Social Desirability Bias and Engagement in Systems Designed for Long-Term Health Tracking

A dissertation presented

by

Laura M. Vardoulakis

to the Faculty of the Graduate School
of the College of Computer and Information Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Northeastern University
Boston, Massachusetts
March, 2013

NORTHEASTERN UNIVERSITY
GRADUATE SCHOOL OF COMPUTER SCIENCE
Ph.D. THESIS APPROVAL FORM

THESIS TITLE: Social Desirability bias + engagement in systems designed for long-term health tracking

AUTHOR: Laura Vardoulakis

Ph.D. Thesis Approved to complete all degree requirements for the Ph.D. Degree in Computer Science.

Thesis Advisor	4/8/13 *Date*
Thesis Reader	March 28, 2013 *Date*
Thesis Reader	March 28, 2013 *Date*
Thesis Reader	March 28/2013 *Date*

GRADUATE SCHOOL APPROVAL:

Director, Graduate School	4/8/13 *Date*

COPY RECEIVED IN GRADUATE SCHOOL OFFICE:

Recipient's Signature	*Date*

Distribution: Once completed, this form should be scanned and attached to the front of the electronic dissertation document (page 1). An electronic version of the document can then be uploaded to the Northeastern University-UMI website.

Abstract

In the coming years, remote health monitoring is an area that is expected to grow significantly. Systems designed to follow-up with patients at home can be used not only to reduce visits to the doctor but also to augment the face-to-face interactions between patients and physicians. These systems could also provide much-needed care to the millions of people living in rural areas.

While many researchers are investigating remote sensing technologies, the use of self-report in technological systems for long-term health monitoring remains a relatively understudied area. In this thesis, we investigate two main challenges in building systems designed for the collection of self-reported health data: 1) maximizing the accuracy of the reported data, and 2) maintaining user engagement with the system over potentially long periods of time.

We describe results from three field trials of systems designed to collect self-reported health data. Results indicate that personified interfaces and designs that include personalized health messages may negatively impact data quality. Results also indicated that, despite incentives designed to promote use, the time commitment needed to interact with the system predicts the likelihood of continued use.

Acknowledgments

Many people have supported me throughout the PhD process and I would like to express my sincere gratitude for their guidance and encouragement.

I want to thank my advisor, Tim Bickmore, for introducing me to the world of research in patient-facing health technologies, as well as Stephen Intille, Harriet Fell, and Joseph Kvedar for being on my PhD committee and providing valuable feedback on my work.

My lab mates have assisted me in countless ways over the years. Daniel Schulman provided an incredible soundboard for research ideas. Lazlo Ring helped keep The Drinks Project website up and running on our server. And Juan Fernandez was an instrumental engineering partner in the Virtual Nurse and MyLink2Care projects.

Donna Byron, Barbara Barry and Candy Sidner are inspiring researchers and I learned so much from their work and our collaborations. Agnes Chan, Krissy Mainello and Ashley Kline were always encouraging. Thank you!

At Boston Medical Canter, Brian Jack and Michael Paasche-Orlow helped me gain invaluable experience with clinical research and for that I am forever grateful. A huge number of people from BMC contributed to the Virtual Nurse and MyLink2Care projects and assisted me with experiments. I'm especially thankful for the work by Mary Goodwin, Lynn Schipelliti, Gail Sanchez, Anna Johnson, Julie O'Donnell, Shaula Forsythe, Lindsey Hollister, Caroline Hesko, Sarah Waite, Stephanie Tiwari, Megan Hempstead and Jessica Martin.

Working at Microsoft Research during the summer of 2011 was a thrilling experience. In those twelve weeks, Amy Karlson, Dan Morris, Greg Smith, and Desney Tan taught me more about research than I ever expected. Our collaboration with Justin Gatewood and field trial at Washington Hospital Center was some of the most intense and rewarding research I have done. I will never forget it.

A special thanks goes to my Mom, Dan, Emily and extended family for supporting me. Visits home always lifted my spirits.

Finally, I thank my husband Dimitris for his love and support. None of this would have been possible without it.

Contents

List of Figures

List of Tables

Introduction

Remote health monitoring is a growing field, especially with the rise of chronic disease management [47]. With primary care offices overloaded, patients are bearing a greater responsibility to manage their own care [9, 57]. Technologies to monitor and assist patients at home can not only reduce visits to the doctor but also help patients become active participants in their care. These systems could also provide much-needed care to the millions of people living in rural areas.

Sensor-based technologies such as internet-connected scales and blood-pressure monitors are valuable devices for collaboratively tracking a patient's health over time. Often, self-reported data from patients is not desirable, as the data may be biased to an unknown extent. Self-reported data is often subject to social desirability biases, the desire to present ourselves in a positive light [18]. Despite potential for biases, for many conditions there exists no sensor-based approach to tracking, and the best way to ascertain how a patient is doing is simply to ask them. For example, pain levels and most medication side effects are extremely difficult to assess via a sensor. Furthermore, psychological states such as a patient's attitudes, beliefs, and intentions about their health behavior may never be measureable by sensing technology.

The use of technology to collect longitudinal, self-reported health data is a relatively understudied area. In fact, longitudinal studies in the field of Human-Computer Interaction (HCI) are rare. HCI generally focuses on how people interact with a particular system at a single point in time. Without research exploring how users' interactions with computer systems might change over time, it is difficult to build systems that are specifically designed for longitudinal use.

In this dissertation, we focus on the basic research needed to provide a foundation for designing long-term, patient-facing systems for self-reported health tracking. In particular, we explore two main challenges: 1) How do we design a system to maximize the *quality* of the self-reported data? And 2) How do we keep people *engaged* with such a system, over potentially long periods of time?

Through several observational studies and experiments, we explore these challenges and provide new insight into systems designed for long-term health tracking. This thesis provides novel contributions to both HCI and health informatics.

- We describe the design and implementation of a home-based system for post-hospitalization follow-up, designed to track patients' wellness from their time of hospital discharge to a follow-up appointment with their primary care doctor. We discuss results from an observational study of patients using the system, and provide first-hand accounts from patients about how they would use such a system at home.
- In examining the quality of self-reported health data, we describe the first experiment to examine how social responses to computers change over time. By examining repeated interactions with a system, we now have insight into how a well-studied phenomenon, the Computers as Social Actors paradigm [54], extends into a longitudinal context. There is now evidence that social responses to computers are not static, but change with time. Furthermore, this change can vary, based on the personification of the interface.

- In an experiment with 375 participants spanning four months, we explore how interface designs and incentives for use can impact both data quality and system engagement. We show that despite incentives for use, *interaction time* is the best predictor of repeat system interactions. Shorter first-time interactions predict an increased number of interactions over the following weeks. We also confirm our previous findings that personification of the interface can lead to social desirability biases. Finally, we show that providing personalized health data at the end of the interaction, designed for feedback and reflection, actually *decreased* data quality and predicted higher levels of social desirability effects.

1.1 Dissertation Outline

This dissertation is organized as follows. In Chapter 2, we discuss related work and review background information and theoretical concepts that will be used in later chapters.

In Chapter 3, we describe our collaboration with medical researchers to design, engineer, and evaluate a patient-facing system for post-hospitalization follow-up. We report results from two observational studies and a field trial with recently discharged hospital patients.

In Chapter 4, we explore the accuracy of daily, self-reported, exercise data in six-week long field experiment. This experiment is the first to evaluate how the Computer as Social Actors paradigm extends into a longitudinal context. We explore how personification of the interface may impact data quality, and find that social responses to computers become *stronger* over time for highly personified interfaces and *weaker* over time for non-personified interface designs.

In Chapter 5, we examine how interface designs and incentives can promote data quality and user engagement with self-report systems over time. In an experiment with 375 participants over four months, we explore weekly self-reports of alcohol consumption. We confirm our previous finding that interface personification can lead to greater biases in the self-reported data. Furthermore we find that providing information for self-reflection (charts/graphs based on the self-reported data) predicts *higher* amounts of social desirability biases, not less.

In Chapter 6, we conclude the dissertation and discuss directions for future research.

CHAPTER 2

Background and Related Work

2.1 Embodied Conversational Agents

Embodied conversational agents (ECA's) are animated humanoid computer characters that simulate face-to-face conversation with users [12]. These agents have three qualities that could positively impact longitudinal health interviews. First, they are easy to use, requiring no prior computer experience [4]. Second, they are based on computational models of natural human behavior, allowing for rapport-building and empathy, which may be key to maintaining long-term engagement. Third, ECAs can provide health information that is adapted to the particular needs of a user, and provide an environment where users are free to take as much time as they need to thoroughly understand the information being discussed.

Several ECAs have been used within the field of health e.g., [4, 5], including one that we designed specifically to spend time with patients at the end of their hospital stay and educate them about their discharge instructions. This system was initially evaluated by 19 hospital patients and results showed high levels of trust and satisfaction with the system [6]. Additionally, 14 (74%) of the participants indicated that they preferred hearing their discharge information from the ECA rather than from their doctor. Primary reasons for satisfaction with the system had to do with the system's ability to communicate at a relaxed pace, re-explain information when needed, and provide the patient with as much time and attention as necessary.

The ECAs used throughout the experiments in this dissertation follow the interface shown in Figure 2-1. The dialogue for the agent is created through scripted hierarchical transition networks, and spoken via a text-to-speech engine[1]. At each turn of the dialogue, an output of responses is shown to the participant via large buttons, which they can select via a mouse or touchscreen. Corresponding gestures for the ECA, such as head nods, gaze-aways, or facial displays, are automatically created via BEAT, a toolkit that performs a context and linguistic analysis of the agent's scripted dialogue, and generates nonverbal behavior based on research of human conversation [14]. The agent also has the ability to hold and gesture at artifacts, as shown in Figure 2-1 (b).

Figure 2-1: ECA interface

[1] http://www.loquendo.com/

2.2 Social Desirability Bias and Self-Report

Initially, it may seem as though self-report is an uncomplicated, straightforward method to obtain data, but in reality the process can invoke deep cognitive and emotional processes [42, 43, 68]. In order to respond to a question accurately (at the most basic level), a person must have a *desire* to answer the question, *understand* the question being asked, be able to *recall* necessary information from memory, and *format* their answer [71]. When answering behavioral frequency questions, recalling exact frequency counts is often impossible and respondents must instead utilize a variety of estimation and inference techniques in order to provide an answer [51].

A potential disadvantage with using face-to-face conversation or ECAs for self-reported data collection is the possibility for *social desirability bias effects*: the tendency of an interviewee or questionnaire respondent to put themselves in a favorable light with respect to social norms [18]. This phenomenon is rooted in theory regarding the presentation of one's public self to others [31]. For example, when asked about the frequency of participating in particular behaviors people might under-report socially undesirable behaviors (drug use, illegal activity, etc) and over-report desirable behaviors (voting, charitable activities, etc), depending upon the context in which the data is collected. As one example of these effects, a study examining patients affected by eating disorders found that a paper questionnaire gathered higher quality data regarding patient behaviors than a face-to-face interview with a clinician [22].

In order to reduce the effects of social desirability bias when collecting self-reported data, the most common approach is to make the data collection mechanism as anonymous as possible. For example, anonymous computerized interviews have been shown to collect higher quality data than face-to-face interviews [11, 56].

2.3 Technologies for Health Tracking

Though technologies designed to collect long-term health data is a relatively understudied area, the area of *self-tracking* has grown significantly in recent years. Organizations such as the Quantified Self[2] and research in Personal Informatics [46] have contributed to this growth. A recent report shows that 69% of U.S adults track some health indicator, for either themselves or a loved one [29]. However, most of this tracking is done informally, and only 20% of trackers use technology such as a spreadsheet or app to record their data.

Self-tracking technology often uses Experience Sampling Method (ESM) [44] to collect data. This method prompts for data entry in real-time, possibly several times a day. Smartphone apps often follow this method, and 19% of U.S smartphone owners have at least one health app on their phone [28], though it is unclear how often or how long these apps are used.

2.4 Computers as Social Actors

We know that the main technique for reducing social desirability bias effects is to make the data collection as anonymous as possible, though it is not known whether people will respond to ECAs. People may react to ECAs as if they are computers, providing feelings of anonymity and give high quality data, or as if the ECAs are human and

[2] http://quantifiedself.com/

provide data that may be skewed, presenting themselves in a more favorable light. Cassell and Miller warn that ECAs may be subjected to the effects of social desirability bias [13].

For almost 20 years now, dozens of experiments have examined how people react socially to computers, using a methodology developed by Nass, et al. [63]. The experimental procedure takes a known finding from social science regarding behavior or attitudes toward humans and tests if the social rule also applies to behavior or attitudes towards computers. For example, when people are asked to evaluate a computer they give more positive evaluations when they complete the evaluation on the *same* computer that they are evaluating, and give less polite evaluations if they complete the evaluation on an *entirely different* computer than the computer being evaluated [55]. These experiments have consistently shown that people respond socially to computers, in ways that are similar to people respond socially to other people [63].

This Computers as Social Actors paradigm would suggest that any computer interface would be subject to social desirability bias effects and indeed, one study has found that ECAs were subject to such effects [38].

2.5 Home Health Monitoring Technologies

Automated telephony systems have been used for many years to interview patients about their health [30]. These systems utilize Interactive Voice Response (IVR) technology to allow patients to conduct a simulated conversation, responding to the system using either speech or DTMF (touch-tone) input. A downside of these systems is that they place a large amount of cognitive load on users. At each turn of the conversation, users must remember the list of acceptable responses given by the system, which can often lead to confusion and frustration.

Home-based devices and sensors have also been used by patients to track and report their health status. These devices can range from internet-connected scales and blood-pressure monitors, to systems such as the Health Buddy in which patients answer a series of daily health questions that are automatically reported to a case manager for review [16]. More advanced devices, such as the LifeShirt, incorporate sensors into clothing to create a wearable device that monitors the vital signs of patients during their day-to-day activities [32]. Unfortunately, many of these systems can be prohibitively expensive, and lack long-term empirical evaluations on their effectiveness.

2.6 Engagement with Technology

While building systems that can elicit accurate self-reported health data is a critical first step, it is only the beginning. These health systems must be designed in such a way that people desire to *use* them. This is especially important to consider when a primary goal of the interface is longitudinal use. Researchers within Human-Computer Interaction often refer to this as building *engaging* user interfaces. Engagement holds many definitions within the field, including presence [50], involvement [20, 23] attention [53], connection [65] and system use [10, 15].

Related to engagement is the notion of technology acceptance. The Technology Acceptance Model (TAM) suggests that perceptions of system *usefulness* and *ease of use* determine the intention to use the system, which mediates actual system use [19]. This model focuses on uptake of prescribed technology, but does not specifically examine factors that lead to long-term engagement.

The notion of triggers, context cues, or prompts is often recognized as improving technological engagement [25]. As an example, the notification system built into Facebook is powerful trigger to persuade its users to frequently visit the social network [24].

In two longitudinal studies, engagement with an ECA system was improved by incorporating dynamicity into interactions. For example, participants that experienced conversational variability among the daily interactions, and participants that interacted with an agent that told first-person background stories about itself (vs. third person stories) reported a greater desire to continue using the system [8].

In this dissertation, we will focus on the *system use* aspect of engagement. In Chapter 5 we will draw from prior work, by examining various forms of reminder messages in study emails sent to participants.

2.7 Social Exchange Theory

Social Exchange Theory asserts that the social interactions and relationships between people may be considered as an exchange of goods [39]. The idea is that humans interact with fellow humans based on the calculated costs and benefits associated with those interactions. Not only does the theory describe how relationships are formed, but also makes predictions regarding the longitudinal commitment in human relationships, and when those relationships may dissolve. The theory consists of several propositions [40], including:

> *The success proposition:* for all actions taken by persons, the more often a particular action of a person is rewarded, the more likely the person is to perform that action.

> *The value proposition:* the more valuable to a person is the result of her action, the more likely she is to perform the action.

> *Rationality proposition:* in choosing between alternative actions, a person will choose that one for which, as perceived by her at the time, the value, V of the result, multiplied by the probability, p, of getting the result, is the greater.

Although social exchange theory has not been applied to the interactions between humans and computers, the approach may be a springboard for examining user engagement with computer systems. For example, a person may have several motivations for using a particular computer system, such as monetary incentives (e.g. a stock-trading application or ad-sense for their website), or personal incentives (e.g. a pedometer to track their walking behavior, or using Wikipedia for knowledge gain). If the value of the system is high to the user (the result of the exchange leads to a net benefit for the user), it follows that they will be likely to continue using the system (*the success proposition)*. As the value decreases or remains static, and/or if better alternatives appear, it also follows that the user will become less engaged with the current system and may switch to an alternative choice (*rationality proposition*). Researchers have proposed a similar notion, one of evaluating interfaces based on the utility that they provide [70].

In Chapter 5, we will evaluate and compare the effects of both monetary incentives and personalized feedback data, on data quality and system use over time.

2.8 Behavioral Economics

Behavioral Economics examines how people make economic decisions, and deviates from classical economics by not following the assumption that decisions are always made rationally. While behavioral economics traditionally applies to the field of finance, outcomes can easily be ported to the field of health. One technique often studied is the use of incentives. Incentives and paternalistic “nudges” can lead people to make positive health choices, such as persuading people to choosing healthier foods, simply by placing the healthier foods in a more prominent place within a cafeteria [48]. Financial incentives have also been successfully used to promote health behavior changes such as smoking cessation and weight loss [73, 74]. Behavioral economics persuasion techniques have also been evaluated in a snack-delivering robot designed to promote healthy food choices [45].

We will draw from previous work with incentives in Chapter 5, by evaluating financial incentives designed to promote system use.

2.9 Tailored Health Messages

The research area of health tailoring focuses on giving personalized, targeted health messages to individuals in order to more successfully induce a positive health behavior change [34]. For example, experiments setup to examine healthy behavior promotion (such as smoking cessation or healthy diet promotion) more positive health outcomes occur when interventions provide personally tailored information rather than static content [64, 69, 76]. When tailored health material is processed, many cognitive processes are triggered, including: attention, deep thinking, emotion and self-reflection

[33]. Tailored information also leads to positive expectations about the health message itself [75].

In Chapter 5, we will use and evaluate tailored health messages in our system designed to track alcohol consumption.

CHAPTER 3

Post-Hospitalization Follow-Up by Embodied Conversational Agents

After a hospitalization, approximately 1 out of 5 patients will suffer from an adverse event, and one-third of these complications are preventable. Having a clinical pharmacist follow-up with patients a few days after leaving the hospital has been shown to significantly reduce re-hospitalizations and adverse drug events. In this chapter, I will describe our design for an Embodied Conversational Agent system for longitudinal, post-hospitalization follow-up. I will discuss the design process we followed – basing our system on best-practice follow-up interactions between patients and clinical pharmacists. I will also describe an observational study of patients interacting with the system, and finally, discuss at-home use by patients in a small field trial. Portions of this work were presented at the 2011 AAAI Spring Symposium [61].

3.1 Introduction

The patient transition from hospital, to home, to first follow-up with a primary care provider represents a gap in the U.S. healthcare system that is largely neglected, highly error prone, and, until recently, non-standardized. Because of these shortcomings, 1 in 5 patients are readmitted to the hospital within 30 days of discharge, and studies have shown that one-third of these readmissions are typically preventable [26]. These unnecessary readmissions represent a significant burden to our health care system in terms of costs and resulting morbidity and mortality to patients. Indeed, as a part of the recent Patient Protection and Affordable Care Act, if a Medicare patient is re-hospitalized within 30 days of discharge, Medicare will *significantly* reduce the payment the hospital receives for that visit [1]. Thus, hospitals are now extremely motivated to reduce preventable readmissions.

A few interventions, developed and evaluated in randomized clinical trials, show promise for reducing the 30-day hospital readmission rate. These interventions typically involve a nurse or pharmacist calling patients a few days after discharge to determine if they are experiencing any problems or complications that can be resolved, or if they have questions or uncertainties about their self-care regimens, particularly regarding their medications. Many issues can be resolved over the phone, potentially preventing and adverse event or a visit to the Emergency Department.

The Re-Engineered Discharge (RED) project at Boston Medical Center is one such intervention that was shown to reduce re-hospitalizations by 30% [41]. In 2007, the National Quality Forum "Safe Practice" update highlighted hospital discharge as a critical area of improvement, and outlined safe practice guidelines based largely on components of the RED program [27]. Important elements of this protocol include: 1) printing discharge instructions in a format that patients (including those with low health literacy) can understand; 2) reviewing these instructions in detail with patients prior to discharge; 3) ensuring that patients comprehend the instructions; and 4) having a nurse or pharmacist call patients a few days post discharge to resolve any issues.

While in-hospital education is unquestionably important and beneficial to patients, a critical factor in achieving positive health outcomes is the last step of the Project RED protocol: post-hospitalization follow-up with patients. Follow-up phone calls by a nurse or pharmacist may be essential for a safe transition from hospital-to-home. One European study of community pharmacists reported that 64% of recently discharged patients evaluated had medication issues [59]. In Project RED, the study pharmacist performed at least one corrective action for 59% of the patients reached, and found that 65% of patients who completed a medication review on the phone had at least one medication problem [41]. Several studies have shown that post-discharge interventions, specifically by pharmacists, can reduce Emergency Department (ED) visits and re-hospitalizations, and also reduce preventable adverse drug events [2, 21, 67].

In collaboration with researchers from Boston Medical Center (BMC), we developed an automated implementation of the Project RED protocol, using an ECA to simulate the effects of face-to-face patient education by a nurse at the time of hospital discharge (Figure 3-1). The ECA had a digital rendering of the patient's personalized discharge booklet, called and After Hospital Care Plan (AHCP), could bring up pages of the booklet on the screen, and teach the patient about their personalized plan, with the patient being able to follow along in their paper copy of the AHCP. In a pilot evaluation with nineteen hospital patients, participants indicated high levels of trust in and satisfaction with the system, reported that the interaction helped prepare them to leave the hospital, and only 16% of them indicated they would have preferred receiving their discharge instructions from a doctor or nurse in the hospital [6].

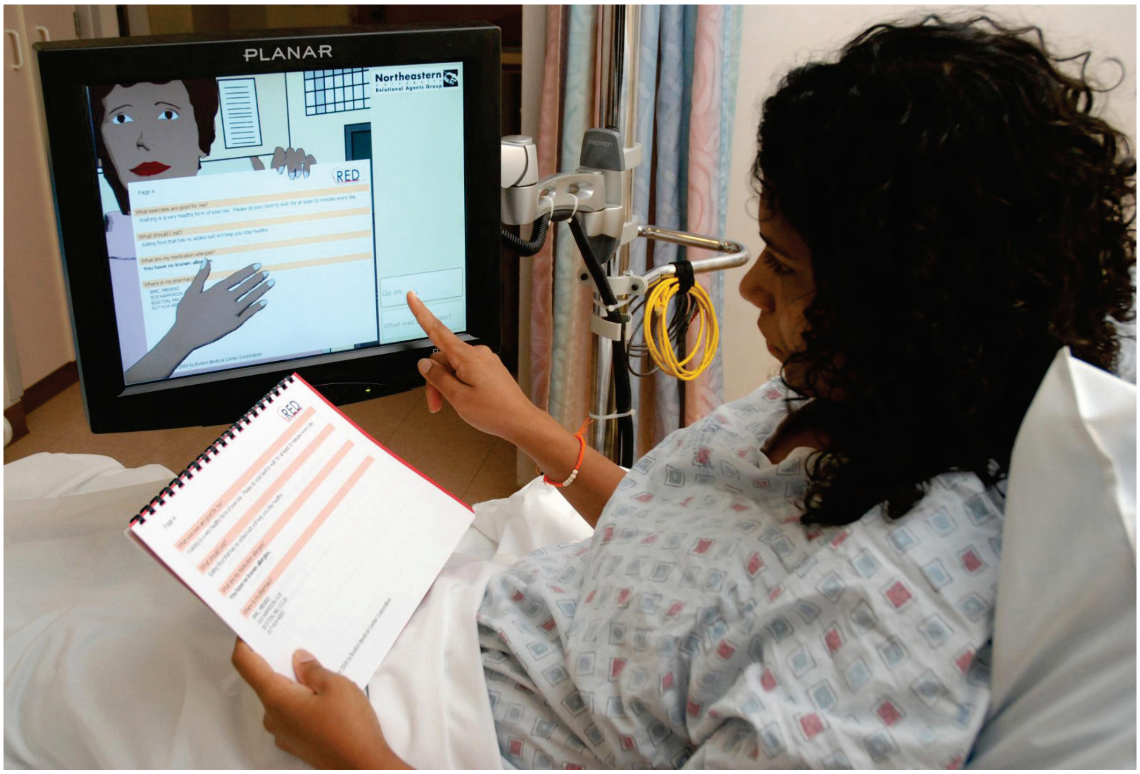


Figure 3-1: Patient holding the AHCP and interacting with the Virtual Nurse

With post-hospitalization follow-up being an important factor in health outcomes, our goal was to create a natural extension to the Virtual Nurse: an intelligent at-home system that can work with patients to prevent and detect adverse events *after* a patient has left the hospital. In this chapter, I will discuss the design of our ECA system, called MyLink2Care, modeled after follow-up interactions between patients and BMC's Project RED clinical pharmacist. I will also report results from an observational study of recently discharged patients using the system in a laboratory setting. Finally, I will discuss at-home usage by hospital patients enrolled in a small field trial.

