract

Recall assistance methods are among the key aspects that improve the accuracy of online dietary assessment surveys. These methods still mainly rely on experience of trained interviewers with nutritional background, but data driven approaches could improve cost-efficiency and scalability of automated dietary assessment. This chapter evaluates the effectiveness of a recommender algorithm developed for an online dietary assessment system called Intake24 that automates the multiple-pass 24-hour recall method. The recommender builds a model of eating behaviour from recalls collected in past surveys. The model is used to remind respondents of associated foods that they may have omitted to report based on foods they have already selected. The performance of prompts generated by the model was compared to that of prompts hand-coded by nutritionists in two dietary studies. The results of this research demonstrate that the recommender system is able to capture a higher number of foods omitted by respondents of online dietary surveys than prompts hand-coded by nutritionists. However, the considerably lower precision of generated prompts indicates an opportunity for further improvement of the system. The work described in this chapter has been undertaken during RA3 and informs the answer to RQ2. The content of this chapter has been rewritten as an (accepted) conference paper for the PervasiveHealth 2019: Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., and Foster, E. (2019b). Validation of a recommender system for prompting omitted foods in online dietary assessment surveys. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 208–215. ACM [123].

5.2 Introduction

The involvement of skilled professionals in the collection and analysis of complex dietary data has both economic implications and implications for scalability. To address these issues a number of systems have been developed that replace an interviewer in the 24-hour recall method with an online survey [39, 148, 154]. For researchers, such systems provide tools to administer surveys and collect dietary records from participants with a detailed breakdown of the foods and drinks consumed, to enable the estimation of energy intake and intakes of macro- and micro-nutrients. Respondents are asked to go through a survey in a form of a web-based interface and record their dietary intake for the previous day. The survey comprises a series of questions about each meal and all its constituent foods and drinks. The collected information includes names of foods/drinks and their portion sizes. Portion sizes are self-estimated by a respondent using validated photographs of weighed servings of foods [59]. Each respondent normally records their meals for the previous day on three separate occasions. A single day (i.e. submission) typically consists of four to seven meals (e.g. breakfast, morning snack, lunch, evening meal etc.). Each reported meal may include a selection of foods, drinks, desserts, condiments and such (referred to generically as foods).

Participants commonly omit foods and under-estimate portion sizes in dietary assessment surveys due to a range of human factors including poor human memory or lack of attention [100, 159]. A study of 83 adults between 20 and 60 years old demonstrated that interviewer-led 24-hour recall underestimated energy intake on average by 33% of that measured with DLW [97]. Various state-of-the-art techniques have been proposed by the research community to aid respondents' self-reported intake recall and minimize under-estimation. For example, images of food captured by respondents using handheld devices or wearable cameras can enhance the quality of self-reported dietary surveys by revealing unreported foods and misreporting errors [65, 13]. In future, image recognition algorithms may be able to automatically identify foods and drinks captured on those images, estimate their portion sizes and even their energy and nutrient content [175, 92]. Another direction for improving the accuracy of dietary assessment is monitoring intake behaviour using wearable sensors and sensors embedded into the environment [79]. Detected eating and drinking occasions can be used, for example, to provide additional cues for respondents during their recall, and to remind respondents to record their meal shortly after it, when they still remember all the details. However, wearable sensors are likely to be logistically challenging and impose high costs when used on a large scale. In the meantime, one of the most widely adopted methods for assisting recall in online dietary assessment surveys in epidemiological studies are associated food questions [154, 30]. An associated food question is a question about two foods that are commonly consumed together and is asked, when the respondent reports one of these foods. For example, when the respondent reports having eaten toast, the system may ask them "Did you have butter on your toast?". These questions aim to remind respondents about omitted foods. Normally each associated food question is manually recorded into the system's database by a nutritionist or a dietitian along with a corresponding relation between an antecedent food (e.g. white bread, toast) and a consequent food (e.g. butter) [122]. The system returns a prompt containing the question once a respondent reports the antecedent food. However, identifying relations for thousands of foods stored in the system seems impractical. Moreover, even an experienced nutritionist may not always be able to identify all the factors and aspects of a diet of a specific population.

The previous chapter (chapter 4) describes the development of a recommender system based on pairwise association rules that aids respondent's recall. The main contribution of this chapter is the evaluation of the recommender system in online dietary assessment surveys, in real-life settings. Further, the chapter briefly reviews the recommender system developed for Intake24. The chapter then presents a report and discussion for the design and results of two studies, in order to compare the efficiency of the recommender to that of prompts added by nutritionists.

5.3 Recommender system

Eating habits depend on a range of factors, including region, culture and a specific diet, which makes them hard to predict. Thus, manual extraction of associated food questions and keeping them up-to-date is prone to omissions and is a time-consuming task. Those challenges motivated the development of a recommender system that extracts food associations in an automated manner. To build a model of eating behaviour the recommender system takes a dataset of all meals reported by a given population, where a meal is a group of foods (e.g. toast, butter, cheese, orange juice) reported in a single intake. Each meal is split into unique pairs of foods (e.g. toast and butter, toast and cheese, toast and orange juice, butter and cheese, etc.). For every pair the resulting model stores the number of meals that contain this pair. The model also stores the number of meals that contain each individual food. To produce a recommendation the algorithm takes foods reported by the current user to find pairs stored in the model that contain those foods. Foods from the filtered pairs that have not been reported by the user yet are used as recommendations. The list of recommended foods is sorted by their likelihood of being observed in a meal having observed the reported foods in descending order. In other words, the algorithm recommends foods that are more likely to be observed with any of reported foods in pairs.

The likelihood for a recommended food f is calculated as $R_f = C_f \times W_f$, where C_f is the aggregation of conditional probabilities of observing that food in a pair with one of the reported foods f_i in a meal having observed that reported food; and W_f is the weight of the aggregation. C_f is calculated as $C_f = \sum_{i=0}^{N} \frac{count(f,f_i)}{count(f_i)}$, where N is the number of reported foods, $count(f,f_i)$ is the number of meals in the dataset that contain the pair (f,f_i) and $count(f_i)$ is the number of meals in the dataset that contain the reported food f_i individually. W_f is calculated as $W_f = \sum_{i=0}^{K} count(f_i)$, where K is the number of reported foods f_i, for which a pair containing the recommended food f being considered has been found. Thus, the reported foods that were previously observed with the recommended food and that were reported more often in the dataset give a higher weight to the recommendation.

The previous chapter analyses the performance and demonstrates the effectiveness of the recommender system over food associations hand-coded by nutritionists in a simulation of respondents omitting foods with data collected from past real-life dietary surveys. This chapter describes the deployment and comparison of the recommender system to the hand-coded food associations in conditions of real-life dietary surveys.

5.4 Methods

5.4.1 Interface design

The existing associated food prompts that are based on links between two foods manually added into Intake24 are returned immediately after one of the foods has been reported (Fig. 5.1, 5.2). So, a respondent is typically prompted with multiple foods while reporting a single meal. In contrast to that, recommendations produced by the recommender system are based on a meal (selection of foods) and have a form of a list. For that reason, a screen was designed with generated food prompts in the form of a checkbox list that is returned at the end of reporting a meal (Fig. 5.3, 5.4). A respondent can accept multiple foods as with the hand-coded prompts. A list of recommendations is limited to 15 foods, which is a slightly larger number than the number of search results displayed to most users by online search websites [33, 162].

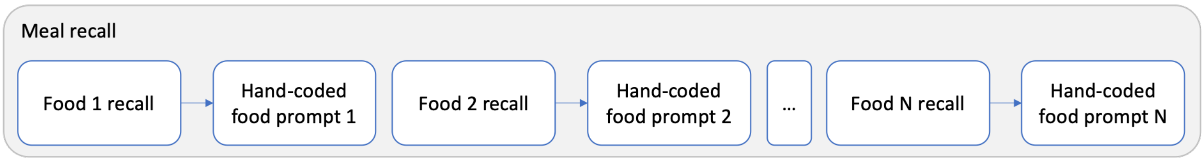


Fig. 5.1 Interface flow diagram for hand-coded associated food prompts in Intake24.

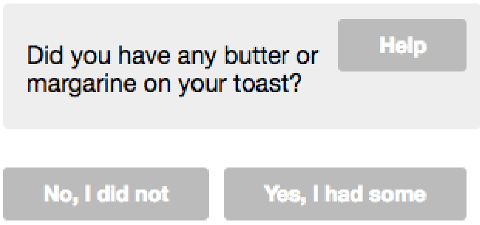


Fig. 5.2 Hand-coded associated food prompt in Intake24.

Fig. 5.3 Interface flow diagram for generated associated food prompts in Intake24.

5.4.2 Recruitment and procedure

5.4.2.1 First study.

The recommender system for prompting omitted foods was evaluated in a study with 49 participants ($n = 24$ females, $n = 26$ males, 18-64 years) that was conducted during RA2 that is described in chapter 3. In this study the performance of associated food prompts generated with

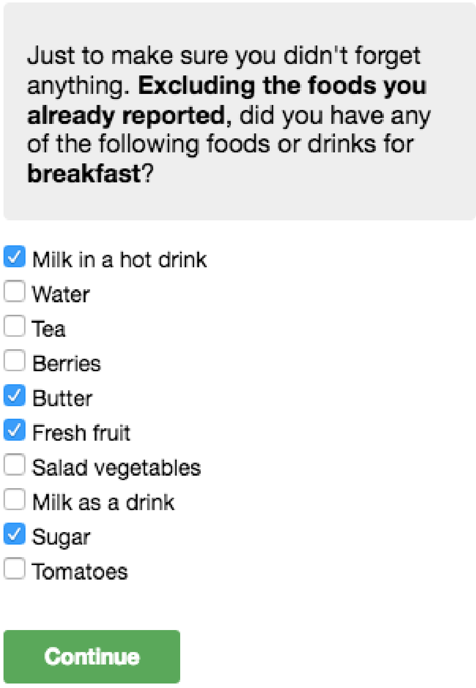


Fig. 5.4 Generated associated food prompt in Intake24.

the algorithm was compared to that of hand-coded food prompts. The overview of this study that is related to the current research is provided in this chapter as well. Participants were asked to record their intake for five consecutive days from Monday to Friday. To help respondents to follow the procedure they were asked about their preferred morning times to receive reminders about completing a dietary recall. Every morning they received automated email and SMS reminders to log onto Intake24 and report the meals they had for the previous day. Participants were asked to avoid changing their diet during the study and not to record their meals elsewhere (e.g. notepads) to aid their recalls. During the first three days (Monday - Wednesday) the system presented one type of associated food prompt and in the remaining two days (Thursday - Friday) the other type of prompt. The system presented hand-coded food prompts in the first three days to $n = 19$ participants. For $n = 30$ participants the system presented hand-coded food prompts in the last two days. Monday recalls were used as an introduction to minimize the learning effect by familiarizing participants with the system's interface and the 24-recall procedure. Those recalls were discarded during the analysis of the results and only diet recalls of weekdays (Monday - Thursday) were used.

5.4.2.2 Second study.

The second study was collected during the Newcastle Can campaign that aims to help to reduce obesity levels in the North East of England during RA3 that is described in chapter 3 [5]. The campaign used various methods to motivate its participants to improve their diets. One of the methods involved participants recording their meals for a previous day using Intake24 to receive feedback on their diet and links to NHS web pages with more information about healthy eating. Compared to the first study, the second did not pose any specific requirements to the time of a day or days of the week, when participants needed to record their intake. The study did not require respondents to use any specific device or to submit a certain number of recalls either. Participants could join the campaign at any point during the study. This study collected recalls with hand-coded prompts for a week and recalls with generated prompts for another week. During this period the second study attracted $n = 91$ respondents to complete at least one recall, of whom $n = 77$ were women and $n = 14$ were men. The age of participants in this study ranged from 18 to 82 years.

5.4.3 Statistical analysis

The analysis measures the proportion of recalls with generated and hand-coded food prompts, where respondents accepted at least one food. This research also analyses the mean number of foods accepted per recall in these two settings. In addition, precision of food prompts is measured, that is the number of accepted foods divided by the number of returned foods. Many participants who reject associated food prompts that are still relevant for other participants genuinely believe that they reported all foods. For example, although someone may drink coffee without milk and sugar, a food prompt querying about milk and sugar in coffee is still relevant for many. For that reason, the mean number of accepted foods is analysed only for recalls, where at least one prompted food was accepted. The analysis compares the coverage of two types of prompts through the number of unique foods that were returned and accepted. The analysis also compares the estimates of energy reported with two types of prompts. Submissions with abnormally low-calorie content are excluded from the analysis (below 250 kcal for the whole day). To examine changes in usability of the system this work compares the mean duration of recalls, i.e. time it took respondents to complete a survey. This comparison assumes that longer recalls may indicate that food prompts generated by the recommender system negatively affect the usability of the survey interface. Previous research shows that participants complete their recalls using Intake24 in 14 minutes on average [139]. For that reason, the analysis of the mean duration of recalls ignores those that took longer than 60 minutes, since that could indicate that respondents took a break while completing their recall. To analyse the significance between two given means the Mann-Whitney U test is used. Mann-Whitney U test is selected for this research since it does not require data being normally distributed [103, 114].

5.4.4 Results

The first study resulted in 96 submissions with hand-coded and 97 with model-generated food prompts. In the survey with hand-coded prompts 66 (69%) recalls were completed on a desktop computer, 29 (30%) on a mobile device, and 1 (1%) on a tablet. In the survey with generated prompts 67 (69%) recalls were completed on a desktop computer, 27 (28%) on a mobile device, and 3 (3%) on a tablet. Hand-coded food prompts were displayed at least once in 86 recalls and accepted in 57 recalls (66%). Generated prompts were returned in 97 recalls at least once and accepted in 61 recalls (63%). In recalls, where respondents accepted at least one associated food prompt, 1.7 foods per recall on average were accepted from the hand-coded prompts. For the generated prompts that is 2.3 foods, which is significantly higher ($P < 0.001$). Precision of the hand-coded and generated food prompts for the first study across all recalls are 24% and 2% respectively.

Participants in the second study submitted 133 recalls with hand-coded and 119 with model-generated food prompts, of which 41 (31%) and 57 (48%) recalls respectively were submitted between Saturday morning and Monday night. Intake records assisted by the hand-coded prompts were submitted on average between 1:20pm and 9pm. Similarly, recalls assisted by the generated prompts were submitted on between 1:25pm and 8:30pm. In the survey with hand-coded prompts 49 (37%) recalls were completed on a mobile device, 44 (33%) on a desktop, and 40 (30%) on a tablet. In the survey with generated prompts 60 (50%) recalls were completed on a mobile device, 26 (22%) on a desktop, and 33 (28%) on a tablet. Hand-coded food prompts were displayed at least once in 122 recalls and accepted in 61 recalls (50%). Generated prompts were returned in 115 recalls and accepted in 83 recalls (72%). As in the first study, there is a significant difference between the mean acceptance rates. Participants on average accepted 1.5 foods per recall from hand-coded prompts and 2.1 from generated prompts ($P = 0.002$) in recalls, where participants accepted at least one associated food prompt. Precision of the hand-coded and generated food prompts for the second study across all recalls are 16% and 2% respectively. Histograms of acceptance rates for both types of prompts in two studies are provided in Fig. 5.5, 5.6.

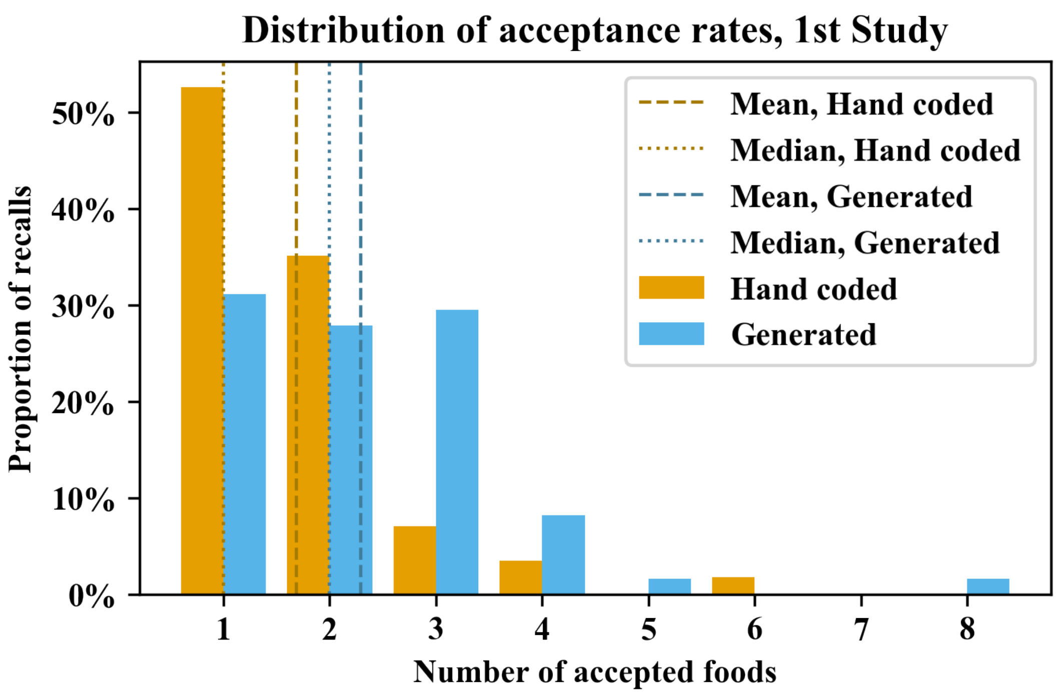


Fig. 5.5 Distribution of accepted foods per recall during the first study.

In the first and the second studies with hand-coded prompts the number of unique foods accepted by participants was 15 (9%) and 16 (9%) out of 164 and 186 unique reported foods respectively (Fig. 5.7, 5.8). In surveys with generated prompts, this number was found to be at least twice as high with 30 (18%) and 35 (19%) out of 165 and 189 unique reported foods (Fig. 5.7, 5.8).

