Online Engagement for a Healthier You: A Case Study of Web-based Supermarket Health Program

Ling Luo[1,2,*], Bin Li[2], Shlomo Berkovsky[2], Irena Koprinska[1], Fang Chen[2]
[1]School of IT, University of Sydney, Sydney, NSW 2006, Australia
{ling.luo, irena.koprinska}@sydney.edu.au
[2]Data61, CSIRO, Eveleigh, NSW 2015, Australia
{bin.li, shlomo.berkovsky, fang.chen}@data61.csiro.au

ABSTRACT

Obesity is a growing problem affecting millions of people. Various behavior change programs have been designed to reduce its prevalence. An Australian supermarket has recently run a web-based health program to motivate people to eat healthily and do more physical activity. The program offered discounts on fresh products and a website, HealthierU, providing interactive support tools for participants. The stakeholders desire to evaluate if the program is effective and if the supporting website is useful to facilitate behavior changes. To answer these questions, in this work we propose a method to: (1) model individual purchase rate from sparse recorded transactions through a mixture of Non-Homogeneous Poisson Processes (NHPP), (2) design criteria for partitioning participants based on their interactions with the HealthierU website, (3) evaluate the program impact by comparing behavior changes across different groups of participants. Our case study shows that during the program the participants significantly increased their purchases of some fresh products. Both the distribution of behavior patterns and impact scores show that the program imposed relatively strong impact on the participants who logged activities and tracked weights. Our method can facilitate the enhancement of personalized health programs, especially aiming to maximize the program impact and targeting participants through web or mobile applications.

Keywords

supermarket health program; activity logs; online support; purchase behavior model; non-homogeneous Poisson process

1. INTRODUCTION

The World Health Organization estimated that the prevalence of obesity more than doubled in the past three decades, with over 600 million obese adults worldwide in 2014 [25]. Overweight and obese people have higher risks of cardiovascular problems, diabetes, and musculoskeletal disorders.

.

To reduce the overweight and obesity problems, various behavior change programs have been designed to encourage participants to adopt a healthier lifestyle, e.g. change their diet and increase physical activity level [3, 4]. There are also numerous websites and mobile applications created for health and fitness, which can support users to easily manage diet plans, count calorie intake, keep activity journals and communicate with others. The behavior change programs increasingly utilize websites, mobile applications and wearable devices to maximize the program benefits and motivate participants to achieve their goals. For example, an Australian supermarket recently run a program that encourages people to do more physical activity[1]. By linking activity trackers with loyalty cards, participants can get discounts on fresh products once reaching the milestones, e.g. 100,000 steps. These technologies can provide interactive and personalized services to participants and allow program stakeholders to manage and analyze behavioral data.

After collecting the behavioral data, the program stakeholders often want to evaluate if the program is effective, and whether the supporting websites and applications are helpful, so that they can improve the program to better motivate the participants and benefit a broader population. Previous studies mainly use descriptive statistics, rating-based surveys and statistical tests to evaluate the impact of programs and supporting web applications [3, 6, 8]. Although these methods can evaluate the overall impact of the health programs, they can not track the behavior changes of the individual participants over time and compare the program impact on participants with different types and levels of engagement with the web applications.

Our Approach. We provide a solution for evaluating the impact of a supermarket health program on different participants via purchase behavior modeling. The health program was delivered by an Australian supermarket chain, which offered a supporting website and 10% discount on fresh products to encourage participants to change lifestyle. Our method tracks the individual purchase behavior changes based on a mixture of NHPP, which adds the regularization to prevent overfitting the sparse and noisy individual records. Then, we measure the program impact on different participants, considering their interactions with the program website. The main **contributions** of our work are:

1) We construct reliable individual purchase rate curves, which act as the backbone of the program impact analysis. To overcome overfitting problems, the individual purchase rate curve is modeled as a mixture of pur-

---

[1]https://www.flybuys.com.au/collect#/health/stepstosavings

chase behavior models for latent customer segments, weighted by the individual soft memberships in all segments.

2) We design four criteria for partitioning participants by the usage of the program website and quantify the health program impact based on the changes of individual purchase rate.

3) We conduct a case study on a supermarket health program and evaluate its impact on participants with different levels of interactions with the program website.

Through the case study, we establish that: 1) during the program, the participants increased their purchase of fresh products; 2) the distributions of behavior patterns are different for participants with different levels of interactions with the program website; 3) the impact of the program is stronger for the participants with higher levels of interactions with the program website.

Related Work. Existing studies have explored the impact of using support websites and applications in the health program. Wieland et al. [24] showed that the digital systems provided an effective support for weight loss and maintenance compared to the offline intervention. Brindal et al. [6] demonstrated the positive role played by the support applications in their meal replacement program. However, only using descriptive statistics of the weight loss and surveys can be subjective and insufficient to capture how the behavior gradually changes due to the program participation. Tracking the behavior changes can provide valuable information to objectively evaluate the impact of a program [5].

Given the context of the supermarket health program, we build purchase behavior models and measure the changes in purchase rates of different products. The purchase behaviors have been studied extensively to provide decision support for businesses [1, 2, 12]. Popular techniques for detecting behavior changes include rule-based methods [7, 18], temporal collaborative filtering [13, 15, 16, 17, 20] and stochastic processes [9, 11, 22]. Considering the accuracy and interpretability, we use stochastic process, a mixture of NHPP models [19] as the base of our purchase behavior model. It is worth noting that the model in [19] focused on the customer segmentation and segment-level responses to promotions, whereas in this paper we have a different purpose, new methodology and evaluation. We construct individual purchase rate curves based on the segment-level model and propose criteria for partitioning customers, which enable us to track individual behavior changes and evaluate the impact of the program and the website.

2. SUPERMARKET HEALTH PROGRAM

Our health program, HealthierU, was conducted by an Australian national supermarket chain, which aimed to encourage participants to maintain a healthier diet. There were 931 participants and the duration of the program was 24 weeks: two cycles of a 12-week diet program. The program included two components: 1) online intervention – the HealthierU website; 2) offline intervention – 10% discount on fresh fruits and vegetables purchased in the supermarket.