3.2 System Design

In collaboration with researchers at Boston Medical Center, we set out to extend the Virtual Nurse system into one that could implement the final element of the Project RED protocol: following up with patients after they have left the hospital and resolving any issues that may have occurred. The goal for the at-home MyLink2Care system was to emulate the post-discharge phone call by the clinical pharmacist in Project RED, with a focus on medication and follow-up appointment adherence, as well as screening for post-hospitalization adverse events.

To inform the design of our system, we studied follow-up conversations between patients and the Project RED clinical pharmacist at BMC. We investigated the distinct techniques used by the pharmacist to detect issues that the patient might be experiencing, post-hospitalization. Several research questions were of particular interest:

R1: How did the pharmacist structure her conversation with patients?

R2: What problems did the pharmacist uncover and how were they resolved?

R3: Did the patients ask the pharmacist any questions? If so, what information did the patients want to know?

In order to answer these questions, we observed and analyzed five conversations between a clinical pharmacist and a recently discharged hospital patient. We describe our analysis of those conversations and discuss how the findings were incorporated into the MyLink2Care system designed for post-hospitalization follow-up.

3.2.1 Patient-Pharmacist Conversations

Three patients participated in the observational study; recruited during their hospitalization, Table 3-1. Participants were asked to return to the hospital a few days after discharge, and meet one-on-one with a pharmacist to discuss how they are doing at home. In order to understand how the pharmacist might change her approach with the patient over time, we asked participants to schedule a second follow-up visit with the pharmacist, and two participants were able to do this. Participants were paid $25 for each visit. All conversations between the patients and pharmacist took place in a small hospital conference room, were audiotaped, and were fully transcribed. We also conducted a separate interview with the pharmacist, to review the transcriptions from the patient sessions, and discuss her motivation and rationale behind particular topics discussed with the patients.

Patient ID	Gender	Age	Primary Diagnosis	Number of Medications
1	M	67	Diverticulitis	14
2	M	43	Cardiomyopathy	15
3	F	50	Asthma	20

Table 3-1: Patient demographics

	Conversation Length	Percent of talking done by the patient	Number of medications discussed	Number of issues discovered	Number of questions the patient asked
Patient 1					
First conversation	74 minutes	49%	11	5	10
Patient 2					
First conversation	68 minutes	41%	11	4	0
Second conversation	66 minutes	47%	11	2	0
Patient 3					
First conversation	58 minutes	49%	15	4	0
Second conversation	29 minutes	50%	16	2	0

Table 3-2: Patient-Pharmacist conversation detail

3.2.1.1 Conversational Structure

The conversations between the patient and pharmacist followed a structured plan, and were generally pharmacist-driven. A summary of the conversations is listed in Table 3-2 and a typical outline is shown in Figure 3-2. Prior to the first conversation, the pharmacist reviewed the patient's hospital discharge summary to familiarize herself with the patient's case and discharge instructions. Upon meeting the patient, the conversation began with an introduction and quickly moved to a discussion about the patient's hospitalization and medical condition. In this portion of the conversation, the pharmacist sought to ascertain the patient's point of view on what led to their hospitalization, as well as to find out if the patient knew their discharge diagnosis. She also asked if the patient had returned to the hospital, Emergency Department or to any clinical appointments since leaving the hospital, in order to determine whether their prescribed medications had been changed since their hospitalization, so that they could be accurately reviewed later on in the conversation.

Condition Review

Ok, so first can you tell me the main reason why you were in the hospital? *Ummm, I was having shortness of breath…and there was also, they found fluid on my lungs which might have been caused by a virus and it might have affected my heart.* Perfect, yep, right that's exactly the information that I got. Because of that virus around your heart, maybe your heart wasn't pumping as efficiently and your blood pressure was high, so they said that maybe you had some cardiomyopathy. That would be the diagnosis.

Medication Review

So how do you remember to take your medicines? Are pill boxes usually... *I just line 'em up. What I do is I put my diabetes medicine on one side, and then the others I just line 'em up and take them one-by-one.* OK, and that system seems to be working for you? *Yeah.* So whenever you're ready I'll have you just take one medicine at a time and we'll go through 'em. I'll compare it to the list I have here and I'll ask you a couple questions about each medicine. So, in any order that you want... *Fffffurosemide….*Yep Furosemide good. *I think this the fluid pill.* That is the fluid pill. *I take it in the morning.* OK and how many tablets do you take in the morning? *One.*

Side-Effect Discussion

Any side effects from this one? This is probably causing your headache, yeah. *It's not that bad it's like in the back here.* OK, how bad is the headache and how often does it come? *It's not too bad, it's tolerable, just annoying.* OK, so on a scale of say zero to ten, zero is no pain and ten is like the worst headache of your life, where would you put it? *Three.* You would put a three, ok and when you get the headache what do you usually do?

Appointment Discussion

Now when are your upcoming appointments? *I have one with the heart specialist on the 9th, and one with my primary care on the 20th.* Perfect so on the 9th you're going to see Dr. _________ the cardiology doctor at nine in the morning. Do you know where to go for that? *Yep.* Are you going to be able to make that appointment? *Yes.*

Figure 3-2: Sample patient-pharmacist dialogue for routine aspects of the conversation (edited for grammar). Patient utterances are in italics.

After reviewing the patient's medical condition, the conversation moved to a discussion about the patient's medications. Patients were asked to bring in all of their prescription medications each session and place them on a table at the start of the session. The pharmacist began by asking the patient about the method they used for remembering to take their medicines and, specifically, whether or not they used a pill-box. Next, each medication was reviewed one by one, with the patient choosing the order in which the medications were discussed. For each prescription, the pharmacist had the patient read the name of the medication out loud, describe how often they took the medication each day, and how much they took at one time. The pharmacist reconciled the patient's information with the information listed on the patient's discharge summary and clarified and corrected any misunderstandings by the patient. This portion of the conversation was typically the longest, taking up 55% of the conversation, on average.

When reviewing medications, the pharmacist would often bring up the subject of side effects. If a patient reported or endorsed a side effect, the pharmacist would find out when it started happening, how severe the patient thought it was, how often it was occurring, and whether or not the patient had taken any action to deal with the side-effect. She would then give advice to the patient on how the side-effect could be handled or avoided, and what action the patient should take if it worsens.

Following the medication discussion, the pharmacist would review the patient's post-hospitalization follow-up appointment with their primary care physician (PCP), and any specialist appointments, if necessary. During this portion of conversations, the pharmacist discovered if the patient understood when and where every appointment was going to take place, who the appointment was with, what it was for, and whether or not the patient was still able to go to the appointment. The pharmacist also discussed emergency situations with the patient, and counseled the patient on situations when they should go to the Emergency Department, and situations when it would be better to contact their primary care physician's office or pharmacy.

Finally, the pharmacist discussed condition self-management with patients. For two of the patients, diabetes self-management was reviewed in detail, discussing how often they should check their blood sugar levels, what their goal level should be, medical terminology related to diabetes, signs of hypoglycemia, and explaining what do in an emergency. For another patient, blood pressure was reviewed in detail, including recent lab test results and goals for the patient.

During the course of the conversation, the pharmacist also discussed topics that were unique to each patient. For example, one patient had recently lost his health insurance and had trouble filling his prescriptions. The pharmacist listened to the patient's background on the situation and made any necessary arrangements to ensure the patient was receiving all available assistance.

Two of the three patients in our study were able to return for a second conversation with the pharmacist. These follow-up conversations followed a similar structure to the initial interaction, with the amount of time spent on each topic allocated differently. For both patients, the pharmacist spent 13% of the second conversation explicitly following up on issues that were discovered during their previous session. For Patient 3, who was not able to bring her medications to the first session, but did bring them to the second session, the pharmacist followed almost the same structure the second time around, spending 54% of the time reviewing medications and 8% of the time on education regarding the patient's medical condition. For Patient 2, the pharmacist altered her approach during the second session, changing the time spent discussing medications from 52% to 25% and increasing the amount of time spend on condition education from 7% to 26%.

3.2.2 Issues Detected by the Pharmacist

During each session the pharmacist detected, on average, 3.4 problems. These included misunderstandings about how often patients were supposed to take their medications, experiences of medication side effects, confusion about dates/times of follow-up appointments, and lack of disease self-management.

With our goal of building an at-home system for the detection and monitoring of adverse events, we were particularly interested in how the pharmacist uncovered these issues, how she attempted to resolve them, and if the patient was compliant in following the pharmacist's recommendations. In this section, we discuss the different classes of problems detected and the various courses of action taken by the pharmacist.

3.2.2.1 Patients Following a Different Medication Regimen than Prescribed

The most common problem detected by the pharmacist was the patient taking their medicine differently than prescribed. This issue is deeply complex, and cannot be attributed to one simple cause. Previous work has shown that a wide-variety of factors can influence medication adherence, including forgetfulness, deciding to omit doses, lack of information, and emotional factors [58]. In our observations of the patient-pharmacist conversations, two examples of non-adherence emerged.

In the first example of non-adherence, the patient had their prescribed medicine at home, was taking the medicine, but was not taking it according to the physician's orders. For Patient 1, this seemed to be a case of non-intentional non-adherence: the patient simply misunderstood how often to take three of his medicines. This patient had seven medicines that were prescribed for two times/day, two medicines to be taken once/day, and one medicine to be taken three times/day. It turned out that patient was taking all medicines twice daily. For this situation, the pharmacist corrected the patient, and checked for patient understanding by having the patient repeat back the correct times of day for the medications that were not being taken correctly. At the end of the conversation, the pharmacist reviewed the correct times to take each medicine, to reiterate the prescribed plan.

Patient 2 had a similar situation, with a medication prescribed for twice a day, but the patient was only taking it once a day. However, in this instance, the *patient* was correct, and the *discharge summary* was incorrect. The particular medication was for diabetes, and prescribed according to the patient's blood sugar levels. When leaving the hospital, the patient was told to take the medicine once a day, and was following that order. The information in the discharge summary listed the medication as twice per day, and was either entered incorrectly or had not been updated to reflect the most recent information. After discussing the patient's blood sugar levels with the patient, the pharmacist realized that the error was most likely a mistake in the hospital's record, not a mistake by the patient. The pharmacist recommended continuing to take the medicine one time per day, called the patient's primary care office and had an appointment made for the patient in order for the PCP to test the patients blood sugar levels and assess the correct medication level for that patient. When the patient returned for their second session, the pharmacist asked the patient to review what the primary care physician recommended, and discovered that indeed the medication should only be taken once per day.

In the last example of non-adherence, the patient did not have their prescribed medication, and thus was not taking it. This included new prescriptions made during the recent hospitalization, as well as standing prescriptions that were never refilled. Patient 3 was not able to bring in her medications during the first session with the pharmacist, but the pharmacist still went through each medication on the discharge summary one-by-one to discuss it with the patient, determining if the patient recognized the medication by name, and whether or not the patient was taking it as prescribed. During that conversation, the patient stated that they never received the paper prescriptions for two of their medications prescribed during their hospitalization and that for another previously prescribed medication, she had not refilled it for over a year. The pharmacist had discovered early in the conversation that the patient had a follow-up appointment with a nurse practitioner that same afternoon, so she gave the patient a detailed printout listing the medications for which the patient needed new prescriptions, for the patient to bring with to her appointment. During the second session with the pharmacist, Patient 3 was able to bring in her medications and the pharmacist and patient were able to review them together more thoroughly than during the previous session. During this follow-up conversation, the pharmacist discovered that for one of the medications that the patient thought they didn't have, they in fact did have it and were taking it as prescribed. For the other two medications, they had still not picked them up from the pharmacy and were not yet taking them. In addition, a few days earlier, this patient had been re-hospitalized for breathing problems, and upon discharge was prescribed a steroid to begin taking immediately. Unfortunately, the patient had not filled this prescription either.

3.2.2.2 Medication Side-Effects

The pharmacist also frequently detected side effects that the patient had experienced at home. For fifty-seven percent of all medications, the pharmacist specifically asked about possible side effects. Each patient endorsed at least 1 side effect during the conversations. Of the 5 total side effects detected, one was detected by the patient self-reporting the issue after the open-ended question, “Do you think you are having any side-effects from this medication?” another was detected by a closed-ended question ("Any dizziness?") and the remaining 3 side effects were detected by mentioning that a specific side effect is possible and then asking if the patient had experienced it, such as, “Sometimes when people start taking this they feel tired, are you feeling tired?”

The pharmacist’s choice for framing and asking about side effects seemed to vary by patient. For example, the technique of asking the open-ended question “Are you having any side effects?” followed by mentioning and teaching about a specific side effect that can occur with the medication and then asking if the patient had experienced that specific side effect, was almost exclusively used with Patient 2. During the beginning of the medication discussion, the pharmacist would ask each patient if they knew why a particular medicine being discussed was prescribed for them, and Patient 2 was the only one of the three who indicated he did not. Thus, for this patient, the pharmacist approached the discussion as teachable-moment, taking the opportunity to explain not only what each medication is for, but also how it works, and what side effects to be aware of.

As mentioned earlier, anytime a patient endorsed a side effect, the pharmacist would find out how often it was occurring and how severe the patient thought it was, in order to help inform her recommended course of action. For all of the side effects that the patients endorsed, the pharmacist encouraged monitoring and follow-up within a few days. For some, she also recommended a specific course of action, such as an over-the-counter remedy (for the headache) or switching their medication from morning to the evening (for drowsiness).

When interviewing the pharmacist about the patient sessions, we were particularly interested in the 43% of medications in which she did not ask about any side effects. It was often the case that these were over-the-counter (OTC) medicines rather than prescription medicines. If a side effect for an OTC medicine was potentially serious, such as bleeding with aspirin, the pharmacist did mention it to the patient, but most often side effects were not mentioned with OTC medicines. In other scenarios, the pharmacist often grouped medications together by indication, and if for example, the patient was on several medications for blood pressure, and dizziness was the most common or serious side effect for all of those medications, the pharmacist would ask about it one time, for one of the medicines, and not bring it up for the rest.

If the pharmacist did bring up the topic of side effects, she almost always mentioned only one of the several potential side effects for a particular medication. As most conversations were over an hour long, the pharmacist explained her decision to keep things as brief as possible and prioritize the most important side effects for discussion.

> I think some of it has to be on the onus of the patient to say ‘I think this [side effect] is going on, and I think it might be attributed to a med, do you agree?’ If they don’t [bring anything up], I just try to go through the things that are life-threatening, that would send them back to the ED, or where I wouldn’t want them to continue to take the medication.

3.2.2.3 Patient Self-Care Regimen

Two of the problems detected by the pharmacist regarded self-care and management of patients’ health conditions. Patient 2 was instructed to weigh himself daily in order to monitor the effects of his blood pressure medication, however the patient did not own a scale. In this situation, the pharmacist called and left a message with his primary care doctor's office, on behalf of the patient, to see if they would be able to give him a scale prior to his appointment.

In her first conversation with Patient 3, the pharmacist discovered that the patient was not monitoring her blood glucose levels. This patient did not want to experience the pain of pricking herself, had an aversion to needles, and did not want to be thought of as a "junkie". The pharmacist reviewed the importance of self-monitoring with the patient, educated her about glucose goals, and most importantly for this patient, how to recognize signs of hypoglycemia and what to do in an emergency. The pharmacist encouraged the patient to try to check her blood sugar once per day

3.2.3 Patient Questions

Of the three patients in our study, only one asked the pharmacist questions during their session. This patient asked several questions throughout the conversation, mostly for clarification or additional information from the pharmacist. For example, the patient asked the pharmacist to point out which medication name was the generic name and which was the brand name. In another example, the patient asked the pharmacist to explain which side effects could be caused by one of her medications. Other questions included asking whether or not a medication should be taken with food, and about the causes of particular side effects.

3.2.4 Design of the MyLink2Care system

After observing and analyzing the pharmacist-patient follow-up interactions, we began to design our post-hospitalization follow-up system, called MyLink2Care. The system is a natural extension to the in-hospital Virtual Discharge Nurse, using the same agent characters and interaction look-and-feel, ported to a web environment (Figure 3-3). The MyLink2Care system is designed to pickup where the Virtual Discharge Nurse interaction ends. Once a patient is at home, they can log into the MyLink2Care website and interact with the same agent that they saw in the hospital. The agent is designed to interact with a patient each day until the first follow-up appointment with their primary care physician. The goals of the agent are to: promote the contents of the AHCP (especially medication and appointment adherence), screen for potential adverse events, and mediate communication with a team of BMC nurses. The agent is not limited to a subset of patients or certain medical conditions, but designed to provide information to any patient leaving the hospital. The agent can discuss information from 248 medical diagnoses, 1751 distinct medications, and 207 potential adverse events and medication side effects.

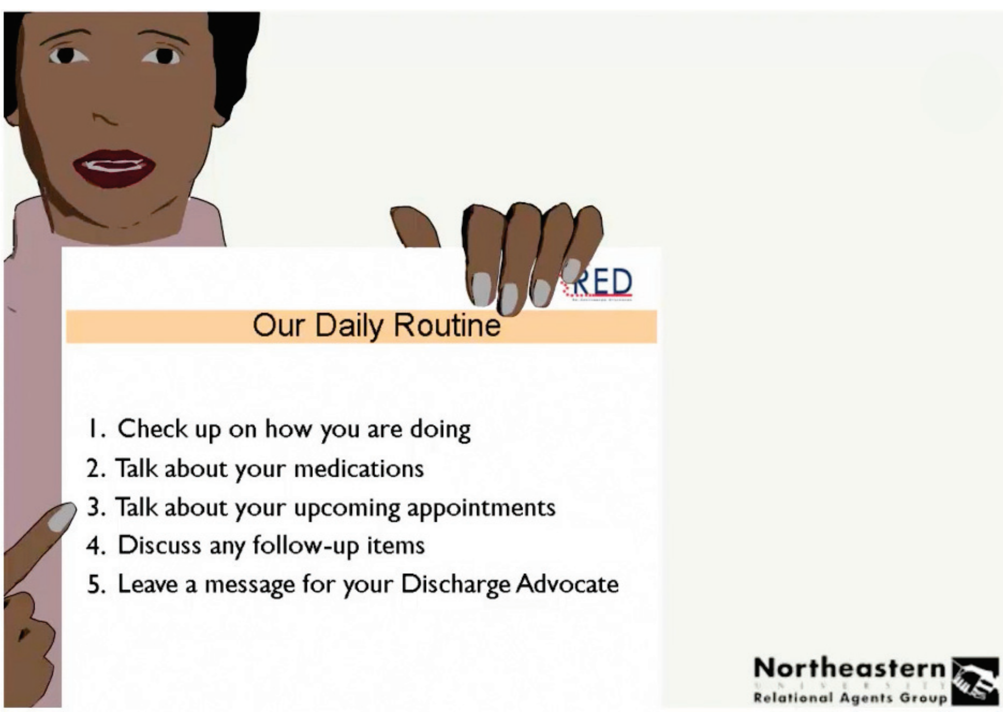


Figure 3-3: The MyLink2Care System

After analyzing the interactions with the Project RED clinical pharmacist, we incorporated the pharmacist's conversational structure into the design of the MyLink2Care system. The agent begins by asking the patient how they are feeling since they left the hospital. The agent also checks to see if the patient has visited the Emergency Department, a doctor's office, or were re-admitted to the hospital, because if so, there is a possibility of changes to their health plan (medications or follow-up appointments may have changed). If so, the agent creates an alert for a hospital nurse to contact the patient and reconcile their health plan before continuing. Once the agent has determined that the health information hasn't changed, it checks with the patient to see if any adverse events have occurred (Figure 3-4). Next, it discusses the patient's medication regimen: checking for prescription acquisition, asking how often the patient is taking their medications, and asking about any potential side effects. Throughout this process the agent will troubleshoot issues such as, lost prescriptions or trouble getting to the pharmacy and create alerts for hospital staff as necessary. Next, the agent will review any upcoming medical appointments, and ensure that the patient can still make the appointment and that they know where to go. Finally, the agent gives the patient the option of reviewing information related to their medical condition, and also allows the patient to request a specific follow-up phone call from a nurse at the hospital.

Figure 3-4: The MyLink2Care agent reviewing potential adverse events.

3.2.4.1 Side-Effect Discussion

When discussing medication side effects, we found that the pharmacist usually asked the patient about one particular side effect. Having an automated system determine which side effect to discuss is a challenging problem. On the one hand, the system should be as accurate as possible: one approach would be to list and discuss all possible side effects for each medication. On the other hand, the system should also be as relevant as possible, and not discuss superfluous information with the patient. Another factor is patient engagement: if the conversation becomes long or irrelevant, the patient may become disinterested and stop using it all together.

In our approach, we seek to strike a balance between providing accurate and relevant information, while reducing the chances of overwhelming the patients (Figure 3-5). For each medication in our database, we had clinicians enumerate the top-five side effects for the system to discuss. These side effects are the most common, or the most likely to be life threatening. Prior to discussing medications with the patient, the ECA displays 20 common adverse events, determined by [26] in a checklist format, and allows patients to report if they have experienced any of those events since leaving the hospital. This information allows the ECA to reduce the number of side effects discussed for each medication. For example, if the patient denies that they have experienced any headaches during the initial adverse event checklist, and the patient is taking a medication with headaches as a potential side effect, then the ECA will not need to ask about that side effect when reviewing those medications. Likewise, as the ECA discusses each medication and we acquire more knowledge about side effects that the patient is or is not experiencing, this will influence the side effects we need to discuss with different medications later in the conversation. We have also incorporated a mechanism for the patient to self-report any side effect that they believe they are having, whether it is tied to a medication in our database or not. This allows us to keep the side effect conversation short and relevant, while also maintaining expressivity by the patient.

1. For each medication M_i load the associated side effects into a list SE_i
2. Display the 20-item adverse event checklist to the patient and collect responses
   a. For each item f from the checklist that was not endorsed by the patient
      i. For each medication's side effect list SE_i check to see if f is a member and if so, remove it from SE_i
   b. For each item f that was endorsed by the patient, do full side effect assessment
3. For each medication M_i that the patient has acquired and is taking, present the list SE_i to the patient and collect responses
   a. For each side effect j in SE_i that was not endorsed by the patient
      i. For each medication side effect list $SE_{(i+1) \ldots n}$ check to see if j is a member and if so, remove it from SE
   b. For each side effect j that was endorsed, do full side effect assessment

Figure 3-5: Algorithm for ECA side effect discussion.

3.2.4.2 Repeated, Adaptive Interactions

The system is designed for daily interactions to transition patients smoothly from their hospitalization to their primary care follow-up appointment, with the behavior of the ECA continuously adapted based on prior interactions with the patient and the actions of clinicians monitoring the system. In designing the conversational structure for repeated interactions with the ECA, we are following the approach of the clinical pharmacist to keep the interactions short, and focus heavily on issues that need follow-up. We also included the ability for the patient to ask questions and find out more information if they so desire.

In order to for the system to effectively discuss follow-up issues with patients, we designed a back-end alert management system for a nurse on the clinical team to resolve any issues detected by the agent. The MyLink2Care system was not designed to provide traditional medical care to the patient, and it was important that a medical expert reviewed the issues uncovered by the agent, worked to resolve them (e.g. calling the patient's physician to clarify any misunderstanding about medication dosage), and provided feedback to the patient on the status of that issue. We designed the system to be aware of if/how an issue was resolved, and have the ability to discuss the resolution or recommended course of action with the patient during their next conversation.