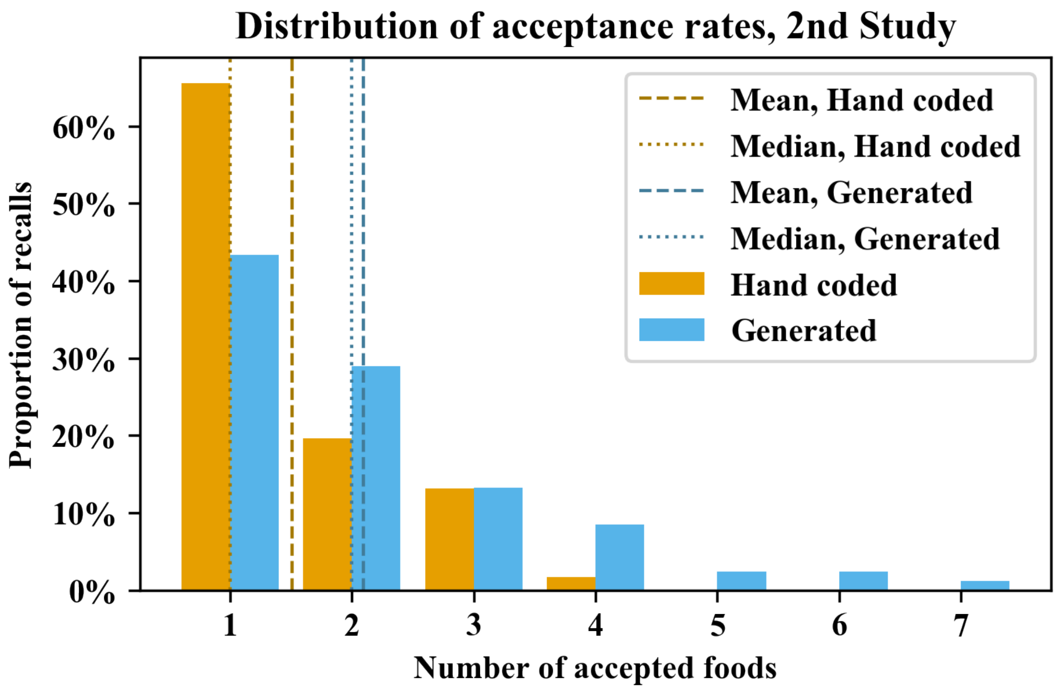


Fig. 5.6 Distribution of accepted foods per recall during the second study.

No significant difference was found between the mean energy reported in recalls with hand-coded food prompts (1911.8 kcal) and generated food prompts (1790.6 kcal) during the first study (P = 0.159). However, the effect was observed during the second study, where the mean reported energy for hand-coded prompts was 1461.7 kcal and 1545.7 kcal for generated prompts (P = 0.02).

While examining the mean duration of recalls, 7 (4%) recalls were excluded from the first study and 15 (6%) from the second study because these were longer than 60 minutes. No significant difference was observed in the duration of recalls in either study. The mean duration of recalls during the first study was 15.9 minutes for recalls with hand-coded prompts and 13.3 minutes for recalls with generated prompts (P = 0.108). During the second study the mean duration of recalls was 15.9 and 16.3 minutes respectively (P = 0.297).

5.5 Discussion

5.5.1 Principal findings

The results of this research demonstrate that associated food prompts generated by the recommender system based on pairwise association rules are an effective and scalable alternative to prompts hand-coded by nutritionists in online dietary assessment surveys. This supports findings from our previous research [122]. In the meantime, according to the recommendations from the NHS UK a man needs around 2,500 kcal a day to maintain his weight given a healthy and a balanced diet; for a woman, that figure is around 2,000 kcal a day [9]. Compared to that, the mean energy intake estimated in the first study of this research is considerably low, which suggests

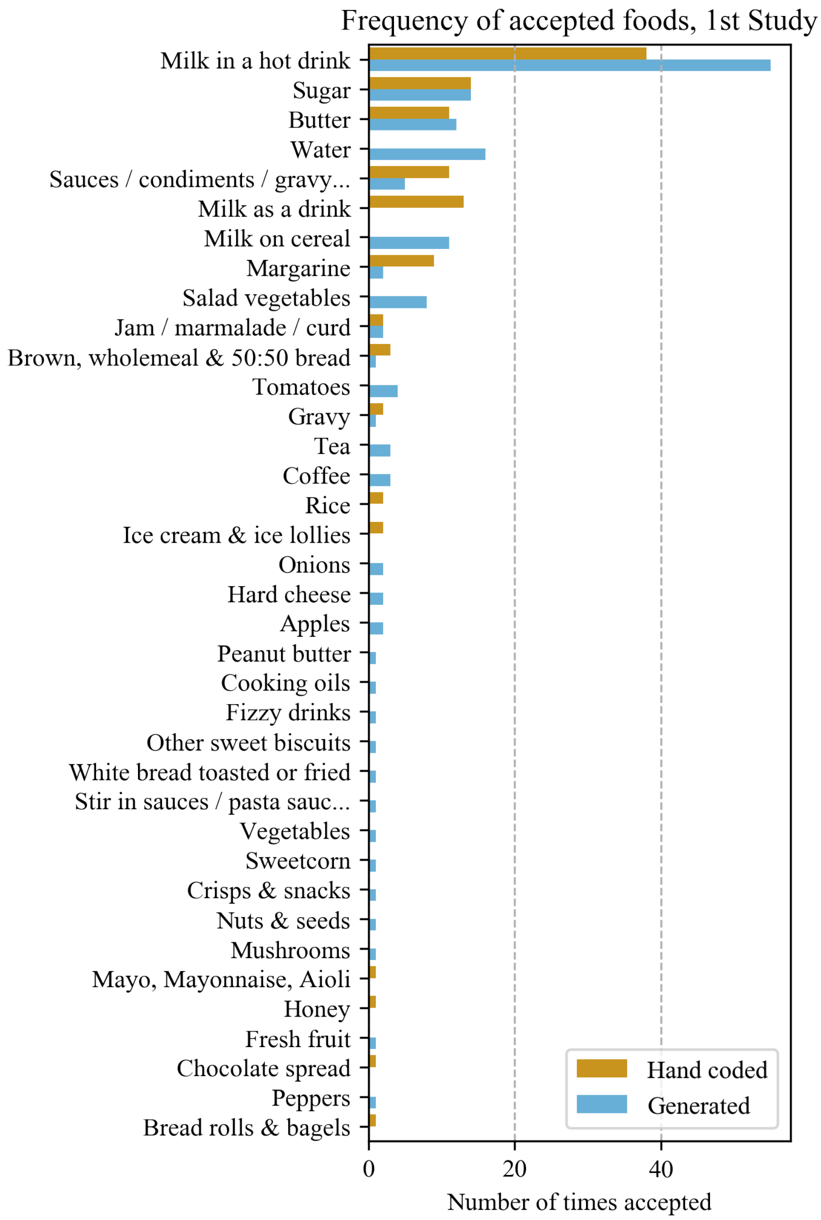


Fig. 5.7 Frequency of accepted foods during the first study.

that there were cases of omissions and under-estimation. This is in agreement with previous research that found Intake24 to underestimate energy intake by 1% on average compared to the interviewer-led recall [30]; and that the interviewer-led recall underestimates energy intake by 33% on average compared to that measured with DLW [97]. The second study was conducted as part of a weight loss campaign, which could explain even lower energy estimates. At the same time, participants of the second study reported higher energy intake in recalls assisted by generated food prompts, which may indicate an improved accuracy of assessment.

Meanwhile, generated food prompts have shown significantly lower precision compared to that of hand-coded prompts. This is due to a much longer list of recommendations produced by the system. At the same time, only 30 and 35 unique foods that had been recommended by the system were accepted across all recalls in the first and second studies. This is in agreement with previous work that indicates that some foods are more likely to be omitted [30]. This also explains

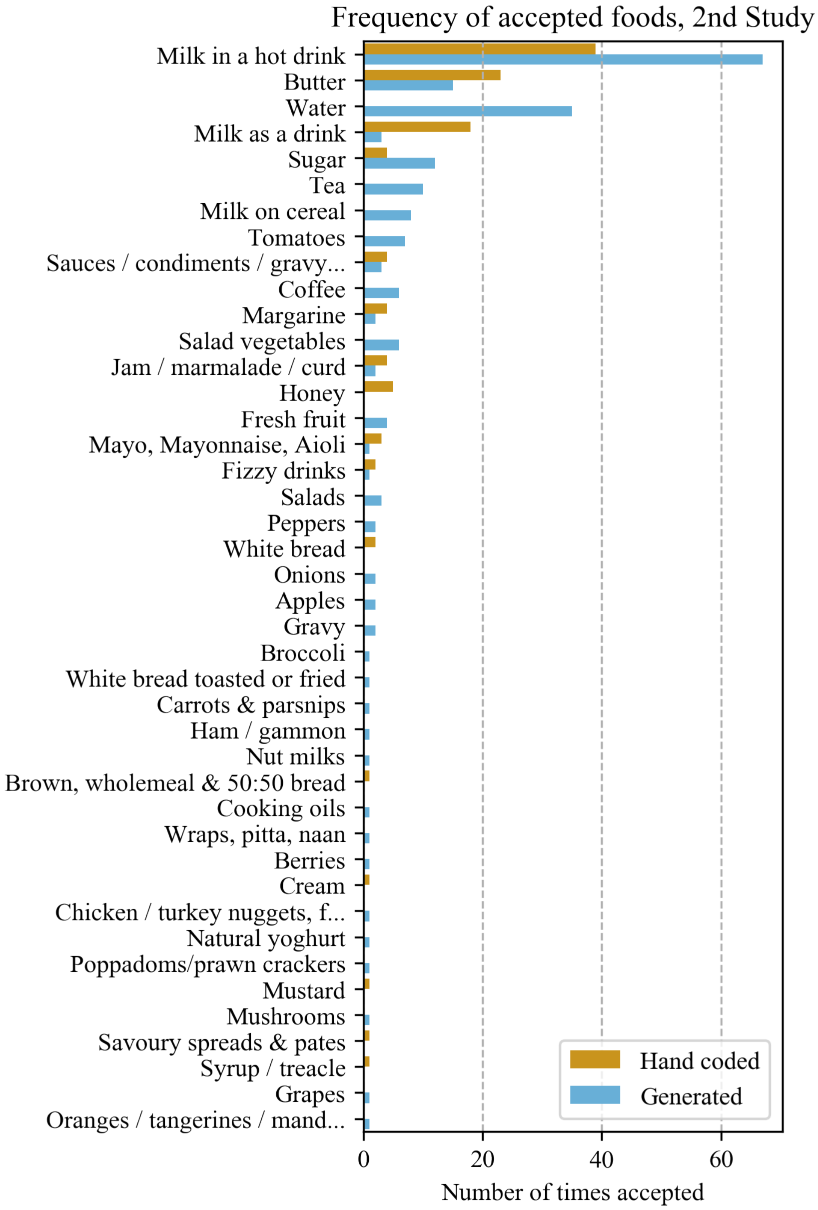


Fig. 5.8 Frequency of accepted foods during the second study.

a considerably higher performance of the recommender in a simulation that was conducted in the previous chapter, where the evaluation procedure assumed that any food can be omitted. To shorten the list of recommendations, a future version of the algorithm is planed to incorporate the acceptance rate of recommended foods into the recommender system. This will potentially allow placing foods that are more likely to be forgotten higher up the list of recommendations. The observed proportion of desktop and touch devices used in recalls assisted with hand-coded and generated prompts is comparable in two studies. In the first study, respondents submitted their recalls in the morning. In the second study, time of the day when respondents recorded their intake remained similar in both conditions. Thus, a comparable duration of recalls assisted with hand-coded and generated prompts may indicate that the usability of the system was not affected by the longer list of generated food prompts.

It should be noted that two types of prompts have a different presentation format. Hand-coded prompts are presented as questions that can be accepted or rejected by a respondent during reporting a meal. Generated prompts are presented in a form of a list after the meal that has been just reported. Thus, the observed greater number of accepted food items from the recommender system may be an effect of the presentation format rather than an effect of the better fit of the suggested food items. In addition to that, the recommender system was trained on data collected in the past surveys, where hand-coded associated food prompts were used. Therefore, there is a chance that would there be no hand-coded prompts, respondents might have reported some foods associations less often or not reported them at all. Hence, the recommender system potentially would not pick those associations from the data in the training process. Nevertheless, in our two studies, more than a half of foods captured by the recommender system were not defined in the database of hand-coded prompts. Thus, the two methods of prompting could potentially complete each other at least when the system has been deployed for a new population and there is no representative data for training the recommender system yet.

5.5.2 Limitations

The current studies involved a relatively small number of participants, and the recruitment method used in the first study meant that the demographics profile of respondents was limited which may limit the applicability of the results. Furthermore, in the second study, participants were members of a weight loss campaign which may imply that they are more health-conscious than the general population and potentially more invested in providing an accurate picture of their diet. Finally, the age and gender distribution of participants was not balanced in both studies. Thus, further research is needed to confirm findings from this research using a wider range and balanced distribution of demographical backgrounds.

In the first study, participants did not submit recalls for Friday and weekend days, when people commonly consume more energy and diet is less structured than on the week days. That is somewhat addressed in the second study that included weekends. However, in the second study participants could submit their meals at any time of the day, which contradicts with the original multiple pass 24-hour recall method, where participants are asked to record their meals in the morning. To address these concerns a similar study should follow the procedure of the 24-hour recall method offering respondents to complete their recalls for three non-consecutive work days and on one day over the weekend.

Lastly, in both studies, in cases when participants rejected food prompts, it was not possible to verify whether they genuinely reported all foods, ignored prompts or prompts did not contain any relevant foods. Similarly, when respondents accepted prompts, there is no evidence that they actually consumed those foods. In future, meals reported by participants could be verified against direct observations.

5.6 Conclusions

This chapter aimed to improve methods used in online dietary surveys to assist respondent's memory and improve the accuracy of dietary assessment. This work validated a recommender system that was designed in chapter 4 and that prompts about foods potentially omitted by respondents during a dietary survey as an alternative to prompts hand-coded by nutritionists and dietitians. In contrast to other contexts, where recommender systems are used (e.g. online retail, entertainment), dietary assessment systems are limited in their ability to collect enough data for each individual to build a model of their personal preferences. For that reason, the recommender system described in this work mines a model of eating behaviour of a population from data collected in previous dietary surveys. To validate associated food prompts generated by the recommender system the chapter compared their performance to prompts that were hand-coded by nutritionists in two real-life online dietary assessment surveys. The validation demonstrated the ability of the recommender system to capture significantly larger number of foods omitted by a respondent per recall than that by the hand-coded prompts. Moreover, the number of distinct foods accepted by respondents from the generated prompts was at least twice as high as that from the hand-coded prompts in both studies. That indicates that more than a half of omitted foods captured by the recommender system were not foreseen by nutritionists. In the meantime, to ensure that a dataset that is used for training the recommender system contains associations of foods commonly omitted by respondents it should be collected in surveys assisted by prompts hand-coded by professionals with nutritional background. Judging by the average time it took respondents of both studies to complete their recalls no difference was observed in complexity of completing a survey with the two prompting methods. At the same time, considerably low precision of generated prompts indicates an opportunity to further improve the relevance of recommendations produced by the system. The results produced in this chapter indicate the effectiveness of using the recommender system over hand-coded rules in online dietary assessment surveys, where there is a rich dataset of past surveys for training the system.

Chapter 6

Progressive 24-hour recall: Feasibility and acceptability of short retention intervals in online dietary assessment surveys

6.1 Abstract

Background: Under-reporting due to limitations of human memory is one of the key challenges in dietary assessment surveys that use the multiple-pass 24-hour recall. Research indicates that shortening a retention interval (i.e. the time between eating event and recall) reduces the burden on memory and may increase the accuracy of assessment.

Objective: This chapter explores the accuracy and acceptability of online dietary assessment surveys based on a progressive recall, where a respondent is asked to record multiple recalls throughout a 24-hour period using the multiple-pass protocol and portion size estimation methods of the 24-hour recall.

Methods: The experiment is conducted with a dietary assessment system Intake24 that typically implements the multiple-pass 24-hour recall method where respondents record all meals they had for the previous day on a single occasion. This chapter modified the system to allow respondents to add multiple recalls throughout the day using the multiple-pass protocol and portion size estimation methods of the 24-hour recall (progressive recall). This chapter conducted a dietary assessment survey with 49 participants, where they were asked to record dietary intake using both 24-hour and progressive recall methods for weekdays only. To examine accuracy, this work compares mean energy estimates and the mean number of reported foods. 24 of these participants were interviewed to examine the acceptability of the progressive recall.

Results: This research found that the mean number of foods reported for evening meals for progressive recalls (5.2 foods) was significantly higher (P = .001) than that for 24-hour recalls (4.3 foods). The mean energy for evening meals reported using progressive recalls (745.7 kcal) was also significantly higher (P = .02) than that for 24-hour recalls (726.4 kcal). The number of foods and the amount of energy reported for other meals remained similar across the two methods. In interviews, 63% of respondents indicated that they remembered meal contents and

portion sizes better with the progressive recall. However, 67% participants said that the 24-hour recall is more convenient in terms of fitting in with their daily lifestyles.

Conclusions: The analysis of interviews and data from this research study indicates that progressive recalls provide minor improvements to the accuracy of dietary assessment in Intake24. Additional work is needed to improve the acceptability of progressive recalls in this system.

The work described in this chapter has been undertaken during RA2 and informs the answer to RQ1. The content of this chapter has been rewritten and is accepted for publication in a peer-reviewed Journal of Medical Internet Research: Osadchiy, T., Poliakov, I., Olivier, P., Foster, E., and Rowland, M. Progressive 24-hour recall: Feasibility and acceptability of short retention intervals in online dietary assessment surveys. *Journal of Medical Internet Research*.

6.2 Introduction

6.2.1 Background

There are a number of different methods for assessing dietary intake of a population by either measuring markers of nutrient intake (e.g. DLW for measuring energy expenditure) or by surveying the intake of foods and drinks [27, 147]. A successful method is expected not only to be cost-effective, scalable and to estimate dietary intake with acceptable accuracy, but also to impose a low subject burden to reduce the likelihood of changes in respondents' diets, misreporting and participant dropouts [100]. One of most widely adopted approaches is the multiple pass 24-hour recall, which is considered to offer a favourable balance of those characteristics [77]. However, in a validation with adults 20-60 years old, Lopes et al. found the interviewer-led multiple-pass 24-hour recall method to underestimate habitual energy intake by 33% compared to energy expenditure measured using the gold-standard method, DLW [97]. The estimation error may, in part, be associated with recall bias since the accuracy of the 24-hour recall method relies on respondents being able to retain details about intake for a relatively long period of time [100, 159].

According to Macdiarmid and Blundell recalling intake even for the previous day is a challenging task for some individuals [100]. Dietary assessment is especially difficult with certain population groups, for example, with people with reduced cognitive and memory abilities (e.g. fading memory, reduced attention span) [83]. Human memory introduces such errors as unintentional food omissions which can contribute significantly to under-reporting of dietary intake. Memory errors may also reduce the accuracy of a method used for portion-size self-estimation, for example, photographs of various food serving sizes presented to respondents [161, 154, 30]. The serving size that a respondent remembers that they ate, the portion size consumed in reality and the portion size presented in the photograph may be different [66, 116, 117]. Additionally, misreporting may occur when respondents are asked about specific details of recipes used for cooking of the reported foods. Especially if the meal was not cooked by the respondent, they can easily misreport its ingredients [66].