HealthierU was designed to motivate participants to change lifestyle and improve their eating habits. The website provides a comprehensive set of interactive and personalized tools for the diet program, based on individual BMI and weight loss goals. The set of tools includes:
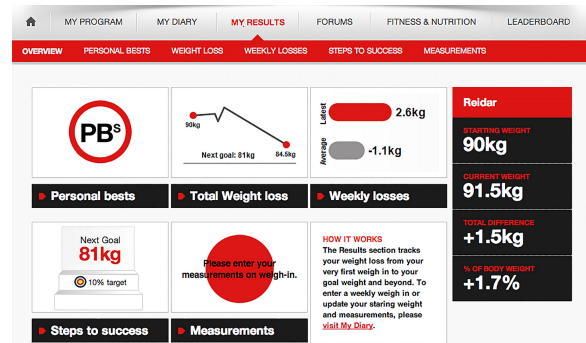

Figure 1: Personal records page of HealthierU

1) Static program content: overview of the diet program, nutrition tips, meal recipes, instructions for exercises, tutorials on the food groups and guides of the website.

2) Personalized diet plan: compliant diet plan with recipes tailored to the BMI, food intake, physical activity, weight loss goals and dietary requirements.

3) Personal diary: the main self-monitoring tool to plan and track daily food intake and exercising.

4) Weight tracker: the weight recording and progress visualization tool to track the weight changes.

5) Personal records and results: a set of interactive tools to view the records, measurements and progress towards the weight goals (Figure 1).

6) Weekly messages: the weekly emails to motivate interactions with the website and a summary of the program.

Overall, the website logged 49,489 actions from all the participants over the course of the program. Each entry contains the user id, timestamp and the type of the action e.g. visit personal page, view a menu, or read a recipe.

In terms of the offline purchase behaviors, although the program was from May to November in 2014, the transaction records were collected through loyalty cards between January 1 and December 31, 2014, which allows us to compare purchase behaviors before, during, and after the program. The participants received 10% discount on fresh products. The transaction record includes the participant ID, product ID, timestamp, product metadata (category, brand, name and bar code), purchased quantity and cost.

The program stakeholders desire to identify which participants are motivated by the program to buy more fresh products, and whether the website can effectively trigger stronger behavior changes towards a healthier lifestyle. To this end, we take both online and offline components of the program into consideration and explore the program impact on different types of customers[2], who are grouped based on their interaction with HealthierU, such as website logon, number of diaries they wrote and the usage of the weight tracker (the groups will be detailed in Section 3.2).

3. METHODOLOGY

Our method consists of three modules (Figure 2), where we: 1) segment customers based on their purchase behavior patterns via a mixture of NHPP; 2) model the individual-

---

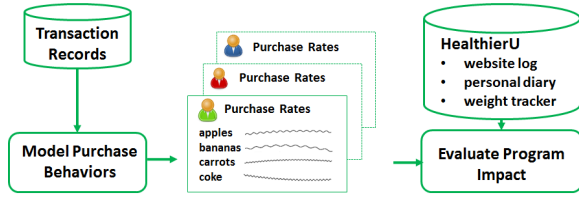[2]We use 'customers' and 'participants' interchangeably.

Figure 2: Flowchart of our methodology

level purchase rates of different products for all the customers; 3) evaluate the program impact by considering both the usage of HealthierU website and the fluctuations of individual purchase rates during the program.

3.1 Model Individual Purchase Rate

Our first task is to accurately model individual purchase behavior. One option would be to use the actual number of purchase events of a customer to characterize their purchase behavior. However, the main problem with this idea is that the purchase events of an individual customer are often too sparse and noisy to extract patterns or examine how the behavior changes within a period of time. *To prevent overfitting the sparse individual purchase records, we include additional information to regularize the individual models.* Thus, we first segment customers by constructing a mixture of NHPP models to identify the typical behavior patterns from the purchase events of all customers. Then, each customer is a member of multiple segments with soft memberships and the individual purchase rate curve is the weighted sum of segment-level models.

The advantages of modeling individual purchase rate via a mixture of NHPP models are: 1) it can mitigate the influence of sparsity and noise in the actual purchase records; 2) the individual purchase rate curve can be easily interpreted as a weighted combination of typical latent purchase behavior patterns shared by all customers; 3) it can show the individual purchase behavior changes and facilitate the analysis of program impact.

Identify Segment-level Purchase Behaviors. In this step, we segment customers and model segment-level purchase behavior patterns, which inform the construction of the individual purchase rate curves. Given a transaction data set with U customers and M products, for a product m, our tasks are to: 1) extract all customers $\{u_i\}_m$ ($i \in \{1, 2, ..., U\}$), who bought m, 2) identify K latent groups of $\{u_i\}_m$, based on each customer's purchase decisions $N(T)$ during the observation period $[0, T]$, where $N_{im}(T)$ is the total number of purchases of m by u_i. We omit the subscript m when the description is in the context of a product m. The purchase behaviors of customers in the latent group $k \in \{1, 2, ..., K\}$ share the same purchase model. We use the NHPP to describe the behavior of each customer group [19].

The number of purchases until time t can be modeled by a counting process $\{N(t)\}$, which has a Poisson distribution, so that $P(N(t) = n) = e^{-\Lambda} \frac{\Lambda^n}{n!}$ and $E(N(t)) = \Lambda$. As the intensity of the purchase process is varying and affected by many factors, we use function $\lambda(x)$ to capture the dynamic purchase behavior. We define $\lambda(x)$ as:

$$\lambda(x) = \sum_{d=0}^{D} w_d x^d + a \sin(bx + c) \quad (1)$$

with the restriction that $\lambda(x) \geq 0$ for any x. $\lambda(x)$ acts as an intensity function for purchase events, so higher $\lambda(x)$ values correspond to more frequent purchases, and vice versa. The polynomial component fits the trends of the purchase rate, with the degree of the polynomial component D tuned to the data. For the sine component, a, b, c are the amplitude, frequency and phase of the short-term periodic purchases.

The customers in a latent group k share the same $\lambda_k(x)$. The behavior of an individual customer could belong to multiple groups at the same time, so that u_i has soft membership π_{ik} in group k, and $\sum_{k=1}^{K} \pi_{ik} = 1$. We need to discover K latent groups and estimate the following parameters:

$\mathbf{\Theta}$: coefficients of $\lambda_k(x)$ for K groups. The coefficient θ_k for group k includes $\{w_{dk}\ (d \in \{1, 2, ..., D\}), a_k, b_k, c_k\}$.

$\mathbf{\Phi} \in [0, 1]^{K \times 1}$: relative sizes of K groups, where ϕ_k is the relative size of mixture component k, $\sum_{k=1}^{K} \phi_k = 1$.