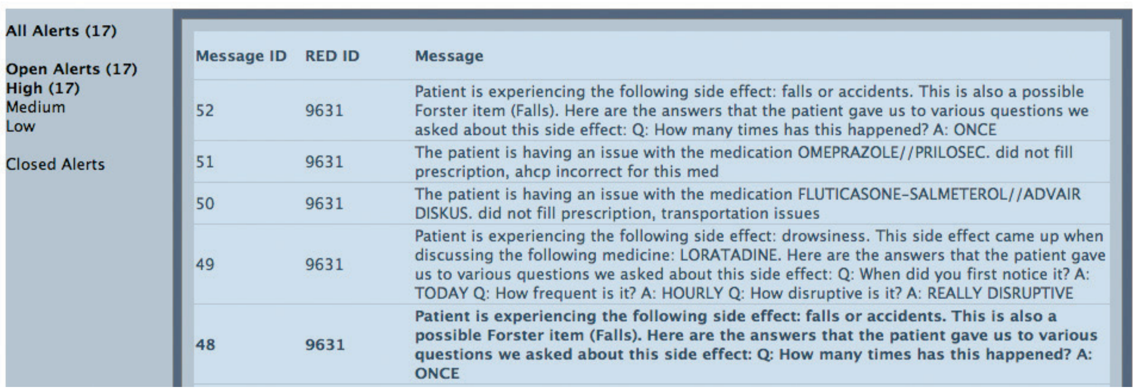


Figure 3-6: Sample alerts generated by the MyLink2Care system

3.3 Observational Study with Patients

To assess the usability of the MyLink2Care system, we conducted a lab-based observation study. Participants were recruited by Boston Medical Center research staff, and consented during their stay in the hospital. About a week after discharge, patients returned to an office at the hospital to interact with the MyLink2Care system. Patients were observed throughout the interaction and encouraged to *think-aloud* if they had any comments or concerns. After the interactions, patients completed a short questionnaire about their experience and I conducted a semi-structured interview. The interview was audio-recorded, and later transcribed and coded for themes regarding their experience.

Four participants interacted with the system, all women (Table 3-3). Participants were middle-to-older aged and were discharged from the hospital with various diagnoses and a large range of medications.

Participant ID	Age	Days in the Hospital	Diagnosis	Number of Medications Prescribed
1	56	1	Chest Pain	2
2	67	2	COPD	9
3	48	1	Angioneurotic Edema	18
4	58	5	Pancreatitis	6

Table 3-3: Observational Study Participants

3.3.1 Results

All participants completed the session, though some had an easier time than others. Participants 1 and 2 were easily able to use the mouse and select options when prompted by the agent. Participant 3 did not have clear vision, but came to the session with a friend. The participant and her friend used the system together, with the friend reading options to the participant, and the participant indicating which reply she would like her friend to select with the mouse. Participant 4 did not have any computer experience and found difficulty in using the mouse. Participant 4 was often confused and drowsy during the session.

Sessions lasted approximately 5-15 minutes. At least two participants reported issues that generated alerts for follow-up by the hospital nurse. Alerts included issues such as pain, side effects such as dry mouth, and problems filling prescriptions.

Participants reported high levels of satisfaction with the agent (Figure 3-7) and with the capabilities of the system (Figure 3-8), but during the interviews participants expressed varying degrees of enthusiasm and satisfaction.

Participant 2 was enthusiastic about the interaction and the potential for future interactions with the agent. She was feeling well after her hospitalization and not experiencing any health issues or troubles with her medications. Despite needing little medical follow-up, she still enjoyed the interaction.

> *"It was awesome! Interesting! I enjoyed the conversation with her very much. … I felt like it was one-on-one, very personal."*
>
> – Participant 2

When asked how often she would envision talking with the agent at home, participant 2 felt that once-a-week or monthly interactions would be appropriate, as she was doing well. The participant indicated that since she had gone through things with the ECA once and became familiar with the potential for medication side effects, she could use the MyLink2Care system later on, if a problem ever arose.

Participant 1, who was also doing well after her hospitalization, was not enthusiastic about the interaction. She felt that it did not help her, and that it simply functioned as a reminder, rather than an interactive system with capabilities to troubleshoot problems.

> *You know, she wasn't really helpful to me, because I already knew everything ... Well, see I knew there was really nothing that she could do, but report it to the discharge advocate, and that person would at that point call me. … I would rather talk with my doctor versus even the discharge advocate. I'd rather talk directly with [doctor's name].*
>
> – Participant 1

Participant 3, and the friend who accompanied her to the session, were excited about the interaction, and like the idea of having such a system at home. This patient was experiencing many chronic health issues, and was enthusiastic about the potential to troubleshoot health issues with ECA, as well as mediate information to/from her doctor. When asked about privacy and sharing, this participant indicated that she would want all the information from her MyLink2Care interactions sent to her doctor.

> *Yes, yes, keep [my doctor] updated yes.* – Participant 3

> *Yeah, cause it's like a little frustrating when you go to the doctor and they say, "Ok what's the problem?" and you know, they don't know. "Oh, I just had this surgery and you don't know I'm doing a follow-up?!" You know what I'm saying?* – Participant 3's friend

Participant 4 had some concerns about receiving health information from the ECA. She indicated that while she would want information about her interactions with the ECA going to her doctor, when it came to *receiving* health information, she would want to speak directly with her doctor or nurse, rather than having it mediated through the ECA.

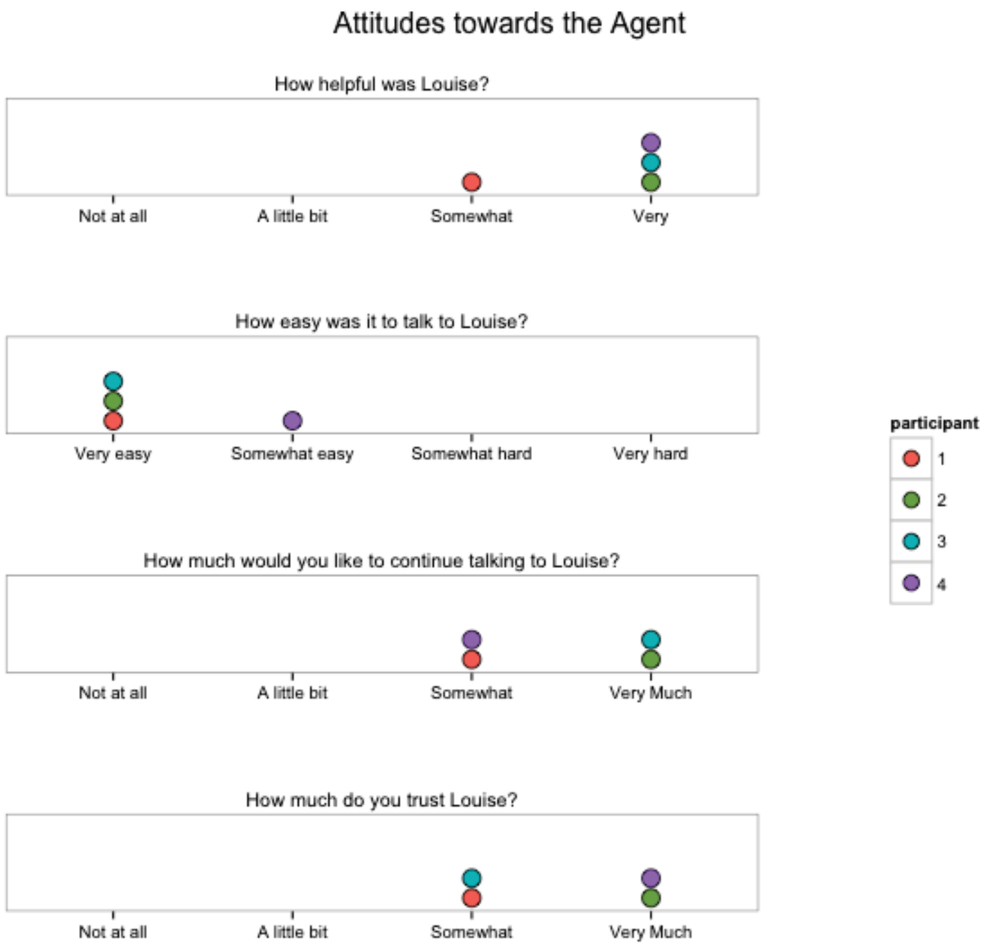


Figure 3-7: Observational Study - Attitudes towards the Agent

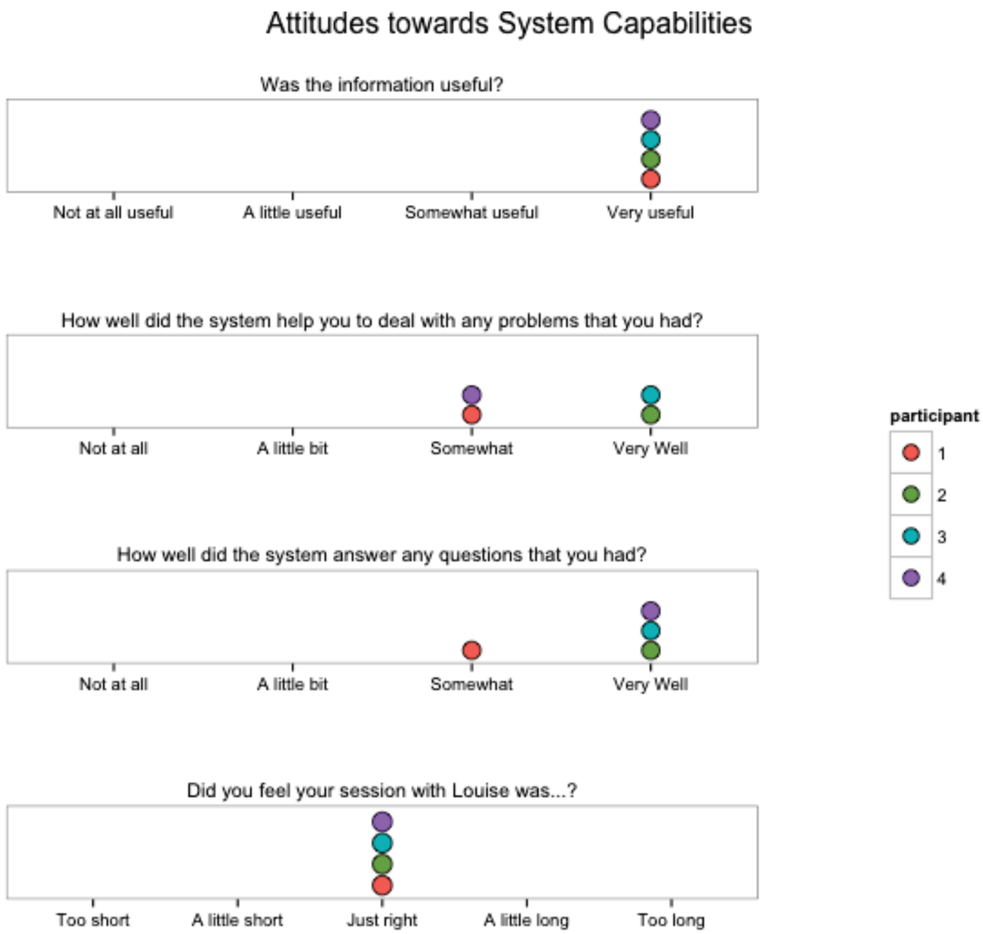


Figure 3-8: Observational Study - Attitudes towards System Capabilities

3.4 At Home Use by Patients

To fully understand the efficacy of this system, we, along with clinician researchers at Boston Medical Center, designed a field experiment to test its effectiveness at preventing, detecting, and alerting hospital staff about adverse effects patients were experiencing after leaving the hospital.

Patients were enrolled and consented during their stay in the hospital, by BMC research staff. To be eligible, participants had to be over 18, speak English, be discharged to home (not to a nursing care facility, etc.), have a valid email account, and have access to a computer with an Internet connection. After consent, participants were randomized into either the intervention or control group. The intervention group interacted with the virtual nurse at the time of hospital discharge (Section 3.1) and setup a username and password for the at-home MyLink2Care system. Participants were instructed to log into the system each day, until their first follow-up appointment with their primary care physician, and were given a handout with instructions for connecting and logging into the website. The control group received the current-standard of care. Approximately two weeks after hospital discharge, all participants received a follow-up phone call to administer questionnaires regarding doctor-patient satisfaction [17], self-efficacy in their medication regimen [66], medication adherence [52], occurrences of adverse events [26], and satisfaction with the ECA (if applicable). All re-hospitalizations and Emergency Department visits between hospital discharge and follow-up appointments were also logged. Patients received at $10 gift card for participating in the study, and intervention participants received an extra $1/day for each day they logged in and used the ECA system at home.

3.4.1 Results

Fifty-two participants were enrolled in the experiment, with five participants dropped from analysis, as they were found ineligible, giving a total of 47 participants (24 intervention).

The main result from this experiment was that only *four* (16.6%) of the intervention participants logged into the MyLink2Care system from home. Of those participants, none had more than two sessions with the ECA, far from the design of one interaction per day.

What might have caused the vast majority of patients to never use the system? Ten intervention participants were reached for the follow-up phone call, and the most frequently reported reason was that participants were unsure of what to do or how to log on. This is certainly a possibility, though research staff emailed each participant instructions for logging into the system, and called participants to resolve any trouble they might be having. Often, research staff suspected that participants might not have had a valid email account, or easy access to a computer.

Other possibilities for the low level of participation might be that patients simply did not feel well enough to interact with the MyLink2Care website, forgot about using the system, did not want to use the system, or did not feel that they needed the system.

One thing is for certain; a system designed for follow-up with patients cannot be successful if patients simply do not elect to use it. The issue of long-term system engagement is an understudied area of human-computer interaction, and one that I will explore in the following two chapters.

CHAPTER 4

Interface Embodiment and Self-Reported Health Data

4.1 Introduction

After lackluster participation by patients in the MyLink2Care trial described in the previous chapter, I understood that in order for at-home, self-report health systems to be successful, especially long-term, we first must explore two important areas of research: 1) How do we improve data accuracy, help people be comfortable providing data, and remove any obstacles to providing data? and 2) How do we keep people engaged with such systems over (potentially long periods of) time? Over the next two chapters, I will describe experiments that explore both of these research questions.

In this chapter, I take two well-understood phenomena regarding self-report – social desirability bias effects, and the computers as social actors paradigm – and examine how these findings extend into a longitudinal context, both with regard to self-reported data quality, and quantity (number of system interactions). The results of this experiment were presented at CHI 2011 [60].

As discussed in section 2.1, ECAs provide an ideal mechanism for building self-report health systems: they have been shown to establish trust and rapport with their users and they provide users with a low-pressure environment. We also know that a primary issue with self-reported data, is the effect of social desirability (section 2.2). While these effects can be minimized by using computerized systems (rather than face-to-face interviews or paper questionnaires), it is not know how people will respond to ECAs designed to collect health data. Will people consider ECAs as computers, and appreciate an anonymous environment? Or will people react to ECAs with engrained social responses, and as a result, self-report data that might be skewed by the effects of social desirability?

The Computers as Social Actors paradigm would suggest that any computer interface would be subject to social desirability bias effects. One downside of the many Computers as Social Actors experiments is that they often only assess participants' first reactions from a single lab session. Do these social reactions hold up over time? A reasonable hypothesis is that when people interact with a social interface repeatedly, over long periods of time, their reactions to the interface might change. They may realize that the interface is not as capable or competent as a human at being a social actor, or they may become conscious of their misattribution, and as a result, their social responses to the computer may diminish with time.

In a system designed for users to self-report their exercise, one concern would be usage bias: participants choosing only to use the system when they have exercised.

Another concern is that when participants do interact with the system, participants may over-report the amount of exercise they actually performed.

In this experiment, I explore how social desirability bias effects, as witnessed in human-computer interactions, hold up over time. I conducted a randomized experiment, in which participants completed an assessment about their exercise behavior (walking), each day, for six weeks. To fully explore how the computers as social actors paradigm affects social desirability bias, I manipulated the embodiment of the interface: the assessment was either conducted through dialogue with an ECA or via displayed text (Figure 4-1). I analyzed how interface embodiment effected both system engagement: how often participants used the system, and data quality: the accuracy of the self-reported data. The self-reported data was corroborated with data collected via a wearable sensor worn by participants throughout the course of the experiment.

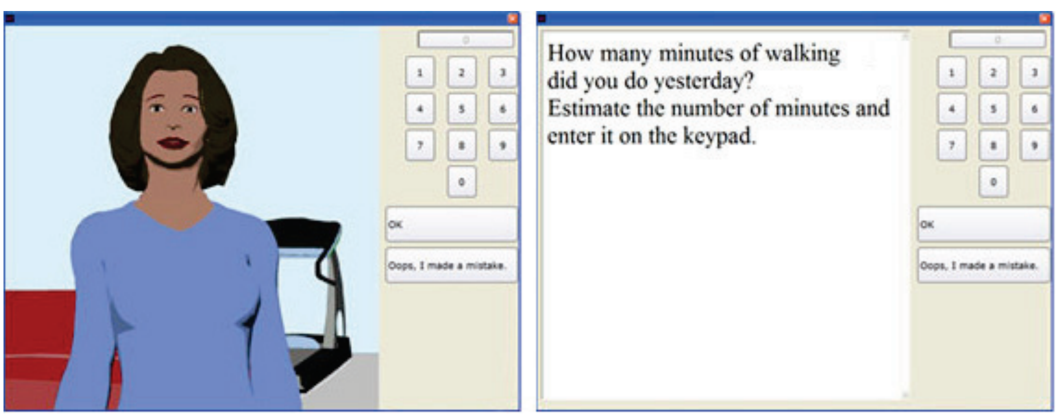


Figure 4-1: AGENT and TEXT conditions

4.2 Experimental Design

Participants in this study were part of the "Virtual Laboratory" system, in which a standing group of participants interact with an ECA from their home computers, and the ECA can be manipulated remotely to periodically implement new experiments [7]. This experiment ran for six weeks. The ECA designed for this experiment emulated an exercise counselor and promoted walking behavior. The daily conversation with the virtual exercise counselor took place on participants' home computers and lasted approximately 10 minutes. It included a discussion about the participant's walking behavior, and problem-solved any barriers to exercise. After the conversation with the virtual counselor was over, participants were given the opportunity to complete a single-item assessment, self-reporting the number of minutes they walked on the previous day.

For the self-report assessment, participants were randomized into one of two conditions, either completing the assessment via a continued conversation with their virtual counselor (AGENT) or via displayed text (TEXT). At the end of the exercise counseling session with the ECA, the participants were asked - either spoken by the virtual counselor (AGENT) or via displayed text (TEXT) - "Do you have time for one more question?" If the participant responded yes, the participant was asked "How many minutes of walking did you do yesterday?" and the system displayed a keypad for the participant to enter their answer (Figure 4-1). This completed the participant's daily interaction with the system. A total of 25 participants (23 women, ages 55-68) took part in the study, and participants were paid a dollar each day that they interacted with the system.

4.3 Results

I analyzed engagement (system usage) and data quality (self-report accuracy) as binary outcomes, by fitting a logistic mixed- effect regression model to the data. Logistic mixed-effect regression is a generalization of logistic regression, suitable for analyzing repeated binary measurements [35]. All analyses were performed using R 2.9.1 with the lme4 package [3, 62].

The experiment produced three main findings. The first is that actual walking behavior significantly impacted engagement with the system. The more a participant walked, the more likely they were to initiate a system interaction. Furthermore, when I examined the longitudinal nature of this effect, I found that it *decreases* for the TEXT condition and *increases* for the AGENT condition.

For this analysis, I used a logistic mixed-effect regression model that included fixed effects of study day, condition and the previous day's pedometer step count, along with corresponding interaction effects (Table 4-1). There was a significant three-way interaction between the pedometer step count, condition and study day on system use. For both conditions throughout the study, the more a participant walked the more likely they were to interact with the system the following day. However, the longitudinal strength of this effect was significantly different between conditions: it increased for members of the AGENT condition and decreased for members of the TEXT condition. This finding is visualized in Figure 4-2. By the end of the study, the amount of walking done by a participant in the TEXT condition had little effect on their likelihood to use the system. In contrast, the amount of walking done by participants in the AGENT condition greatly determined the likelihood that they would interact with the system.

Random Effects	Std. Dev.
Intercept	0.8457

Fixed Effects	Estimate	Std. Error	p
Intercept	0.5858	0.3306	0.0764
Condition	-0.6632	0.4699	0.1581
Steps	0.1788	0.3055	0.5583
Study Day	0.0125	0.0103	0.2243
Cond*Steps	0.5917	0.4062	0.1453
Cond*Day	0.0074	0.0139	0.5940
Steps*Day	0.0300	0.0139	0.0306
Cond*Steps*Day	-0.0422	0.0172	0.0140

Table 4-1: Mixed-Effect Regression Estimate of Effects of Condition, Previous Day's Step Count and Study Day on System Use. Condition 0=AGENT, 1=TEXT

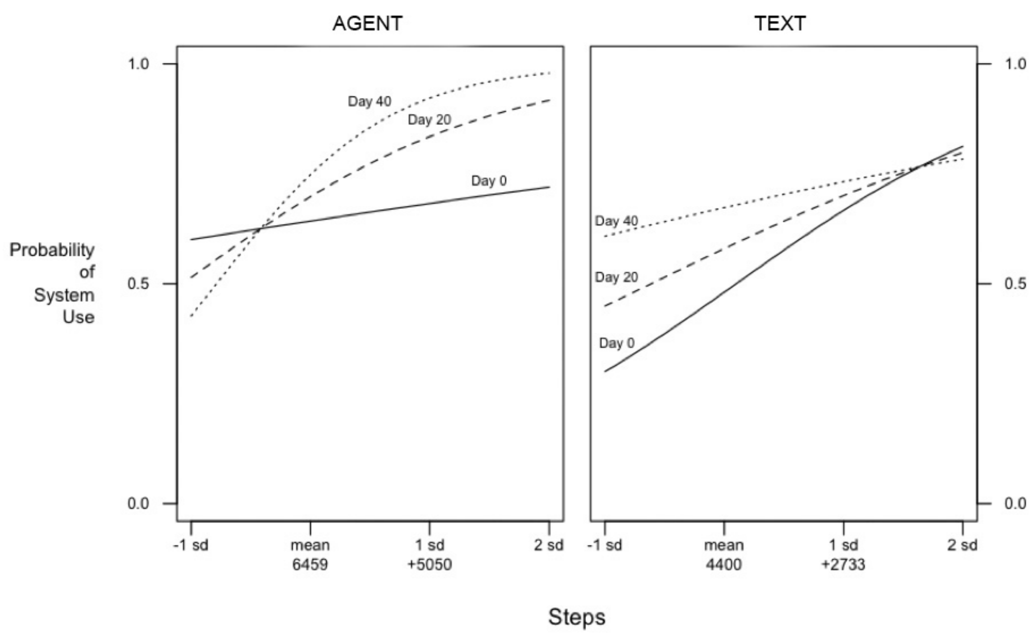


Figure 4-2: Effects of steps walked (pedometer-measured) on system use over time. Steps are mean-centered, by condition. Actual walking behavior influenced system use. This effect diminished over time for participants in the TEXT condition but grew stronger over time for participants in the AGENT condition.

The second finding is that when participants did initiate a session with the system, members of the TEXT condition were more likely than members of the AGENT condition to complete the self-report assessment at the end of the session.

As with the previous analysis, I used a model that included fixed effects of study day, condition and the previous day's pedometer step count (no interaction effects were present). As shown in Table 4-2, members of the TEXT condition were more likely than members of the AGENT condition to self-report their minutes of walking. The amount of actual time walking and the day within the study did not have an effect.

Random Effects	Std. Dev		
Intercept	1.4963		
Fixed Effects	**Estimate**	**Std. Error**	**p**
Intercept	0.8751	0.4938	0.0764
Condition	1.8173	0.6987	0.0093
Steps	0.0716	0.1787	0.6886
Study Day	0.0044	0.0096	0.6439

Table 4-2: Mixed-Effect Regression Estimate of Effects of Condition, Previous Day's Step Count, and Study Day on Self-Report. Condition 0 = AGENT, 1 = TEXT.

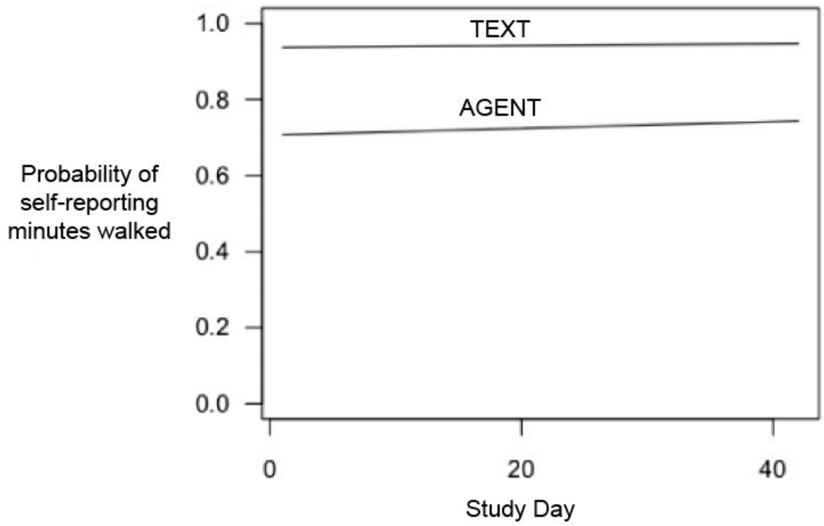


Figure 4-3: Probability, over time, of self-reporting data, once a session has started

Finally, the analysis showed that when participants did complete self-report assessments, data given by members of the AGENT condition were more accurate than data given by members of the TEXT condition.