Memories of eating and drinking start deteriorating even an hour after a meal [84, 24]. Indeed, research by Baxter et al. indicates that shortening the retention interval may increase the accuracy of a dietary intake recall among school age children [22, 23]. In two studies children were observed eating two school-provided meals, and interviewed to obtain a 24-hour recall. In the first study children were interviewed using one of six interview conditions achieved by crossing two target periods (prior-24-hours; previous-day) with three interview times (morning; afternoon; evening) [22]. In the second study the interviews were conducted either the same day in the afternoon (shorter retention interval) or in the morning for the previous day (longer retention interval) [23]. In both cases the correspondence rates for the observed/reported energy and the number of reported food items was higher when interviews were conducted after a shorter period of time. At the same time, the advantages of a reduced stress on human memory can be observed in other dietary assessment methods. For example, although it has the potential disadvantage of bias in intake reports and even changing respondents' diets due to the burden of recording [132] the WFD method where respondents are asked to record their foods as the day progresses is less prone to memory errors [159]. In contrast to that, the involvement of an interviewer in the 24-hour recall makes recording the intake progressively highly impractical.

The emergence of dietary assessment systems that automate the multiple-pass 24-hour recall method offers a multitude of benefits including cost efficiency and scalability [154, 148, 158, 39]. However, online dietary assessment surveys mostly implement the multiple-pass 24-hour recall interviewer-led procedure. With some of its methodological elements these systems inherit its limitations including errors related to human memory [30, 147]. Specifically, the 24-hour interval has been likely imposed by practical constraints of the interviewer-led approach. This time frame is short enough for the respondent to retain their eating memories and allows the interviewer to collect that information on a single occasion. Meanwhile, the self-administered manner of online surveys allows shorter retention intervals that could potentially improve the accuracy of dietary assessment. This chapter proposes a *progressive* recall method, where a respondent is asked to record multiple recalls of meals throughout the day instead of reporting all meals for a day in one occasion. This research hypothesizes that respondents would need to remember significantly less information over shorter periods of time, which reduces the burden on their memory and potentially increases the accuracy of dietary assessment. This chapter provides an overview of the implementation of the multiple-pass 24-hour recall in Intake24 and modifications added to that system to enable the new progressive recall method. The chapter then describes the design and reports the results of a dietary survey conducted in those two settings to support the hypothesis. This study also examines the effects of the modification on the usability and acceptability of the progressive recall in online surveys by interviewing survey respondents.

6.2.2 Multiple-pass 24-hour recall in Intake24

Intake24 is designed to administer large-scale dietary surveys. The system automates a multiple-pass 24-hour recall method [30]. Typically, respondents of these surveys perform a recall in the morning on three or four non-consecutive days to capture a wider variety of foods eaten. To avoid changes in diets subjects are ideally not aware about specific dates of surveys. Respondents are asked to answer a series of questions about meals they had for a previous day via a web-based interface. For every meal they are asked to provide its name and time, list of foods and drinks.

Respondents select the name of a meal from a list of suggestions (breakfast, lunch, evening meal, early / afternoon / late snack or drink) or they can type a new name for the meal. Contents of a meal are searched and selected from a food taxonomy, organized in a tree-based, multi-level structure. As the method of portion size estimation, Intake24 uses validated photographs of weighed servings. For every reported food and drink respondents are asked to select a photograph that most closely resembles the serving size they had (Fig. 6.1). A single submission typically includes four to seven eating occasions (e.g. breakfast, morning snack, lunch etc.). At the end of a study Intake24 produces a report that contains an estimated portion size, energy and nutrient intake for each reported food and drink. Energy and intakes of macro- and micro-nutrients are calculated using the national food composition table from a region, where the population was surveyed.

Fig. 6.1 Food serving size estimation with photographs used in Intake24.

6.2.3 Progressive recall

To explore the potential of improving the accuracy of dietary assessment results produced by Intake24 further by reducing the retention interval (i.e. time between an intake and a recall) this research implemented a modified version of the system that allows recording intake as the day progresses. While using the same multiple-pass procedure and portion size estimation methods of the multiple-pass 24-hour recall described in the previous section, progressive recall asks respondents to make three submissions on the day of a survey and one submission the next morning. In the first three submissions, subjects report morning, afternoon and evening meals. On the next morning, they report late meals or snacks for the previous day. The system alerts the respondent, if they select time of a meal that is later than the current time on their machine and does not allow submission of that meal (Fig. 6.2). Contrary to the weighed food diary method, the progressive recall does not require recording intake at the time of consumption and uses food photographs for portions size estimation instead of weighing foods and drinks using scales.

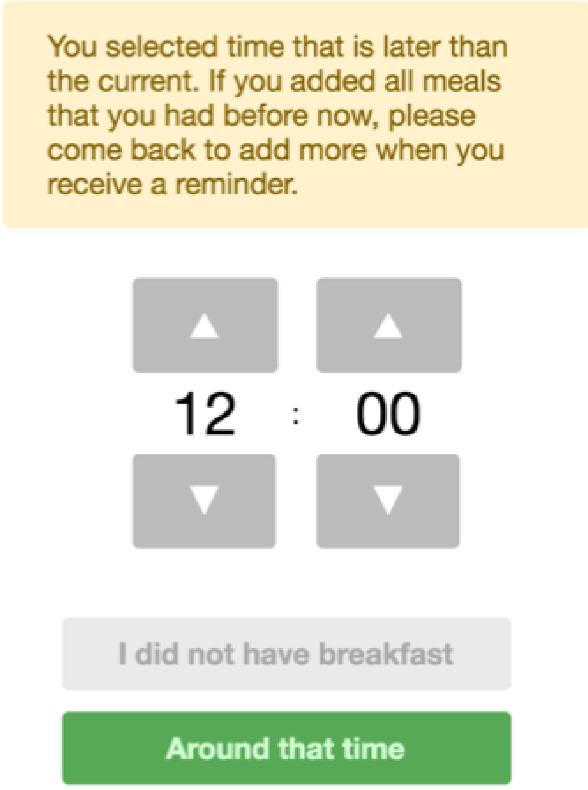


Fig. 6.2 Warning message in Intake24 when a user tries to log meals before the actual intake.

6.3 Methods

6.3.1 Recruitment

The progressive recall method was analysed in a study with 49 participants ($n = 24$ females, $n = 26$ males, 18-64 years) that was conducted during RA2 that is described in chapter 3. To investigate the effectiveness of using the progressive recall in automated dietary assessment systems as part of this study the new method was compared to the 24-hour recall. The overview of this study that is related to the current research is provided in this chapter as well. As part of the recruitment process respondents were asked about their preferred times for recording their diet to set up automated reminders for them. For the 24-hour recall candidates were asked to specify preferred morning times (before 10am) for recording their meals for a previous day. For the progressive type of recall participants were asked to provide three time points to record meals for the same day. The first time point, before 12pm, for recording breakfast, morning snacks and drinks; the second, between 12pm and 4pm, for lunch, afternoon snacks and drinks; the third, between 4pm and 12am, for dinner, evening snacks and drinks.

6.3.2 Procedure

The study was conducted over a six-week period with a separate group of participants in every two weeks. Each group of participants was asked to complete their recalls on two consecutive weeks. During each week participants were asked to login to Intake24 and complete three dietary recalls on three consecutive days between Monday and Friday. Participants were surveyed using the 24-hour recall during one week and the progressive method during another. For $n = 34$ participants the 24-hour recall was used on the first week and the progressive recall was used on the following week. With the remaining $n = 15$ participants the two methods were applied in the reverse order. The first of each type of recall was used to minimize the learning effect by familiarizing participants with the interface of the system and the procedure. For that reason, the first day of each type of recall was excluded from analysis, leaving four days of recalls from every individual. Participants were asked to avoid changes in their diets and not to record their meals elsewhere (e.g. notepads) to aid their recalls. Respondents were informed about the schedule of their recalls two days before the first one, which could affect their diets. However, the procedure assumes that this factor affects the accuracy of both types of recall. Thus, if there is a difference in the accuracy of the methods, it still can be observed.

Every morning, when participants were surveyed using the 24-hour recall method, they received automated reminders by email and SMS to submit their intake for the previous day. On the days of progressive recalls, participants received three reminders to add meals into the system as the day progressed. The next morning, they received a reminder to record late snacks and drinks, if they had any, and complete their submission. Reminders were circulated at the preferred times specified by respondents during registration for the study.

6.3.3 Statistical analysis

The analysis compared the mean number of foods and energy reported for a single day and for individual meals reported using the progressive recall to those reported with the 24-hour recall. Food items that can be reported by respondents include drinks and condiments (e.g. water, pear juice, ketchup, sour cream in soup). Each food item can be reported more than once for a single day and for a single eating occasion (meal). A study with 83 adults between 20 and 60 years of age shows that the interviewer-led multiple-pass 24-hour recall underestimates energy intake on average by 33% compared to that measured with DLW [97]. Meanwhile, validations of Intake24 estimated participants under-reporting energy intake compared to the interviewer-led method on average by 1% [30]. For those reasons, the analysis assumes that an increase in the number of reported foods and energy is a likely indication of an increase in accuracy of the method. The significance of difference between the means is analysed using Mann-Whitney U test, which is selected since it does not assume data being normally distributed [103, 114]. The analysis provides histogram and probability density plots to visualize that difference.

6.3.4 User interviews

To analyse usability and acceptability of the progressive recall a subset of participants was asked to share their experience of the two methods in an interview after their last recall. This resulted in interviews with N=24 participants [P1-24] (19 men and 5 women) aged from 18 to 44. The interview asked respondents which type of recall, if any, was more convenient for them; and which type of recall, if any, helped them to remember foods better. Respondents were asked to elaborate on these two topics. The interviews were audio recorded and transcribed. The transcripts were thematically analysed and this chapter discusses the topics that emerged during the analysis [31].

6.4 Results

6.4.1 Statistical analysis

The study resulted in 96 and 98 submissions surveyed with 24-hour and progressive recall methods respectively. The mean number of foods recorded for a single day was not significantly different for the two methods (P = .11). In the 24-hour and progressive recall methods respondents on average reported 15.9 and 16.8 foods respectively per a single submission. The mean energy reported with the two methods also remained similar (P = .47) with 1964.2 kcal and 1897.1 kcal for the 24-hour and progressive types of recall correspondingly. The same trend remained across all individual meals except for the evening meal (Table 6.1). The mean number of foods reported for evening meals during progressive recalls (5.2 foods) was significantly higher (P = .001) than during 24-hour recalls (4.3 foods). The same stands for energy (P = .02). The mean energy measured for progressive recalls was 745.4 kcal, whereas for the 24-hour recall the mean energy was 726.4 kcal. The distribution of submission sizes (number of foods per submission) and of reported energy for evening meals are presented in Fig. 6.3 and Fig. 6.4.

Meal	Number of foods			Energy		
	24-hour	Progressive	P-value	24-hour	Progressive	P-value
Evening meal	4.3	5.2	.001	726.4	745.4	.02
Lunch	4.0	4.1	.21	553.5	511.1	.29
Breakfast	3.6	3.7	.29	355.7	307.0	.45
Afternoon snack or drink	2.5	2.8	.12	250.3	257.0	.45
Early snack or drink	2.5	2.2	.19	148.6	146.3	.47
Late snack or drink	2.6	2.4	.18	291.2	288.2	.40
Full day	15.9	16.8	.11	1964.2	1897.1	.47

Table 6.1 Mean sizes and energy contents of meals reported with conventional and progressive recall methods.

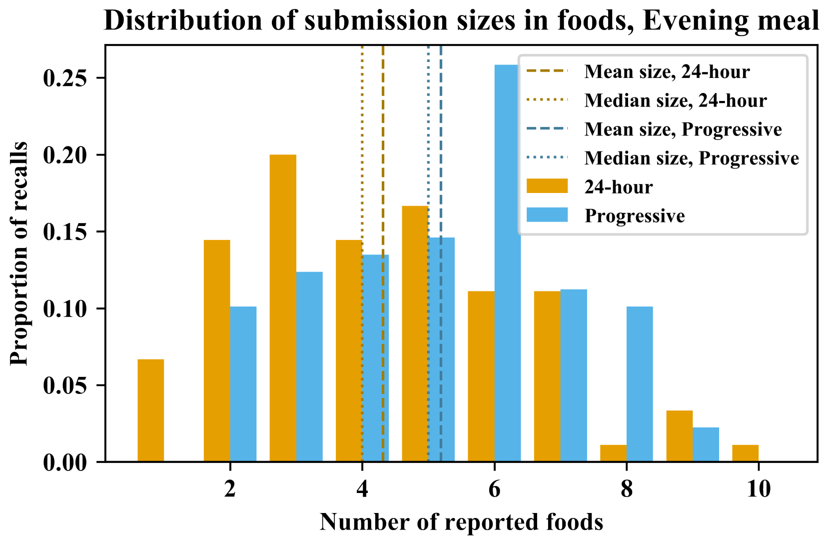


Fig. 6.3 Distribution of submission sizes for evening meals measured in foods.

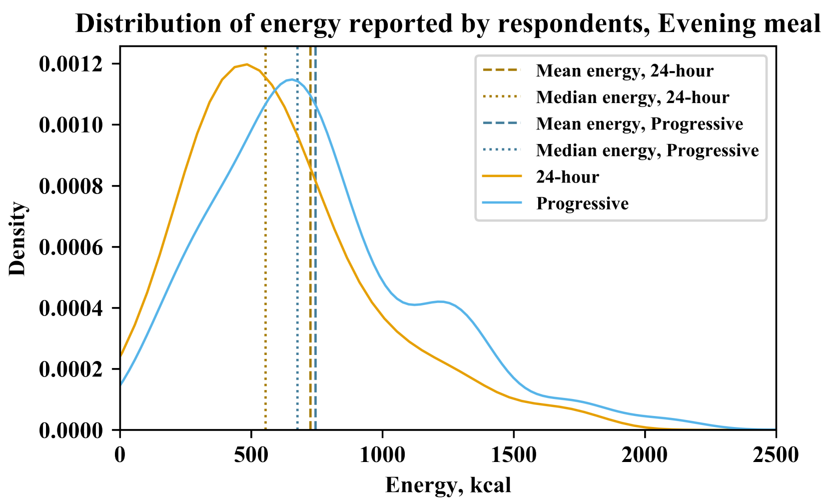


Fig. 6.4 Distribution of energy reported by respondents for evening meals.

As can be seen from the Table 6.2 evening meals had the largest number of distinct foods reported over the study by a single respondent (i.e. mean variety). At the same time evening meals had the largest mean number of reported foods per a single submission. In other words, evening meals were the largest in size but foods in those meals were the least repetitive.

6.4.2 User interviews

In the interviews, exploring participants' experiences of the two different types of recall,16 participants (67%) stated that they preferred the 24-hour recall method, 7 (29%) preferred the progressive method and one participant (4%) remained neutral. Two topics that were picked by thematic analysis were *"the challenge of fitting dietary recall into daily plans"* and *"details about intake that were easier to remember during the progressive recall"*.

Meal	Mean variety	Mean size
Evening meal	14.3	4.8
Lunch	12.5	4.1
Breakfast	7.0	3.7
Afternoon snack or drink	6.1	2.7
Early snack or drink	4.5	2.3
Late snack or drink	4.3	2.5

Table 6.2 Mean varieties and sizes of meals reported during the study.

The challenge of fitting dietary recall into daily plans. The major advantage of the 24-hour recall described by respondents was them being able to record meals on a single occasion without, as one participant said, *"changing my life routine too much"* [P2]. Despite notifications being sent at the preferred times specified by respondents, these often did not fit into their daily plans as, for example, for participant [P10]: *"If I'm really busy in a day and I've not really had a break between breakfast and lunch, I won't necessarily get a chance to record what I had for breakfast until like 2 o'clock"* She then added that being able to change previously defined notification preferences would help addressing that issue: *"I think you should give an option for changing the times of the prompts... I set down time for my breakfast and then I realized that the prompt that I was getting was actually when I was travelling to work"*. Three respondents [P2, 9, 14] stated that doing their recalls in the evenings was especially difficult for them. For example, respondent [P14] said: *"I find it really difficult to do any work at night... Usually you have food, you have dessert, then you're in relaxation mode. So, to bring yourself to do work is really difficult at like 10:00-10:30 p.m. You're getting ready for bed... So the last thing you want to do is do a study form."* Nonetheless, another three respondents [P6, 12, 19] suggested replacing the morning recall with an evening recall after the last meal in the 24-hour method.

Details about intake that were easier to remember during the progressive recall. In this study, interviews indicated that some information about dietary intake was easier to remember for respondents, when they recorded their intake using progressive recall. 15 out of all interviewed participants (63%) suggested that the progressive recall helped them to better remember the foods and drinks they had consumed. Participant [P1], who said that 24-hour recall fitted better into his lifestyle, experienced the following issue with this recall method: *"I think I must have eaten something cause I didn't have lunch until like two o'clock. But I don't really remember. I was actually guessing today. I was guessing about yesterday."* Respondents who expressed in favor of the progressive method said that short retention intervals assisted them recalling more details about their meals. For example, participant [P18] noticed that she remembered serving sizes better during the progressive type of recall: *"I think the portion size in general was hard especially with foods like where there were multiple components and they were all mixed together. So, how do you remember exactly how much something was? So, I think I was more accurate when I did it after every meal."* While explaining that respondent [P17] pointed out that memorizing foods is not a casual task and for that reason recording his diet as the day progresses worked better for him: *"The previous day was a bit of a task because I couldn't remember the small details and I relied more on the Intake24 to actually remind me like butter and bread. . . The small thing I would forget. Looking back for the previous day there was a lot of information that I tried to hold considering it's not something that you normally commit to memory. However, I've really actually enjoyed this week just going through it [diet] as the day progresses."*

Some respondents stated that short retention intervals were helpful in recalling irregular eating patterns. For example, this is how [P14] compared the two types of recall: *"The second one [24-hour] obviously relies on a lot more memory, which is difficult, especially when you had days when you've eaten out and you had a few different types of snacks. . . The days, I had consistent meals, my regular lunch and dinner, it was really easy next day because I have three coffees and . . . the same soup, but then . . . I ate a Lebanese food one evening and I had food outside during the afternoon as well and the next day I was like, 'Ah, so many different ingredients to remember!' "* This experience is supported by another respondent [P12]: *"One day when the school had put on like a buffet, and I had some things from the buffet, and the next morning I couldn't remember exactly what I had. So, yeah, I think it's definitely easier to remember in the moment"*

6.5 Discussion

6.5.1 Principal findings

More than half of the respondents in this study preferred the 24-hour recall method for the previous day since it was easier to integrate into their daily routine. At the same time, the interviews indicated that in many cases respondents did not have time to complete a recall when they received a reminder. The reminders were customized by the administrators at the beginning of the study to fit a normal eating pattern of each respondent. However, the actual timing of eating events for some respondents was different during the study. For other respondents notifications did not account for their plans for those days and distracted them. These factors could cause negative reaction to the progressive recall captured in the interviews. Thus, giving respondents the ability to change their notification preferences in the survey interface of Intake24, for example, postpone the received reminders, could potentially improve the acceptability of the progressive recall method. Another potential option is to give respondents the ability to decide the number of recalls they want to make during the day. That could help to identify a comfortable number of recalls that helps memory of respondents without being intrusive.