$\mathbf{\Pi} \in [0, 1]^{U \times K}$: soft memberships over K groups for all the customers, where $\sum_{k=1}^{K} \pi_{ik} = 1$.

We use the Expectation Maximization (EM) algorithm [10] to construct a mixture of NHPP for the observations and infer the parameters iteratively. The algorithm input includes the number of groups K, the parametric form of $\lambda_k(x)$, and the observations of u_i's $N_i(T)$ purchase events $\boldsymbol{x}_i \in (0, T]^{1 \times N_i(T)}$, where the element x_{ij} is the time of u_i's j^{th} purchase. Then the algorithm starts from the E-step and iterates between the E- and M-steps until convergence.

In the **E-step**, we assign u_i to a group randomly or based on a predefined rule for the first iteration. From the second iteration, we use the estimation of Θ and Φ from the M-step in the previous iteration to update Π. The posterior probability of u_i in the latent group k is:

$$p(z_i = k | \boldsymbol{x}_i, \Theta, \Phi) = \frac{p(\boldsymbol{x}_i | z_i = k, \Theta) p(z_i = k, \Phi)}{\sum_{s=1}^{K} p(\boldsymbol{x}_i | z_i = s, \Theta) p(z_i = s, \Phi)} \quad (2)$$

where $z_i \in \{1, ..., K\}$ is the latent group variable for u_i. The likelihood of n ordered and independent observations at $\{x_{ij}\}$ ($j \in \{1, ..., n\}$) is [21]:

$$p(\boldsymbol{x}_i | z_i = k, \Theta) = \Big(\prod_{j=1}^{n} \lambda_k(x_{ij})\Big) e^{-\int_0^T \lambda_k(x) dx} \quad (3)$$

Particularly, $p(z_i = k | \boldsymbol{x}_i, \Theta, \Phi)$ in Equation 2 is u_i's soft membership in group k, which is also denoted by π_{ik}. When the group memberships for all the customers are successfully updated for the current iteration, the M-step starts.

In the **M-step**, we estimate the values of Θ and Φ based on daily purchase log $\{\Delta N(t)\}$ ($t \in \{1, ..., T\}$) and Π obtained in the E-step. For group k, the daily log $\{\Delta N_k(t)\}$ includes the purchase events of all customers in k. It is computed by $\Delta N_k(t) = \sum_{i=1}^{U} \pi_{ik} \Delta N_i(t)$, which considers the purchase event increment $\Delta N_i(t)$ of u_i on day t, and u_i's soft membership in k, π_{ik}. The $\{\Delta N_k(t)\}$ is used to estimate θ_k by maximizing the likelihood of $\lambda_k(x)$. Finally, we update the size of group k by summing individual soft memberships of group k, $\phi_k = \sum_{i=1}^{U} \pi_{ik}$. The algorithm iterates between the E- and M-steps until convergence and outputs the estimations of Φ, Π and Θ in the final iteration.

Construct Individual Purchase Rate Curve. Based on Φ, Π and Θ of the mixture NHPP models, we derive the individual purchase rate curve $\lambda_{u_i m}(x)$, which represents the estimated u_i's daily purchase rate of product m. The
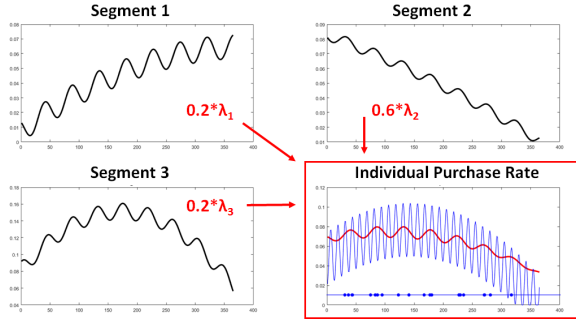
Figure 3: Using sparse individual records (blue circles) can cause overfitting problems (blue curve), while our purchase rate curve (red curve) formed by the weighted sum of segment-level models (black curves) is more reliable.

equation of $\lambda_{u_i}(x)$ (omitting m) is:

$$\lambda_{u_i}(x) = \sum_{k=1}^{K} \pi_{ik}\lambda_k(x) \quad (4)$$

which is a linear combination of purchase behavior models for K segments, weighted by u_i's soft membership in each segment. The polynomial component of $\lambda_{u_i}(x)$ can capture behavior patterns such as increase, decrease, U-shape and inverse U-shape, depending on the coefficients $\sum_{k=1}^{K} \pi_{ik}w_{dk}$. Consider the example in Figure 3, where there are 3 customer segments with the increase, decrease and inverse U-shape patterns respectively, and the purchase rate curve for a customer with soft membership $\pi_i = [0.2, 0.6, 0.2]$ in these segments is the red curve in the bottom right plot. Compared to the blue curve which fits sparse records (blue circles) directly, our model prevents overfitting.

Given $\lambda_{u_i m}(x)$, we know the customer preference on any day and can track the behavior changes. Most importantly, the derivative of $\lambda_{u_i}(x)$ reflects the rate of change of u_i's purchase behavior, and the integral of $\lambda_{u_i}(x)$ over a time period shows the estimated number of purchases u_i would make over that period.

3.2 Program Impact Evaluation

As mentioned above, $\lambda_{u_i}(x)$ allows to evaluate the program impact, so we utilize $\lambda_{u_i}(x)$ to answer three questions.

Q1: How do the purchase rates during the program differ from those observed before and after the program? We firstly obtain the mean purchase rates in three phases of the observation period: before, during and after the program. The mean purchase rate r_{ph} of phase ph is computed by:

$$r_{ph} = \frac{1}{t_2 - t_1 + 1} \int_{t_1}^{t_2} \lambda_{u_i}(x)dx \quad (5)$$

where t_1 and t_2 are the start and end time of ph. Then, we conduct statistical tests to check if the purchase rates across the three phases are significantly different. The test we use is the one-way ANOVA [23], which is an extension of the t-test to compare more than two groups based on one independent variable, the mean purchase rate in our case.