For both conditions, I found a significant positive correlation between the pedometer readings and the self-reported minutes of walking. A stronger correlation was present for the AGENT condition, $r = 0.75$, $p < 0.001$; [95% CI, 0.69-0.81], vs. the TEXT condition: $r = 0.50$, $p < 0.001$; [95% CI, 0.40-0.59].

4.4 Discussion

This study is the first to examine how social responses to computers change over time. By examining repeated interactions with a system, we now have an insight into how the Computers as Social Actors paradigm extends into a longitudinal context. There is now evidence that social responses to computers are not static, but change with time.

This study leaves some questions unanswered, and provides directions for further research. Why were people more likely to self-report to the TEXT system? Since the TEXT system took a slightly shorter amount of time to complete, was it simply a matter of time or effort? Also, participants did not benefit in any way by completing the self-report assessment. There was no incentive to provide the data. Do the results change if the participant receives a direct benefit? What types of incentives have the greatest impact on system use? Furthermore, do incentives have impacts beyond system use, i.e., can they improve data quality?

I will examine these directions in an additional longitudinal study, specifically designed to examine the effect of incentives on system engagement and data quality.

CHAPTER 5

Interaction Value and Self-Reported Health Data

5.1 Introduction

The results from the previous chapter might suggest that ECAs are a less-than-ideal choice for use in longitudinal health systems designed to collect self-reported data. I have shown that ECAs are susceptible to social desirability bias effects, and for this second study, I will explore if these bias effects can be overcome by other manipulations within the interface. One of the most common approaches to increase engagement in long-term system interactions is the use of incentives.

I am motivated by ideas from Behavioral Economics and the study of incentives (Section 2.8). Financial incentives are often powerful motivators, and heavily applied in the field of Behavioral Economics. For this study, I will experiment with applying financial incentives to using the technology itself, with the goal of promoting long-term engagement with the system.

Since this work is an early exploration into how incentives impact system use, I will explore not only financial incentives, but also personal incentives - tailored health messages (Section 2.9). These incentives will be in the form of personally relevant feedback given to users at the end of each session. By examining the theories behind behavioral economics and tailored health messages side-by-side, I will be able to see the effects of each approach, and discover which approach holds a more powerful incentive to continue to engage with a technology.

I am also motivated by the application of economic ideas to sociology, in particular, I am interested in exploring if the theories describing how people form and continue relationships with other people – social exchange theory (Section 2.7) – are also applicable to the manner in which people build and form relationships with technological artifacts. In this second experiment, I will examine a cost/benefit manipulation, examining if the personal benefits (monetary and personal feedback) can outweigh perceived costs (efficiency of the interaction).

For this experiment, I will be recruiting from the university student population and as such, focus on a health issue extremely relevant to students: alcohol consumption [37]. Participants will have an online interview about their alcohol consumption, once a week, for up to 16 weeks. Each interview will consist of up to 28 questions regarding alcohol consumption, consequences of alcohol use (missing a class, etc.), and protective behaviors taken (pacing drinks, etc.). This experiment extends the previous study in the following ways:

1. The main focus of this experiment will be to systematically manipulate incentives for completing the self-report assessment. Furthermore, two distinct incentives will be evaluated: monetary reward and personalized feedback.

2. Secondarily, I will also track the time/effort that it takes to interact with the system and will measure its effect, and examine the impact of reminders or triggers to use the system. Also, I will study a younger population (college students rather than older adults) and examine a different frequency of contact with the system (weekly instead of daily).

5.2 Experimental Design

This will be a 2 (AGENT vs. TEXT FORM) x 2 (INCENTIVE REMINDERS vs. NO INCENTIVE REMINDERS) x 2 (FEEDBACK vs. NO FEEDBACK), between-subjects design. Participants will be randomized to either have the online assessment with an animated agent, or receive a standard web survey (Figures Figure 5-1,Figure 5-2). The agent for this project was pre-selected by a convenience sample of thirteen participants, choosing from four possible agents (Appendix C).

Participants will also be randomized to either receive reminders at the time of each login about the weekly prize drawing (a $20 gift certificate to amazon.com), or will not receive any reminders about the drawing after the initial disclosure during consent and enrollment (FiguresFigure 5-3, Figure 5-4).

Lastly, participants will be randomized to either receive personal feedback about their alcohol consumption (calories consumed, dollars spent) at the end of each interaction, or receive no feedback (Figures Figure 5-5,Figure 5-6). The content for the personalized feedback was iteratively developed in collaboration with a domain expert at the Office of Prevention and Education at Northeastern (OPEN)[3].

[3] http://www.northeastern.edu/open/

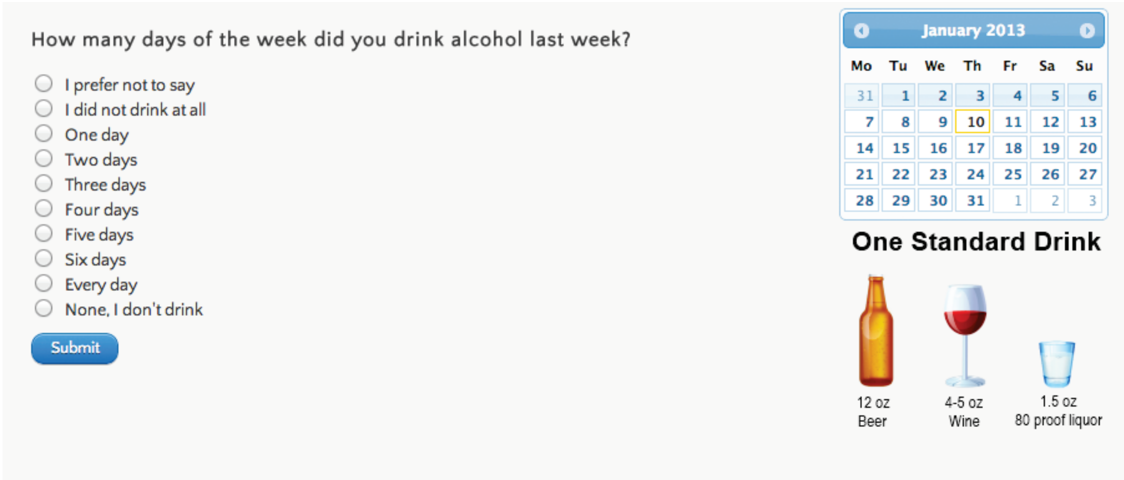


Figure 5-1: TEXT interface

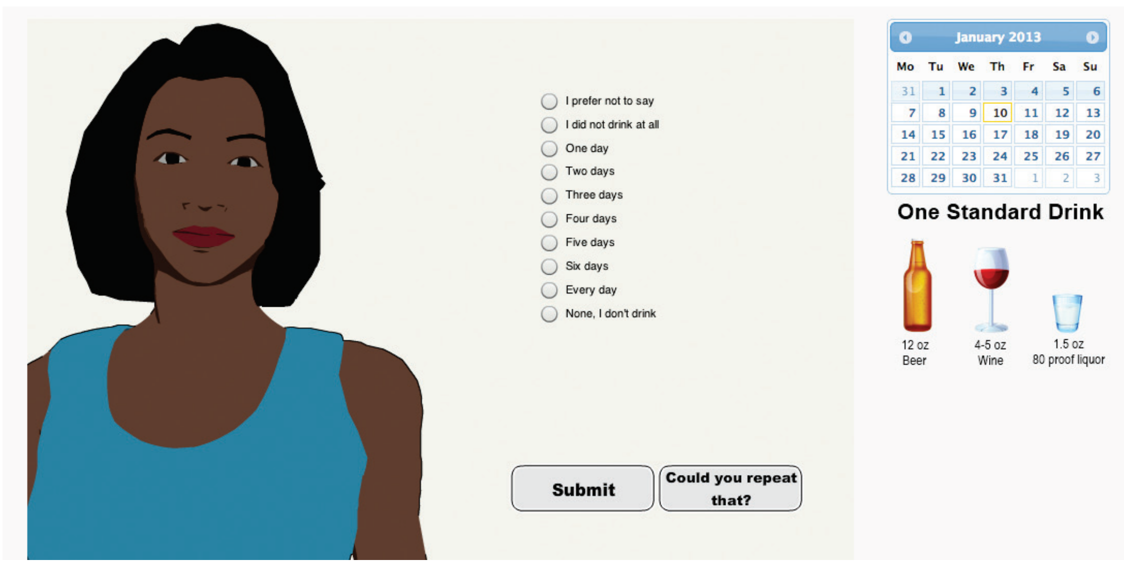


Figure 5-2: AGENT interface

Thank you for participating in the study. As a reminder, go to http://wonder.ccs.neu.edu/survey to complete the survey. You can participate up to once-a-week, with new weeks starting each Monday. And for each completed survey, you will be entered in a drawing for a $25 gift certificate to Amazon.com!

The Drinks Project team,

http://wonder.ccs.neu.edu/survey

Figure 5-3: Monetary incentive reminder, included in the weekly study email.

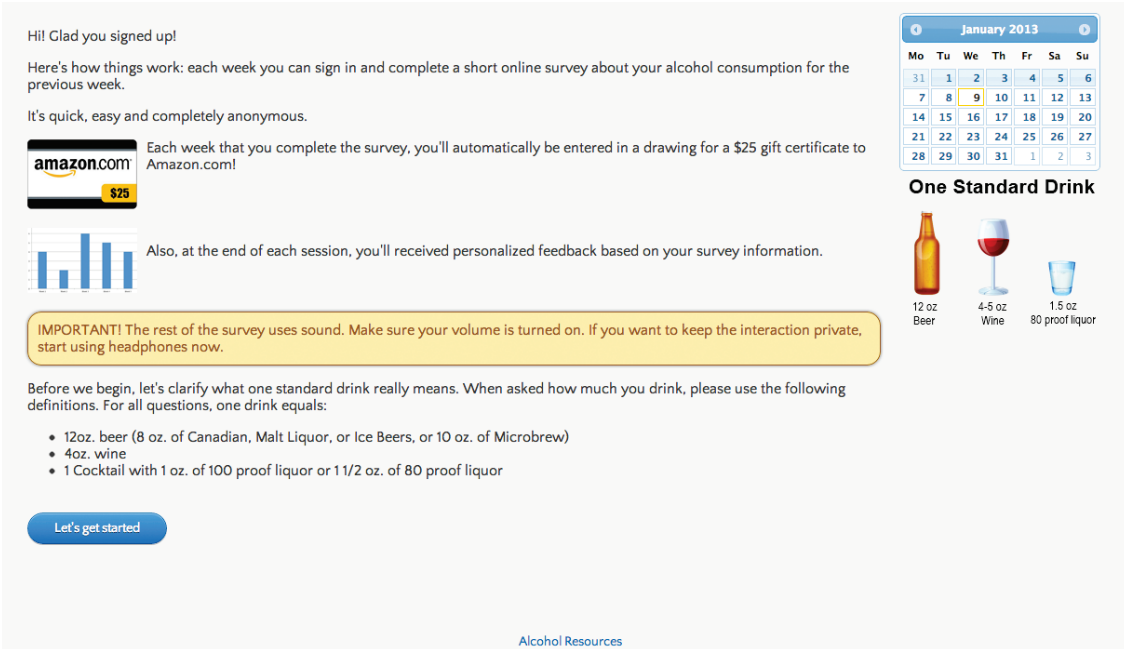


Figure 5-4: Monetary incentive reminders, present at the beginning of each session.

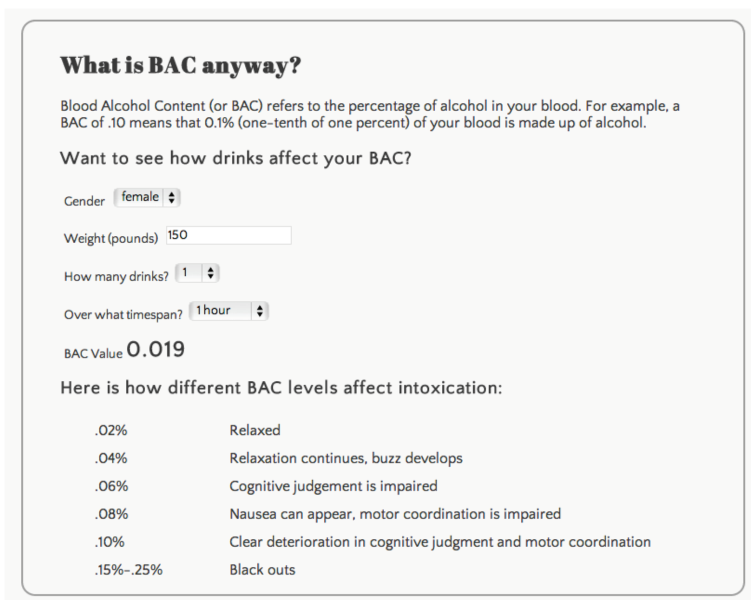


Figure 5-5: Sample end-of-session feedback (this information varied each week).

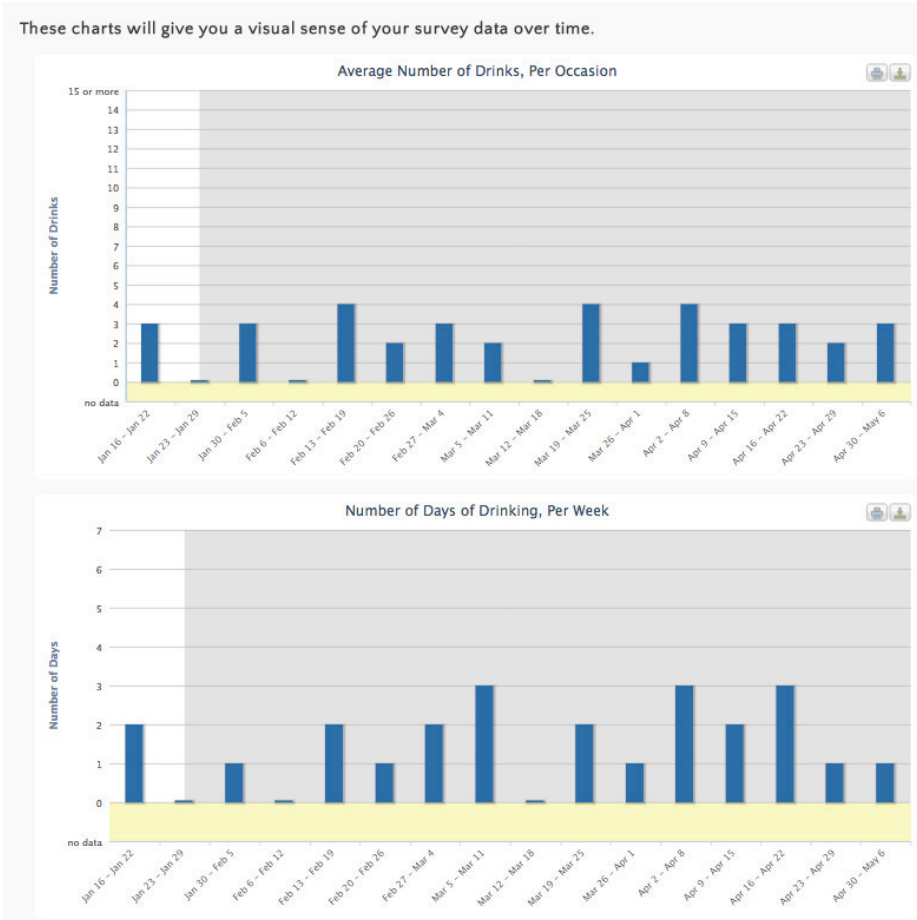


Figure 5-6: Sample feedback chart data, updated weekly with the latest participant-reported data.

5.2.1 Measures

The experiment included a number of outcome measures to examine the factors of engagement and self-report quality. The first measure was the set of questions used for each interaction, developed and used by the Office of Prevention and Education at Northeastern (OPEN). Unlike the previous study described in CHAPTER 4 , I did not have a sensor-based source of truth with which I can compare the participants' self-report answers. Instead, I focused on the amounts of alcohol consumed, the number of negative consequences and the number of protective behaviors. With randomized study conditions, the self-report behaviors should not be significantly different. I followed the standard approach that increased levels of self-reporting indicate more accurate answers.

Another outcome measure was the number of interactions that each participant has over the course of the 16-week study. This provides a measure of engagement. I also modeled and predicted how the likelihood of a participant completing the weekly survey changes over time.

During each interaction, I measured the amount of time it takes for participants to answer individual questions and complete the entire session. This provides a notion of cost to the user, which may interact with the various study conditions and effect system engagement.

Finally, I assessed attitudes towards the system through a follow-up questionnaire sent to all participants at the end of the study (6.1Appendix F).

5.2.1.1 Power Analysis

Based on engagement data from the experiment described in Chapter 4, in order to show the full statistical model of longitudinal usage differences (intercept and slope differences) a parametric bootstrap analysis [49] indicates that enrolling 375 participants will provide 94% power with a 2-sided α of 0.05, and 84% power with a 2-sided α of 0.01 (Appendix E).

5.3 Participants

Participants were recruited via emails, student newspaper ads, handouts and flyers displayed around Northeastern University's campus (Appendix B). In order to be eligible for the study, participants needed to be:

1. Age 18 or older
2. Currently enrolled as a student at Northeastern University
3. Have a valid husky.neu.edu email account
4. Have access to a computer with Internet and audio capabilities, usable for private purposes.

Three hundred and seventy-five participants enrolled in the study and were randomized into one of the eight possible conditions. All participants completed an unsigned consent process (Appendix A). Participant demographics are shown in Table 5-1.

		Agent	No				Yes			
		Incentive Reminder	No		Yes		No		Yes	
		Personalized Feedback	No	Yes	No	Yes	No	Yes	No	Yes
Women	249		33	30	33	27	33	32	31	30
Men	126		13	17	14	19	14	16	16	17
Freshman	103		15	16	14	13	11	19	14	10
Sophomore	69		5	9	7	8	9	16	4	11
Middler	50		7	9	4	6	7	7	5	4
Junior	39		9	3	5	4	7	2	5	3
Senior	62		6	5	13	8	9	4	9	8
Grad Student	52		4	5	4	7	4	9	8	11
Ages 18-20	206		25	29	23	26	25	29	24	25
21-23	134		19	16	20	16	19	11	19	14
24+	35		2	2	4	4	3	8	4	8
Drinkers	251		34	31	38	36	28	31	25	28
Non-drinkers	124		12	16	9	10	19	17	22	19

Table 5-1: Participant Demographics

5.4 Results

5.4.1 Survey Completion Activity

We will first examine differences in the overall number of completed sessions among experimental groups. For the analysis of count data (total number of completed sessions), a Poisson model is often considered. However, after examining the data from this experiment, the conditional variances are much higher than the conditional means, suggesting that a Poisson model would not be appropriate. Instead a negative binomial regression will be used. Negative binomial regression is a generalization of a Poisson regression – it has the same structure with an extra parameter to model the over-dispersion (the conditional variances not equaling the conditional means) [36]. All negative binomial regression analysis was performed in R 2.15.2, using the MASS 7.3-22 package [62, 72].

Overall completion rates are shown in Table 5-2. Eighty-one percent of participants completed the first weekly survey. An additional 52 participants (14%) logged in, but never completed the first weekly survey. Subsequent weeks had lower rates of completion, with only 5% (n=18) of participants completing the survey at the final week of the study.

Week	Number of Completed Surveys	%
1	302	81
2	118	31
3	119	32
4	98	26
5	76	20
6	71	19
7	72	19
8	68	18
9	59	16
10	55	15
11	43	11
12	39	10
13	35	9
14	39	10
15	21	6
16	18	5

Table 5-2: Overall survey completion rates, per week.

5.4.1.1 Survey Presentation: Agent vs. Text

Based on previous findings, I hypothesized that participants in the TEXT condition will complete more weekly sessions than participants in the AGENT condition. Furthermore, I hypothesized that this effect could be mediated by the time it takes to complete each session.

Differences in the time needed to complete sessions between the TEXT/AGENT groups are shown in Table 5-3.

	Agent	
	No	Yes
Time to complete a session (in seconds) M (SD)	218.18 (121.73)	246.63 (125.45)

Table 5-3: Average time to complete a session, in seconds. Agent vs. Text

Summary statistics in Table 5-4 indicate differences in completion rates between the AGENT and TEXT groups. A histogram and density plot also show visual differences between groups (Figures Figure 5-7 and Figure 5-8).

	Agent	
	No	Yes
Completed Sessions M (SD)	3.74 (4.27)	2.79 (3.86)

Table 5-4: Average number of completed sessions: Agent vs. Text (max possible sessions = 16)

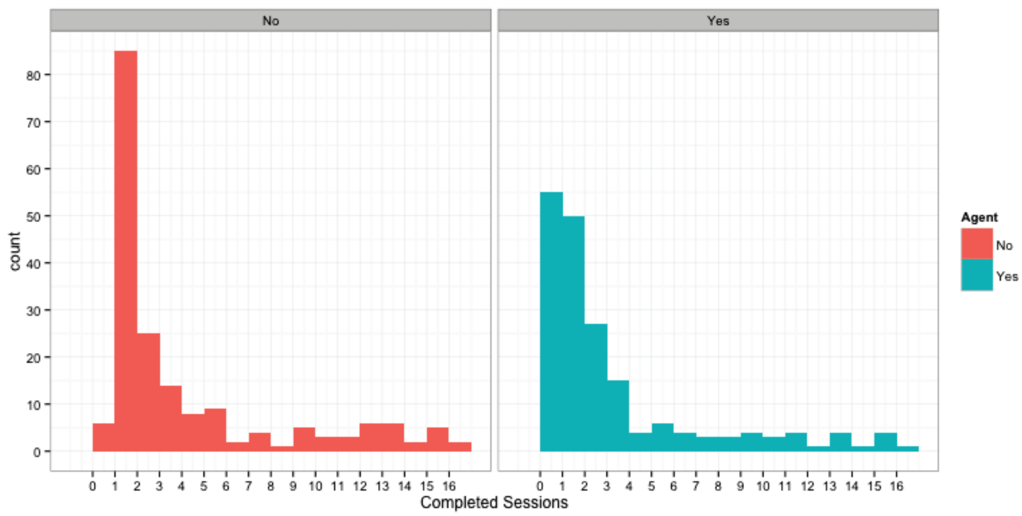


Figure 5-7: Histogram showing number of completed sessions: Agent vs. Text.

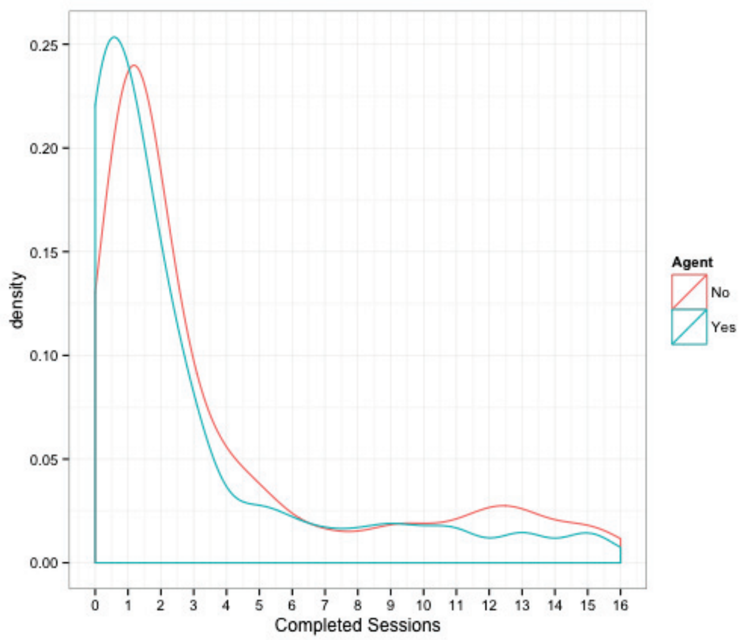


Figure 5-8: Density plot showing number of completed sessions: Agent vs. Text

We model a negative binomial regression with survey presentation (AGENT/TEXT) and time taken to complete the first week's survey (in seconds) as predictors. Table 5-5 shows the estimated negative binomial regression. The variable *AgentYes* is the expected difference in log count between the group receiving the AGENT interface, and the group receiving the TEXT interface. The expected log count is 0.429 times higher for the AGENT group, not significant. There is a significant interaction between survey presentation and the time it took participants to complete their first survey. Table 5-6 expresses these variables as Incidence Rate Ratios, IRR, rather than model coefficients. For each second increase in the time needed to complete the survey, the AGENT participants had a lower rate of completion than the TEXT participants (incidence rate ratio, 0.994 [95% CI, 0.996 to 0.999], p=0.0190).