Future research could potentially find improvements to the acceptability of progressive recalls in Intake24 and similar dietary assessment systems by examining user experience implemented in popular mobile applications for personal dietary assessment (e.g. MyFitnessPal, Lose It!) [34]. Such applications allow respondents recording their intake progressively. An audience of millions of users voluntarily tracking their diet on a daily basis demonstrates a certain level of acceptability of the progressive method used in these dietary applications. At the same time, recording intake in a mobile dietary application is comparable in terms of tasks and difficulty to that in a dietary survey. Thus, user experience of mobile dietary applications could be used as a source of inspiration for addressing acceptability issues identified in this research.

The statistical analysis of data collected in this study shows that retention intervals for meals reported during progressive recalls are significantly shorter compared to those for meals reported during 24-hour recalls. A significant difference in the number of foods reported with the two methods was observed only for evening meals, where respondents reported more foods during progressive recalls. The size and energy contents of other meals, and of the day's intake overall, remained comparable to that reported during 24-hour recalls. A larger variety of foods in evening meals that was found in the course of the analysis could make this type of meal harder to recall the next morning but easy shortly after intake. Indeed, irregular eating patterns were suggested to be difficult to remember by some participants in our interviews. In contrast with our study design, 24-hour recall surveys often include longer time gaps between recall days and a mixture of week and weekend days, aiming to capture more variety in individual dietary intake [30]. Such variety is likely to increase the burden on human memory and it is possible that the advantages of the progressive recall for other meals and snacks would be more salient in studies conducted over longer periods of time. That is supported by those participants in our study who suggested that shorter retention intervals helped them to remember more details about their intake such as portion sizes.

6.5.2 Limitations

The current study involved a relatively small number of participants. The recruitment method meant that the demographics of our respondents were limited, which may mean that the results do not generalize to a wider population. Due to the study design the analysis compared one day's intake against intake from another day and therefore it is impossible to determine whether the observed difference is due to the method or to day-to-day variation in intake. The benefits of shorter retention intervals have already been demonstrated in interviewer-led studies with school age children [22, 23]. The progressive recall has the potential to address the issue of deteriorating memory in dietary assessment in population groups with limited cognitive function [83]. Still, the use of technology in online dietary surveys with a wider variety of population groups could face digital literacy and accessibility issues [81]. Therefore, findings of this research cannot be transferred to all population groups and further studies with larger and more diverse samples are needed. For a more reliable judgement of the accuracy of energy intakes estimated with the progressive recall they could be compared against true intake measured by direct meal observation or using objective biomarkers of dietary intake.

6.6 Conclusions

This chapter aimed to address one of the key challenges in dietary assessment, which is unintentional under-reporting due to poor human memory [100]. Previous research has demonstrated that the burden on memory can be minimised by reducing the amount of information that needs to be remembered along with the period of time it needs to be retained [22, 23]. This chapter proposed a modified procedure of the multiple-pass 24-hour recall that is referred to as a progressive recall. The modified method instead of requiring respondents to report their intake for the prior 24 hours or a previous day on a single occasion offers recording meals progressively, shortly after intake.

The method was compared to the multiple-pass 24-hour recall that is also implemented in Intake24. This research did not find a significant difference in the numbers of foods or the amounts of energy reported during progressive and 24-hour recalls for a single day in Intake24. Progressive recalls were found to capture more foods and energy for evening meals. More than half of the interviewed respondents in our study found fitting multiple intake recalls into their daily lifestyles to be difficult and preferred the 24-hour recall method. To address concerns raised by respondents the chapter proposed methods for improving the acceptability of progressive recalls in Intake24 that could be investigated in future. At the same time, a similar number of respondents pointed out that they remembered their intake better with the progressive method.

Chapter 7

Tailored dietary feedback as an incentive in large-scale dietary assessment surveys

7.1 Abstract

Dietary assessment using online surveys faces a range of challenges including motivation of respondents to take part in such surveys. Despite relatively developed research into designing engaging experience in online surveys, little is known about its application to online dietary assessment surveys. This chapter explores the feasibility of using people's interest in personal informatics as the driver for engagement with such surveys. In collaboration with qualified nutritionists, a personal dietary dashboard was designed that presents dietary feedback to respondents after completion of a dietary assessment survey. User experience with two styles of feedback was explored in an interview study with 24 respondents. The analysis of the interviews informed the refined version of the dashboard, which was deployed in a dietary assessment survey with 1381 respondents. The analysis of behavioural data identified a correlation between respondents receiving dietary feedback and the completion rates of this survey. The work described in this chapter has been undertaken during RA2 and RA4. This chapter informs the answer to RQ3.

7.2 Introduction

Most online dietary assessment surveys automate interviewer-led methods, for example, food frequency questionnaires or 24-hour recalls [147]. As with interviewer-led methods respondents commonly omit foods and under-report their intake in online surveys, which reduces the accuracy of automated dietary assessment due to a number of factors including limitations of memory of respondents [30]. Dietary assessment systems are able to emulate the behaviour of an interviewer to a certain extent and assist respondents' recall, for example, by asking follow-up prompting questions where necessary [148, 122, 123]. However, there are human factors that are challenging to detect using these systems as they lack the intelligence of a trained interviewer. One such factor is social desirability bias, when respondents under-report foods that are perceived as unhealthy by the society [100]. Another factor is the motivation of respondents to provide accurate answers, since most variations of dietary assessment surveys require respondents to invest a considerable amount of time and effort, which results not only in misreporting, but also in participants dropping out [63, 101, 147]. Misreporting in such cases has to be identified at the stage of data quality assessment. This often involves manual analysis of individual dietary records with abnormally low or high energy content or short completion time of a recall. Moreover, misreporting can be highly problematic to identify. Respondents who fail to record all foods they consumed during a study period or underestimate their portion sizes are commonly confused with genuine low-energy intake reports caused by illness, diet or irregular eating patterns [127].

The motivation of participants is one of the key factors affecting the quality of data gathered from crowdsourcing tasks, including online surveys. Even if participants receive a monetary incentive there is little guarantee that they will provide reliable and honest answers. This especially affects the outcomes of long questionnaires, which are known to have lower completion rates [63]. For example, a 24-hour recall, which requires respondents to record every food and drink they had on the previous day and to select portion sizes of those foods, is a highly complex task which may mean participants skip items, leading to necessary details being missed [58, 147, 63, 101]. For these reasons, collection of high-quality data in dietary assessment surveys is a challenging task that often involves high financial costs from rewarding respondents [17, 53]. Research exploring participatory engagement with online surveys has shown that respondents may feel more motivated to provide reliable answers if they see a personal gain in accurate outcomes of a survey [96, 115, 130]. The wide adoption of physical activity trackers (e.g. Fitbit) and dietary mobile applications (e.g. MyFitnessPal) indicates a strong user interest in quantifying and understanding various aspects of life and well-being including their diet and nutrition [45, 62]. This interest in personal data could potentially be used as a driver for engagement with online dietary surveys. For example, information about the intake of respondents could be presented back to them in a form of feedback about their dietary intake at the end of a survey.

This chapter designs a personal dietary dashboard that presents feedback about the intake of nutrients and the quality of diet to respondents of online dietary assessment surveys. The aim of this work is to analyse the feasibility of appealing to people's potential interest in personal informatics, with an intention to motivate their participation in dietary assessment surveys, in addition to educating respondents and encouraging positive health behaviours. The design of the dashboard is inspired by the previous research in the area of interaction design in online surveys and persuasive technology. However, the scope of this work does not extend to the correlation between the information presented to participants and changes to their dietary behaviour. The design of dietary feedback is itself a complex and sensitive problem, as the appropriate choice and form of nutritional information is highly dependent on contextual factors ranging from the demographics of the user (e.g. age, ethnicity, nationality, literacy) to the context of the intervention; whether well-being (e.g. weight self-management) or even for clinical issues (e.g. diabetes, heart disease). For that reason, before deploying on a larger scale, the two styles of the dietary dashboard were presented to respondents taking part in a small-scale dietary survey. Interviews with respondents explore the acceptability of each style, characteristics of dietary feedback that facilitate usability and issues with presenting such information. The design of the dietary dashboard was then refined and deployed in a population dietary assessment survey, where a correlation between respondents receiving dietary feedback and completion rates of this survey was analysed. This paper reports on qualitative and quantitative findings from these two studies.

7.3 Related work

In a typical dietary assessment study that is conducted using Intake24, respondents would use the survey interface to record all foods and drinks they had consumed the previous day. For each meal/snack, respondents are asked to provide the time of intake and the list of foods in that meal with portion sizes. On average respondents spend 14 minutes to complete a record for a single day, which is a considerable commitment of time [138]. Validations of Intake24 show that respondents commonly omit some food items (e.g. drinks, vegetables) or underestimate their intake [148]. For that reason, this research aimed to motivate respondents to provide more reliable answers during their recalls.

Nakamura and Csikszentmihalyi suggest that individuals are more likely to be engaged in an activity, if they see a personal value in it [115]. In the context of an online dietary assessment survey, that could be dietary feedback about the nutrient intake of respondents. Such feedback could help respondent to make informed choices related to their diet. Research demonstrates the effectiveness of personalized feedback advice over generic information for changing health related behaviours including eating habits [52, 86, 99, 111, 124, 151, 156]. The quality of the feedback will directly depend on the accuracy of the information participants provide; this can motivate them to put more effort into recording every necessary detail about their intake. This hypothesis is supported by a growing number of people using personal informatics for collecting and analysing their well-being (including dietary) data [45, 62, 98]. For example, in case of mobile dietary applications (e.g. MyFitnessPal) people are motivated mainly by receiving quantified information about their intake that helps them to improve or maintain their diet. At the same time, it takes a comparable amount of time and effort to record meals in a mobile dietary application for the whole day as completing a 24-hour recall survey online.

There are examples of dietary assessment systems that already provide nutrient feedback for participants, for example, ASA24 and MyFood24 [39, 154]. At the time of writing this paper, dietary feedback in these systems went little beyond a simplified presentation of a subset of the full nutritional analysis that is presented to the administrator. For example, MyFood24 provides a single table containing estimated intake of energy, calories, protein, fat, saturated fat, carbohydrates, sugars, fibre and salt along with recommended intakes for an *"average sized woman doing an average amount of physical activity"* (Fig. 7.1). ASA24 in contrast provides more dietary metrics, both estimated and recommended, taking into account user's age and gender. The feedback includes daily calorie intake, intake of 28 nutrients and even splits the diet into five food groups - grains, fruits, dairy, protein and vegetables (Fig. 7.2). The feedback in both systems could potentially be made more personalized by, for example, taking into account physical parameters of an individual (i.e. weight, height) and the level of physical activity that defines their energy needs and the recommended intake for some nutrients (e.g. fat, non-milk extrinsic sugars, protein). Dietary feedback could also provide more information and educate respondents about healthy eating practices.

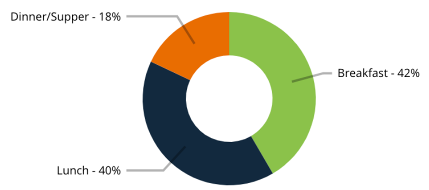
Your Nutrient Summary	Project Consumption	Adult Reference Intake*
Energy	9,366kj	8400KJ
Fibre (AOAC)	9g	24g
Calories	2,238kcal	2000kcal
Protein	98g	44g
Fat	129g	70g
of which saturates	46g	20g
Carbohydrates	181g	250g
of which sugars	35g	90g
Salt	5g	6g

The above nutrient summary information is provided for guidance only and is only as accurate as the food and drink information specified

* Adult reference intake values are based on an average sized woman, doing an average amount of physical activity

Fig. 7.1 Nutrient feedback in Myfood24

Caloric Intake By Meal



Day(s)
From: Nov 5, 2018 12:00:00 AM
To: Nov 5, 2018 11:59:59 PM

Daily Calories More Info

Allowance	2,600
Eaten	994

Daily Food Group Targets More Info

	Grains	Vegetables	Fruits	Dairy	Protein Foods
Target	9.0 ounces	3.5 cups	2.0 cups	3.0 cups	6.5 ounces
Eaten	3.7 ounces	0.0 cups	0.2 cups	0.8 cups	6.7 ounces
Status	Under	Under	Under	Under	Achieved

Fig. 7.2 Feedback on caloric intake by meal in ASA24

Lessons for developing an engaging experience in online dietary assessment can be drawn from research that concerns gamification of online surveys. In a study with teenagers and young adults (14 - 26 year old), Harms et al. demonstrated that the user experience of an online survey can be improved with minimal change and cost via a single gamification element. In this case, virtual badges were awarded for user achievements and no other incentive for participation was offered [71]. The study mainly revealed that the gamification of the survey was connected with positive qualitative feedback from the respondents. Meanwhile, the use of feedback did not affect the respondents' completion time or their responses. E. Law et al. explored the feeling of satisfied curiosity as a motivational driver to incentivise crowd workers [87]. In a study on Amazon's Mechanical Turk they successfully demonstrated that crowdworkers were more likely to complete tasks with higher accuracy when those tasks were designed with elements of puzzles [87]. The authors explained the success of this type of motivation through information gap theory [95]. According to this theory, individuals who are aware of a gap in their knowledge will aim to complete their knowledge and resolve the uncertainty. Similarly, the development of the personal dietary dashboard pursues the drive for respondent's curiosity about their diet to make the completion of an online dietary assessment survey more attractive.

In a number of projects, personal informatics has been complemented with elements of gamification to implement more effective persuasion; examples include "Playful bottle" [44], "Fish'n'steps" [90] and "OrderUP!"" [68]. Specifically, the "Playful bottle" [44] compared the effectiveness of interventions for improving water-drinking habits with and without gamification. The study indicated that gamification positively affected persuasion in this context. An interesting design choice shared by the authors of "Playful bottle" [44] and "Fish'n'steps" [90] is the visualization of healthy habits of participants via the well-being and emotional state of virtual characters (a tree or fish). Contrary to these projects and HCI research that focuses on behaviour change, RQ3 presents a unique challenge for the development of a tailored dietary feedback experience in online surveys for large-scale population dietary assessment. Such feedback needs to be delivered in a way that does not change the eating behaviour of respondents. Changes in diets of respondents pose risks to the reliability of the outcomes of a study, which aims to capture habitual intake of a population [157, 100].

7.4 Designing the dietary dashboard

7.4.1 General layout

One of the goals of this research was to preserve the workflow of the questionnaire implemented in Intake24, which is validated to demonstrate a comparable accuracy to that of an interviewer-led recall [30, 139, 56]. Thus, as a simple change, a personal dietary dashboard is added at the end of the questionnaire. The dashboard comprises three elements: a consent form to receive the dietary feedback, a form to fill in the physical characteristics necessary to calculate nutrient feedback, and the feedback itself. The dashboard was designed in collaboration with qualified nutritionists based in Newcastle University Human Nutrition Research Centre.

The feedback presented in the dashboard relies on a set of rules that were designed based on the Eatwell Guide and the Public Health England dietary recommendations [35, 1]. To estimate energy needs of respondents the Harris-Benedict equation was used [72]. The rules are implemented as numeric ranges representing low, slightly low, good, slightly high and high intakes for various nutrients, taking into account the individual's physical characteristics. Respondents' physical characteristics, which are voluntarily provided when they consent to receive the feedback, include gender, age, height, weight, weight target (to maintain, lose or gain weight) and self-assessed physical activity level. Nutrients for which the system displays feedback include energy, fat, saturated fat, non-milk extrinsic sugars, protein, Englyst fibre, carbohydrate, calcium, vitamins C and A. These nutrients were selected by nutritionists as fundamental in an individual's diet, and are those for which the system can calculate the most accurate feedback from a completed recall. The feedback assumes an individual without any specific diet related health conditions. The dashboard gives no medical advice to respondents, but rather informs them of directions for improvement. To explicitly indicate this to respondents, the dashboard includes a message at the top: *"The feedback is based on the food intake information you have given, the Eatwell Guide and UK dietary reference values and is suitable for healthy individuals. For dietary advice relating to a specific health condition (such as diabetes or hypertension) please see your GP."*

The general idea of the personal dietary dashboard is to present respondents with information about their diet in a form that is easy to understand for them. As validated by van Weert et al., using visual cues and lower language complexity in presenting digital content to patients enhances their satisfaction and interaction [164]. Taking lessons from this research, personal nutrient feedback was designed in the form of cards, each visualizing the measured intake of one of the nutrients (Fig. 7.3, 7.4). Every card contains a short line of text describing the range that fits the measured intake. The card also contains a colour-coded numeric value for the measured intake and a target range. Respondents can press a "Tell me more" button at the bottom of the card that opens a modal window with extracts of important information about that nutrient and its role in human nutrition. This information references links to relevant resources that the user can explore such as the NHS website. High consumption of saturated fat is related to a higher risk of coronary heart disease [18], while non-milk extrinsic sugars are linked with weight gain, dental caries and cardiovascular disease [102, 113, 168]. Participants given dietary feedback may be interested in making dietary changes based on the information provided to them, and therefore they are able to view the dashboard that lists the top five foods that contribute to their calorie, non-milk extrinsic sugars and saturated fat intake. (Fig. 7.5).

7.4.2 Two styles of the personal dietary dashboard

In the course of this work two contrasting styles for the dashboard were designed. The first style takes inspiration from the Fish'n'steps study with 19 participants (11 females and 8 males, aged from 23 to 63), which demonstrates the positive effect of appealing to respondents' care for and attachment to virtual creatures to improve their physical activity [90]. In this version of the personal dietary dashboard, each card contains a virtual character that represents a particular nutrient; for example, a strawberry for vitamin C intake or a loaf of bread for carbohydrate intake (Fig. 7.3). The characters demonstrate different sentiments (e.g. sad, disappointed or happy) with facial expressions and backgrounds depending on the range that fits the nutrient

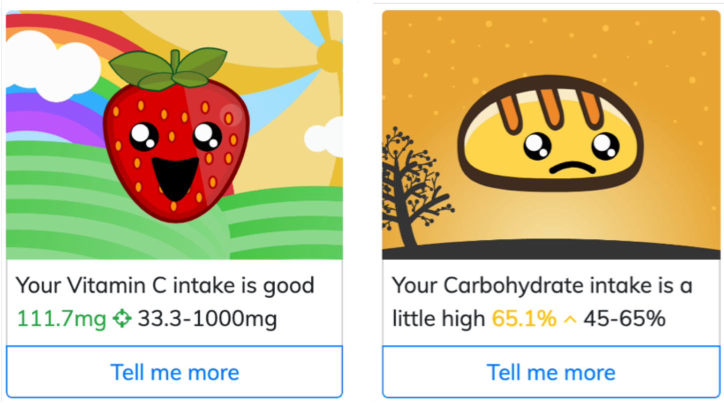


Fig. 7.3 Examples of nutrient cards for vitamin C and carbohydrate intake in the personal dietary dashboard with virtual characters.