Q2: How does the purchase rate change during the program? To gain insights of the behavior changes during the program, we analyze the behavior patterns for different products and the distribution of various patterns among all customers. Specifically, the coefficients of $\sum_{d=0}^{D} w_d x^d$ in Equation 1 can capture different patterns of behavior changes. Setting $d = 2$, based on whether the parabola opens upward ($w_2 > 0$) or downward ($w_2 < 0$) and on the location of the turning point x_{tp} ($-w_1/2w_2$), we observe five long-term patterns: *increase, decrease, U-shape, inverse U-shape*, and *stable*. More details about the specific conditions used to determine different types of long-term patters are given in [19]. For any product, the distribution of these patterns can show the trend of the customer preferences and indicate the proportion of customers whose purchase behaviors have been influenced by the program.

Q3: Are the active users of the website more receptive to the program than the others? Being receptive to the program refers to increasing the purchase rate of fresh products. The main challenge is that the purchase behaviors could also be impacted by the other factors such as promotions and seasons, due to the variations of product availability and price among different seasons. Particularly, the purchase of the fresh products may be significantly affected by seasonality. For example, although the purchase rate of grapes may decrease in winter for most customers during the program, we desire to identify which customers have larger increases in their purchase rate compared to others. We evaluate the program impact in two steps.

In the first step, we use the derivative of $\lambda(x)$ to measure the rate of changes in the purchase rate curve, where the positive value means the preference is increasing at that time. To measure the increase of the purchase rate of a product during the program for u_i, we define:

$$\sigma_{u_i} = \sum_{x \in S} \lambda'(x), \ S = \{x | \lambda'(x) > 0, x \in \{t_1, t_1+1, ..., t_2\}\} \quad (6)$$

which is the sum of all positive $\lambda'(x)$ for u_i from t_1 to t_2. σ_{u_i} reflects both the duration and rate of the increases of the purchase behavior during the program.

In the second step, we rank all σ_{u_i} in a descending order, so that the high-ranked customers have a larger increase in their purchase rates. We examine if certain groups of customers, e.g. those who used the website frequently or those who lost more weight, are more receptive to the program. The group of customers we are interested in is labeled as the *target group* (i.e. treatment group), and we compare their ranks to other customers (i.e. control group). We quantify the *impact* of the program for the target group as:

$$impact = \frac{1}{U_{target}U} \sum_{\alpha=1}^{U} n_\alpha \quad (7)$$

where n_α is the number of customers from the target group and also in the set of top-α customers with the highest σ. $\sum_{\alpha=1}^{U} n_\alpha$ is the cumulative sum of the number of customers in the target group and who are also in top-α ($\alpha \in \{1, ..., U\}$), after sorting all customers by σ. U_{target} is the total number of customers in the target group, and U is the total number of customers. The baseline of *impact* is 0.5, and the greater *impact* value implies that the ranks of the target group customers are higher than the others, indicating that the program is more effective for the target group.
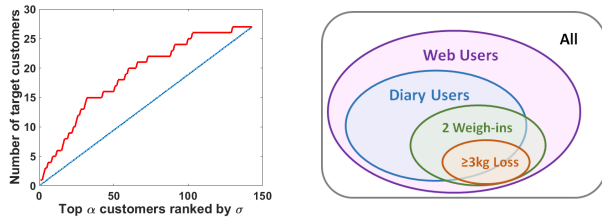
Figure 4: Left: illustration of the impact score; right: relationships of 4 partition methods

In Figure 4-left, the blue line corresponds to the baseline, whereas the red line shows the number of target group customers among top-α customers, and the area under the red line is the impact score.

In the analysis of purchase pattern distribution and impact scores, we partition customers into exclusive groups in four ways, considering the frequency and type of interactions with HealthierU. These are four illustrative examples and other ways may be considered. Our goal is to understand whether customers from different groups[3] have different purchase behaviors in fresh products and other categories. The four ways to partition the customers are:

PT1 *active vs. inactive web users*: if the customer used HealthierU 12 times or more during the program;

PT2 *active vs. inactive diary users*: if the customer wrote 12 diary entries or more during the program;

PT3 *2 vs. <2 valid weigh-ins*: if the customer reported weight at least twice and these were at least 6 weeks apart;

PT4 *≥3kg vs. <3kg weight loss*: if the customer with at least 2 valid weight reports lost ≥3kg.

These thresholds are set to generate comparable groups and prevent impact of the noise. The relationships among the customer groups formed by different partitions are illustrated in Figure 4-right. The specific sizes of customer groups, e.g. active diary users, are different for different products. The partitions enable us to compare the program impact on customers with different levels of interactions with HealthierU.

4. CASE STUDY

We select 38 popular products (listed in Table 1) based on the purchase records. Specifically, 21 products are fresh fruits and vegetables, and 17 are from other categories such as soft drinks and confectionery. The own brand refers to the supermarket's home brand.

The program started on May 26 and ran for 24 weeks. The participants were able to access the HealthierU website at any time during the program, and they were required to complete two surveys of basic information and health conditions. The first survey was before the start of the program and the second one was 12 week after the start, at the end of the first cycle. The participants who did not complete the second survey were not eligible for the second cycle, so the

---

[3]Note that partitioning based on the website usage and customer segmentation based on purchase behavior (as in Section 3.1) are different concepts. We use the term 'customer segment' for the segmentation result, and the term 'customer group' for the partitioning result.

program duration for them was 12 weeks. Among 931 participants, 198 customers attended both cycles of the program. For all the participants, the three phases in the analysis are:

1) *before* the program (from Jan 1 to May 25);
2) *during* the program (from May 26 to the end of program determined by the second survey completion);
3) *after* the program (from the end to Dec 31).

For each product, the active customers are those who bought products more than 10 times. The construction of NHPP model is based on the active customers of each product. In terms of the other parameters, the number of segments for each product is configured empirically based on the data fitness; the degree of polynomial component of Equation 1 is 2, which is adequate to capture the patterns within the one-year duration of the logs.

4.1 Purchase Rate Analysis

We conduct ANOVA tests to check if the mean purchase rates in the 3 phases – *before*, *during* and *after* the programs are different. The *p-values* and the mean purchase rates of all products in 3 phases are listed in the last four columns of Table 1. There are 12 products ($p < 0.05$ and names in bold) with statistically significantly different purchase rates across the 3 phases. Among these 12 products, 10 are from the Fruits & Vegetables category, 1 is Biscuits, and 1 is Confectionery. In more detail, 7 out of the 10 fresh products (*p-values* are underlined) have significantly higher mean purchase rates in Phase 2 than one or both of the other two phases, which shows the increase in preference towards these products during the program. For the other 3 fresh products, truss tomatoes, blueberries and grapes, the purchase rates in Phase 2 are not the largest, although rates of the 3 phases are significantly different. We notice that the purchase rates of these 3 products are strongly negatively correlated with price (Pearson's correlations are -0.28, -0.59 and -0.64, respectively), so the lower purchase rates in Phase 2 may be caused by the strong seasonal effects.