<table>
<tr><td colspan="6"><code>glm.nb(completed.total ~ Agent * week0time, data=plotdata)</code></td></tr>
<tr><td>Coefficients:</td><td>Estimate</td><td>95% CI</td><td>Std. Error</td><td>z value</td><td>Pr(>|z|)</td></tr>
<tr><td>(Intercept)</td><td>1.489</td><td>1.20 – 1.78</td><td>0.163</td><td>9.157</td><td><2e-16 ***</td></tr>
<tr><td>AgentYes</td><td>0.429</td><td>-0.04 – 0.91</td><td>0.249</td><td>1.719</td><td>0.0855 .</td></tr>
<tr><td>week0time</td><td>-0.001</td><td>-0.002 – 0.001</td><td>0.001</td><td>-0.875</td><td>0.3815</td></tr>
<tr><td>AgentYes:week0time</td><td>-0.002</td><td>-0.004 – 0.00</td><td>0.001</td><td>-2.345</td><td>0.0190 *</td></tr>
<tr><td colspan="6">Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1</td></tr>
</table>

Table 5-5: Negative binomial regression examining survey presentation and time needed for completion as predictors of completion rates.

Incidence Rate Ratios:	Estimate	95% CI
(Intercept)	4.43	3.31 – 5.95
AgentYes	1.54	0.96 – 2.48
week0time	0.994	0.998 – 1.00
AgentYes:week0time	0.998	0.996 – 0.999

Table 5-6: Regression coefficients from Table 5-5, expressed as Incidence Rate Ratios

To further understand this effect, we can use this model to predict the number of completed surveys, for our AGENT/TEXT groups, based on the time needed to complete the survey. The model predictions, along with 95% confidence intervals, are visualized in Figure 5-9.

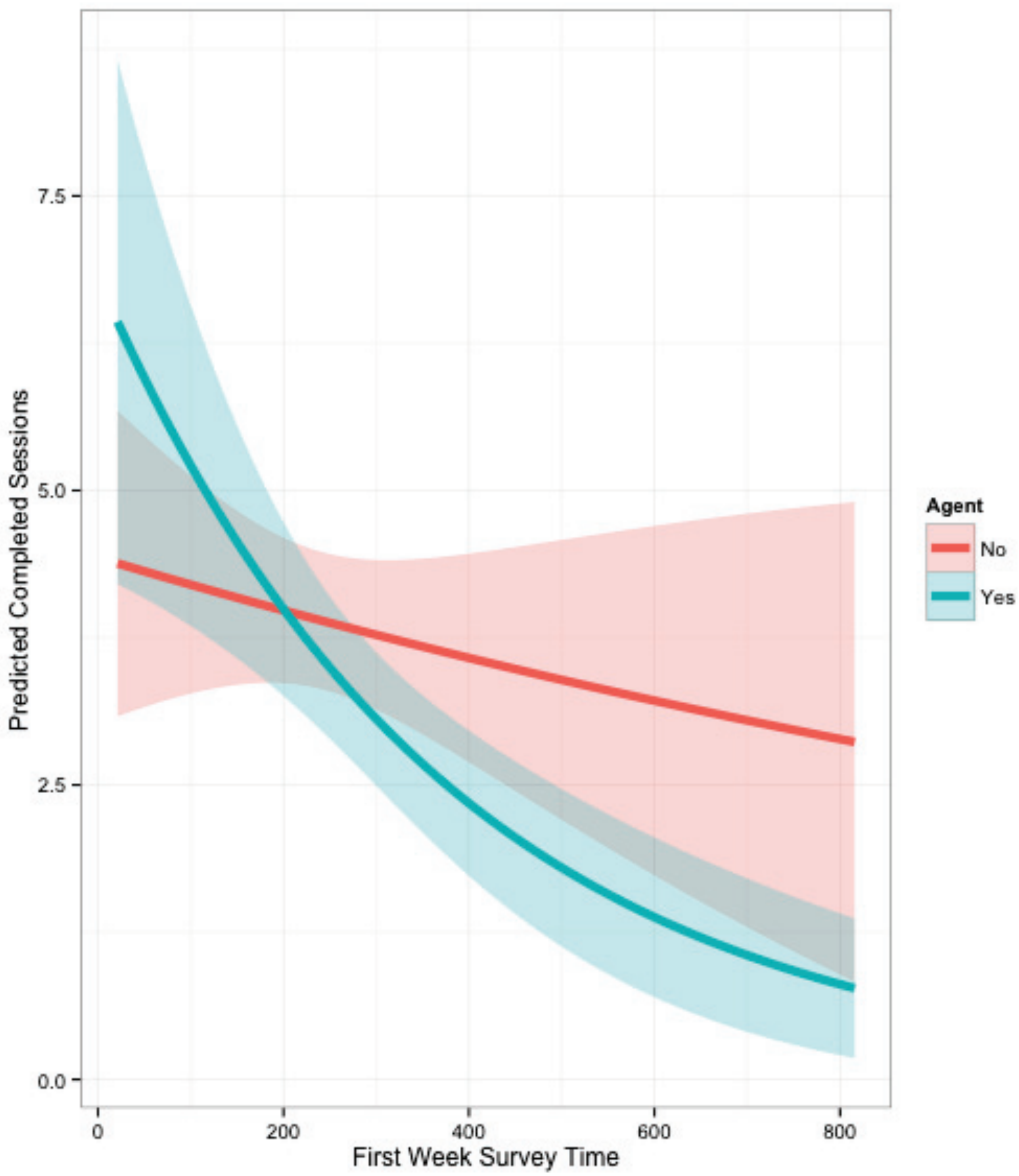


Figure 5-9: Predicted number of completed sessions and 95% confidence intervals, by survey presentation and survey time (in seconds). Note that the lines are curved because this is a log linear model. Expected values are plotted, not the log of expected values.

We find support for our hypothesis that the differences in AGENT/TEXT completion rates are moderated by the time needed to complete the survey. We see that for sessions under 200 seconds, there is essentially no difference in completion rates between the AGENT and TEXT groups. However, for the AGENT group, as the survey time increases, the total number of completed surveys throughout the study decreases. Put another way, the longer it takes a participant in the AGENT group to complete the survey during their first week, fewer total sessions will be completed over the course of the study.

5.4.1.2 Feedback Incentives

I hypothesized that participants in the FEEDBACK condition (receiving personalized feedback at the end of each session, based on their survey data) will complete more weekly sessions than participants in the NO-FEEDBACK condition. This hypothesis is informed by previous work in tailored health messages (section 2.9).

Summary statistics in Table 5-7 do not indicate strong differences in total completed sessions between those receiving and those not receiving personalized feedback at the end of each session. A histogram and density plot examining total completed sessions, by feedback group are shown in Figures Figure 5-10: Histogram showing number of completed sessions: Feedback vs. No Feedback Figure 5-11: Density plot showing number of completed sessions: Feedback vs. No Feedback.

	Feedback	
	No	Yes
Completed Sessions M (SD)	3.14 (3.98)	3.38 (4.2)

Table 5-7: Summary Statistics: Feedback vs. No Feedback (max possible sessions = 16)

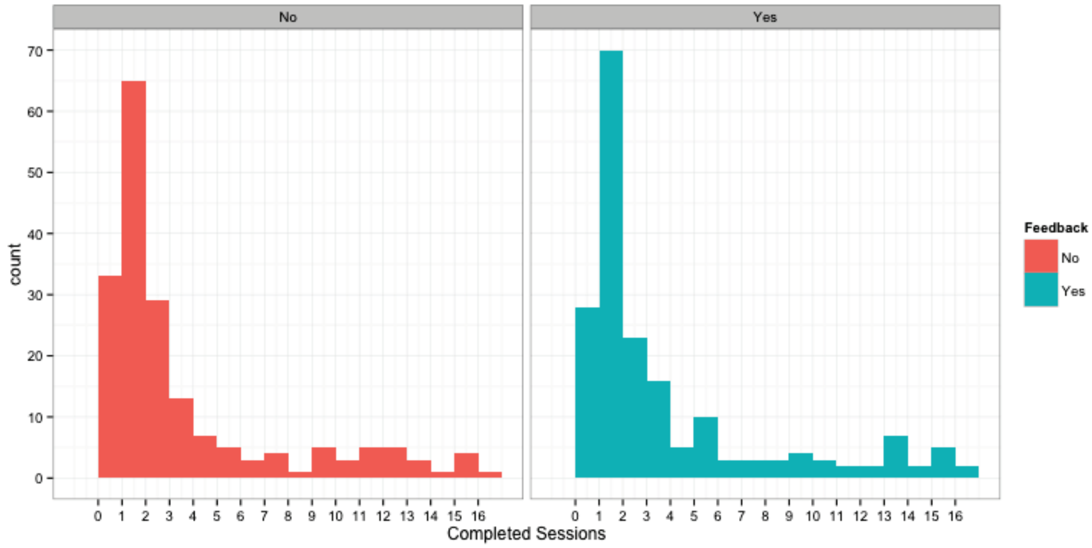


Figure 5-10: Histogram showing number of completed sessions: Feedback vs. No Feedback

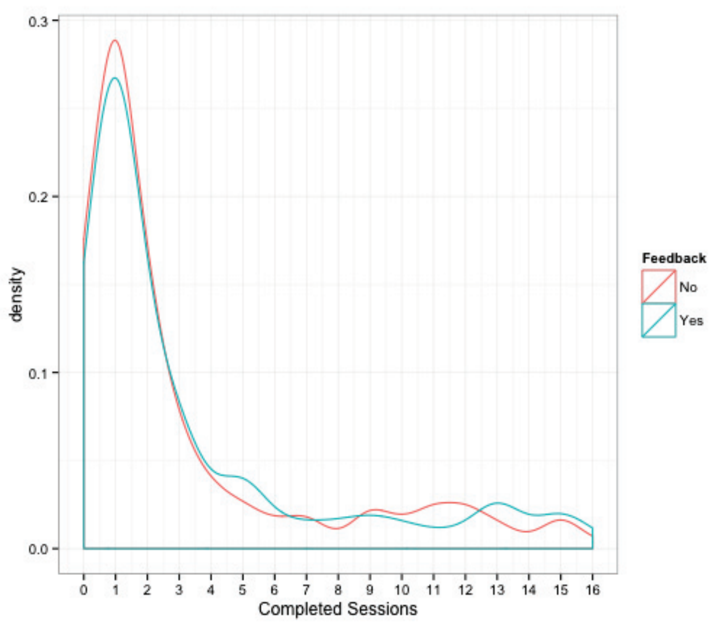


Figure 5-11: Density plot showing number of completed sessions: Feedback vs. No Feedback

Table 5-8 shows the estimated negative binomial regression and Table 5-9 the corresponding Incidence Rate Ratios. The expected log count for the number of completed sessions is 0.07 times higher for the FEEDBACK group, not significant.

glm.nb(completed.total ~ Feedback, data=plotdata)					
Coefficients:	Estimate	95% CI	Std. Error	z value	Pr(>|z|)
(Intercept)	1.15562	0.98 – 1.32	0.08599	13.322	<2e-16 ***
FeedbackYes	0.07156	-0.17 – 0.31	0.12097	0.592	0.554
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 5-8: Negative binomial regression examining feedback as a predictor of completion rates.

Incidence Rate Ratios:	Estimate	95% CI
(Intercept)	3.14	2.66 – 3.73
FeedbackYes	1.07	0.85 – 1.36

Table 5-9: Regression coefficients from Table 5-8, expressed as Incidence Rate Ratios

Participants receiving personalized feedback based on their survey answers did not complete a greater number of sessions than participants receiving no feedback at all. Thus, we do not see support for our hypothesis.

5.4.1.3 Monetary Incentives

We hypothesize that participants in the MONETARY condition (those being reminded about the weekly drawing for a $25 gift cart to Amazon.com) will complete more weekly sessions than participants in the NO-MONETARY condition. This hypothesis is based on the findings regarding financial incentives in behavioral economics (section 2.8), combined with past research on triggers for technological use (section 2.6).

Summary statistics in Table 5-10 do not indicate strong differences in total completed sessions between monetary groups. A histogram and density plot examining total completed sessions, by monetary group are shown in Figures Figure 5-12 and Figure 5-13.

	Monetary Incentive Reminders	
	No	Yes
Completed Sessions M (SD)	3.24 (4.06)	3.28 (4.13)

Table 5-10: Summary statistics: Monetary Incentives vs. No Monetary Incentives (max possible sessions = 16)

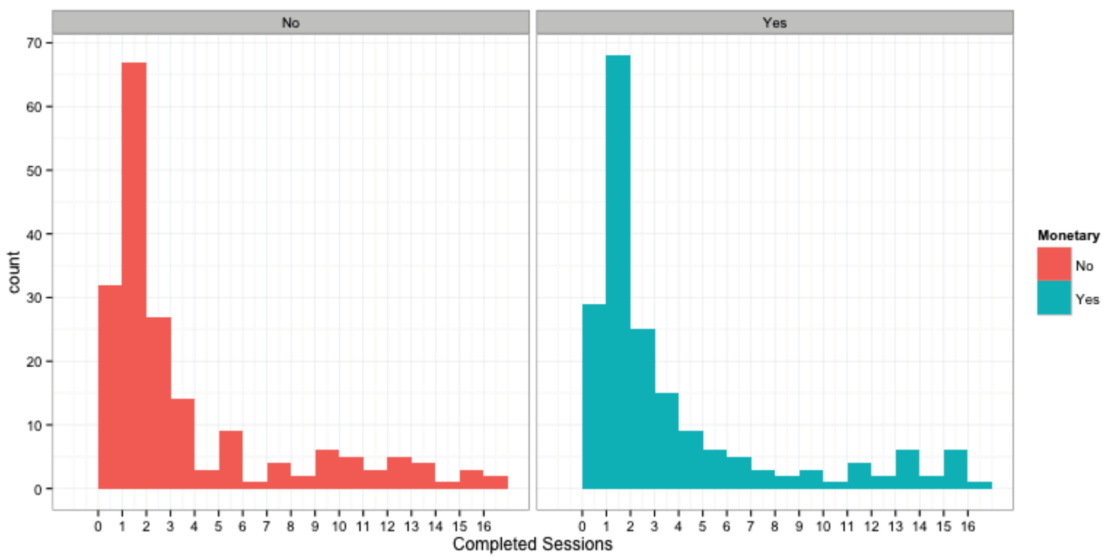


Figure 5-12: Histogram showing number of completed sessions: Agent vs. Text

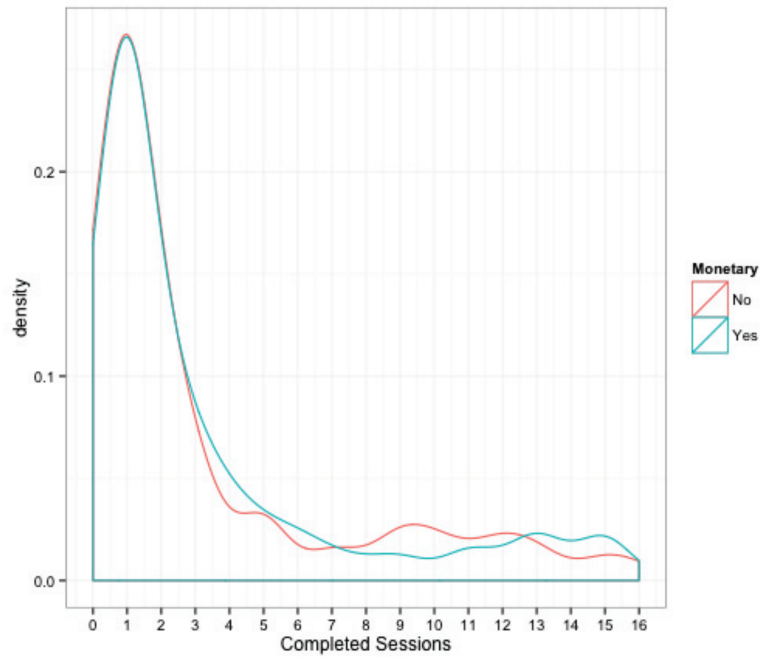


Figure 5-13: Density plot showing number of completed sessions: Agent vs. Text

Tables Table 5-11 and Table 5-12 show the estimated negative binomial regression and the corresponding Incidence Rate Ratios. The expected log count for the number of completed sessions is 0.01 times higher for the MONETARY group, not significant.

Participants who were reminded about the weekly monetary incentive did not complete a greater number of sessions than participants receiving no mention of the monetary incentive (other than during the enrollment process). Thus, we do not see support for our hypothesis.

`glm.nb(completed.total ~ Monetary, data=plotdata)`					
Coefficients:	Estimate	95% CI	Std. Error	z value	Pr(>|z|)
(Intercept)	1.17702	1.01 – 1.35	0.08550	13.766	<2e-16 ***
MonetaryYes	0.01024	-0.23 – 0.25	0.12101	0.085	0.933
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 5-11: Negative binomial regression examing monetary incentives as a predictor of completion rates.

Incidence Rate Ratios:	Estimate	95% CI

(Intercept)	3.25	2.75 – 3.84
MonetaryYes	1.01	0.80 – 1.28

Table 5-12. Regression coefficients from Table 5-11, expressed as Incidence Rate Ratios

5.4.1.4 Interactions Among Experimental Conditions

I hypothesized that participants in the TEXT-FEEDBACK-MONETARY condition will complete the most weekly sessions (due to low-cost / high-benefit classification) and participants in the AGENT-NO-FEEDBACK-NO-MONETARY condition will complete the least number of weekly sessions (due to high- cost / low-benefit classification). This hypothesis is informed by social exchange theory (section 2.7).

Summary statistics in Table 5-13 show completion rates among all experimental conditions. We can see that for the AGENT group, those receiving personal feedback did have the highest number of sessions, on average. However, an estimated negative binomial regression shows no significant differences among groups. When analyzing the AGENT group separately, there was no relationship between the addition of incentives and a higher number of total sessions. The same was true for a sub-analysis of the TEXT group. Thus, we do not see support for our hypothesis.

Agent	No				Yes			
Monetary	No		Yes		No		Yes	
Feedback	No	Yes	No	Yes	No	Yes	No	Yes
Completed Sessions M (SD)	3.48 (3.86)	4.00 (4.68)	3.68 (4.05)	3.78 (4.57)	3.00 (4.37)	2.52 (3.18)	2.43 (3.60)	3.23 (4.23)

Table 5-13: Summary statistics: interactions among experimental conditions (max possible sessions = 16)

5.4.1.5 *Drinkers vs. Non-Drinkers*

We are also able to examine the effect that being a drinker or non-drinker may have on number of completed sessions. A participant was classified as a drinker if they reported at least one incidence of drinking throughout their 16 weeks in the study. Forty-eight participants could not be classified, due to their lack of responses.

Summary statistics in Table 5-14 indicate differences in total completed sessions between groups, with non-drinkers completing more sessions than drinkers, on average. Furthermore, this effect appears to be moderated by survey presentation (Table 5-15). A histogram and density plot examining total completed sessions, by drinker categorization, are shown in Figures Figure 5-14 and Figure 5-15.

	Drinker	
	No (n=80)	Yes (n=247)
Completed Sessions M (SD)	4.31 (4.73)	3.55 (3.97)

Table 5-14: Average number of completed sessions: Drinker vs. Non-Drinker (max possible sessions = 16)

Agent	No		Yes	
Drinker	No (n=42)	Yes (n=138)	No (n=38)	Yes (n=109)
Completed Sessions M (SD)	3.00 (4.07)	4.12 (4.33)	5.76 (5.03)	2.82 (3.34)

Table 5-15: Average number of completed sessions: Drinker x Agent (max possible sessions = 16)

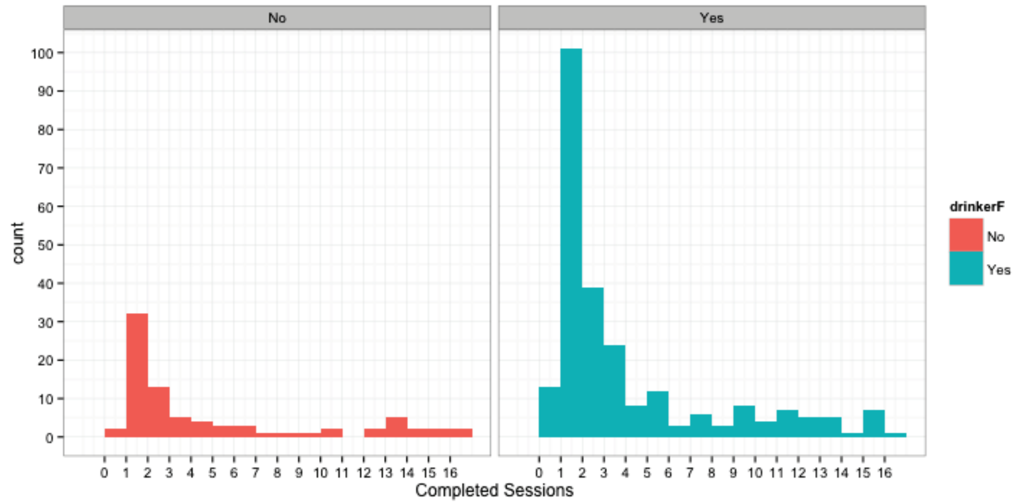


Figure 5-14: Histogram showing number of completed sessions: Drinker vs. Non-Drinker

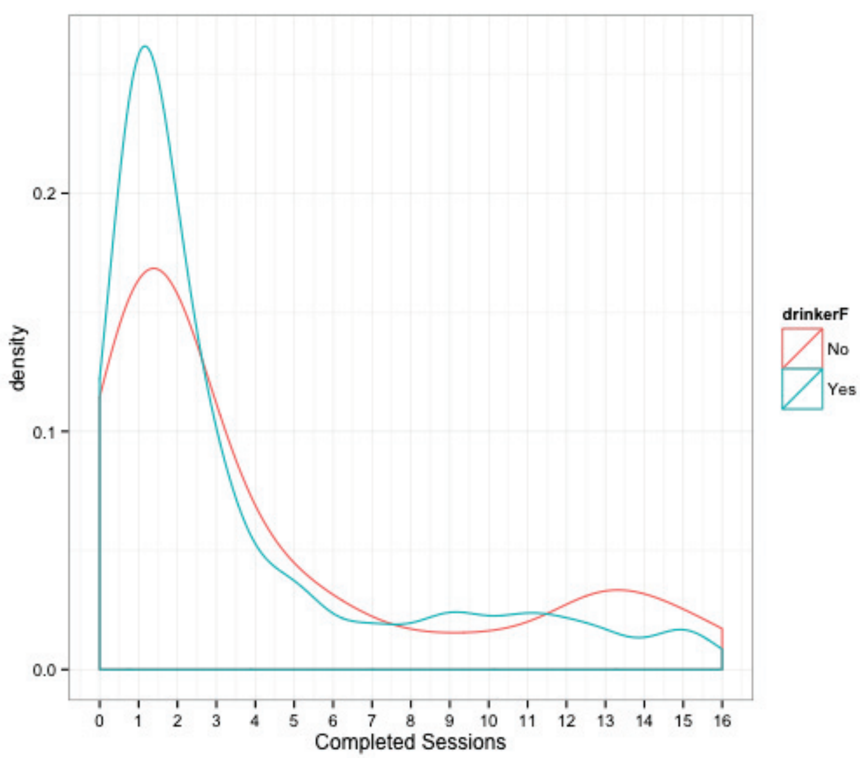


Figure 5-15: Density Plot showing number of completed sessions: Drinker vs. Non-Drinker

We model a negative binomial regression with drinker categorization (yes/no) and survey presentation (AGENT/TEXT) as predictors. Table 5-16: Negative binomial regression examining Drinker, Agent as predictors of completion rates shows the estimated negative binomial regression and Table 5-17 shows the Incidence Rate Ratios. There is a significant interaction between survey presentation and drinker categorization. For drinkers, the AGENT participants had a lower number of completed surveys than the TEXT participants (incidence rate ratio, 0.3555 [95% CI, 0.2153 to 0.5858], $p < 0.001$).

We can use this model to predict the number of completed surveys, for our AGENT/TEXT participants, based on whether or not they are a drinker. The model predictions, along with 95% confidence intervals, are visualized in Figure 5-16.