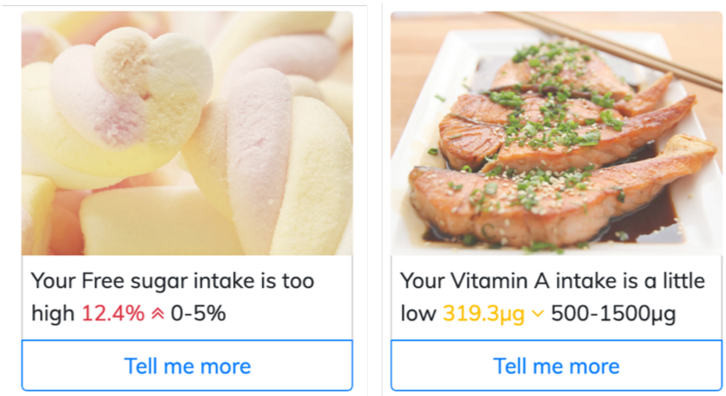


Fig. 7.4 Examples of nutrient cards for non-milk extrinsic sugar and vitamin A intake in the neutral style of the personal dietary dashboard.

intake (low, slightly low, high, slightly high or good, respectively). The motivation for this style of dashboard is to display a 'mood' or 'emotion' related to the diet of the respondent in an engaging, friendly and non-abusive manner. The design also aimed to associate the presented information with striking and unusual images to help respondents to remember the feedback [172]. However, there were concerns that some respondents might feel uncomfortable with this method of presenting information about such a sensitive topic as their diet. In the Fish'n'steps project, the authors observed that "punishing" participants with a sad facial expression of a virtual fish negatively affected their engagement [90]. For that reason, a 'neutral' style for the personal dietary dashboard was developed, in which instead of characters the cards contained generic pictures of foods that represent sources of each nutrient (Fig. 7.4). In this case, the image did not change depending on the intake of the respondent.

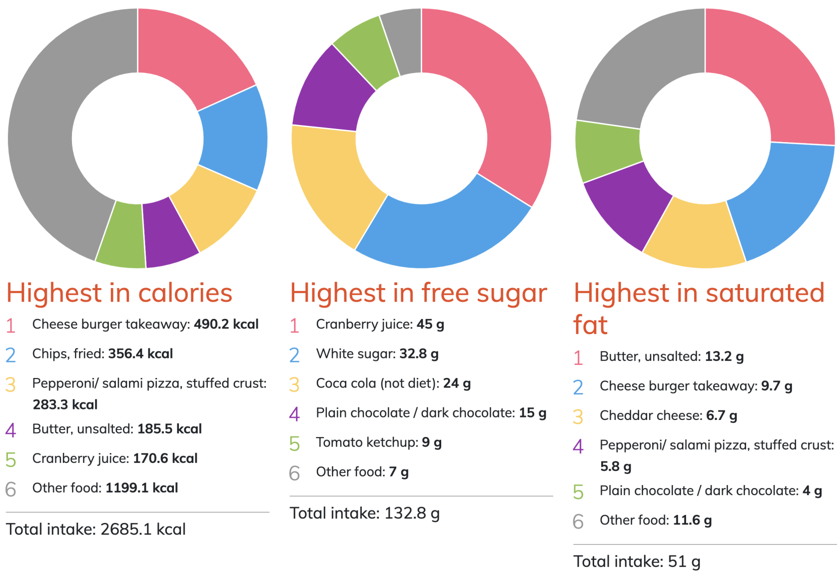


Fig. 7.5 Feedback on five foods highest in calories, non-milk extrinsic sugars and saturated fat in the personal dietary dashboard.

7.5 Study 1: Interview study

7.5.1 Procedure

To gain in-depth insights into the experience of respondents with two styles of the personal dietary dashboard (virtual characters and neutral pictures of foods), this research conducted an interview study with 24 (5 females and 19 males) participants (P1-24) aged from 18 to 44. Participants of this study were mostly students and staff of Newcastle University. Each style of the dashboard was displayed after completing one dietary recall in Intake24. For taking part in this study each respondent received a £30 Amazon voucher. After respondents completed a survey and received dietary feedback with both styles, they were invited for a semi-structured interview conducted in person or via means of remote video communication. Participants were asked to reflect on their experiences with the two styles of the dietary dashboard. This included questions such as: *"Did you learn anything new from the dietary dashboard?"*; *"Which style of the dashboard is more appealing to you and why?"*; *"What other information do you want to get from your dietary feedback?"*. With permission from respondents interviews were audio recorded and transcribed. Transcripts were coded according to the topics around feedback engagement characteristics that emerged during the analysis. This study received ethical approval from Newcastle University Research Office (Ref: 4971/2018).

7.5.2 Findings

This section explores characteristics of the dietary dashboard that improved user experience with personal dietary feedback in this study. These characteristics represent themes that were picked by thematically analysing interviews with participants of this study. Overall, 17 participants stated that the dashboard with virtual characters felt more appealing to them and 6 favoured the neutral style. One participant did not express a specific preference.

Exploratory. Important information should be easy to discover. For many respondents the card design of the dashboard made issues in their diets more vivid and easier to identify. For example, this is how one of the interviewees described their experience [P12]: *"Cards allowed me to skim really quickly. Like skim-skim-skim, ah OK, here as an error and I have a little read. It really helped to identify, like traffic lights, you know... "* Each of the presented nutrients was mentioned by at least one respondent. When participants were asked about information presented in the dashboard they had not been aware of, the most mentioned nutrient was sugar and its sources. Even though this element was positioned fourth in the dashboard, ten participants mostly drew their attention to it. The next most noticed nutrients, energy and vitamin A, were mentioned five times, half as often as sugar. However, respondents mainly paid attention to nutrient cards that flagged issues with their intake. For example, the participant [P3] said: *"There is something that stood out, which I think was my Vitamin A intake, which I never even considered before and I read up on it last night. It was dramatically high ... "*.

The most revealing items for participants, which they perceived as the most valuable in the dashboard, were the sources of nutrients that were flagged as deficient or excessive. For some respondents, the feedback completely changed their understanding of their favourite foods. For example, here is what the respondents [P14] learned about their sugar intake: [P14] *"I knew how much sugars were there in breakfast bars and smoothies but I didn't realize how quickly a few of those per day add up. It was a shock for me."* Dietary feedback helped respondents to learn about previously unknown sources of some of their nutrients. For example, the respondent [P8] was surprised to discover that jam is a source of sugar intake: *"I have jam in my porridge and that comes consistently as highest in sugar. I was like ah but that's fruit, right?"* Similarly, the interviewee [P9] learned about their source of fat intake: *"I wasn't aware that I eat so much of goat cheese... It consistently came as my number one in fat!"*

Meanwhile, it was not always negative discoveries that caught respondents' attention. A few respondents had negative expectations about some of the nutrients in their diet but became intrigued when the feedback highlighted the opposite as, for example, for [P15]: *"Surprised I was getting really good amount of calcium every day just from having milk in tea and coffee."*

Memorable. Dietary feedback should be easy to remember. When comparing the two styles of the personal dietary dashboard, five interviewees had issues with recalling information displayed with the neutral style of the dashboard. For example, respondent [P7] points out: *"I actually forgot that there was another feedback cause all I can remember is the cartoon characters (laughs), so they were definitely more like salient . . . More in your face . . . Definitely remember them more . . . "* The respondent [P3] described a similar experience: *"I can distinctly remember all the little emojis and I can actually remember generally the whole recall, I can remember roughly what the numbers were underneath as well, the pictures were more vivid and I think that's more useful".*

Personalized. Dietary feedback should account for personal dietary goals. The two styles of the dashboard differed only in types of images used in nutrient cards (virtual characters and neutral images of foods). The way information about nutrients was presented in both cases remained untouched. Nevertheless, some interviewees expressed the feeling of a tailored experience that virtual characters gave to them. This, for example, was pointed out by the respondent [P14]: *"Little characters made it feel more tailored . . . "*, and the respondent [P8] in a separate interview: *"Cartoon [dashboard] definitely more memorable. Definitely more personal."* The respondent [P5], while discussing tailored experience with virtual characters, described the following: *"It was friendly and approachable and it gave quite a lot of nice information about like how I could adjust my diet."*

However, in other cases the dietary dashboard with virtual characters caused negative emotions by presenting information that contradicted personal dietary goals. Some respondents paid attention to that information only when it was presented with virtual characters. This is what the respondent [P22] said: *"There is some literature saying that the saturated fat depending on the quality of it isn't that bad . . . basically that is something I dispute being a bad thing . . . it shouldn't be that negative . . . "* In the discussion of the personal dietary dashboard style with virtual characters, the respondent [P5] gave a good overview of the issue: *"I think that feedback has the danger of being normative; and what works for some people doesn't necessarily works for another person . . . Even in that jovial way . . . and even perhaps more so in that jovial way we risk making people feel bad if they have a different diet."*

Credible. The style of presenting dietary feedback should demonstrate credibility. Six participants who expressed a preference for the neutral style of the personal dietary dashboard considered its main advantage to be the impression of professionalism and appropriateness, matching the seriousness of the topic. Two of the respondents, for instance, stated the following about their experience: [P9] *"I really enjoyed reading through it, looking at it, I thought it was just professional, you know, just something I would expect from food tracking website . . . "*; [P19] *"The tone felt better or more appropriate. Matched the degree of seriousness of the study . . . "*

Active. Feedback should suggest ways to improve the diet. When the respondents were asked what would they add into the dashboard, one shared view was that dietary feedback should not only provide an overview of their diet but also recommend necessary changes to make it healthier. For example, it should recommend alternatives to foods they reported that were high in sugar as, for example, was requested by [P8]: *"Maybe suggestions of alternatives, so instead of* *jam, you know which is high in sugar what realistically would be better . . . Have you considered nuts instead?"*

Some of the respondents wanted their feedback to support specific athletic targets. For example, the respondent [P4] pointed out: *"One thing that would be useful knowing is what I need for building strength . . . I know many people start training for marathons now but they don't know what diet they need . . . "*

7.6 Study 2: Dietary assessment survey

7.6.1 Design refinement

A number of respondents in the interview study questioned the appropriateness of delivering dietary feedback with emotional expressions of virtual characters. For that reason, the style of the personal dietary dashboard with neutral pictures of foods was selected for the use in a larger dietary assessment survey. In the interview study, many respondents expressed that they learned new information about their diet from nutrient cards in the dietary dashboard. To make the dashboard more informative it was extended it with cards that provide feedback on the consumption of red meat, as well as feedback on the number of portions of fruit and vegetables in relation to the five a day guidelines [35, 41]. To collect reflections on user experience from a large cohort a comment widget was added at the bottom of the dashboard. The widget allows respondents reacting to their experience with Intake24 anonymously by pressing a "like" or "dislike" button and optionally leaving a comment (Fig. 7.6).

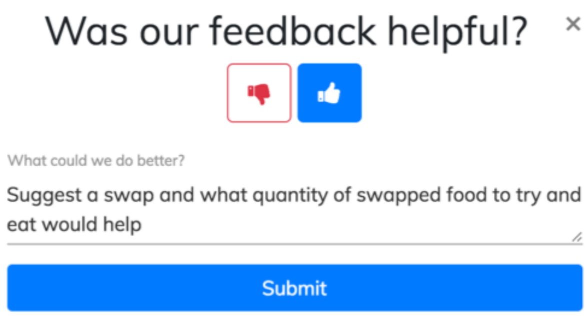


Fig. 7.6 User experience comment widget in personal dietary dashboard.

7.6.2 Procedure

The refined version of the dietary dashboard was deployed in a dietary assessment survey conducted in the UK. Respondents of this survey were paid £20 each and were offered to receive dietary feedback after recording two days of their diet. Participants could also choose to reject dietary feedback in consent form displayed after completing the survey. This survey was completed by 1381 respondents (760 females, 621 males) 12-85 years old. For respondents aged under 16 years, both parental and child consent was collected. In this survey, respondents were asked to submit at least two recalls. This research was performed under ethical approval from Newcastle University Research Office (Ref: 1377/2017).

As with the interview transcripts, anonymous reactions with comments received via the comment widget were coded according to the topics that emerged. The paper also reports the proportion of recalls that provoked positive or negative reactions (like or dislike button presses). In analysing the results of this survey, the primary focus was on the proportion of respondents that opted in for receiving dietary feedback and the comparison of their submission rates to those of respondents that opted out from receiving the feedback. This assumes that if respondents are interested in dietary feedback, they may want to receive feedback on more of their dietary days and, thus, submit more recalls than they initially were asked to.

7.6.3 Findings

The section presents the analysis of the distribution of numbers of submitted recalls with and without presenting dietary feedback to respondents in a population dietary assessment survey.

The survey generated 2846 dietary submissions. It also resulted in 162 likes (91% of reactions) with 20 comments (13% of reactions) and 10 dislikes (9% of reactions) with 6 comments (7% of reactions). Most of the comments described positive and negative experiences with the user interface of Intake24 rather than with the dietary dashboard, for example: *"Took me a while to complete but I will get used to it. Would prefer to add today's food as I eat it so I don't miss anything."* A number of comments suggested that the nutrient feedback should support dietary goals of respondents. For example, one common request was to add recommendations about the ways to improve the diet of an individual, which correlates with the results of the interviews: *"Suggest a swap and what quantity of swapped food to try and eat would help."* As in the interviews, respondents also asked for recommendations of alternative foods; but a more interesting suggestion learned from these comments was the idea of recommending changes to portion sizes: *"It's absolutely brilliant - a real eye opener. Suggesting serving sizes at the conclusion of the data would be brilliant such as "instead of bowl A, try to use bowl C" or "instead of filling to line 5, try filling to line 3" etc."*

Another similarity with the interviews was found in the comments that requested dietary feedback to aid the sporting targets of respondents: *"Relate diet intake more specifically to exercise - need more than x3 categories, but otherwise very interesting."*

In this study, only 469 (34%) respondents accepted receiving dietary feedback. The group of respondents that rejected dietary feedback submitted on average 1.92 recalls, whereas the group of respondents that accepted feedback submitted 2.33 recalls (Fig. 7.7), which is significantly higher under a a Mann-Whitney U test ($P < 0.001$).

7.7 Discussion

Reflections from respondents in both studies indicate their interest in dietary feedback that they received from the personal dietary dashboard. In most of the interviews in the first study, respondents expressed that they discovered new knowledge about their diet. Thematic analysis of

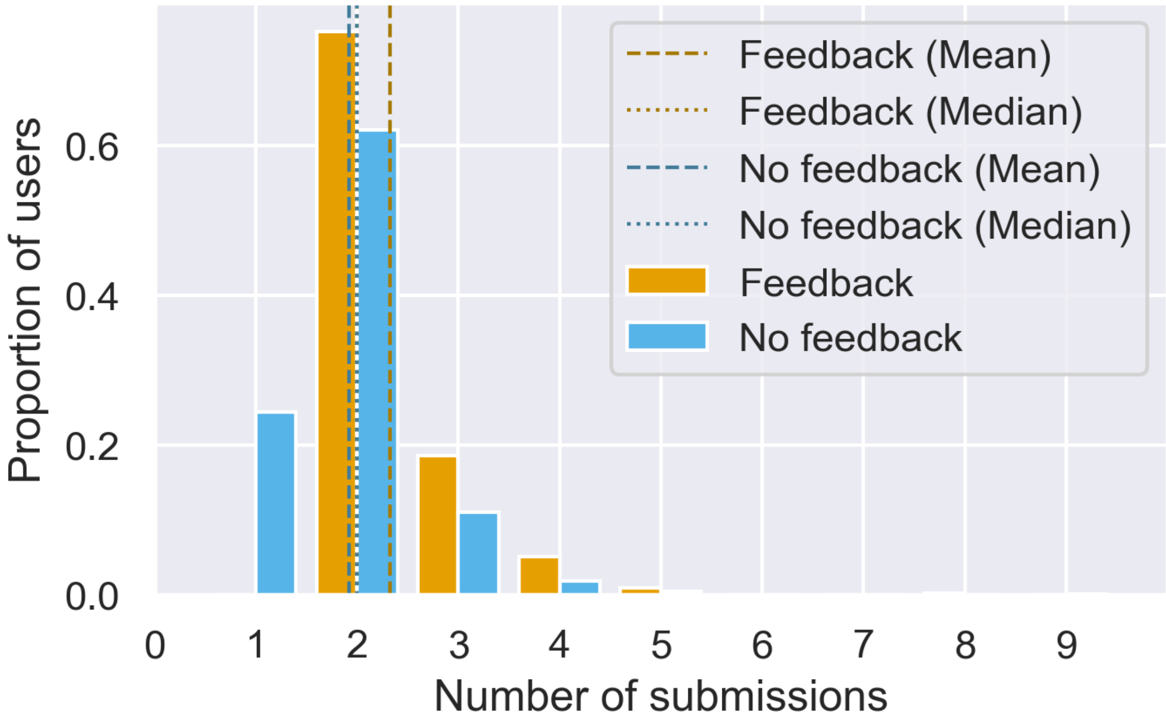


Fig. 7.7 The distribution of submission rates for the group of respondents that requested dietary feedback and for the group of respondents that rejected dietary feedback. Both medians are equal to 2 on this figure.

reflections also highlighted the following characteristics that facilitated positive user experience with the personal dietary dashboard.

Exploratory. Important information should be easy to discover. The card design that presented information in the form of an overview was found to help users to navigate it quickly and explore it gradually. In this case, nutrient cards with colour-coded nutrient values were used to flag positive or negative results with a short line of text describing it. Users could get a detailed explanation in cards that caught their attention by pressing a "Tell me more" button. The presented information helped them to identify points of action. For example, they were able not only to identify excessive intake of some nutrients (e.g. sugar), but also the sources of that excess.

Memorable. Dietary feedback should be easy to remember. Interviews indicated that the style of the personal dietary dashboard in some cases affected the ease of remembering information presented with it. Some respondents pointed out that the use of virtual characters in the dashboard helped them to memorize presented information better. However, the interview study described in this paper was not designed to compare the memorability of dietary feedback presented with the two styles and this could be an interesting question for future research. At the same time, this observation demonstrates the importance of dietary feedback being easy to remember.

Personalized. Dietary feedback should account for personal dietary goals. Feedback should be generated based on personal user characteristics and allow customization to their needs. For example, despite dietary feedback provided to respondents in the interview study was based on Eatwell Guide and UK dietary reference values [35, 1], some respondents reacted negatively, when it contradicted their personal dietary targets (e.g. sports, specific diets).

Credible. The style of presenting dietary feedback should demonstrate credibility. The style of the feedback should demonstrate the trustworthiness of the resource. The neutral style of the dashboard was found to represent this characteristic more effectively in the interview study. At the same time, most interviewees found virtual characters more engaging and memorable. Moreover, the interviews did not involve children who might had also found the style with virtual characters as more engaging. This conflict creates an opportunity to explore the balance between these two directions.

Active. Feedback should suggest ways to improve the diet. Respondents in both studies suggested that their dietary feedback could be more useful, if it provided guidance on specific changes to their intake that would improve their diet and maintain it according to a defined set of their personal goals. For example, as prompted by some respondents, it could recommend foods and their portion sizes to support a balanced diet or specific nutrient goals (e.g. sports). This characteristic should be considered in conjunction with personalization of dietary feedback. However, the development of active dietary feedback that is capable of suggesting meaningful changes to one's intake is a challenging and sensitive research task. To avoid harmful implications for human health active dietary feedback and personal dietary goals that such feedback is based on should be supervised by a qualified dietitian.