There are 7 out of 21 (33.3%) fresh products with significantly higher purchase rates in Phase 2, while this ratio is 2 out of 17 (11.8%) for other products. This partially demonstrates the impact of the program. However, the limitations of mean purchase rates are that they cannot demonstrate how the purchase behavior changes over time, and it is hard to distinguish the program impact from the other factors. Therefore, we investigate the pattern distributions and relative increase of purchase rates.

4.2 Distributions of Behavior Patterns

We investigate the pattern of individual purchase rate curve $\lambda_{u_i}(x)$, which allows us to track how the behavior changes during the program. For each user and product, we assign one of the five patterns – increase, decrease, U-shape, inverse U-shape or stable. Then, we divide customers by the four partitions PT1 – PT4 defined in Section 3.2 and explore the distributions of the five behavior patterns for different types of customers. The proportion of each pattern is an average over customers in a certain group. In order to check if the distributions of patterns are correlated to the categories of products, we also compute the average proportions of each pattern across fresh products and other products.

The distributions of the behavior patterns for various cases are shown in Table 2. There are four sections in the table, corresponding to the four partitions. Each customer group

Table 1: Product information with the estimated number of purchases for customer segments generated by NHPP models and ANOVA test results of the mean purchase rates in the 3 phases (before, during and after the program)

	Category	Product Name	Estimated N for Segments				p	Phase 1	Phase 2	Phase 3
			S1	S2	S3	S4				
1	Vegetables	Carrots	14.35	14.64	34.84	24.23	.102	0.0202	0.0223	0.0206
2	Vegetables	Cup Mushrooms	13.49	15.57	22.4	43.34	.16	0.0167	0.0169	0.0152
3	Vegetables	Red Onions	15.87	14.45	12.12	32.06	.583	0.015	0.0141	0.0145
4	Vegetables	**Sweet Potatoes**	50.21	14.66	17.09	29.03	**.022**	0.0124	0.0139	0.012
5	Vegetables	Red Capsicum	13.4	20.9	28.1	46.34	.184	0.024	0.0224	0.0218
6	Vegetables	**Broccoli**	12.86	17.98	29.8	41.48	**.000**	0.0201	0.0242	0.0199
7	Vegetables	Cucumbers	13.73	14.35	24.52	45.1	.181	0.0189	0.018	0.02
8	Vegetables	**Zucchini**	14.18	16.7	33.62	81.31	**.007**	0.016	0.0179	0.0146
9	Vegetables	**Field Tomatoes**	12.79	15.45	18.89	33.83	**.000**	0.014	0.0175	0.0155
10	Vegetables	Lettuce	14.44	12.53	21.48	39.28	.186	0.0133	0.0119	0.0126
11	Vegetables	**Truss Tomatoes**	13.91	14.15	29.48	57.41	**.001**	0.015	0.012	0.0132
12	Vegetables	Leb. Cucumbers	13.22	15.83	19.07	46.6	.067	0.014	0.0124	0.0119
13	Vegetables	Grape Tomatoes	13.44	49.81	21.27	112.09	.062	0.0138	0.0121	0.0115
14	Vegetables	Spring Onions	13.53	13.8	19.95	33.16	.617	0.0112	0.0107	0.0108
15	Vegetables	Mix Salad	18.22	15.99	13.99	41.03	.87	0.0088	0.0085	0.0086
16	Fruits	Apples	31.64	15.48	16.77	58.11	.051	0.0139	0.0162	0.0162
17	Fruits	**Bananas**	15.91	29.44	46.78	77.26	**.004**	0.0638	0.0701	0.062
18	Fruits	**Strawberries**	15.64	19.8	37.21	64.03	**.000**	0.0216	0.0296	0.0323
19	Fruits	**Avocado**	15.52	14.68	26.21	41.47	**.000**	0.0126	0.0285	0.0302
20	Fruits	**Blueberries**	28.24	12.85	14.45	43.07	**.000**	0.0104	0.0093	0.0144
21	Fruits	**Grapes**	12.28	12.76	14.26	18.36	**.000**	0.0158	0.0084	0.0092
22	Soft Drinks	Coca-cola	15.28	37.05	73.39	159.69	.976	0.0304	0.03	0.0299
23	Soft Drinks	own drinks	26.01	15.65	44.84	94.22	.517	0.0249	0.0228	0.0238
24	Soft Drinks	Schweppes	15.16	14.77	41.03		.127	0.0108	0.0103	0.0119
25	Biscuits	**Arnotts**	29.2	15.18	46.37	84.24	**.006**	0.0429	0.0478	0.0425
26	Biscuits	own biscuits	13.53	15.12	29.09	67.51	.158	0.0133	0.0149	0.014
27	Snacks	Smiths	15.04	15.04	32	58.51	.207	0.0246	0.0254	0.023
28	Snacks	own snacks	12.41	17.79	13.53	22.87	.832	0.0104	0.0107	0.0108
29	Snacks	Doritos	48.5	12.39	14.63		.555	0.009	0.0093	0.0096
30	Confectionery	**Cadbury**	13.25	17.22	33.05	75.57	**.018**	0.0187	0.022	0.0193
31	Confectionery	own confectionery	15.61	50.77	18.78		.679	0.0083	0.0087	0.0082
32	Confectionery	Mars	12.07	13.51	23.81		.127	0.0087	0.0097	0.0089
33	Yogurt	own yogurt	23.6	14.03	34.85	77.33	.164	0.0269	0.0294	0.0277
34	Yogurt	Yoplait	14.63	18.39	19.43	35.13	.264	0.0118	0.0124	0.011
35	Yogurt	Vaalia	16.23	18.22	45.47		.289	0.0096	0.0097	0.0085
36	Cereal	Kelloggs	14.83	13.39	17.34	32.08	.126	0.015	0.0147	0.0134
37	Cereal	Uncle Toby	20.8	12.94	14.47	35.84	.052	0.0108	0.0111	0.0096
38	Cereal	Sanitarium	13.88	14.38	19.7	37.5	.304	0.0128	0.0134	0.0123

has three rows, which show the distributions across different categories. The most dominant pattern for each case is in bold. For example, rows 1, 3 and 5 are for active web users, and they show average results over all the 38 products, fresh products and other products respectively.