`glm.nb(completed.total ~ Drinker*Agent, data=drinkerdata)`					
Coefficients:	Estimate	95% CI	Std. Error	z value	Pr(>|z|)
(Intercept)	1.0956	0.7942 – 1.4146	0.1579	6.956	3.49e-12 ***
DrinkerYes	0.3180	-0.0365 – 0.6643	0.1785	1.781	0.07487 .
AgentYes	0.6529	0.2228 – 1.0857	0.2198	2.971	0.00297 **
DrinkerYes:AgentYes	-1.0340	-1.5353 – -0.5346	0.2550	-4.054	5.03e-05 ***
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 5-16: Negative binomial regression examining Drinker, Agent as predictors of completion rates

Incidence Rate Ratios:	Estimate	95% CI
(Intercept)	3.0000	2.2128 – 4.1149
DrinkerYes	1.3743	0.9641 – 1.9432
AgentYes	1.9210	1.2495 – 2.9616
DrinkerYes:AgentYes	0.3555	0.2153 – 0.5858

Table 5-17: Regression coefficients from Table 5-16, expressed as Incidence Rate Ratios

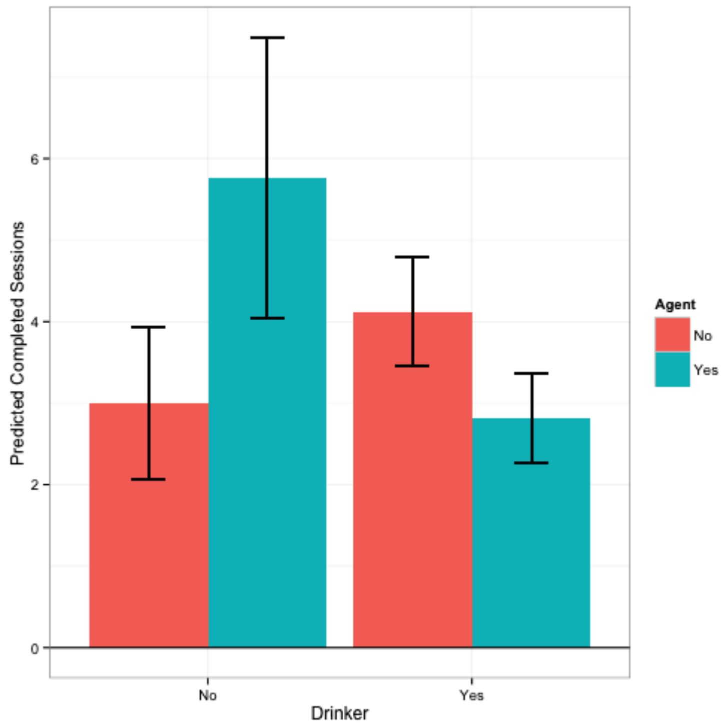


Figure 5-16: Predicted number of completed sessions and 95% confidence intervals, by survey presentation and drinker categorization.

As non-drinkers receive a significantly shorter survey than drinkers, the binary classification of drinker/non-drinker might actually be a proxy for time taken to complete the survey (average times are shown in Table 5-18). In order to exclude that possibility, we can examine finer-grained categories of drinking behavior. We can further sub-classify participants as below-average or above-average drinkers, based on their average drinking rates throughout the study compared to the sample population's average drinking rates (m=4.95 drinks per week).

	Drinker	
	No	Yes
Time to complete a session, in seconds M (SD)	111.26 (73.94)	267.30 (111.79)

Table 5-18: Time need to complete a session, in seconds. Drinker vs. Non-Drinker.

Summary statistics in Table 5-19 indicate differences in total completed sessions among the three drinker classifications, and the effect also appears to be moderated by survey presentation (Table 5-20). A histogram and density plot examining total completed sessions, by drinker categorization, are shown in Figures Figure 5-17 and Figure 5-18.

	Drinker		
	No (n=80)	Below Avg (n=83)	Above Avg (n=164)
Completed Sessions M (SD)	4.31 (4.73)	4.93 (4.43)	2.85 (3.52)

Table 5-19: Average number of completed sessions: Non-Drinker vs. Below Average Drinker vs. Above Average Drinker (max possible sessions = 16)

Agent	No			Yes		
Drinker	No (n=42)	Below Avg (n=43)	Above Avg (n=95)	No (n=38)	Below Avg (n=40)	Above Avg (n=69)
Completed Sessions M (SD)	3.00 (4.07)	5.37 (4.73)	3.56 (4.04)	5.76 (5.03)	4.45 (4.10)	1.87 (2.36)

Table 5-20: Average number of completed sessions: Agent x Drinker Category (max possible sessions = 16)

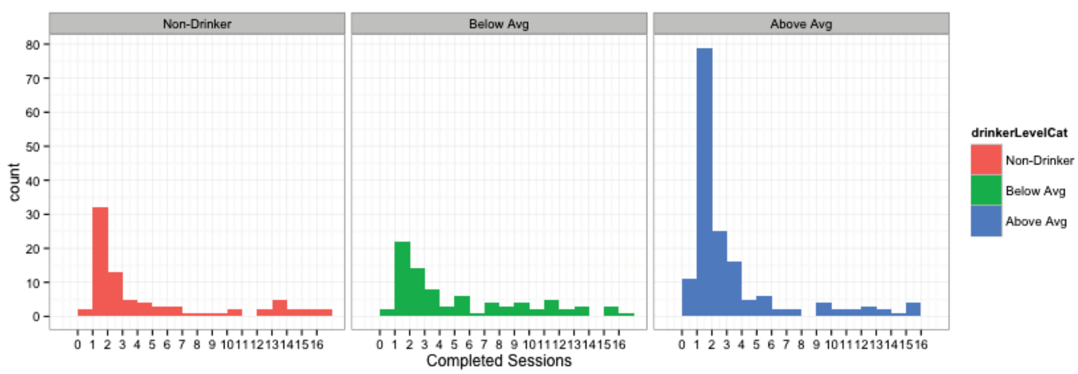


Figure 5-17: Histogram showing number of completed sessions: Non-Drinker vs. Below Average Drinker vs. Above Average Drinker

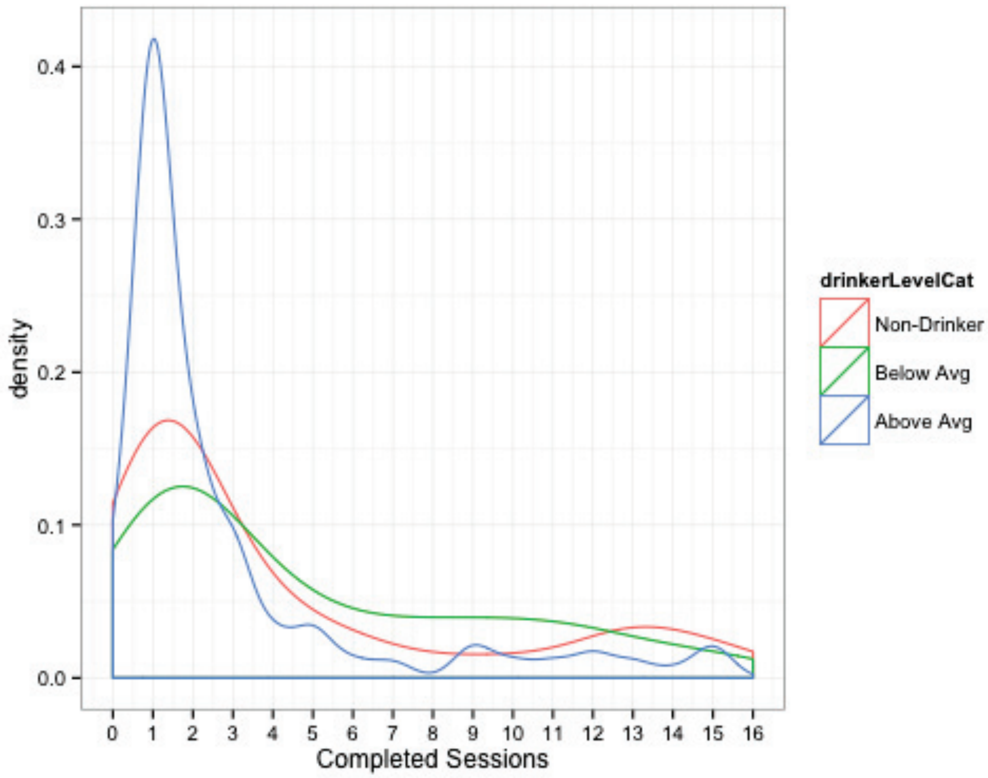


Figure 5-18: Density Plot showing number of completed sessions: Non-Drinker vs. Below Average Drinker vs. Above Average Drinker

We model a negative binomial regression with drinker level categorization (non-drinker, below-average, above-average) and survey presentation (AGENT/TEXT) as predictors. Table 5-21 shows the estimated negative binomial regression and Table 5-22 shows the Incidence Rate Ratios. There is a significant interaction between survey presentation and drinker level categorization. For AGENT participants, below-average drinkers had a lower number of completed sessions than the non-drinkers participants (incidence rate ratio, -0.8411 [95% CI, -1.418 to -0.2654], $p < 0.01$), and above-average drinkers had a lower number of completed sessions than below-average drinkers (incidence rate ratio, -1.2963 [95% CI, -1.8236 to -0.771], $p < 0.001$).

```
glm.nb(completed.total ~ DrinkerCat*Agent, data=drinkerdata)
```

Coefficients:	Estimate	95% CI	Std. Error	z value	Pr(>|z|)
(Intercept)	1.09861	0.8031 — 1.4038	0.15296	7.1824	6.849e-13 ***
DrinkerCatBelowAvg	0.58261	0.1769 — 0.9892	0.20694	2.8153	0.004873 **
DrinkerCatAboveAvg	0.17056	-0.1896 — 0.5254	0.18218	0.9362	0.349173
AgentYes	0.65287	0.2376 — 1.0708	0.21225	3.0760	0.002098 **
DrinkerCatBelowAvg:AgentYes	-0.84119	-1.418 — -0.2654	0.29382	-2.8629	0.004197 **
DrinkerCatAboveAvg:AgentYes	-1.29634	-1.8236 — -0.771	0.26834	-4.8310	1.359e-06 ***
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 5-21: Negative binomial regression examining Drinker Category, Agent as predictors of completion rates.

Incidence Rate Ratios:	Estimate	95% CI
(Intercept)	1.0986	0.8031 — 1.4038
DrinkerCatBelowAvg	0.5826	0.1769 — 0.9892
DrinkerCatAboveAvg	0.1705	-0.1896 — 0.5254
AgentYes	0.6528	0.2376 — 1.0708
DrinkerCatBelowAvg:AgentYes	-0.8411	-1.418 — -0.2654
DrinkerCatAboveAvg:AgentYes	-1.2963	-1.8236 — -0.771

Table 5-22: Regression coefficients from Table 5-21, expressed as Incidence Rate Ratios

We can use this model to predict the number of completed surveys, for our AGENT/TEXT participants, based on their level of average drinking behavior. The model predictions, along with 95% confidence intervals, are visualized in Figure 5-19. For participants in the AGENT group, non-drinkers completed the most sessions, followed by below-average drinkers and then above-average drinkers. Time/effort needed to complete the survey and/or social desirability biases may play a role in that finding. For members of the TEXT group, below-average drinkers completed the most weekly sessions, followed by above-average drinkers and then non-drinkers. I hypothesize that the reason we see a significant difference between the AGENT/TEXT groups with regard to non-drinkers, is a combination of social desirability and relevance. For members of the AGENT group, non-drinkers may be more inclined to use the system if they interpret the AGENT as reacting more positively towards non-drinkers. For members of the TEXT condition, the same bias does not appear. Instead, non-drinkers in the TEXT condition may feel as though the survey is less applicable to their life.

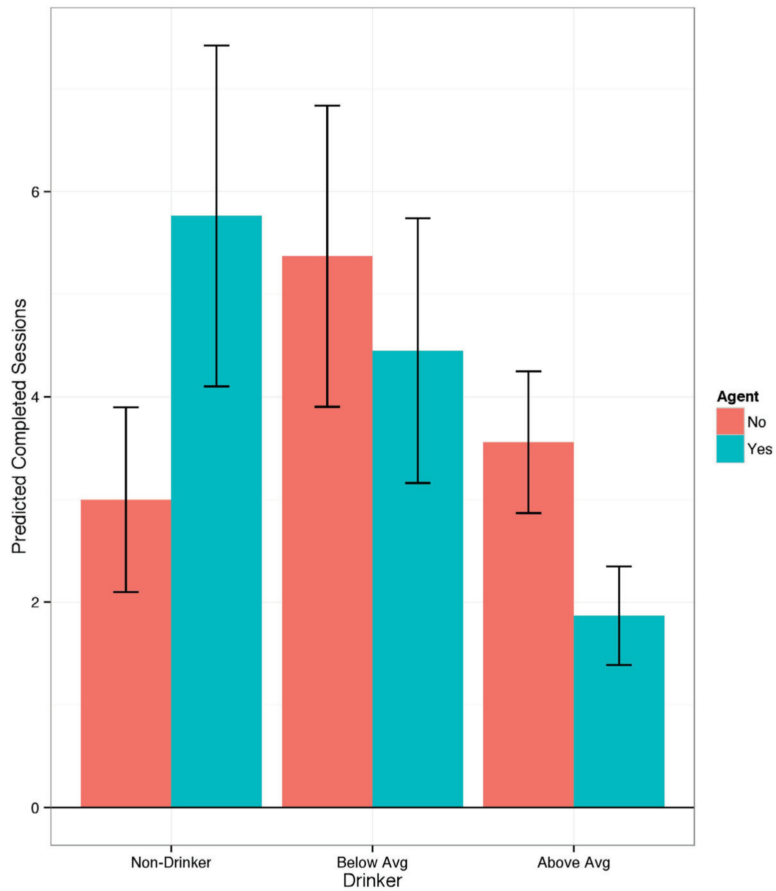


Figure 5-19: Predicted number of completed sessions and 95% confidence intervals, by survey presentation and drinker level.

5.4.1.6 Summary

We see two main factors that influenced the total number of completed surveys. First, we see an interaction between the survey presentation (AGENT/TEXT) and the amount of time participants needed to complete their first survey. If the experience during a participant's first week was quick (under 200 seconds), there is essentially no difference in total completion rates between the AGENT and TEXT groups. However, for the

AGENT group, the longer it took a participant complete the survey during their first week, fewer total sessions were completed over the course of the study.

We also see an interaction effect between survey presentation (AGENT/TEXT) and drinkers vs. non-drinkers. Drinkers in the AGENT group completed fewer sessions than those in the TEXT group. We see this effect played out further when classifying drinkers as below-average or above-average drinkers. Non-drinkers completed the most sessions in the AGENT group, followed by below-average drinkers and then above-average drinkers. In the TEXT group, we see below-average drinkers completing the most weekly sessions, followed by above-average drinkers and then non-drinkers.

5.4.2 Longitudinal Changes in Completion Rates

In this section, we examine how completion rates among experimental groups *change over time*. In the previous section, we examined predictors of the *total* number of completed surveys, statically, without longitudinal consideration. By examining the data longitudinally, we can explore differences in change curves and see what factors (if any) predict a slower drop in survey completion rates over time.

I will use linear (and also, logistic) mixed-effect regression (LMER, pronounced “elmer”) for this analysis. LMER is a generalization of logistic regression, appropriate for analyzing repeated binary measures[35, 49]. In contrast to traditional regression, LMER associates participants with their repeated measures (as these measures are correlated and certainly not independent). LMER is also more flexible in accommodating missing data, allows for varying distance between measured time-points, and allows for various types of predictors, thus providing a more robust and accurate analysis. All LMER analysis was performed in R 2.15.2, using the lme4 package [3, 62].

Also in contrast to traditional methods, null-hypothesis testing is generally not preferred in longitudinal analysis [49]. Rather, the preferred method is to present a set of models for consideration, compare the models and order them in terms of likelihood. The best model(s) among the set is then examined in detail. To predict survey completion over time, we consider the set of models shown in Table 5-23.

Model	Working Hypothesis	Static Predictors
1	Intercept and slope differences between agent and text groups	agent
2	Intercept and slope differences between those receiving and not receiving reminders about the monetary incentive	monetary
3	Intercept and slope differences between those receiving and not receiving personalized feedback	feedback
4	Intercept and slope differences between those whose first survey took below-average time, and those whose first survey took above average time	week0time
5	Intercept and slope difference between a long vs. short previous survey completion time	prevtime
6	Intercept and slope difference among non-drinkers, below-average drinkers, and above average drinkers	drinkerLevel
7	Intercept and slope differences among non-drinkers, below-average drinkers, and above average drinkers, controlling for the amount of time	drinkerLevel, week0time

	taken to complete the first week's survey	
8	Intercept and slope differences between agent and text groups, controlling for the amount of time taken to complete the first week's survey	agent, week0time
9	Intercept and slope differences between agent and text groups, controlling for feedback conditions	agent, feedback
10	Intercept and slope differences between agent and text groups, controlling for monetary conditions	agent, monetary

Table 5-23: Set of LMER models considered as predictors of survey completion over time. Intercept differences refer to differences at the onset of the study, and slope differences refer to different rates of change over time.

The models are compared using the AIC (Akaike's information criterion term). AIC is an estimate of a model's predictive accuracy. Raw AIC values are not meaningful by themselves. Rather, they are used to compare the predictive accuracy of each model to the other models in the set. Results are shown in Table 5-24. The K column lists the number of estimated parameters. AICc is the AIC-corrected value (a standard adjustment to AIC, to protect against possible small-sample size bias, recommended for both large and small sample sizes). A lower AICc value is better. Delta_AICc is the difference between each AICc and the smallest AICc (a measure of effect size). AICcWt is the weight of evidence for each model (with the sum of all models in the set equal to one). Finally, Eratio is the evidence ratio, which expresses the difference between the best-fitting model and another model in terms of odds (another measure of effect size).

We see that model 7 is the best-fitting model of the set, with a Delta of 0 and a Weight of 1. In fact, model 7 is the only candidate model, as none of the other models hold weight.

Model	K	AICc	Delta_AICc	AICcWt	Eratio
1	7	2695.81	310.98	0	3.37E+67
2	7	2699.25	314.41	0	1.88E+68
3	7	2699.13	314.3	0	1.77E+68
4	7	2410.51	25.67	0	375110.11

5	7	2527.79	142.95	0	1.10E+31
6	9	2611.85	227.01	0	1.97E+49
7	11	2384.84	0	1	1
8	9	2414.35	29.51	0	2564893.15
9	9	2699.02	314.18	0	1.67E+68
10	9	2699.11	314.28	0	1.75E+68

Table 5-24: AIC Comparison of models listed in Table 5-23

```
Generalized linear mixed model fit by the Laplace approximation

Formula: completed ~ study_week * drinkerLevelCat +study_week * week0timecat + (1 +
study_week | participant_id)
```

	Estimate	95% CI	Std. Error	Pr(>|z|)
(Intercept)	-0.4492	-1.347 — 0.4486	0.4489	0.317
study_week	-0.4284	-0.5334 — -0.3234	0.0525	0 ***
drinkerLevelCatBelow Avg	2.3342	0.9626 — 3.7058	0.6858	7.00E-04 ***
drinkerLevelCatAbove Avg	-0.3856	-1.6939 — 0.9228	0.6542	0.5556
week0timecatlong	-1.2833	-2.3592 — -0.2073	0.538	0.0171 *
study_week:drinkerLevelCatBelow Avg	-0.1126	-0.271 — 0.0458	0.0792	0.1552
study_week:drinkerLevelCatAbove Avg	-0.0548	-0.2187 — 0.1091	0.0819	0.5036
study_week:week0timecatlong	-0.0635	-0.2015 — 0.0746	0.069	0.3578

Table 5-25: Estimate details of model 7

Details of model 7 are shown in Table 5-25. Note that this model includes data from week 1 of the study as predictors, thus examining data from weeks 2-16. The output indicates that the likelihood of a participant completing the survey diminishes with time (β_1 = -0.4284, [95% CI, -0.5334 to -0.3234], $p < 0.001$). It also indicates that at week 2 below-average drinkers have a higher likelihood of completing the survey over non-drinkers and that this effect remains consistent over time, (β_2 = 2.3342, [95% CI, 0.9626 to 3.7058], $p < 0.001$). Finally, it indicates that those who experienced above average times to complete the survey during the first week, had consistently lower completion rates than participants who experienced below average times needed to complete the survey during the first week, (β_4 = -1.2833, [95% CI, -2.3592 to -0.2073], $p < 0.05$). These effects are visualized in Figures Figure 5-20 and Figure 5-21.

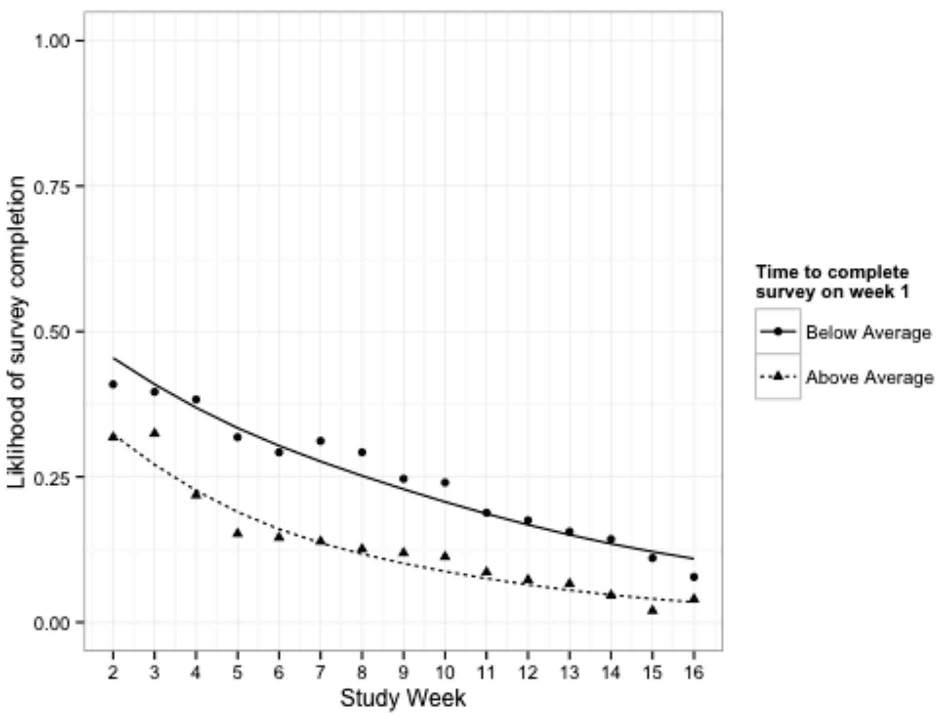


Figure 5-20: Plot of Model 7, likelihood of survey completion over time: Below average initial survey time vs. above average initial survey time. Means (points) and fitted values (lines).

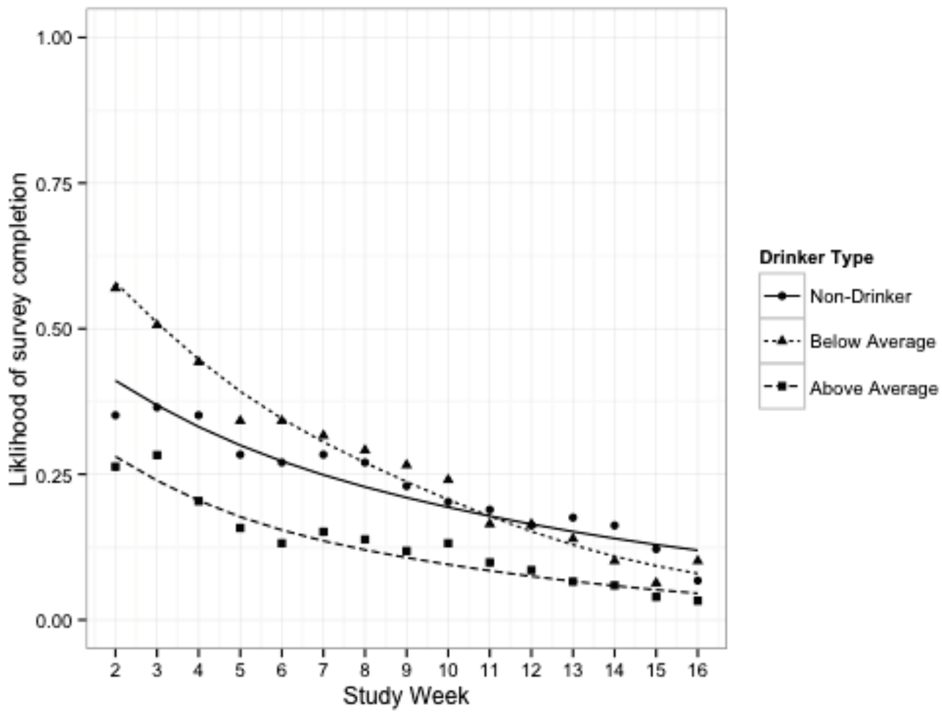


Figure 5-21: Plot of Model 7, likelihood of survey completion over time: Non-drinkers vs. below-average drinkers vs. above-average drinkers. Means (points) and fitted values (lines).