A considerably low number of respondents that requested dietary feedback in the dietary assessment survey indicates that the feedback was not the main driver of participation in this case. Nevertheless, respondents that did not request dietary feedback on average submitted less recalls per user than they were asked to, which indicates some level of attrition. In the meantime, respondents who requested the feedback submitted a significantly higher number of recalls. This may indicate an increased level of engagement of this group of respondents with the survey and that could potentially be associated with the presence of dietary feedback. It is worth noting that recalls collected after respondents received dietary feedback should not be used for the analysis in a conventional dietary assessment study, since there is a chance that respondents changed their diet after seeing it [157].

7.8 Limitations

The results of the research described in this paper are limited by the demographics profile of respondents that took part in the interviews. Results from the analysis of data collected in the dietary assessment survey cannot confirm whether offering respondent's dietary feedback led them to submitting more recalls or if the respondents who were more engaged with the survey were more likely to request dietary feedback. To support the findings in this work and to analyse the effect of presenting nutrient feedback to respondents on the accuracy of online dietary assessment surveys, a future study could compare the mean error rate between estimated energy intake in a survey with / without dietary feedback and true intake measured by direct meal observation or using objective biomarkers of dietary intake (e.g. DLW) [94].

7.9 Conclusion

This paper presents the design of a personal dietary dashboard for respondents of online dietary assessment surveys. The dashboard provides participants with feedback about their nutrient well-being derived from their answers. The aim of this work is to address the long-standing challenge of participatory engagement with online surveys in the context of dietary assessment by appealing to a potential interest of respondents in personal informatics. Reflections from participants in two studies highlighted characteristics that improved their user experience with dietary feedback. The analysis of submitted recalls shows that respondents who requested to receive the feedback on average recorded more days of their intake, which may indicate an increased level of engagement with a dietary survey. The results presented in this paper are based on the deployment of the personal dietary dashboard in Intake24 - an automated dietary assessment system that implements a multiple pass 24-hour recall method. Nevertheless, the described methods and findings could potentially be useful for other types of dietary assessment surveys (e.g. food frequency questionnaires).

Chapter 8

Discussion, Relevance, and Conclusion

8.1 Abstract

This chapter concludes this thesis by revisiting research question presented in chapter 1. This chapter then summarises key contributions to knowledge resulted from the research activities used in answering those questions. The chapter describes the limitations of this thesis and indicates how those limitations may be improved. Finally, directions and opportunities for the future research that emerge from the findings of this thesis are presented.

8.2 Answering research questions

The main argument of this thesis is that the next step for automated dietary assessment is to go beyond automation of interviewer-led protocols and address the remaining challenges by taking advantage of other fields of computing including data science and HCI. Currently automated dietary assessment systems mostly translate interviewer-led approaches into a digital form and inherit some methodological problems of original methods that lead to misreporting of dietary intake and lower the accuracy of assessment [147]. To support this argument the thesis selected three research directions that assemble three research questions defined in chapter 1. Namely, the thesis explores the new ways to assist memory of respondents in dietary assessment surveys using (1) data driven methods, (2) shorter retention intervals (i.e. time period between an intake and a recall), and (3) tailored dietary feedback as an incentive for participating in dietary assessment surveys. To build a theoretical basis for answering those questions chapter 2 provides the anatomy of automated dietary assessment systems. Following that the chapter discusses challenges for these systems inherited from interviewer-led methods and dictated by the self-administered nature of recalls. These include limitations of memory of respondents and social desirability bias. Chapter 2 also identifies challenges that are specific to the use of technology in dietary assessment studies such as the usability of online surveys, the motivation of respondents to take part in them and technical literacy of respondents. All of those challenges cause intentional and unintentional misreporting of respondents' intake (e.g. omissions, under- or over-reporting) and affect the accuracy of dietary assessment [100]. Chapters 4- 7 propose new methods for addressing these challenges and improving the accuracy of dietary assessment. These methods were implemented and evaluated in Intake24, a system for conducting large scale population dietary assessment surveys based on the multiple pass 24-hour recall method [30]. Chapter 3 reviews activities performed in the course of answering the research questions and describes the development of new modules for the system that supported the collection of data and the analysis of results for the research activities. The remainder of the thesis provides answers to the defined research questions.

8.2.1 Research question 1 (RQ1). How can recall assistance be improved and evaluated in automated dietary assessment?

The answer to the first research question of this thesis is informed from the results of RA1-4 described in chapters 3-6. The research described in these chapters aimed to address unintentional under-reporting of intake by respondents during automated dietary assessment surveys. The thesis identifies two opportunities to assist memory of respondents in automated dietary assessment surveys. The first opportunity is to employ such data driven methods as recommender systems that are already being used in online entertainment and retail services (e.g. YouTube, Amazon) [46, 91]. Chapters 4 and 5 demonstrate a successful application of a novel recommender system for reminding respondents about potentially omitted foods.

Chapter 6 demonstrates another approach to assisting dietary recall by reducing burden on memory of respondents through shortening the interval between an intake and a recall (i.e. retention interval). The chapter argues that the 24-hour time frame in the original 24-hour multiple-pass recall is likely forced by practical constraints of the interviewer-led method. Whereas an automated dietary assessment system does not have to require respondents to recall foods they ate in the past 24-hours on a single occasion. Instead, as it was suggested in the previous validation of Intake24 [30] respondents could record their intake progressively shortly after it. The increased accuracy of a dietary recall with a shorter retention interval has been demonstrated by Baxter et al. among school age children [22, 23]. Thus, chapter 6 proposes changes to the original protocol of the 24-hour multiple-pass recall method implemented in Intake24 [69, 30] and offers respondents to make multiple records of their intake throughout the day. The modified procedure is referred to as a progressive recall. The chapter covers the implementation of the progressive recall in Intake24 and describes a study that compared the new method to the original 24-hour recall. In the study, respondents reported a larger number of foods and more energy for evening meals with the progressive method, which offers some support to the original hypothesis. In addition to that some of the respondents were interviewed to get a better understanding of acceptability and limitations of the progressive recall. More than a half of interviewed respondents suggested that it was easier for them to recall their intake with the progressive method. However, this method may have practical limitations since fitting multiple dietary recalls in a day was indicated to be challenging by a similar number of respondents.

The analysis of results from research activities that answered the three research questions of this thesis is based upon a usability framework developed in chapter 3. The framework provides explicit and implicit quantitative metrics for measuring the impact of changes in a previously validated dietary assessment system on its accuracy through its usability. For example, in chapter 5 the number of accepted food prompts by respondents was used to judge the effectiveness of those prompts; and in chapter 6 the duration of a survey was used as an indicator of changes in its complexity.

8.2.2 Research question 2 (RQ2). Can data driven methods facilitate the accuracy of dietary assessment?

The answer to RQ2 of this thesis is informed from the results of RA1-3 described in chapters 3-5. The goal of this research question was to demonstrate the application of data driven methods to recall assistance in dietary assessment systems. The specific challenge that was addressed with data science in that work was mining of associated food prompts (i.e. reminders). In a dietary assessment survey, when respondents select a food or a drink (e.g. coffee), the system prompts them with foods and drinks that are commonly consumed with the selected item (associated foods, e.g. sugar, milk). In Intake24, these prompts are hand-coded manually by researchers with nutritional background. Chapter 4 describes the development of a new type of a recommender system that builds a model of eating behaviour of a population from dietary recalls collected in the past surveys. Based on the foods selected by the respondent in a dietary assessment survey this model recommends foods that the respondent has potentially omitted. The chapter also demonstrates how the recommender system can improve user experience with searching and selecting foods in the system's database of foods during the survey. Chapter 5 describes the deployment of the algorithm in two real-life dietary surveys, where the algorithm was found to capture a considerably larger number of foods omitted by respondents without any observed effects on the usability of the survey interface when compared to hand-coded prompts. Thus, the analysis of results in these two chapters concludes that the recommender algorithm can facilitate the accuracy of dietary assessment, where there is a large dataset of past survey responses.

In the meantime, the work described in chapters 4 and 5 highlights opportunities for future research in the area of data driven dietary assessment. For example, the study in chapter 5 demonstrated that secondary foods (e.g. sauces, spreads) are more likely to be omitted than main foods (e.g. steak, chips). That is reflected in a low proportion of foods suggested by the recommender algorithm being accepted by respondents (i.e. precision). That proportion is significantly larger for the hand-coded prompts. That can be explained by the fact that the recommender algorithm by its design suggests all food items that are statistically likely to appear with foods reported by the respondent. In contrast to that, hand-coded food prompts include only foods that are commonly omitted based on experience of nutritionist from past surveys and interviews. Thus, the improved version of the algorithm could learn, which foods are more likely to be omitted, and use that knowledge to improve the precision of recommendations. Future research, could also look into alternative machine learning methods for building the recommender system for prompting omitted foods that would potentially offer higher precision, for example, the Skip-gram with negative sampling (SGNS) [21, 107]. Another limitation of the algorithm discussed in chapter 5 is that it was trained on data collected in surveys that used hand-coded food prompts. In a study described in chapter 5 the algorithm captured twice as many unique foods as hand-coded prompts. However, the algorithm might not have learnt some food associations would respondents not reported them without hand-coded prompts in surveys that were used for training. In addition to that, the algorithm needs a rich dataset of recalls from past surveys to be trained on, which might be not available for new contexts the system is deployed in (regional, cultural, intervention etc.). Hence, at the moment the recommender evidently improves the performance of the system's recall assistance but it should be used as an extension for hand-coded prompts and not as their replacement. To move towards fully automated associated food prompts the future research needs to (1) analyse the performance of the recommender algorithm that is trained on a dataset that was collected without hand-coded associated food prompts; (2) develop and validate a methodology for training the algorithm for new contexts. For example, when the system is deployed in a new country, recalls could be assisted by hand-coded prompts to collect enough of high-quality data to train the recommender model and switch the system to using this model afterwards.

Another challenge that could potentially be addressed with data driven methods in dietary assessment in future is the detection of misreporting. At the moment, data collected from a survey is analysed manually based on such metrics as the duration of a recall or the amount of reported energy. That can be a challenging process given a large sample size. Instead, the system could harvest additional information about the quality of answers from collected bahavioural data. For example, a user that spends too little time on searching and selecting a food may potentially lie about their intake. Such metrics can be aggreagated with a behaviour data collection module that is described in chapter 3. The outliers can then be flagged by the system to be further examined by researchers. This may significantly reduce time and effort involved in data quality assurance in dietary assessment studies.

8.2.3 Research question 3 (RQ3). Can tailored dietary feedback improve participatory engagement in online dietary assessment surveys?

RA2 and RA4 described in chapters 3 and 7 of this thesis aim to address intentional misreporting in online dietary assessment surveys. As one of the major sources of intentional under-reporting chapter 2 identifies social-desirability bias that causes some respondents of dietary surveys to alter their reported intake towards accepted societal norms. The chapter identifies another reason for intentional misreporting in low motivation of some respondents to go through a long survey [139, 100]. Taking lessons from research in participatory engagement in crowdsourcing tasks and online surveys chapter 7 proposes to address these issues by adding personal interest in accurate survey results for respondents [96, 115, 130]. This chapter examines the feasibility of using tailored dietary feedback as such an incentive and identifies characteristics of the feedback that facilitate the engagement of respondents with it. The chapter describes the development of a personal dietary dashboard for respondents of dietary assessment surveys. The visual design of the dashboard is informed by the previous HCI research in the area of participatory engagement with online surveys, persuasive technology and personal informatics [71, 90, 44, 68]. Dietary recommendations presented in the dashboard are based on the Eatwell Guide and UK dietary reference values [35, 1]. To collect the data that informs the answer to the RQ3 the dashboard was deployed in two dietary surveys and some of the respondents who used the feedback were interviewed. This study has demonstrated that respondents who requested receiving dietary feedback submitted on average considerably more recalls and reported more energy per a single recall, which may indicate an improved quality of collected data. This chapter also discusses the challenge of delivering dietary feedback without changes to diets of respondents that may limit the ability to capture habitual intake of a population [157, 100]. This, for example, contrasts with the focus of technology for personal dietary assessment (e.g. mobile dietary tracking application) that mainly aims to change and improve diets of its users. For that reason, in this chapter, the respondents of the second study received dietary feedback only after completing the required number of recalls. Future research could analyse and estimate the degree, to which presenting dietary feedback to respondents during online dietary assessment surveys changes their diets and the results of the assessment.

In interviews with respondents, five key engagement characteristics of dietary feedback emerged that are described in detail in chapter 7. One experience valued by respondents was their ability to grasp information quickly from a glance through the dashboard and to gradually explore the areas of their interest in more detail. For example, if the dashboard highlights a lack of some nutrient, respondents can open a modal window that contains extracts of information about the role of that nutrient in their diet, foods that contain it and links to trusted resources (e.g. NHS website) that respondents can explore. Interviews also indicated that information presented in the feedback should be easy to remember. The use of striking and unusual images could potentially facilitate that [172]. At the same time, visual design of nutrient feedback should demonstrate credibility and trustworthiness of presented information. In case of this study, for example, the use of cartoon characters resulted in a negative attitude of some respondents to the design of the dietary dashboard. The next two engagement characteristics identified from interviews highlighted a clear research direction for the development of nutrient feedback not only for dietary assessment surveys but for other contexts as well (e.g. personal mobile dietary tracking applications). Many respondents wanted the nutrient feedback presented to them to go beyond general recommendations and to be customizable towards their personal dietary targets. Above that they suggested that the feedback should provide guidance on shaping eating habits toward those goals (e.g. changing portion sizes, offering food substitutions). That is a great research opportunity that presents another challenge in dietary assessment that could be tackled with data science. For example, a recommender system could be developed that offers substitutions for foods recorded by a respondent that do not fit into their nutrient targets. However, since diet is a key life-factor in human health and well-being, an inexperienced individual should not have an unsupervised access to such a tool. Instead, this system could assist dieticians and make their job even more effective. So, they could discuss intervention targets with their patients and customise the tool to provide recommendations for the patient according to those targets.

8.3 Limitations

As with most research studies, there are limitations to results presented in this thesis. It is important to acknowledge these limitations and provide context to the validity of claims stated in this research. The limitations presented in this thesis can also inform future studies in the area of automated dietary assessment.

Throughout the history of Intake24 there have been hundreds of thousands of meals reported in the system. That includes such information as food contents of those meals, meal names, nutrient values and system usage data (e.g. duration of recall, user search query). This dataset presents an invaluable resource for developing and refining data driven methods for improving the accuracy of dietary assessment systems. For example, some of that data was used for the evaluation purposes in the development of the recommender algorithm covered in chapter 4 and for analysing the engagement of respondents with dietary feedback described in chapter 7. User behaviour data collected with Intake24 also helped to analyse the performance of a progressive recall method for recording dietary intake described in chapter 6. However, a large proportion of data collected from dietary assessment surveys that use Intake24 is owned by researchers behind those surveys and sharing of that data is limited by non-disclosure agreements. That factor limits the reproducibility of some of the results described in this thesis. In other words, researchers that want to reproduce these results would have to obtain such data from other resources or, otherwise, collect it themselves. In the meantime, the system collects no personally identifiable or any sensitive information about respondents. The development of Intake24 has always followed principles of open source and I cannot think of any sensible reason for the dataset to be not publicly available for the research community. Public access to the dataset would not only help the reproducibility of the current research but would also facilitate the development of alternative solutions for challenges described in this work.

Every study that informed answers to questions of this thesis compared the performance of Intake24 extended with one of proposed recall assistance methods to that of the existing previously validated version of the system [30]. The primary focus of these activities was not on the accuracy of results produced by each version of the system individually but on deviations in performance metrics of the updated system from its validated state. The aim of these activities was to find initial support to hypotheses behind research questions with a view to validate the results on a larger scale and/or against alternative methods of measuring dietary intake (e.g. direct meal observation, boilermakers of dietary intake) [77]. For those reasons, studies in research activities aimed to be designed so that both versions of the system were equally affected by limitations of that design. One limitation of this research is that the user study performed in RA2 that is described in chapter 3 and informed answers to research questions of this thesis involved a relatively small sample size (N=49) with a limited demographic profile. That may mean that findings from that study do not translate to a wider population. Nevertheless, findings related to the use of a recommender system for reminding respondents about potentially omitted foods that are described in chapter 5 received additional support from findings from a larger survey conducted during the RA3. Similarly, findings from interviews from RA2 related to the use of tailored dietary feedback as an incentive for participation in dietary assessment surveys described in chapter 7 were supported by the results of the analysis of usage data collected in a survey during RA4. The other limitation that should also be noted is that in all studies the protocol of a multiple pass 24-hour recall was not strictly followed [69, 157]. The method recommends that respondents should complete at least three recalls on non-consecutive weekdays and one recall on a weekend day to capture their habitual intake [77]. In the user study performed in RA2 respondents submitted only week day recalls and in the surveys during the RA3 there were no specific requirements to the days of the week for respondents for completing their recalls. Although respondents of the study in the RA2 were explicitly asked not to change their diet, they were informed about the schedule of their recalls, which potentially could affect their diets.

8.4 Conclusion

This thesis identifies and provides some support to three novel research directions in improving the accuracy of data aggregated in large-scale population online dietary assessment surveys. Firstly, the thesis developed a recommender system for assisting dietary recalls and analysed its performance in real-life dietary assessment surveys. As the second direction this thesis explored using short retention intervals in a multiple-pass 24-hour recall to reduce burden on human memory. Lastly, this research looked into improving participatory engagement with online dietary assessment surveys by providing dietary feedback to respondents based on their answers. These three directions were explored within a usability framework for dietary assessment systems that was also developed in this thesis. Applications of the new methods were demonstrated and analysed in a system that is based on a multiple-pass 24-hour recall. At the same time, the proposed methods can likely find applications in online surveys that use different protocols for collecting dietary data. For example, online surveys that are based on the FFQ method could potentially benefit from the usability framework that was used for the evaluation of new recall assistance methods and was described in chapter 3, the recommender system for prompting omitted foods that was developed in chapter 4 and evaluated in chapter 5, as well as from the use of tailored dietary feedback as an incentive that was explored in chapter 7. The new approach to building recommender systems that was initially developed for prompting survey respondents about potentially omitted foods and that is described in chapters 4, 5 can potentially find applications outside the scope of online dietary assessment surveys. Contrary to other recommender systems that rely heavily on historical data about personal choice of an individual, the new method builds a collective model from choices of a large group of people. Such approach makes the recommender system less prone to the cold-start problem, when the system is not able to produce recommendations for a new user who does not have an established history of preferences in the system. That makes the recommender system developed in this thesis suitable for contexts, where an extensive history of user behaviour is not aggregated or stored and/or where the range of indicators representing user interests is limited (e.g. systems with high privacy concerns and/or irregular usage). Lastly, consumer mobile applications for personal dietary assessment (e.g. MyFitnessPal) could potentially benefit of some of the approaches used in validated large-scale dietary assessment systems covered in this thesis [43].