Active vs. inactive web users (PT1). For these two groups, we notice that the average proportions of U-shape, inverse U-shape and stable patterns for all products are significantly different, while the proportions of increase and decrease patterns are similar. Specifically, the active web users have significantly less stable ($p < 0.05$), but more U-shape ($p < 0.001$) and inverse U-shape ($p < 0.01$) patterns than the inactive web users. This means that the purchase rates of active web users have more fluctuations during the program than of the other customers. After splitting the products into fresh and other categories, for the active web users, the amount of U-shape is 14.19% for fresh products, and 4.87% for other products. As for the inactive web users, the amount of U-shape is 6.25% for fresh products, and 1.09% for other products. For the active web users, the amounts

of the other four patterns between fresh and other products are close to each other, with the differences between 0.75% and 4.15%. The increase in U-shape of fresh products for both groups could be caused by the seasonal effect, as the purchase rates of fresh products may first decrease when the program started in winter and increase when the program impact became stronger than the seasonal effect.

Active vs. inactive diary users (PT2). The amount of inverse U-shape pattern for the active diary users is 20.44%, which is significantly higher than the inactive diary users, 8.13% ($p < 0.001$). For the active diary users, the amount of stable pattern is 35.80%, which is about 20% lower than for the inactive diary users, 54.04% ($p < 0.01$). The differences between the two groups in other patterns are not significant. For the active diary users, the proportions of first 4 patterns for fresh products range from 15% to 18%, whereas the amount of stable pattern is 34%. Similar to the active web users, the active diary users also have significantly more U-shape patterns for fresh products (15.35%) than other products (6.09%). Except for the U-shape pat-

Table 2: Distributions of the 5 individual purchase behavior patterns during the program

Partition	Products	Customer Group	Increase	Decrease	U-shape	Inverse U	Stable
PT1	all	active web users	17.12%	15.85%	10.02%	18.54%	**38.47%**
		inactive web users	17.64%	14.91%	3.94%	4.90%	**58.61%**
	fresh	active web users	15.60%	14.00%	14.19%	18.08%	**38.14%**
		inactive web users	18.91%	13.52%	6.25%	4.59%	**56.72%**
	others	active web users	19.00%	18.15%	4.87%	19.10%	**38.89%**
		inactive web users	16.07%	16.62%	1.09%	5.28%	**60.94%**
PT2	all	active diary users	16.90%	15.65%	11.21%	20.44%	**35.80%**
		inactive diary users	17.42%	15.22%	5.20%	8.13%	**54.04%**
	fresh	active diary users	16.06%	15.33%	15.35%	18.51%	**34.75%**
		inactive diary users	17.77%	13.15%	8.04%	8.31%	**52.73%**
	others	active diary users	17.94%	16.03%	6.09%	22.83%	**37.11%**
		inactive diary users	16.98%	17.78%	1.68%	7.90%	**55.65%**
PT3	all	2 valid weigh-ins	19.81%	15.04%	15.81%	**36.39%**	12.95%
		<2 valid weigh-ins	17.17%	15.42%	5.73%	9.05%	**52.64%**
	fresh	2 valid weigh-ins	12.16%	14.06%	21.80%	**35.60%**	16.39%
		<2 valid weigh-ins	17.83%	13.66%	8.58%	8.79%	**51.14%**
	others	2 valid weigh-ins	29.26%	16.25%	8.40%	**37.37%**	8.71%
		<2 valid weigh-ins	16.35%	17.59%	2.21%	9.36%	**54.49%**
PT4	all	≥3kg weight loss	18.89%	14.75%	14.82%	**43.64%**	7.90%
		<3kg weight loss	19.55%	14.99%	16.73%	**30.01%**	16.09%
	fresh	≥3kg weight loss	10.20%	15.12%	21.38%	**42.91%**	10.39%
		<3kg weight loss	13.72%	13.04%	21.33%	**32.10%**	19.82%
	others	≥3kg weight loss	29.62%	14.29%	6.72%	**44.55%**	4.82%
		<3kg weight loss	28.44%	18.49%	11.74%	**29.14%**	12.20%

tern, the differences between fresh and other products for either customer group (e.g. comparing row 3 with 5 or comparing row 4 with 6 of the second section in Table 2) are less than 5%. This shows that the difference between fresh and other products for one customer group is less pronounced than the difference between the two customer groups. The possible reasons include that customers may buy different products together and shop for the family when they visit supermarket.

2 vs. <2 valid weigh-ins (PT3). The customers with 2 valid weigh-ins have significantly more U-shape ($p < 0.05$), inverse U-shape ($p < 0.001$), and less stable patterns ($p < 0.001$) than other customers. The customers with 2 valid weigh-ins have larger amounts of inverse U-shape on both fresh and other products than their counterparts, which shows that they were motivated by the program at the start, but about 36% of them did not persist until the end of the program. We notice that the gaps between two customer groups for this partition are larger than the previous two partitions. For example, the difference of the proportions of the stable pattern for two groups is about 40%, which doubles the differences between the active and inactive web users. This is a positive result, showing that the purchase behavior of customers with or without 2 valid weigh-ins are more distinguishable. It indicates that the program impacts the customers with higher levels of engagement, such as those using weight trackers, stronger than others.

≥3kg vs. <3kg weight loss (PT4). The groups formed by PT4 are: ≥3kg, <3kg weight loss, all having 2 or more valid weigh-ins. The proportions of all patterns between these two groups are not significantly different. The most dominant pattern for all cases in this section is inverse U-shape. The customers who lost ≥3kg only have 7.9% stable pattern but 43.64% inverse U-shape. On average, they have 13% more inverse U-shape than customers who lost <3kg. Compared to customers with <2 valid weigh-ins in PT3,

both groups in PT4 have significantly more inverse U-shape and less stable patterns. The results mean that the program is more influential for customers who reported weigh-ins at least twice, but whether they lost greater or less than 3kg makes smaller differences.