5.4.2.1 Summary

When longitudinally examining survey completion, we see two main factors come into play: drinker status and survey time. Below-average drinkers are week-after-week more likely than both non-drinkers and above-average drinkers to complete a weekly survey. Social desirability can explain why above-average drinkers are less likely to complete the weekly sessions over time. However, it is interesting that non-drinkers are not more likely to complete weekly sessions. As sessions for non-drinkers are short (they have only four questions to answer), it's reasonable to think that they might be the group most likely to complete a weekly session. There is little effort involved to be entered into the weekly lottery for $25. I suspect this has to do with feelings of *relevance*. I suspect that non-drinkers are more likely to consider the survey topic personally irrelevant, and thus are less likely to complete weekly sessions.

We also see that when examining the data longitudinally, first impressions of the survey matter: those who experienced below-average times needed to complete the survey on the first week, had consistently higher completion rates over time, when compared to those who experienced above-average completion times in the first session. I considered other models of time, for example, a participant's previous session time, which did predict completion rates, but the session time from the first week was the best predictor.

Just as interesting as factors that influenced survey completion, is what *did not* affect survey completion rates over time. Neither feedback nor monetary reminders led to increased rates of survey completion, nor did they decrease the decline of weekly survey completion. It is possible that the monetary manipulation was too week to elicit any differences between groups. The fact that the feedback was given at the end of the survey, also might have contributed to its lack of effect. As time has been shown to be an important factor, it is possible that participants did not want to spend additional time exploring the charts and information presented upon completing the survey. We will see the effect of feedback come into play in the next section, where we will explore participant self-reports.

5.4.3 Self-Reporting Activity

In this section, we explore responses to the survey questions, and examine experimental effects on the levels of reported drinking activity. We'll examine a few survey questions individually and some sections of the survey as composite scores – the entire survey can be found in Appendix D .

5.4.3.1 Results After the First Exposure to the Survey

To begin, we examine responses on the very first exposure to the survey questions: week one. Based on previous results on social desirability biases, we could reasonably expect differences between the AGENT and TEXT conditions. Additionally, participants in the FEEDBACK group were explicitly told that they would be receiving feedback based on their answers (and also shown a sample chart, depicting what the feedback would look like), right before starting the survey. That notification could lead to another possible cause of social desirably in participants' answers. Thus, we should examine differences between the FEEDBACK and NO-FEEDBACK conditions. There is no reason to expect differences between the MONETARY and NO-MONETARY groups, as both groups were enrolled and informed of the monetary incentive seconds earlier.

Results from questions the first five questions of the survey, assessing quantitative levels of drinking activity, are shown in Table 5-26.

	AGENT	No		Yes		Agent x Feedback ANOVA effects
	FEEDBACK	No	Yes	No	Yes	
		M (SD)	M (SD)	M (SD)	M (SD)	
1	How many days of the week did you drink alcohol last week?	1.64 (1.37)	1.28 (1.20)	1.51 (1.34)	1.30 (1.53)	Feedback (.)
2	Please indicate the average number of drinks you consume on one occasion at parties or when socializing last week.	4.80 (2.40)	4.85 (2.34)	4.45 (2.48)	5.70 (3.72)	
3	Think about the occasion you drank the most last week. How much did you drink?	6.82 (3.51)	7.12 (3.55)	6.63 (3.50)	8.54 (6.01)	
4	Think about the occasion you drank the most last week. How many HOURS did you spend drinking on that occasion?	5.04 (1.69)	5.31 (1.66)	4.91 (1.85)	5.00 (1.96)	
5	How many drinks per occasion do you consider moderate (not excessive) for yourself?	4.10 (1.43)	4.20 (1.52)	3.84 (1.51)	3.71 (1.56)	Agent (*)

Table 5-26: Week 1 results from questions 1-5, quantitative levels of drinking activity.

We begin with the first question of the survey: *How many days of the week did you drink alcohol last week?* A two-way, factorial design ANOVA was conducted to compare the effects of AGENT and FEEDBACK on the reported number of days drinking. There was a trending main effect of FEEDBACK on reported number of days drinking, $F(1, 311) = 3.702$, $p = 0.055$. The reported number of days drinking was lower for those receiving feedback (m= 1.29) versus those not receiving feedback (m=1.59).

We also see a significant effect of survey presentation on the fifth survey question: *How many drinks per occasion do you consider moderate (not excessive) for yourself?* In a two-way, factorial design ANOVA comparing the effects of AGENT and FEEDBACK, there was a significant main effect of AGENT on reported number of drinks considered to be moderate, $F(1, 282) = 4.127$, $p < 0.05$. The number of drinks was lower for those AGENT group (m= 3.79) versus the TEXT group (m = 4.15).

The next section of the survey deals with negative consequences that can occur as a result of drinking (e.g., missing a class the next day). For these questions, we have created a composite score for each participant by averaging responses (Table 5-27).

In a two-way, factorial design ANOVA comparing the effects of AGENT and FEEDBACK, there was a trending interaction effect on reported amount of risky behaviors, $F(1, 201) = 2.978$, $p = 0.091$. For participants receiving FEEDBACK, those in the AGENT condition reported more risky behaviors than those in the TEXT condition (0.222 vs. 0.099).

AGENT	No		Yes		Agent x Feedback ANOVA effects
FEEDBACK	No	Yes	No	Yes	
	M (SD)	M (SD)	M (SD)	M (SD)	
Negative Consequences	0.173 (0.288)	0.099 (0.142)	0.145 (0.194)	0.222 (0.544)	Agent * Feedback (.)

Table 5-27: Week 1 negative consequence responses.

The final section of the survey asks about protective behaviors that a person might take while drinking (e.g., not exceeding a predetermined number of drinks). Again, we have created a composite score for each participant by averaging responses (Table 5-28). A two-way, factorial design ANOVA comparing the effects of AGENT and FEEDBACK shows no significant effects.

AGENT	No		Yes		Agent x Feedback ANOVA effects
FEEDBACK	No	Yes	No	Yes	
	M (SD)	M (SD)	M (SD)	M (SD)	
Protective Behaviors	1.944 (0.845)	1.722 (0.854)	1.867 (0.950)	1.760 (1.064)	none

Table 5-28: Week 1 protective behaviors.

After the first exposure to the survey, we are seeing two examples of social desirability taking place. There is a trend where participants in the FEEDBACK condition report fewer days of drinking than those in the NO-FEEDBACK condition. Participants may be self-censoring, not wanting to receive feedback that could self-viewed as negative. Though, it is important to note that this trend only occurs for the first question of the survey. We also see social desirability towards the agent. Participants interacting with the agent reported that they consider fewer drinks per occasion as moderate, compared to participants interacting with the text interface. While we do not (and cannot) have a sensor-based form of ground truth for this question, mostly likely those in the TEXT condition are being more forthcoming in their answers to this question.

5.4.3.2 Changes in Survey Responses Over Time

In this section, we examine changes in self-reported levels of drinking activity over time. Unlike the longitudinal analysis of completion rates reported in Section 5.4.2, where we had a binary data point for each participant each week (they either did or did not complete the survey), for this section we are limited by missing data. More than half of participants never completed more than one week of the study, thus severely limiting our power when analyzing self-reported data over time. As a result, we will limit the amount of weeks included in our analysis (only using the first six weeks of the study) and only include participants that completed at least 3 sessions (n=127). *Results in this section are to be interpreted with caution, and considered on par with pilot data.*

We consider four possible models for all our longitudinal analyses of self-reported behaviors (Table 5-29). As there is no theoretical reason why the MONETARY condition may affect self-reported data, it will not be a factor in considered models. We are also not considering potential interaction effects due to limited power.

Model	Working Hypothesis	Static Predictors
1	Intercept and slope differences between agent and text groups	agent
2	Intercept and slope differences between those receiving and not receiving personalized feedback	feedback
3	Intercept differences between agent and text groups	agent
4	Intercept differences between those receiving and not receiving personalized feedback	feedback

Table 5-29: LMER models considered for self-reported changes over time.

We begin by looking at changes in quantitative reports of alcohol consumption over time. The first question of the survey asks, *How many days of the week did you drink alcohol, last week?* A comparison of models is shown in Table 5-30. We see that model 3 (intercept differences between AGENT and TEXT groups) is the best fitting model. With its weight of 0.53, model 3 accounts for the majority of probability, but not by much. As the evidence ratios of the other models are rather low, they should not be ruled out of consideration as possible models to explain the data.

Model	K	AICc	Delta_AICc	AICcWt	Eratio
1	8	1663.53	1.80	0.21	2.46
2	8	1664.33	2.61	0.14	3.68
3	7	1661.73	0.00	0.53	1.00
4	7	1664.71	2.98	0.12	4.43

Table 5-30: AIC comparison of models listed in Table 5-29 – fitted to question 1: “How many days of the week did you drink alcohol last week?”

Details of model 3 are shown in Table 5-31. It indicates that those in the AGENT group consistently reported a lower number of days drinking per week, as compared to the TEXT group (β_2 = -0.3383, [95% CI, -0.7263 to -0.0497]). This effect is visualized in Figure 5-22.

Linear Mixed model fit by maximum likelihood Formula: Q1 ~ study_week + agent + (1 + study_week | participant_id)				
Fixed effects:	Estimate	95% CI	Std. Error	t value
(Intercept)	1.1983	0.9191 – 1.475	0.1396	8.583
study_week	0.0091	-0.0478 – 0.0659	0.0284	0.318
agentYes	-0.3383	-0.7263 – 0.0497	0.1939	-1.744

Table 5-31: Estimate details of model 3 (intercept differences between AGENT and TEXT groups).

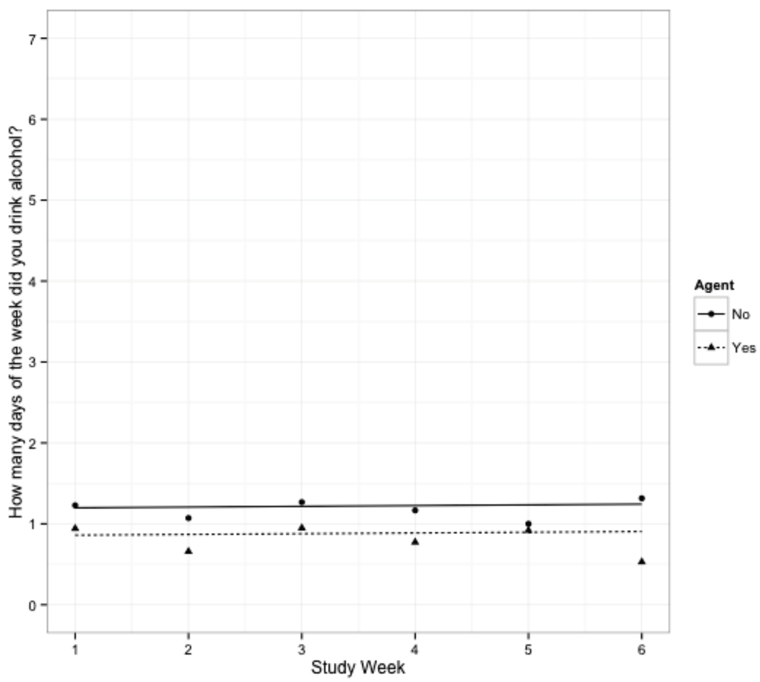


Figure 5-22 Plot of model 3, days of drinking over time: AGENT vs. TEXT. Means (points) and fitted values (lines).

Next, we examine the fifth survey question, one that indicted significant differences between AGENT and TEXT groups in the initial exposure to the survey: *How many drinks per occasion do you consider moderate (not excessive) for yourself?*

Models are compared in Table 5-32, and we see that model 3 (intercept differences between agent and text groups) is the best fitting model. With its weight of 0.69, model 3 accounts for the majority of probability, with model 1 also holding 0.25 percent of the weight. With higher evidence ratios, the two models that include feedback as a predictor are most likely implausible.

Model	K	AICc	Delta_AICc	AICcWt	Eratio
1	8	1311.90	2.02	0.25	2.74
2	8	1316.64	6.75	0.02	29.27
3	7	1309.89	0.00	0.69	1.00
4	7	1315.93	6.04	0.03	20.47

Table 5-32: AIC comparison of models listed in Table 5-29 – fitted to question 5: "How many drinks per occasion do you consider moderate (not excessive) for yourself?"

Details of model 3 are shown in Table 5-33. It indicates that those in the AGENT group consistently reported a lower number of drinks that they considered moderate for themselves, as compared to the TEXT group (β_2 = -0.72, [95% CI, -1.285 to -0.155]). This effect is visualized in Figure 5-23.

```
Linear Mixed model fit by maximum likelihood

Formula: Q5 ~ study_week + agent + (1 + study_week | participant_id)
```

Fixed effects:	Estimate	95% CI	Std. Error	t value
(Intercept)	3.790	3.407 – 4.173	0.191	19.797
study_week	0.0438	0.007 – 0.080	0.018	2.399
agentYes	-0.720	-1.285 – -0.155	0.282	-2.550

Table 5-33: Estimate details of model 3.

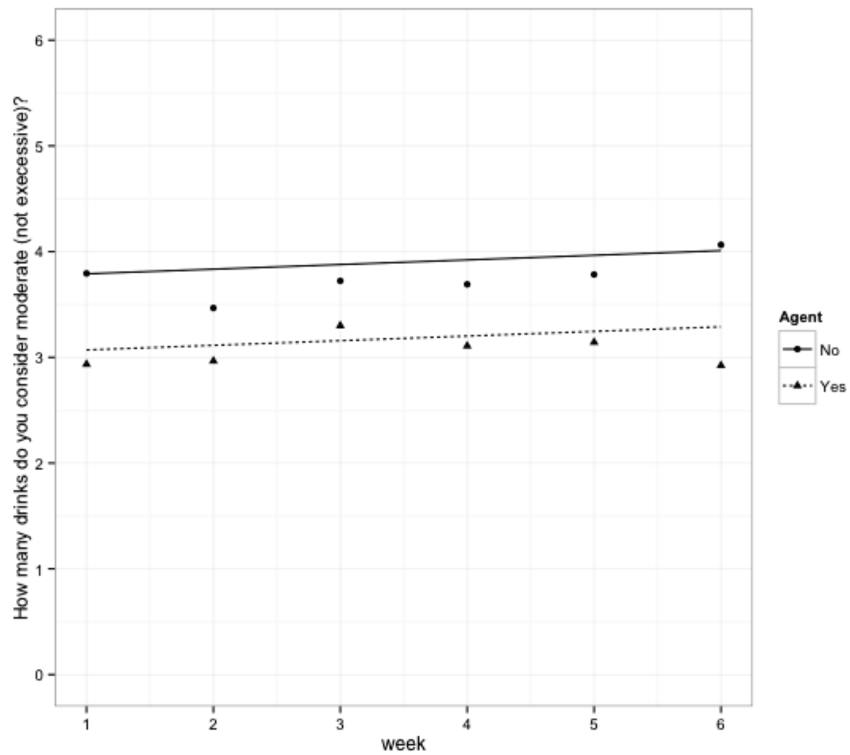


Figure 5-23: Estimate details of model 3.

Next, we examine changes in reports of negative consequences that occur as a result of drinking (e.g., having a hangover, missing a class). Models are compared in Table 5-34, and model 4 (intercept differences between the feedback and no feedback groups) is the best fitting model. It accounts for the majority of probability, with a weight of 0.69, and the other model with feedback as a predictor also holds 0.24 of the weight. Models 1 and 3, which consider agent/text groups as predictors, have a low weight and higher evidence ratios, thus are most likely implausible.

Model	K	AICc	Delta_AICc	AICcWt	Eratio
1	8	-118.7	6.44	0.03	25.02
2	8	-123.01	2.13	0.24	2.89
3	7	-119.31	5.83	0.04	18.48
4	7	-125.14	0.00	0.69	1.00

Table 5-34: AIC comparison of models listed in Table 5-29 – fitted to composite score of negative consequences.

The details of model 4 indicate that those in the FEEDBACK group consistently reported a lower number of negative consequences then those in the NO-FEEDBACK group (β_2 = -0.061, [95% CI, -0.109 to -0.011]), Table 5-35. This effect is visualized in Figure 5-24.

Linear Mixed model fit by maximum likelihood Formula: NegConseq ~ study_week + feedback + (1 + study_week | participant_id)				
Fixed effects:	Estimate	95% CI	Std. Error	t value
(Intercept)	0.110	0.070 – 0.150	0.020	5.507
study_week	0.004	-0.014 – 0.023	0.009	0.479
feedbackYes	-0.061	-0.109 – -0.011	0.025	-2.461

Table 5-35: Estimate details of model 4.

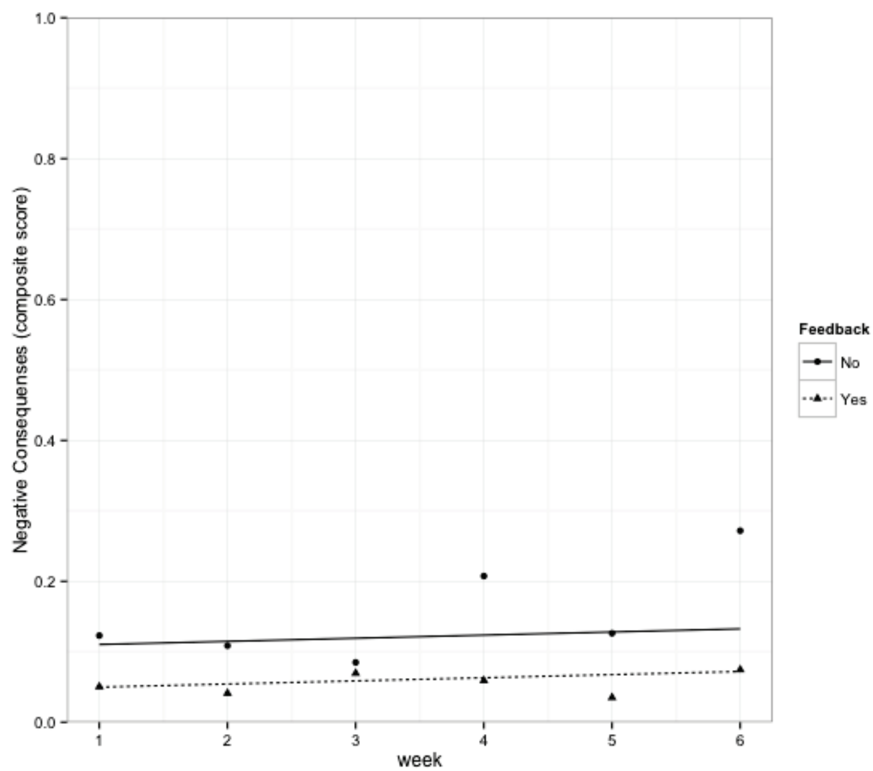


Figure 5-24: Plot of model 4, changes in reports of negative consequences: Feedback vs. No Feedback. Means (points and fitted values (lines).

Finally, we examine changes in reports of protective behaviors taken by participants when drinking (e.g., pacing drinks, eating before/during drinking). The models are shown in Table 5-36, and model 1 (intercept and slope differences between the agent and text groups) is the best fitting model. However, with a weight of 0.39, it does not account for the majority of probability. All other models hold some weight and have low evidence ratios, indicating that all models might be candidates to explain the data.

Model	K	AICc	Delta_AICc	AICcWt	Eratio
1	8	1034.2	0	0.39	1
2	8	1036.24	2.04	0.14	2.78
3	7	1035.03	0.83	0.26	1.51
4	7	1035.51	1.31	0.2	1.93

Table 5-36: AIC comparison of models listed in Table 5-29 – fitted to composite score of protective behaviors.

The details of model 1 indicate that as the study progressed, the AGENT group reported taking more protective behaviors over time, compared to the TEXT group (β_3 = 0.099, [95% CI, -0.016 to 0.215]), Table 5-37. This effect is visualized in Figure 5-25.

```
Linear Mixed model fit by maximum likelihood

Formula: Protect ~ study_week * agent + (1 + study_week | participant_id)
```

Fixed effects:	Estimate	95% CI	Std. Error	t value
(Intercept)	2.009	1.744 – 2.274	0.132	15.169
study_week	-0.046	-0.123 – 0.030	0.038	-1.210
agentYes	0.032	-0.372 – 0.437	0.202	0.163
study_week:agentYes	0.099	-0.016 – 0.215	0.058	1.712

Table 5-37: Estimate details of model 1.

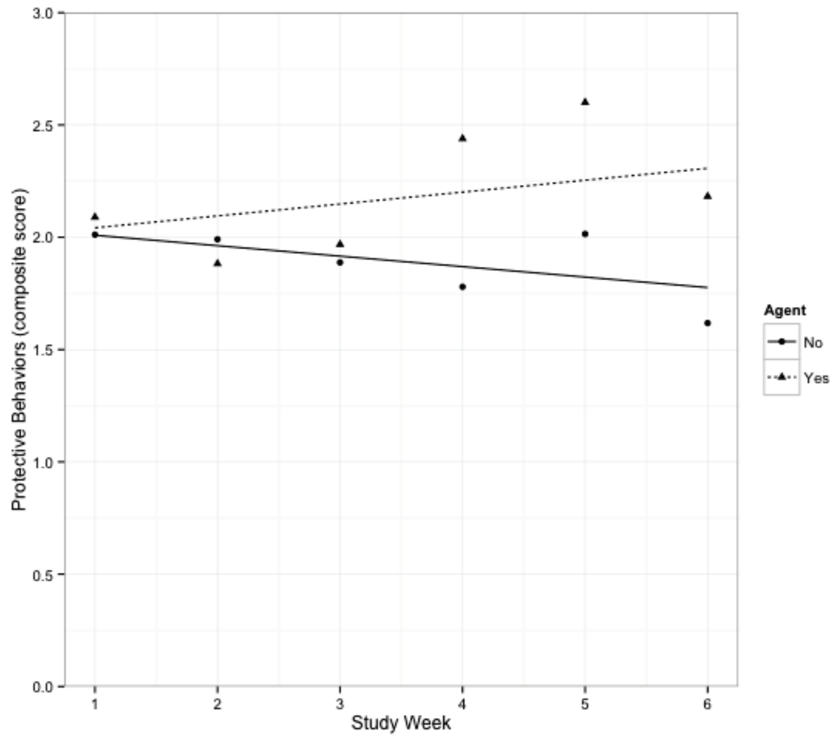


Figure 5-25: Plot of model 1, changes in reports of protective behaviors: Agent vs. Text. Means (points) and fitted values (lines).

When examining the data longitudinally, we see that social desirability effects remain over time. Participants in the AGENT group consistently reported fewer days of drinking per week and fewer drinks per occasion to be moderate for themselves. When examining reports of negative consequences from drinking, we see a reliable social desirability effect from participants in the FEEDBACK condition. Those in the FEEDBACK condition reported fewer instances of negative consequences as a result of drinking, and the effect remained consistent over time. Finally, we see changes over time in reports of protective behaviors taken while drinking. Reports of protective behaviors increased over time for participants in the AGENT condition and decreased over time for participants in the TEXT condition.

5.4.4 Participant Attitudes Regarding the Study

At the end of the 16 weeks, all participants were asked (via email) to complete an online questionnaire regarding their attitudes towards the study (Appendix F). As an incentive, participants who completed the questionnaire were entered into a final drawing for a $25 gift card to Amazon.com. Fifty-eight participants (15.47%) completed the follow-up questionnaire. These participants completed, on average, 8.7 (sd = 5.7) weekly sessions throughout the study.

All participants received the first six questions, shown in Table 5-38 and Figure 5-26. Enjoyment and usefulness of the weekly surveys received low ratings. Participants reported that the surveys were not time-consuming, and that they answered honestly. Finally, participants reported moderate-to-high feelings of confidentiality and moderate influences of the monetary incentive.