There are a number of challenges in large-scale automated dietary assessment that remain to be addressed with the development of this technology. One of the biggest limitations in assessing diets in low-income populations is the reliance on respondents owning a personal mobile device and having an access to the Internet. For example, only 15% of Africans owned a smartphone in 2015 [129]. Dietary assessment methods based on interviewer-led recalls and direct observations are still widely spread in this context [42, 66]. A potential solution could be learned from research by Mitra et al. who used shared public computers for self-organized education for children [109, 110]. Similar to that the feasibility of using shared computer devices for self-administered dietary recalls could be explored.

In the meantime, research directions in technology for improving the accuracy of dietary assessment are not limited to the methods discussed in this paper. There are a number of state-of-the-art methods that could benefit the accuracy of assessment or add more details to data collected in dietary studies. For example, machine learning can be used for detecting eating behaviour including hand-to-mouth actions, chewing, and swallowing of food from data collected with wearable sensors [136]. Applied to geolocation data from a mobile phone machine learning could potentially infer a respondent being at a restaurant. One way these clues could be beneficial is if they are used for prompting respondents about reporting their intake in the right moment. Wearable sensors can also provide some indication of physical activity of respondents, which can be useful, for example, to estimate their energy requirement. Wearable cameras or handheld devices can be used to capture the exact foods on respondent's plate [65, 13]. Research demonstrates the potential for image processing and machine learning technology to be used to identify foods and drinks captured on those images, estimate their portion sizes and even their energy and nutrient content [175, 92]. That could further ease the process of reporting foods and reduce the burden on respondents. The use of these methods in large-scale dietary assessment currently faces economic and scalability concerns. Nonetheless, dietary assessment systems could explore opportunistic use of contextual data, when, for example, respondent agrees to share sensor data from their mobile or wearable device. At the same time, such methods require collecting sensitive data, which means that issues related to privacy of respondents should also be explored. Thus, automated dietary assessment offers a range of unique challenges and opportunities to researchers from various fields of computing.

References

[1] (2016). Government recommendations for energy and nutrients for males and females aged 1 – 18 years and 19+ years. Government Dietary Recommendations.

[2] (2017). Health matters: obesity and the food environment.

[3] (2018). Britain's fat fight with hugh fearnley-whittingstall. https://www.bbc.co.uk/programmes/b0b15qt7 . Archived at: http://www.webcitation.org/77zpd1Drw. Accessed: 2019-04-29.

[4] (2018). NDNS: results from years 7 and 8 (combined).

[5] (2018). Newcastle can. https://www.newcastlecan.com/ . Archived at: http://www.webcitation.org/744EwYZYM. Accessed: 2018-11-20.

[6] (2018). Pilot of INTAKE24 in the Scottish Health Survey.

[7] (2018). Statistics on Obesity, Physical Activity and Diet - England, 2018.

[8] (2019). Statistics on Obesity, Physical Activity and Diet, England, 2019.

[9] (2019). What should my daily intake of calories be? https://www.nhs.uk/common-health-questions/food-and-diet/what-should-my-daily-intake-of-calories-be/ . Archived at: http://www.webcitation.org/76rgYeqgK. Accessed: 2019-03-14.

[10] Abu-Ouf, N. M. and Jan, M. M. (2015). The impact of maternal iron deficiency and iron deficiency anemia on child's health. *Saudi medical journal*, 36(2):146.

[11] Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Acm sigmod record*, volume 22, pages 207–216. ACM.

[12] Agrawal, R., Srikant, R., et al. (1994). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499.

[13] Ahmad, Z., Bosch, M., Khanna, N., Kerr, D. A., Boushey, C. J., Zhu, F., and Delp, E. J. (2016). A mobile food record for integrated dietary assessment. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, pages 53–62. ACM.

[14] Albar, S. A., Alwan, N. A., Evans, C. E., Greenwood, D. C., and Cade, J. E. (2016). Agreement between an online dietary assessment tool (myfood24) and an interviewer-administered 24-h dietary recall in british adolescents aged 11–18 years. *British Journal of Nutrition*, 115(9):1678–1686.

[15] Albar, S. A., Carter, M. C., Alwan, N. A., Evans, C. E., and Cade, J. E. (2015). Formative evaluation of the usability and acceptability of myfood24 among adolescents: a uk online dietary assessments tool. *BMC Nutrition*, 1(1):29.

[16] Arab, L. (2003). Biomarkers of fat and fatty acid intake. *The Journal of nutrition*, 133(3):925S–932S.

[17] Ashwell, M., Barlow, S., Gibson, S., and Harris, C. (2006). National diet and nutrition surveys: the british experience. *Public health nutrition*, 9(4):523–530.

[18] Astrup, A., Dyerberg, J., Elwood, P., Hermansen, K., Hu, F. B., Jakobsen, M. U., Kok, F. J., Krauss, R. M., Lecerf, J. M., LeGrand, P., et al. (2011). The role of reducing intakes of saturated fat in the prevention of cardiovascular disease: where does the evidence stand in 2010? *The American journal of clinical nutrition*, 93(4):684–688.

[19] Atkinson, J., Figueroa, A., and Pérez, C. (2013). A semantically-based lattice approach for assessing patterns in text mining tasks. *Computación y Sistemas*, 17(4).

[20] Aune, D., Chan, D. S., Lau, R., Vieira, R., Greenwood, D. C., Kampman, E., and Norat, T. (2011). Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *Bmj*, 343:d6617.

[21] Barkan, O. and Koenigstein, N. (2016). Item2vec: neural item embedding for collaborative filtering. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE.

[22] Baxter, S. D., Guinn, C. H., Royer, J. A., Hardin, J. W., and Smith, A. F. (2010). Shortening the retention interval of 24-hour dietary recalls increases fourth-grade children's accuracy for reporting energy and macronutrient intake at school meals. *Journal of the American Dietetic Association*, 110(8):1178–1188.

[23] Baxter, S. D., Hitchcock, D. B., Guinn, C. H., Vaadi, K. K., Puryear, M. P., Royer, J. A., McIver, K. L., Dowda, M., Pate, R. R., and Wilson, D. K. (2014). A validation study concerning the effects of interview content, retention interval, and grade on children's recall accuracy for dietary intake and/or physical activity. *Journal of the Academy of Nutrition and Dietetics*, 114(12):1902–1914.

[24] Baxter, S. D., Thompson, W. O., Davis, H. C., and Johnson, M. H. (1997). Impact of gender, ethnicity, meal component, and time interval between eating and reporting on accuracy of fourth-graders' self-reports of school lunch. *Journal of the American Dietetic Association*, 97(11):1293–1298.

[25] Bhatnagar, P., Wickramasinghe, K., Williams, J., Rayner, M., and Townsend, N. (2015). The epidemiology of cardiovascular disease in the uk 2014. *Heart*, 101(15):1182–1189.

[26] Biltoft-Jensen, A., Trolle, E., Christensen, T., Islam, N., Andersen, L., Egenfeldt-Nielsen, S., and Tetens, I. (2014). Webdasc: a web-based dietary assessment software for 8–11-year-old danish children. *Journal of human nutrition and dietetics*, 27:43–53.

[27] Bingham, S. A. (2002). Biomarkers in nutritional epidemiology. *Public health nutrition*, 5(6a):821–827.

[28] Bingham, S. A., Gill, C., Welch, A., Cassidy, A., Runswick, S. A., Oakes, S., Lubin, R., Thurnham, D. I., Key, T., Roe, L., et al. (1997). Validation of dietary assessment methods in the uk arm of epic using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin c and carotenoids as biomarkers. *International journal of epidemiology*, 26(suppl_1):S137.

[29] Boushey, C., Spoden, M., Zhu, F., Delp, E., and Kerr, D. (2017). New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. *Proceedings of the Nutrition Society*, 76(3):283–294.

[30] Bradley, J., Simpson, E., Poliakov, I., Matthews, J. N., Olivier, P., Adamson, A. J., and Foster, E. (2016). Comparison of intake24 (an online 24-h dietary recall tool) with interviewer-led 24-h recall in 11–24 year-old. *Nutrients*, 8(6):358.

[31] Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

[32] Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

[33] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., and Hullender, G. (2005). Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pages 89–96. ACM.

[34] Burrows, T. L., Martin, R. J., and Collins, C. E. (2010). A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *Journal of the American Dietetic Association*, 110(10):1501–1510.

[35] Buttriss, J. (2016). The eatwell guide refreshed. *Nutrition Bulletin*, 41(2):135–141.

[36] Buzzard, I. M., Faucett, C. L., Jeffery, R. W., McBANE, L., McGOVERN, P., Baxter, J. S., Shapiro, A. C., Blackburn, G. L., T CHLEBOWSKI, R., Elashoff, R. M., et al. (1996). Monitoring dietary change in a low-fat diet intervention study: advantages of using 24-hour dietary recalls vs food records. *Journal of the American Dietetic Association*, 96(6):574–579.

[37] Cade, J. E. (2017). Measuring diet in the 21st century: Use of new technologies. *Proceedings of the Nutrition Society*, 76(3):276–282.

[38] Carrer-Neto, W., Hernández-Alcaraz, M. L., Valencia-García, R., and García-Sánchez, F. (2012). Social knowledge-based recommender system. application to the movies domain. *Expert Systems with applications*, 39(12):10990–11000.

[39] Carter, M. C., Albar, S. A., Morris, M. A., Mulla, U. Z., Hancock, N., Evans, C. E., Alwan, N. A., Greenwood, D. C., Hardie, L. J., Frost, G. S., et al. (2015). Development of a UK online 24-h dietary assessment tool: Myfood24. *Nutrients*, 7(6):4016–4032.

[40] Carter, M. C., Hancock, N., Albar, S. A., Brown, H., Greenwood, D. C., Hardie, L. J., Frost, G. S., Wark, P. A., and Cade, J. E. (2016). Development of a new branded uk food composition database for an online dietary assessment tool. *Nutrients*, 8(8):480.

[41] Castiglione, C. and Mazzocchi, M. (2019). Ten years of five-a-day policy in the uk: Nutritional outcomes and environmental effects. *Ecological economics*, 157:185–194.

[42] Caswell, B. L., Talegawkar, S. A., Dyer, B., Siamusantu, W., Klemm, R. D., and Palmer, A. C. (2015). Assessing child nutrient intakes using a tablet-based 24-hour recall tool in rural zambia. *Food and nutrition bulletin*, 36(4):467–480.

[43] Chen, J., Berkman, W., Bardouh, M., Ng, C. Y. K., and Allman-Farinelli, M. (2019). The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. *Nutrition*, 57:208–216.

[44] Chiu, M.-C., Chang, S.-P., Chang, Y.-C., Chu, H.-H., Chen, C. C.-H., Hsiao, F.-H., and Ko, J.-C. (2009). Playful bottle: a mobile social persuasion system to motivate healthy water intake. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 185–194. ACM.

[45] Choe, E. K., Lee, N. B., Lee, B., Pratt, W., and Kientz, J. A. (2014). Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1143–1152. ACM.

[46] Covington, P., Adams, J., and Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 191–198. ACM.

[47] Davis, J. and Goadrich, M. (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240. ACM.

[48] De Bourdeaudhuij, I. and Brug, J. (2000). Tailoring dietary feedback to reduce fat intake: an intervention at the family level. *Health education research*, 15(4):449–462.

[49] DuMouchel, W. and Pregibon, D. (2001). Empirical bayes screening for multi-item associations. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 67–76. ACM.

[50] Elmagarmid, A. K., Ipeirotis, P. G., and Verykios, V. S. (2007). Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):1–16.

[51] Ettienne-Gittens, R., Boushey, C. J., Au, D., Murphy, S. P., Lim, U., and Wilkens, L. (2013). Evaluating the feasibility of utilizing the automated self-administered 24-hour (asa24) dietary recall in a sample of multiethnic older adults. *Procedia food science*, 2:134–144.

[52] Ezendam, N. P., Brug, J., and Oenema, A. (2012). Evaluation of the web-based computer-tailored fataintphat intervention to promote energy balance among adolescents: results from a school cluster randomized trial. *Archives of pediatrics & adolescent medicine*, 166(3):248–255.

[53] Fiedler, J. L., Martin-Prével, Y., and Moursi, M. (2013). Relative costs of 24-hour recall and household consumption and expenditures surveys for nutrition analysis. *Food and nutrition bulletin*, 34(3):318–330.

[54] Fitt, E., Mak, T., Stephen, A., Prynne, C., Roberts, C., Swan, G., and Farron-Wilson, M. (2010). Disaggregating composite food codes in the uk national diet and nutrition survey food composition databank. *European journal of clinical nutrition*, 64(S3):S32.

[55] Foster, E. and Bradley, J. (2018). Methodological considerations and future insights for 24-hour dietary recall assessment in children. *Nutrition Research*, 51:1–11.

[56] Foster, E., Delve, J., Simpson, E., and Breininger, S.-P. (2014a). Comparison study: Intake24 vs interviewer led recall final report.

[57] Foster, E., Hawkins, A., Barton, K. L., Stamp, E., Matthews, J. N., and Adamson, A. J. (2017). Development of food photographs for use with children aged 18 months to 16 years: Comparison against weighed food diaries–the young person's food atlas (uk). *PloS one*, 12(2):e0169084.

[58] Foster, E., Hawkins, A., Delve, J., and Adamson, A. (2014b). Reducing the cost of dietary assessment: Self-completed recall and analysis of nutrition for use with children (scran 24). *Journal of human nutrition and dietetics*, 27:26–35.

[59] Foster, E., Hawkins, A., Stamp, E., and Adamson, A. (2010). Development and validation of an interactive portion size assessment system (ipsas). *Proceedings of the Nutrition Society*, 69(OCE6).

[60] Foster, E., Lee, C., Imamura, F., Hollidge, S. E., Westgate, K. L., Venables, M. C., Poliakov, I., Rowland, M. K., Osadchiy, T., Bradley, J. C., and et al. (2019). Validity and reliability of an online self-report 24-h dietary recall method (intake24): a doubly labelled water study and repeated-measures analysis. *Journal of Nutritional Science*, 8:e29.

[61] Foster, E., Matthews, J., Lloyd, J., Marshall, L., Mathers, J., Nelson, M., Barton, K. L., Wrieden, W., Cornelissen, P., Harris, J., et al. (2008). Children's estimates of food portion size: the development and evaluation of three portion size assessment tools for use with children. *British Journal of Nutrition*, 99(1):175–184.

[62] Fritz, T., Huang, E. M., Murphy, G. C., and Zimmermann, T. (2014). Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 487–496. ACM.

[63] Galesic, M. and Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly*, 73(2):349–360.

[64] Gemming, L., Jiang, Y., Swinburn, B., Utter, J., and Mhurchu, C. N. (2014). Under-reporting remains a key limitation of self-reported dietary intake: an analysis of the 2008/09 new zealand adult nutrition survey. *European journal of clinical nutrition*, 68(2):259.

[65] Gemming, L., Utter, J., and Mhurchu, C. N. (2015). Image-assisted dietary assessment: a systematic review of the evidence. *Journal of the Academy of Nutrition and Dietetics*, 115(1):64–77.

[66] Gibson, R. S., Charrondiere, U. R., and Bell, W. (2017). Measurement errors in dietary assessment using self-reported 24-hour recalls in low-income countries and strategies for their prevention. *Advances in Nutrition*, 8(6):980–991.

[67] Gomez-Uribe, C. A. and Hunt, N. (2016). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):13.

[68] Grimes, A., Kantroo, V., and Grinter, R. E. (2010). Let's play!: mobile health games for adults. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 241–250. ACM.

[69] Guenther, P., DeMaio, T., Ingwersen, L., and Berlin, M. (1996). The multiple-pass approach for the 24-hour recall in the continuing survey of food intakes by individuals (csfii) 1994-96. In *FASEB JOURNAL*, volume 10, pages 1142–1142. FEDERATION AMER SOC EXP BIOL 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998.

[70] Harmon, B. E., Boushey, C. J., Shvetsov, Y. B., Ettienne, R., Reedy, J., Wilkens, L. R., Le Marchand, L., Henderson, B. E., and Kolonel, L. N. (2015). Associations of key diet-quality indexes with mortality in the multiethnic cohort: the dietary patterns methods project–. *The American journal of clinical nutrition*, 101(3):587–597.

[71] Harms, J., Seitz, D., Wimmer, C., Kappel, K., and Grechenig, T. (2015). Low-cost gamification of online surveys: Improving the user experience through achievement badges. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pages 109–113. ACM.

[72] Harris, J. A. and Benedict, F. G. (1918). A biometric study of human basal metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 4(12):370.

[73] Haytowitz, D. B. and Pehrsson, P. R. (2018). Usda's national food and nutrient analysis program (nfnap) produces high-quality data for usda food composition databases: Two decades of collaboration. *Food chemistry*, 238:134–138.

[74] Hebert, J. R., Clemow, L., Pbert, L., Ockene, I. S., and Ockene, J. K. (1995). Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *International journal of epidemiology*, 24(2):389–398.

[75] Hernández, T., Wilder, L., Kuehn, D., Rubotzky, K., Moser-Veillon, P., Godwin, S., Thompson, C., and Wang, C. (2006). Portion size estimation and expectation of accuracy. *Journal of Food Composition and Analysis*, 19:S14–S21.

[76] Ji, C.-R. and Deng, Z.-H. (2007). Mining frequent ordered patterns without candidate generation. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, volume 1, pages 402–406. IEEE.

[77] Johnson, R. K. (2002). Dietary Intake-How Do We Measure What People Are Really Eating? *Obesity*, 10(S11):63S–68S.

[78] Johnson, R. K., Driscoll, P., and Goran, M. I. (1996). Comparison of multiple-pass 24-hour recall estimates of energy intake with total energy expenditure determined by the doubly labeled water method in young children. *Journal of the American Dietetic Association*, 96(11):1140–1144.

[79] Kalantarian, H., Alshurafa, N., and Sarrafzadeh, M. (2017). A survey of diet monitoring technology. *IEEE Pervasive Computing*, 16(1):57–65.

[80] Kirkpatrick, S., Raffoul, A., Sacco, J., Lee, K., Chen, E., Pasha, S., Marcinow, M., Orr, S., and Hobin, E. (2017a). Evaluation of the automated self-administered 24-hour dietary assessment tool (asa24) for use with children: An observational feeding study. *The FASEB Journal*, 31(1_supplement):149–7.

[81] Kirkpatrick, S. I., Gilsing, A. M., Hobin, E., Solbak, N. M., Wallace, A., Haines, J., Mayhew, A. J., Orr, S. K., Raina, P., Robson, P. J., et al. (2017b). Lessons from studies to evaluate an online 24-hour recall for use with children and adults in canada. *Nutrients*, 9(2):100.

[82] Kirkpatrick, S. I., Thompson, F. E., Subar, A. F., Douglass, D., Zimmerman, T. P., Kahle, L. L., George, S. M., and Potischman, N. (2013). Validity of the national cancer institute's automated self-administered 24-hour recall (asa24): results of a feeding study.

[83] Klipstein-Grobusch, K. d., Den Breeijen, J., Goldbohm, R., Geleijnse, J., Hofman, A., Grobbee, D., and Witteman, J. (1998). Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. *European Journal of Clinical Nutrition*, 52(8):588.