4.3 Program Impact Scores

In addition to the pattern distributions, we also use the impact score (defined in Section 3.2) to evaluate if the program is more effective for a certain group of customers. The impact score is designed to compare increase in the purchase rates of customer groups. We use partition criteria PT1 – PT4 again to group the customers. The target groups are: 1) active web users, 2) active diary users, 3) customers with 2 valid weigh-ins and 4) customers with ≥3kg weight loss. The other groups are the control groups. For each product, the higher impact score is preferred, which means the customers from the target group have larger increases in purchase rates during the program than other customers. For example, 0.6 means 20% impact lift from the baseline 0.5. Therefore, we rank impact scores across all products for each partition method, and present the results of top 20 products in Table 3. Each section in the table includes the product names (fruits and vegetables in bold) and the impact scores of top products for that partition method.

When the target group is the active web users (PT1), the scores of top 20 products are between 0.53 and 0.62. There are 12 fresh products in the top 20 list, and 4 of them are fruits: grapes, strawberries, blueberries and apples. The proportion of fruits among top 20 scores is higher than other categories, given there are 6 fruits involved in the study. The proportion of fresh products appearing among top 20 is 57% (12 out of 21), while it is 47% (8 out of 17) for other products. For the active diary users (PT2), there are also 12 fresh products among the top 20 impact scores. Truss tomatoes rank first, with 40% lift from the baseline. The scores for

Table 3: Products with top 20 program impact scores for different partitions

	active web users		active diary users		2 valid weigh-ins		≥ 3kg weight loss	
	Product	Imp.	Product	Imp.	Product	Imp.	Product	Imp.
1	**Grapes**	0.62	**Truss Tomatoes**	0.70	Doritos	0.95	Yoplait	0.99
2	Mars	0.62	**Spring Onions**	0.68	**Truss Tomatoes**	0.80	**Spring Onions**	0.99
3	**Truss Tomatoes**	0.61	Uncle Tobys	0.68	Schweppes	0.80	Doritos	0.93
4	**Spring Onions**	0.60	Yoplait	0.65	**Strawberries**	0.79	Schweppes	0.89
5	**Strawberries**	0.58	Mars	0.65	Yoplait	0.79	**Truss Tomatoes**	0.88
6	Doritos	0.57	Coca-cola	0.64	**Grapes**	0.79	**Strawberries**	0.86
7	Yoplait	0.57	**Grapes**	0.63	own drink	0.78	**Blueberries**	0.79
8	Uncle Tobys	0.56	Doritos	0.62	Coca-cola	0.76	**Red Capsicum**	0.78
9	**Red Capsicum**	0.56	own drink	0.60	Uncle Tobys	0.76	**Grape Tomatoes**	0.78
10	Coca-cola	0.56	**Cucumbers**	0.59	**Spring Onions**	0.73	Uncle Tobys	0.78
11	**Cucumbers**	0.56	**Leb. Cucumbers**	0.58	**Lettuce**	0.70	own drink	0.78
12	own drink	0.56	**Strawberries**	0.58	own confectionery	0.70	**Grapes**	0.75
13	Schweppes	0.55	**Blueberries**	0.57	Vaalia	0.69	Vaalia	0.73
14	**Lettuce**	0.55	**Red Capsicum**	0.57	**Red Capsicum**	0.68	**Cucumbers**	0.70
15	**Blueberries**	0.55	**Field Tomatoes**	0.57	**Grape Tomatoes**	0.68	**Broccoli**	0.67
16	own snacks	0.54	Smiths	0.56	**Cucumbers**	0.67	**Red Onions**	0.67
17	**Leb. Cucumbers**	0.54	**Carrots**	0.56	Smiths	0.65	own confectionery	0.66
18	**Mix Salad**	0.53	**Cup Mushrooms**	0.54	**Red Onions**	0.64	Kelloggs	0.66
19	**Broccoli**	0.53	**Lettuce**	0.54	**Blueberries**	0.63	Smiths	0.65
20	**Apples**	0.53	Schweppes	0.54	**Apples**	0.63	**Lettuce**	0.64

the active diary users are between 0.54 and 0.7, which are higher than the scores for the active web users. This means the increase of purchase rates of active diary users are larger than of their active web users counterparts. For the third section (PT3), since the number of customers reported 2 valid weigh-ins is much lower than the number of active web or diary users, the target group is smaller in this case. The increases of purchase rates of the target group are more distinguishable from others than the previous two cases, so the top 10 scores are all higher than 0.7, and the range of top 20 scores is from 0.63 to 0.95. As for the customers with ≥3kg weight loss (PT4), the top scores are even higher, with 6 scores greater than 0.85. The number of fresh products in top 20 list is 11, which is close to the other cases.

We find that products from the other categories, such as Uncle Tobys from Cereal, Yoplait from Yogurt and Doritos from Snacks, also have high scores in top 20 lists for all four partitions. Although the program did not have discounts on them, this should not be interpreted as negative results. The main reason is that customers may buy them together with fresh products when they visit the supermarket [14], so we could see the purchase rate changes with fresh products simultaneously. This has also been discussed in Section 4.2, as we find the differences of pattern distributions for fresh and other products for the active web users are small.

It is also worth noting that grapes, truss tomatoes and blueberries have significantly lower mean purchase rates during the program (Phase 2) than the other phases as discussed in Section 4.1, but their impact scores appear in top 20 lists for all four partitions. Especially, for customers with ≥3kg weight loss, the scores are 0.88, 0.79 and 0.75 for these three products. The high scores for these products show that the target customers have larger increases in the purchase rates than other customers, even though the overall purchase rates are lower at that time of the year. As discussed, the lower absolute purchase rates could be caused by seasonality or other factors, while the higher increases in the purchase rates indicate that the target customers are more receptive to the program than other customers. This shows the advantage

of using the impact score, when it is hard to isolate the influence of program from the other factors.

In our future work, it would explore the causality between the purchase behaviors and the program; this would require transaction records of customers who had similar pre-intervention purchase behaviors but did not attend the program.

5. CONCLUSIONS

We propose a method for evaluating the impact of a supermarket health program by tracking individual behavior changes and conducting a fine-grained analysis across groups of customers and products. Our method allows to uncover hidden dependencies and impacts, which may be overlooked by the traditional overall analysis methods without behavioral tracking. To better analyze the impact of the health program, we use both purchase data and usage logs from an associated website which encourages behavior changes.

Since the individual purchase records are sparse and noisy, we construct accurate individual purchase rate curves based on a mixture of segment-level NHPP models, which facilitates the evaluation of the program impact. We also design four criteria for partitioning the customers and compare the program impact on customers with different interactions with the program website. The key findings from our case study on an Australian supermarket health program are:

1) For 7 fresh products and 2 other products, the average purchase rates during the program are significantly higher than before or after the program.