[84] Kristal, A. R., Andrilla, C. H. A., D KOEPSELL, T., Diehr, P. H., and Cheadle, A. (1998). Dietary assessment instruments are susceptible to intervention-associated response set bias. *Journal of the American Dietetic Association*, 98(1):40–43.

[85] Kristal, A. R., Peters, U., and Potter, J. D. (2005). Is it time to abandon the food frequency questionnaire?

[86] Kroeze, W., Oenema, A., Campbell, M., and Brug, J. (2008). The efficacy of web-based and print-delivered computer-tailored interventions to reduce fat intake: results of a randomized, controlled trial. *Journal of nutrition education and behavior*, 40(4):226–236.

[87] Law, E., Yin, M., Goh, J., Chen, K., Terry, M. A., and Gajos, K. Z. (2016). Curiosity killed the cat, but makes crowdwork better. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4098–4110. ACM.

[88] Li, H., Wang, Y., Zhang, D., Zhang, M., and Chang, E. Y. (2008). Pfp: parallel fp-growth for query recommendation. In *Proceedings of the 2008 ACM conference on Recommender systems*, pages 107–114. ACM.

[89] Lika, B., Kolomvatsos, K., and Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4):2065–2073.

[90] Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., and Strub, H. B. (2006). Fish'n'steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*, pages 261–278. Springer.

[91] Linden, G., Smith, B., and York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, (1):76–80.

[92] Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., and Ma, Y. (2016). Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In *International Conference on Smart Homes and Health Telematics*, pages 37–48. Springer.

[93] Livingstone, M., Prentice, A., Strain, J., Coward, W., Black, A., Barker, M., McKenna, P., and Whitehead, R. (1990). Accuracy of weighed dietary records in studies of diet and health. *Bmj*, 300(6726):708–712.

[94] Livingstone, M. B. E. and Black, A. E. (2003). Markers of the validity of reported energy intake. *The Journal of nutrition*, 133(3):895S–920S.

[95] Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, 116(1):75.

[96] Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., and Maher, C. (2017). Does gamification increase engagement with online programs? a systematic review. *PloS one*, 12(3):e0173403.

[97] Lopes, T., Luiz, R., Hoffman, D., Ferriolli, E., Pfrimer, K., Moura, A., Sichieri, R., and Pereira, R. (2016). Misreport of energy intake assessed with food records and 24-h recalls compared with total energy expenditure estimated with dlw. *European journal of clinical nutrition*, 70(11):1259.

[98] Lupton, D. (2014). Self-tracking cultures: towards a sociology of personal informatics. In *Proceedings of the 26th Australian computer-human interaction conference on designing futures: The future of design*, pages 77–86. ACM.

[99] Lustria, M. L. A., Noar, S. M., Cortese, J., Van Stee, S. K., Glueckauf, R. L., and Lee, J. (2013). A meta-analysis of web-delivered tailored health behavior change interventions. *Journal of health communication*, 18(9):1039–1069.

[100] Macdiarmid, J. and Blundell, J. (1998). Assessing dietary intake: Who, what and why of under-reporting. *Nutrition research reviews*, 11(2):231–253.

[101] Macdiarmid, J. I. and Blundell, J. (1997). Dietary under-reporting: what people say about recording their food intake. *European Journal of Clinical Nutrition*, 51(3):199.

[102] Malik, V. S., Pan, A., Willett, W. C., and Hu, F. B. (2013). Sugar-sweetened beverages and weight gain in children and adults: a systematic review and meta-analysis. *The American journal of clinical nutrition*, 98(4):1084–1102.

[103] Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60.

[104] Martinez-Lacoba, R., Pardo-Garcia, I., Amo-Saus, E., and Escribano-Sotos, F. (2018). Mediterranean diet and health outcomes: A systematic meta-review. *European journal of public health*, 28(5):955–961.

[105] Mayne, S. T., Playdon, M. C., and Rock, C. L. (2016). Diet, nutrition, and cancer: past, present and future. *Nature reviews Clinical oncology*, 13(8):504.

[106] Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D., Amde, M., Owen, S., et al. (2016). Mllib: Machine learning in apache spark. *The Journal of Machine Learning Research*, 17(1):1235–1241.

[107] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

[108] Miller, L. M. S. and Cassady, D. L. (2015). The effects of nutrition knowledge on food label use. a review of the literature. *Appetite*, 92:207–216.

[109] Mitra, S. (2005). Self organising systems for mass computer literacy: Findings from the 'hole in the wall'experiments. *International Journal of Development Issues*, 4(1):71–81.

[110] Mitra, S., Dangwal, R., Chatterjee, S., Jha, S., Bisht, R. S., and Kapur, P. (2005). Acquisition of computing literacy on shared public computers: Children and the" hole in the wall". *Australasian Journal of Educational Technology*, 21(3).

[111] Moore, T. J., Alsabeeh, N., Apovian, C. M., Murphy, M. C., Coffman, G. A., Cullum-Dugan, D., Jenkins, M., and Cabral, H. (2008). Weight, blood pressure, and dietary benefits after 12 months of a web-based nutrition education program (dash for health): longitudinal observational study. *Journal of medical Internet research*, 10(4).

[112] Moshfegh, A. J., Rhodes, D. G., Baer, D. J., Murayi, T., Clemens, J. C., Rumpler, W. V., Paul, D. R., Sebastian, R. S., Kuczynski, K. J., Ingwersen, L. A., et al. (2008). The us department of agriculture automated multiple-pass method reduces bias in the collection of energy intakes. *The American journal of clinical nutrition*, 88(2):324–332.

[113] Moynihan, P. and Petersen, P. E. (2004). Diet, nutrition and the prevention of dental diseases. *Public health nutrition*, 7(1a):201–226.

[114] Nachar, N. et al. (2008). The mann-whitney u: A test for assessing whether two independent samples come from the same distribution. *Tutorials in quantitative Methods for Psychology*, 4(1):13–20.

[115] Nakamura, J. and Csikszentmihalyi, M. (2003). The construction of meaning through vital engagement. *Flourishing: Positive psychology and the life well-lived*, pages 83–104.

[116] Naska, A., Valanou, E., Peppa, E., Katsoulis, M., Barbouni, A., and Trichopoulou, A. (2016). Evaluation of a digital food photography atlas used as portion size measurement aid in dietary surveys in greece. *Public health nutrition*, 19(13):2369–2376.

[117] Nelson, M., Atkinson, M., and Darbyshire, S. (1994). Food photography i: the perception of food portion size from photographs. *British Journal of Nutrition*, 72(5):649–663.

[118] Nelson, M., Atkinson, M., and Darbyshire, S. (1996). Food photography ii: use of food photographs for estimating portion size and the nutrient content of meals. *British journal of nutrition*, 76(1):31–49.

[119] Nielsen, J. (1996). Usability metrics: Tracking interface improvements. *Ieee Software*, 13(6):1–2.

[120] Oenema, A., Brug, J., and Lechner, L. (2001). Web-based tailored nutrition education: results of a randomized controlled trial. *Health education research*, 16(6):647–660.

[Osadchiy et al.] Osadchiy, T., Poliakov, I., Olivier, P., Foster, E., and Rowland, M. Progressive 24-hour recall: Feasibility and acceptability of short retention intervals in online dietary assessment surveys. *Journal of Medical Internet Research*.

[122] Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., and Foster, E. (2019a). Recommender system based on pairwise association rules. *Expert Systems with Applications*, 115:535–542.

[123] Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., and Foster, E. (2019b). Validation of a recommender system for prompting omitted foods in online dietary assessment surveys. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 208–215. ACM.

[124] Patrick, K., Calfas, K. J., Norman, G. J., Rosenberg, D., Zabinski, M. F., Sallis, J. F., Rock, C. L., and Dillon, L. W. (2011). Outcomes of a 12-month web-based intervention for overweight and obese men. *Annals of Behavioral Medicine*, 42(3):391–401.

[125] Pazzani, M. J. and Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer.

[126] Popescul, A., Pennock, D. M., and Lawrence, S. (2001). Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 437–444. Morgan Kaufmann Publishers Inc.

[127] Poslusna, K., Ruprich, J., de Vries, J. H., Jakubikova, M., and van't Veer, P. (2009). Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *British Journal of Nutrition*, 101(S2):S73–S85.

[128] Potischman, N. (2003). Biologic and methodologic issues for nutritional biomarkers. *The Journal of nutrition*, 133(3):875S–880S.

[129] Poushter, J. and Oates, R. (2015). Cell phones in africa: communication lifeline. *Washingston DC: Pew Research Centre*.

[130] Powers, D. E. and Alderman, D. L. (1982). Feedback as an incentive for responding to a mail questionnaire. *Research in Higher Education*, 17(3):207–211.

[131] Raeder, T. and Chawla, N. V. (2011). Market basket analysis with networks. *Social network analysis and mining*, 1(2):97–113.

[132] Rebro, S. M., Patterson, R. E., Kristal, A. R., and Cheney, C. L. (1998). The effect of keeping food records on eating patterns. *Journal of the Academy of Nutrition and Dietetics*, 98(10):1163.

[133] Reedy, J., Krebs-Smith, S. M., Miller, P. E., Liese, A. D., Kahle, L. L., Park, Y., and Subar, A. F. (2014). Higher diet quality is associated with decreased risk of all-cause, cardiovascular disease, and cancer mortality among older adults1, 2. *The Journal of nutrition*, 144(6):881–889.

[134] Roberts, C., Steer, T., Maplethorpe, N., Cox, L., Meadows, S., Nicholson, S., Page, P., and Swan, G. (2018). National diet and nutrition survey: results from years 7 and 8 (combined) of the rolling programme (2014/2015–2015/2016).

[135] Robinson, F., Morritz, W., McGuiness, P., and Hackett, A. (1997). A study of the use of a photographic food atlas to estimate served and self-served portion sizes. *Journal of Human Nutrition and Dietetics*, 10(2):117–124.

[136] Rollo, M. E., Williams, R. L., Burrows, T., Kirkpatrick, S. I., Bucher, T., and Collins, C. E. (2016). What are they really eating? a review on new approaches to dietary intake assessment and validation. *Current nutrition reports*, 5(4):307–314.

[137] Roth, M., Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., Leiser, N., Matias, Y., and Merom, R. (2010). Suggesting friends using the implicit social graph. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 233–242. ACM.

[138] Rowland, M., Adamson, A., Poliakov, I., Bradley, J., Simpson, E., Olivier, P., and Foster, E. (2018). Field testing of the use of intake24—an online 24-hour dietary recall system. *Nutrients*, 10(11):1690.

[139] Rowland, M., Poliakov, I., Christie, S., Simpson, E., and Foster, E. (2016). Field testing of the use of intake24 in a sample of young people and adults living in scotland.

[140] Rudin, C., Letham, B., Salleb-Aouissi, A., Kogan, E., and Madigan, D. (2011). Sequential event prediction with association rules. In *Proceedings of the 24th annual conference on learning theory*, pages 615–634.

[141] Ruiz, P. P., Foguem, B. K., and Grabot, B. (2014). Generating knowledge in maintenance from experience feedback. *Knowledge-Based Systems*, 68:4–20.

[142] Salzberg, S. L. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data mining and knowledge discovery*, 1(3):317–328.

[143] Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM.

[144] Schulz, K. U. and Mihov, S. (2002). Fast string correction with levenshtein automata. *International Journal on Document Analysis and Recognition*, 5(1):67–85.

[145] Sene, A., Kamsu-Foguem, B., and Rumeau, P. (2018). Discovering frequent patterns for in-flight incidents. *Cognitive Systems Research*, 49:97–113.

[146] Shaw, G., Xu, Y., and Geva, S. (2010). Using association rules to solve the cold-start problem in recommender systems. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 340–347. Springer.

[147] Shim, J.-S., Oh, K., and Kim, H. C. (2014). Dietary assessment methods in epidemiologic studies. *Epidemiology and Health*.

[148] Simpson, E., Bradley, J., Poliakov, I., Jackson, D., Olivier, P., Adamson, A. J., and Foster, E. (2017). Iterative development of an online dietary recall tool: Intake24. *Nutrients*, 9(2):118.

[149] Slimani, N., Casagrande, C., Nicolas, G., Freisling, H., Huybrechts, I., Ocké, M., Niekerk, E., Van Rossum, C., Bellemans, M., De Maeyer, M., et al. (2011). The standardized computerized 24-h dietary recall method epic-soft adapted for pan-european dietary monitoring. *European journal of clinical nutrition*, 65(S1):S5.

[150] Steele, R. (2015). An overview of the state of the art of automated capture of dietary intake information. *Critical reviews in food science and nutrition*, 55(13):1929–1938.

[151] Sternfeld, B., Block, C., Quesenberry Jr, C. P., Block, T. J., Husson, G., Norris, J. C., Nelson, M., and Block, G. (2009). Improving diet and physical activity with alive: a worksite randomized trial. *American journal of preventive medicine*, 36(6):475–483.

[152] Subar, A., Mittl, B., Zimmerman, T., Kirkpatrick, S., Schap, T., Wilson, M., and Potischman, N. (2015). Use of the automated self-administered 24-hour recall (asa24) in the real world. *The FASEB Journal*, 29(1_supplement):131–6.

[153] Subar, A. F., Crafts, J., Zimmerman, T. P., Wilson, M., Mittl, B., Islam, N. G., McNutt, S., Potischman, N., Buday, R., Hull, S. G., et al. (2010). Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. *Journal of the Academy of Nutrition and Dietetics*, 110(1):55–64.

[154] Subar, A. F., Kirkpatrick, S. I., Mittl, B., Zimmerman, T. P., Thompson, F. E., Bingley, C., Willis, G., Islam, N. G., Baranowski, T., McNutt, S., and Potischman, N. (2012). The Automated Self-Administered 24-Hour Dietary Recall (ASA24): A Resource for Researchers, Clinicians and Educators from the National Cancer Institute. *Journal of the Academy of Nutrition and Dietetics*, 112(8):1134–1137.

[155] Tantamango-Bartley, Y., Jaceldo-Siegl, K., Fan, J., and Fraser, G. (2013). Vegetarian diets and the incidence of cancer in a low-risk population. *Cancer Epidemiology and Prevention Biomarkers*, 22(2):286–294.

[156] Tate, D. F., Jackvony, E. H., and Wing, R. R. (2006). A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an internet weight loss program. *Archives of internal medicine*, 166(15):1620–1625.

[157] Thompson, F. E. and Byers, T. (1994). Dietary assessment resource manual. *The Journal of nutrition*, 124(suppl_11):2245s–2317s.

[158] Thompson, F. E., Dixit-Joshi, S., Potischman, N., Dodd, K. W., Kirkpatrick, S. I., Kushi, L. H., Alexander, G. L., Coleman, L. A., Zimmerman, T. P., Sundaram, M. E., et al. (2015). Comparison of interviewer-administered and automated self-administered 24-hour dietary recalls in 3 diverse integrated health systems. *American journal of epidemiology*, 181(12):970–978.

[159] Thompson, F. E., Subar, A. F., Loria, C. M., Reedy, J. L., and Baranowski, T. (2010). Need for technological innovation in dietary assessment. *Journal of the American Dietetic Association*, 110(1):48–51.

[160] Tooze, J. A., Midthune, D., Dodd, K. W., Freedman, L. S., Krebs-Smith, S. M., Subar, A. F., Guenther, P. M., Carroll, R. J., and Kipnis, V. (2006). A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *Journal of the American Dietetic Association*, 106(10):1575–1587.

[161] Turconi, G., Guarcello, M., Berzolari, F. G., Carolei, A., Bazzano, R., and Roggi, C. (2005). An evaluation of a colour food photography atlas as a tool for quantifying food portion size in epidemiological dietary surveys. *European journal of clinical nutrition*, 59(8):923.

[162] Van Deursen, A. J. and Van Dijk, J. A. (2009). Using the internet: Skill related problems in users' online behavior. *Interacting with computers*, 21(5-6):393–402.

[163] Van Gaal, L. F., Mertens, I. L., and Christophe, E. (2006). Mechanisms linking obesity with cardiovascular disease. *Nature*, 444(7121):875.

[164] van Weert, J. C., van Noort, G., Bol, N., van Dijk, L., Tates, K., and Jansen, J. (2011). Tailored information for cancer patients on the internet: effects of visual cues and language complexity on information recall and satisfaction. *Patient Education and Counseling*, 84(3):368–378.

[165] Voigt, P. and Von dem Bussche, A. (2017). The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*.

[166] Vuckovic, N., Ritenbaugh, C., Taren, D. L., and Tobar, M. (2000). A qualitative study of participants' experiences with dietary assessment. *Journal of the American Dietetic Association*, 100(9):1023–1028.

[167] Wark, P. A., Hardie, L. J., Frost, G. S., Alwan, N. A., Carter, M., Elliott, P., Ford, H. E., Hancock, N., Morris, M. A., Mulla, U. Z., et al. (2018). Validity of an online 24-h recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews. *BMC medicine*, 16(1):136.

[168] Welsh, J. A., Sharma, A., Cunningham, S. A., and Vos, M. B. (2011). Consumption of added sugars and indicators of cardiovascular disease risk among us adolescents. *Circulation*, 123(3):249–257.

[169] Whitton, C., Nicholson, S. K., Roberts, C., Prynne, C. J., Pot, G. K., Olson, A., Fitt, E., Cole, D., Teucher, B., Bates, B., et al. (2011). National diet and nutrition survey: Uk food consumption and nutrient intakes from the first year of the rolling programme and comparisons with previous surveys. *British journal of nutrition*, 106(12):1899–1914.

[170] Wrieden, W., Peace, H., Armstrong, J., and Barton, K. (2003). A short review of dietary assessment methods used in national and scottish research studies. In *Briefing Paper Prepared for: Working Group on Monitoring Scottish Dietary Targets Workshop. Edinburgh*.

[171] Wright, J. L., Sherriff, J. L., Dhaliwal, S. S., and Mamo, J. C. (2011). Tailored, iterative, printed dietary feedback is as effective as group education in improving dietary behaviours: results from a randomised control trial in middle-aged adults with cardiovascular risk factors. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1):43.

[172] Yates, F. A. (2013). *Art of Memory*. Routledge.

[173] Zack, R. M., Irema, K., Kazonda, P., Leyna, G. H., Liu, E., Gilbert, S., Lukmanji, Z., Spiegelman, D., Fawzi, W., Njelekela, M., et al. (2018). Validity of an ffq to measure nutrient and food intakes in tanzania. *Public health nutrition*, 21(12):2211–2220.

[174] Zheng, Z., Kohavi, R., and Mason, L. (2001). Real world performance of association rule algorithms. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 401–406. ACM.

[175] Zhu, F., Bosch, M., Khanna, N., Boushey, C. J., and Delp, E. J. (2015). Multiple hypotheses image segmentation and classification with application to dietary assessment. *IEEE Journal of Biomedical and Health Informatics*, 19(1):377–388.

[176] Zuniga, K. and McAuley, E. (2015). Considerations in selection of diet assessment methods for examining the effect of nutrition on cognition. *The journal of nutrition, health & aging*, 19(3):333–340.