2) The active web users, active diary users, customers with 2 valid weigh-ins and customers with ≥3kg weight loss have been motivated to a greater extent to change purchase behaviors. They have significantly less of the stable pattern, and more of the inverse U-shape pattern than others. Based on the impact scores, they have larger increases in the purchase rates for some fresh products, especially fruits.

3) Among the four partition methods, when the criterion considers higher levels of interactions such as using the

weight tracker (PT3 and PT4), the program impact is more pronounced on the target groups than the rest. The customers who reported at least 2 valid weigh-ins and who lost more than 3kg achieved larger behavior changes than others.

The results show the importance of the program website, which provides interactive and personalized tools and motivates behavior changes. The interactions with the participants and impact of the program can be improved by using mobile applications, e.g. for food diary and tracking weight loss. Our method can be used by health program designers to understand the behavior of the participants, increase the participant engagement, target different types of participants and improve future health programs.

6. REFERENCES

[1] P. Adamopoulos and V. Todri. The effectiveness of marketing strategies in social media: Evidence from promotional events. In *Proceedings of the 21th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pages 1641–1650. ACM, 2015.

[2] G. M. Allenby, R. P. Leone, and L. Jen. A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, 94(446):365–374, 1999.

[3] K. Ball, S. A. McNaughton, H. N. Le, L. Gold, C. N. Mhurchu, G. Abbott, C. Pollard, and D. Crawford. Influence of price discounts and skill-building strategies on purchase and consumption of healthy food and beverages: outcomes of the supermarket healthy eating for life randomized controlled trial. *The American Journal of Clinical Nutrition*, 101(5):1055–1064, 2015.

[4] S. Berkovsky, G. Hendrie, J. Freyne, M. Noakes, and K. Usic. The healthieru portal for supporting behaviour change and diet programs. In *Proc. of 23rd Australian National Health Informatics Conf.*, volume 214, page 15. IOS Press, 2015.

[5] E. Brindal, J. Freyne, I. Saunders, S. Berkovsky, G. Smith, and M. Noakes. Features predicting weight loss in overweight or obese participants in a web-based intervention: randomized trial. *Journal of medical Internet research*, 14(6):e173, 2012.

[6] E. Brindal, G. Hendrie, J. Freyne, M. Coombe, S. Berkovsky, and M. Noakes. Design and pilot results of a mobile phone weight-loss application for women starting a meal replacement programme. *Journal of Telemedicine and Telecare*, 19(3):166–174, 2013.

[7] M.-C. Chen, A.-L. Chiu, and H.-H. Chang. Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*, 28(4):773–781, 2005.

[8] W. P. Dickinson, R. E. Glasgow, L. Fisher, L. M. Dickinson, S. M. Christensen, P. A. Estabrooks, and B. F. Miller. Use of a website to accomplish health behavior change: if you build it, will they come? and will it work if they do? *The Journal of the American Board of Family Medicine*, 26(2):168–176, 2013.

[9] A. S. Ehrenberg. The pattern of consumer purchases. *Applied Statistics*, pages 26–41, 1959.

[10] D. Karlis. EM algorithm for mixed poisson and other discrete distributions. *ASTIN Bulletin*, 35(01):3–24, 2005.

[11] H. Kim, N. Takaya, and H. Sawada. Tracking temporal dynamics of purchase decisions via hierarchical time-rescaling model. In *Proc. of the 23rd ACM Int'l Conf. on Information and Knowledge Management*, pages 1389–1398. ACM, 2014.

[12] K. Kopperschmidt and W. Stute. Purchase timing models in marketing: a review. *AStA Advances in Statistical Analysis*, 93(2):123–149, 2009.

[13] Y. Koren. Collaborative filtering with temporal dynamics. *Communications of the ACM*, 53(4):89–97, 2010.

[14] J. S. Larson, E. T. Bradlow, and P. S. Fader. An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing*, 22(4):395 – 414, 2005.

[15] B. Li, X. Zhu, R. Li, and C. Zhang. Rating knowledge sharing in cross-domain collaborative filtering. *IEEE Transactions on Cybernetics*, 45(5):1054–1068, 2015.

[16] B. Li, X. Zhu, R. Li, C. Zhang, X. Xue, and X. Wu. Cross-domain collaborative filtering over time. In *Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence*, pages 2293–2298. AAAI Press, 2011.

[17] R. Li, B. Li, C. Jin, X. Xue, and X. Zhu. Tracking user-preference varying speed in collaborative filtering. In *Proc. of the 25th AAAI Conf. on Artificial Intelligence*, 2011.

[18] B. Liu, W. Hsu, H.-S. Han, and Y. Xia. Mining changes for real-life applications. In *Proc. of the 2nd Int'l Conf. on Data Warehousing and Knowledge Discovery*, volume 1874, page 337. Springer, 2000.

[19] L. Luo, B. Li, I. Koprinska, S. Berkovsky, and F. Chen. Discovering temporal purchase patterns with different responses to promotions. In *Proc. of the 25th ACM Int'l Conf. on Information and Knowledge Management*. ACM, 2016.

[20] L. Luo, B. Li, I. Koprinska, S. Berkovsky, and F. Chen. Who will be affected by supermarket health programs? Tracking customer behavior changes via preference modeling. In *Pacific Asia Conference on Knowledge Discovery and Data Mining*, pages 527–539. Springer, 2016.

[21] S. M. Ross. *Stochastic processes*, volume 2. John Wiley & Sons New York, 1996.

[22] G. Trinh, C. Rungie, M. Wright, C. Driesener, and J. Dawes. Predicting future purchases with the poisson log-normal model. *Marketing Letters*, 25(2):219–234, 2014.

[23] J. R. Turner and J. Thayer. *Introduction to Analysis of Variance: Design, Analyis & Interpretation: Design, Analyis & Interpretation*. Sage, 2001.

[24] L. S. Wieland, L. Falzon, C. N. Sciamanna, K. J. Trudeau, S. Brodney, J. E. Schwartz, and K. W. Davidson. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. *Cochrane Database Systematic Reviews*, 8(8), 2012.

[25] WorldHealthOrganization. World health organization factsheet: obesity and overweight. *http://www.who.int/mediacentre/factsheets/fs311/en/*, 2015.