	Anchors 0= not at all 6=very much	
Question	Mean	SD
Did you enjoy filling out the weekly surveys?	3.32	1.35
Overall, how useful were the surveys to you?	2.53	1.66
How time-consuming was it to complete the weekly surveys?	1.64	1.58
Thinking back, how honest were your answers to the weekly survey questions?	5.63	0.73
How confidential did the surveys feel to you?	4.74	1.35
How much did the $25 weekly drawing influence your decision to fill out the survey each week?	4.21	1.92

Table 5-38: Participant attitudes towards the study

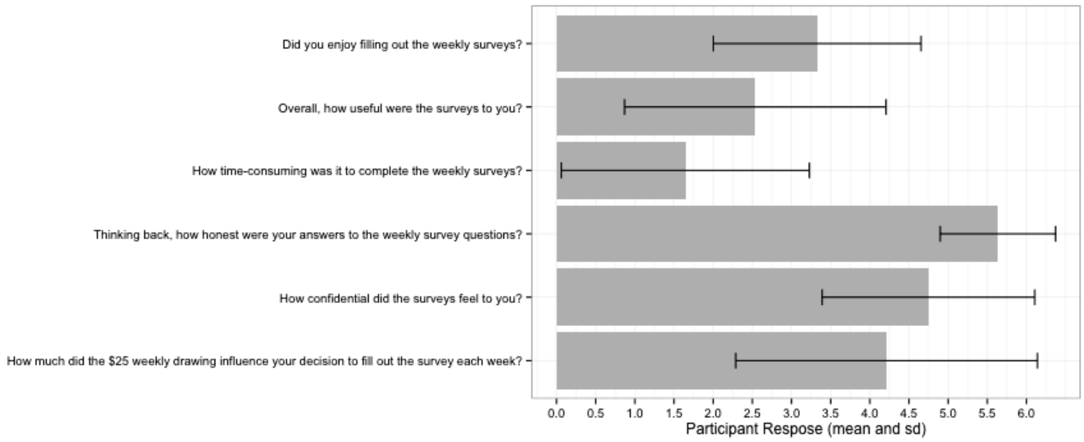


Figure 5-26: Participant attitudes towards the study (0 = not at all, 6 = very much)

Of these questions, only one evoked significant differences between study groups: the fourth question, *Thinking back, how honest were your answers to the weekly survey questions?* In a two-way, factorial design ANOVA comparing the effects of AGENT and FEEDBACK, there was a significant main effect on reported amount of honesty, $F(1, 54) = 5.837$, $p = 0.019$. Participants in the TEXT condition reported being more honest than those in the AGENT condition (5.85 vs. 5.4).

The next set of questions was presented only to participants in the FEEDBACK condition (Table 5-39). Participants did not report that the feedback was useful or influential in their decision to return in subsequent weeks.

	Anchors 0= not at all 6=very much	
Question (n=30)	Mean	SD
How useful was the weekly feedback you received at the end of each weekly survey?	2.47	1.73
How much did the weekly feedback influence your decision to fill out the survey each week?	2.06	1.83
	Feedback	GiftCard
What motivated you more when deciding whether or not to fill out the survey?	20% (n=6)	80% (n=24)

Table 5-39: Participant attitudes towards the personalized feedback.

The next two questions were presented to participants in the AGENT condition (Table 5-40). Participant reported moderate levels of satisfaction and trust towards the agent.

	Anchors 0= not at all 6=very much	
Question (n=28)	Mean	SD
How satisfied were you with Tanya?	3.32	1.65
How much did you trust Tanya?	3.75	1.52

Table 5-40: Participant attitudes toward the agent (Tanya).

Finally, all participants were asked to explain why they either completed few weekly sessions or many weekly sessions. Forty-seven participants (81 %), submitted answers and a summary of responses is shown in Tables Table 5-41 and Table 5-42.

Reasons for Returning	Example	Number of mentions
Entry into the gift card drawing	"I was hoping to receive the $25 gift card."	14
Email reminders	"The weekly reminder emails were very helpful"	8
Altruism	"Originally I filled it out because answering surveys can be really helpful to research…"	6
Time	"The short amount of time needed."	4
Won the drawing	"I won the drawing once. Hoped to win again."	3
Study break	"I started the survey to kill some time/avoid studying and just decided to continue once I had started."	2
Self-track drinking	"To have a record of my drinking activity. I was never a heavy drinker, and I was curious about how the amount of alcohol I drank affected me."	2
Commitment	"I made a commitment to fill it out every week. Even if no one else was holding me to that commitment, I was."	1
Fun	"*shrugs* It was fun."	1
Feedback	"I liked the little info tidbits."	1
Drank the previous week	"Actually having consumed alcohol that week"	1
Averaging out data	"Wanting to average out the data. Some weeks I did drink a lot on the weekends, when other weeks I did not at all. I am almost always responsible with my drinking habits."	1

Table 5-41: Participant reasons for completing the weekly survey many times

Reasons for NOT Returning	Example	Number of mentions
Too much time	"It was too time consuming to do weekly."	6
Too busy	"As the semester wound down, I became busy with school work. Then it was difficult to pick up. "	2
Forgot	"I kept forgetting to complete them, more reminder emails next time."	2
Stopped drinking / Infrequent drinker	"I stopped drinking, so I figured my data wasn't valuable anymore!"	2
Boring	"It was boring."	1

Table 5-42: Participant reasons for not returning / only completing the survey few times

Data from this follow-up questionnaire provide additional details about the experimental conditions. Participants in the AGENT and TEXT groups did not differ in their level of enjoyment of the survey, feeling of the time commitment involved, nor in feelings of confidentiality. However, when asked to reflect on the accuracy of their self-reported data, participants in the TEXT condition reported being more honest than those in the AGENT condition. This corroborates our findings of participants in the AGENT group provided more socially desirable data than those in the TEXT group.

We also see hints that feelings of *relevance* may play a part in survey completion rates over time. For example, participants reported that if they stopped drinking, they felt that the weekly survey was no longer relevant to them, and therefore stopped completing it. Likewise, we see participants reporting that if they *did* drink during the week, that was a reason for them to fill out the survey.

When it comes to personalized feedback, we see that most participants did not find it useful. Furthermore, the feedback might cause negative feelings of judgment. As one participant reported, if they drank a lot during one week, they didn't like the idea that it might appear as if they are irresponsible: *"[I wanted] to average out the data. Some weeks I did drink a lot on the weekends, when other weeks I did not at all. I am almost always responsible with my drinking habits."* This may explain why participants in the FEEDBACK condition were often providing more socially desirable data than those in the NO-FEEDBACK condition.

5.5 Discussion

This study examined voluntary self-reported health data by 375 participants over the course of 16 weeks. By manipulating the presentation of the interface, study incentives reminders, and the presence of personalized feedback, we are able to explore designs that lead to increased system use and more accurate self-reported data.

Results from this study indicate that if we want people to interact with a system repeatedly over time, their first experience with the system should take a short amount of time. Participants whose first-time interactions with the system lasted longer than 200 seconds completed significantly less weekly sessions than those whose interactions were under 200 seconds. Furthermore, this effect was more pronounced for participants who interacted with the AGENT system.

We also see social desirability effects when it comes to repeated use of the system over time. The group of participants in the TEXT condition who completed the most weekly sessions was below-average drinkers, followed by above-average drinkers, followed-by non-drinkers. Comparatively, the group of participants in the AGENT condition who completed the most weekly sessions was: *non-drinkers*, followed by below-average drinkers, followed by above-average drinkers. Drinkers in the AGENT condition self-selected out of the study, and completed significantly fewer weekly sessions.

The monetary incentive conditions (either providing weekly reminders about the monetary incentive, or never mentioning it after study enrollment) proved to be too weak of a manipulation to evoke any differences in study completion. The study advertisements prominently featured the monetary incentive, and we can see high completion rates (81%) during the week of study enrollment. But completion rates immediately dropped to 31% during the second week of the study and continued to decline week after week, with 5% completion rates at week 16. So while the monetary incentive proved effective during the first week of the study, its power appeared to decline with time.

When it came to the self-reported health data, we were indeed able to see social desirability effects in this study. Most questions assessing the quantitative levels of alcohol consumption did not evoke different responses by the AGENT and TEXT groups. However, one question, *How many drinks per occasion do you consider moderate (not excessive) for yourself?* did conjure significantly different responses. This question, assessing a personal attitude, caused participants in the AGENT condition to report a smaller number of drinks than those in the TEXT condition. Furthermore, this effect appeared to be consistently present week after week. We also see that with time, participants in the AGENT condition reported taking more protective behaviors when drinking, compared to participants in the TEXT condition reporting fewer protective behaviors over time.

Finally, I hypothesized that personalized feedback may provide participants with an incentive for repeated interactions over time and providing higher quality data. Unfortunately, this study indicates that feedback may actually *predict* social desirability effects. Despite the fact that the personalized feedback was designed to be neutral and fact-based, and that there was no intervention aspect to the experiment, we saw significant differences in reports by participants in the FEEDBACK and NO-FEEDBACK groups. During the first exposure to the survey, participants in the FEEDBACK condition reported fewer days of drinking than those in the NO-FEEDBACK condition. Though, this effect did not continue past the first week. However, when it came to reports of negative consequences from drinking, participants in the FEEDBACK condition reported fewer negative consequences, and this effect *persisted* over time.

CHAPTER 6

Conclusions

As chronic diseases become increasingly prevalent, and primary care offices continue to be overloaded, patients now face a greater responsibility to manage their own care. Technologies to help patients track their health over time have the potential to reduce visits to the doctor and also help patients become more conscious and active participants in decisions surrounding their health and wellness.

Sensor-based technologies are invaluable and show many promising directions, however, there are many health conditions for which sensors are simply not viable. Though self-reported data can be biased, oftentimes, the only way to assess how well a patient is doing, is simply to ask them.

In this dissertation, I provide a foundation for designing long-term, patient-facing systems for self-reported health tracking. In particular, we explore two main challenges faced by researchers when building these systems: 1) How do we design a system to maximize the *quality* of the self-reported data? and 2) How do we keep people *engaged* with such a system, over potentially long periods of time?

We explored these challenges by studying three systems that we designed for long-term health tracking. In Chapter 3, we discussed our collaboration with researchers from Boston Medical Center, to design and build a system for patients to track their post-hospitalization recovery and report any potential adverse events. We discussed our design-approach to creating our system, and reported findings from a lab-based user study with recently discharged patients. In a field trial, we saw that few patients utilized the system from home, which prompted the additional studies discussed in this dissertation.

In Chapter 4, we took a step back and conducted basic research that examined how different interface presentations may affect the quality of self-reported health data. In a six-week long study, we explored differences in data reported to an embodied agent vs. data reported to a text-based interface. In particular, we examined how the prevalence of social desirability biases in the data changed, as people continued to use the interface over time. The experiment showed that as people interacted with the ECA interface over time, social desirability biases increased, and as people interacted with the text-based interface over time, social desirability biases decreased. This experiment was the first to show that social responses to computers, as shown by the Computers as Social Actors paradigm [54], are not static, but can change over time.

In Chapter 5, we explored the impact of incentives to promote long-term user engagement with self-report systems. In an experiment with 375 participants lasting four months long, we showed that despite incentives, interaction time was the best predictor of repeat system interaction. In addition to replicating previous findings on interface personification and social desirability bias, we also showed that providing personalized feedback, designed for self-reflection and to incentivize data quality, actually *reduced* data quality and predicted higher levels of bias effects.

6.1 Future Work

To further this work, a natural step would be to explore differences between people who *choose* to self-track health indicators, and those that specifically choose *not* to self-track. I would hypothesize that these groups of people have strong differences in personality, motivation, and even overall health. It is a grand challenge to design health-tracking technologies for those who critically need it, but might not want it. For example, certain blood pressure medications may require a person to weigh themselves daily, as weight changes might indicate a problem. But what if a person does not want to weigh themselves daily? What if the feedback from the scale is not something a person is willing to face? Designing technologies to deal with the upfront hurdles and potential shame that may occur when a person begins to track their health is an open problem, and one with large potential research contributions.

References

1. Patient Protection and Affordable Care Act, §412.150 - 412.154, 124 Stat 119, 2010

2. Al-Rashed, S. A., Wright, D. J., Roebuck, N., Sunter, W., and Chrystyn, H. The value of inpatient pharmaceutical counselling to elderly patients prior to discharge. British journal of clinical pharmacology, 54, 6, (2002) 657-664.

3. Bates, D., Maechler, M., and Bolker, B., lme4: Linear mixed-effects models using S4 classes, 2012.

4. Bickmore, T., Caruso, L., and Clough-Gorr, K., Acceptance and usability of a relational agent interface by urban older adults, CHI '05: CHI '05 extended abstracts on Human factors in computing systems, 2005, pp. 1212-1215.

5. Bickmore, T., Gruber, A., and Picard, R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Education and Counseling, 59, 1, (2005) 21-30.

6. Bickmore, T., Pfeifer, L., and Jack, B., Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents, CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, 2009, pp. 1265-1274.

7. Bickmore, T. and Schulman, D., A virtual laboratory for studying long-term relationships between humans and virtual agents, Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1, 2009, pp. 297-304.

8. Bickmore, T., Schulman, D., and Yin, L. Maintaining engagement in long-term interventions with relational agents. Applied artificial intelligence: AAI, 24, 6, (2010) 648.

9. Bodenheimer, T., Lorig, K., Holman, H., and Grumbach, K. Patient self-management of chronic disease in primary care. JAMA : the journal of the American Medical Association, 288, 19, (2002) 2469-2475.

10. Bohus, D. and Horvitz, E., Learning to predict engagement with a spoken dialog system in open-world settings, SIGDIAL '09: Proceedings of the SIGDIAL 2009 Conference, 2009, pp. 244-252.

11. Booth-Kewley, S., Larson, G. E., and Miyoshi, D. K. Social desirability effects on computerized and paper-and-pencil questionnaires. Computers in Human Behavior, 23, 1, (2007) 463-477.

12. Cassell, J., Embodied conversational agents, MIT Press, Cambridge, Mass., 2000.

13. Cassell, J., Miller, P., Conrad, F. G., and Schober, M. F., *Is it Self-Administration if the Computer Gives you Encouraging Looks?* In *Evisioning the Survey Interview of the Future,* John Wiley & Sons, 2007, pp. 161-178.

14. Cassell, J., Vilhjalmsson, H., and Bickmore, T., BEAT: the Behavior Expression Animation Toolkit, Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 2001, pp. 477-486.

15. Charlton, J. and Danforth, I. Distinguishing addiction and high engagement in the context of online game playing. Computers in Human Behavior, 23, 3, (2007) 1531-1548.

16. Cherry, J. C., Moffatt, T., Rodriguez, C., and Dryden, K. Diabetes disease management program for an indigent population empowered by telemedicine technology. Diabetes technology & therapeutics, 4, 6, (2002) 783-791.

17. Clayman, M., Pandit, A., Bergeron, A., Cameron, K., Ross, E., and Wolf, M. Ask, understand, remember: A brief measure of patient communication self-efficacy within clinical encounters. Journal of health communication, 15, S2, (2010) 72-79.

18. Crowne, D. and Marlowe, D., The approval motive; studies in evaluative dependence, Wiley, 1964.

19. Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. User acceptance of computer technology: a comparison of two theoretical models. Management Science, 35, 8, (1989) 982-1003.

20. Dow, S., Mehta, M., Harmon, E., MacIntyre, B., and Mateas, M., Presence and engagement in an interactive drama, Proceedings of the SIGCHI conference on Human factors in computing systems, 2007, pp. 1475-1484.

21. Dudas, V., Bookwalter, T., Kerr, K. M., and Pantilat, S. Z. The impact of follow-up telephone calls to patients after hospitalization. The American journal of medicine, 111, 9B, (2001).

22. Fairburn, C. and Beglin, S. Assessment of eating disorders: Interview or self-report questionnaire? International Journal of Eating Disorders, 16, 4, (1994) 363-370.

23. Fischer, J. and Benford, S., Inferring player engagement in a pervasive experience, CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, 2009, pp. 1903-1906.

24. Fogg, B. and Iizawa, D., *Online Persuasion in Facebook and Mixi: A Cross-Cultural Comparison.* In H. Oinas-Kukkonen, P. Hasle, M. Harjumaa, K. Segerstal, and P. Ohrstrom, Eds., *Persuasive Technology, Volume 5033,* Springer Berlin / Heidelberg, Oulu, Finland, 2008, pp. 35-46.

25. Fogg, B. J., A behavior model for persuasive design, Proceedings of the 4th International Conference on Persuasive Technology, 2009, pp. 1-7.

26. Forster, A., Murff, H., Peterson, J., Gandhi, T., and Bates, D. The Incidence and Severity of Adverse Events Affecting Patients after Discharge from the Hospital. Ann Intern Med, 138, 3, (2003) 161-167.

27. Forum, N. Q. Safe Practices for Better Healthcare 2006 Update. A Consensus Report., National Quality Forum, accessed on

28. Fox, S. and Duggan, M. Mobile Health, Pew Internet and American Life Project, November 8th, 2012, http://www.pewinternet.org/Reports/2012/Mobile-Health.aspx, accessed on February 20, 2013.

29. Fox, S. and Duggan, M. Tracking for Health, Pew Internet and American Life Project, January 28th, 2013, http://www.pewinternet.org/Reports/2013/Tracking-for-Health.aspx, accessed on February 20, 2013.

30. Friedman, R. H., Stollerman, J. E., Mahoney, D. M., and Rozenblyum, L. The virtual visit: using telecommunications technology to take care of patients. J Am Med Inform Assoc, 4, 6, (1997) 413-425.

31. Goffman, E., The presentation of self in everyday life, Harmondsworth, 1978.

32. Grossman, P. The LifeShirt: a multi-function ambulatory system monitoring health, disease, and medical intervention in the real world. Studies in health technology and informatics, 108, 2004) 133-141.

33. Hawkins, R., Kreuter, M., Resnicow, K., Fishbein, M., and Dijkstra, A. Understanding tailoring in communicating about health. Health Education Research, 23, 3, (2008) 454-466.

34. Hawkins, R. P., Kreuter, M., Resnicow, K., Fishbein, M., and Dijkstra, A. Understanding tailoring in communicating about health. Health education research, 23, 3, (2008) 454.

35. Hilbe, J. M., Logistic Regression Models, Chapman & Hall/CRC, 2009.

36. Hilbe, J. M., Negative binomial regression, Cambridge University Press, 2011.

37. Hingson, R., Zha, W., and Weitzman, E. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. Journal of studies on alcohol and drugs. Supplement, 16, (2009) 12-20.

38. Hoffmann, L., Krämer, N., Lam-chi, A., Kopp, S., Ruttkay, Z., Kipp, M., Nijholt, A., and Vilhjálmsson, H., *Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent?* In *Intelligent Virtual Agents, Volume 5773,* Springer Berlin / Heidelberg, 2009, pp. 159-165.

39. Homans, G. Social Behavior as Exchange. The American Journal of Sociology, 63, 6, (1958) 597-606.

40. Homans, G., Social behavior; its elementary forms, Harcourt, Brace, Jovanovich, New York, 1974.

41. Jack, B., Chetty, V., Anthony, D., Greenwald, J., Sanchez, G., Johnson, A., Forsythe, S., O'Donnell, J., Michael, Manasseh, C., Martin, S., and Culpepper, L. A Reengineered Hospital Discharge Program to Decrease Rehospitalization: A Randomized Trial. Ann Intern Med, 150, 3, (2009) 178-187.

42. Jobe, J. B. and Herrmann, D. J. Implications of models of survey cognition for memory theory. Basic and Applied Memory Research: Volume 1: Theory in Context; Volume 2: Practical Applications, 1996) 193.

43. Kihlstrom, J. F., Eich, E., Sandbrand, D., and Tobias, B. A. Emotion and memory: Implications for self-report. The science of self-report: Implications for research and practice, 2000) 81-99.

44. Larson, R. and Csikszentmihalyi, M. The experience sampling method. New Directions for Methodology of Social & Behavioral Science, 1983).

45. Lee, M. K., Kiesler, S., and Forlizzi, J., Mining behavioral economics to design persuasive technology for healthy choices, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Vancouver, BC, Canada, 2011, 325-334.

46. Li, I., Dey, A., and Forlizzi, J., A stage-based model of personal informatics systems, Proceedings of the 28th international conference on Human factors in computing systems, 2010, pp. 557-566.

47. Liederman, E. M. and Morefield, C. S. Web messaging: a new tool for patient-physician communication. J Am Med Inform Assoc, 10, 3, (2003) 260-270.

48. Loewenstein, G., Brennan, T., and Volpp, K. Asymmetric Paternalism to Improve Health Behaviors. JAMA, 298, 20, (2007) 2415-2417.

49. Long, J. D., Longitudinal Data Analysis for the Behavioral Sciences Using R, SAGE Publications, 2011.

50. McQuiggan, S., Rowe, J., and Lester, J., The effects of empathetic virtual characters on presence in narrative-centered learning environments, Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, 2008, pp. 1511-1520.

51. Menon, G. The effects of accessibility of information in memory on judgments of behavioral frequencies. Journal of Consumer Research, 20, 3, (1993) 431-440.

52. Morisky, D., Green, L., and Levine, D. Concurrent and predictive validity of a self-reported measure of medication adherence. Medical Care, 1986) 67-74.

53. Nakano, Y. and Ishii, R., Estimating user's engagement from eye-gaze behaviors in human-agent conversations, Proceedings of the 14th international conference on Intelligent user interfaces, 2010, pp. 139-148.

54. Nass, C., Steuer, J., and Tauber, E., Computers are social actors, Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, 1994, pp. 72-78.

55. Nass, C., Steuer, J., and Tauber, E., Computers are social actors, CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems, 1994, pp. 72-78.

56. Newman, J., Des, Turner, C., Gribble, J., Cooley, P., and Paone, D. The Differential Effects of Face-to-Face and Computer Interview Modes. Am J Public Health, 92, 2, (2002) 294-297.

57. Ostbye, T., Yarnall, K., Krause, K., Pollak, K., Gradison, M., and Michener, L. Is There Time for Management of Patients With Chronic Diseases in Primary Care? Ann Fam Med, 3, 3, (2005) 209-214.

58. Osterberg, L. and Blaschke, T. Adherence to medication. The New England journal of medicine, 353, 5, (2005) 487-497.

59. Paulino, E. I., Bouvy, M. L., Gastelurrutia, M. A., Guerreiro, M., and Buurma, H. Drug related problems identified by European community pharmacists in patients discharged from hospital. Pharmacy world & science, 26, 6, (2004) 353-360.

60. Pfeifer, L. and Bickmore, T., Is the media equation a flash in the pan?: the durability and longevity of social responses to computers, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 777-780.

61. Pfeifer, L. M. and Bickmore, T., Longitudinal Remote Follow-Up by Intelligent Conversational Agents for Post-Hospitalization Care, 2011.

62. R Core Team, R: A Language and Environment for Statistical Computing, 2012.

63. Reeves, B. and Nass, C., The media equation: how people treat computers, television, and new media like real people and places, CSLI Publications, 1996.

64. Resnicow, K., Davis, R., Zhang, N., Strecher, V., Tolsma, D., Calvi, J., Alexander, G., Anderson, J., Wiese, C., and Cross, W. Tailoring a fruit and vegetable intervention on ethnic identity: Results of a randomized study. Health Psychology, 28, 4, (2009) 394-403.

65. Rich, C., Ponsleur, B., Holroyd, A., and Sidner, C., Recognizing engagement in human-robot interaction, Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction, 2010, pp. 375-382.

66. Risser, J., Jacobson, T., and Kripalani, S. Development and psychometric evaluation of the Self-efficacy for Appropriate Medication Use Scale (SEAMS) in low-literacy patients with chronic disease. Journal of nursing measurement, 15, 3, (2007) 203-219.

67. Schnipper, J., Kirwin, J., Cotugno, M., Wahlstrom, S., Brown, B., Tarvin, E., Kachalia, A., Horng, M., Roy, C., McKean, S., and Bates, D. Role of pharmacist counseling in preventing adverse drug events after hospitalization. Archives of internal medicine, 166, 5, (2006) 565-571.

68. Schwarz, N. Cognitive aspects of survey methodology. Appl. Cognit. Psychol., 21, 2, (2007) 277-287.

69. Strecher, V., Shiffman, S., and West, R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. Addiction, 100, 5, (2005) 682-688.

70. Toomim, M., Kriplean, T., Portner, C., and Landay, J., Utility of human-computer interactions: toward a science of preference measurement, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Vancouver, BC, Canada, 2011, 2275-2284.

71. Tourangeau, R., Jabine, T., Straf, M., and Tanur, J., *Cognitive science and survey methods: A cognitive perspective.* In *Cognitive aspects of survey methodology: Building a bridge between disciplines,* National Academy Press, Washington D.C., 1984, pp. 73-100.

72. Venables, W. N. and Ripley, B. D., Modern Applied Statistics with S, Springer, 2002.

73. Volpp, K., John, L., Troxel, A., Norton, L., Fassbender, J., and Loewenstein, G. Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial. JAMA, 300, 22, (2008) 2631-2637.

74. Volpp, K., Troxel, A., Pauly, M., Glick, H., Puig, A., Asch, D., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., Corbett, E., Weiner, J., and Audrain-McGovern, J. A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation. New England Journal of Medicine, 360, 7, (2009) 699-709.

75. Webb, M., Simmons, V., and Brandon, T. Tailored Interventions for Motivating Smoking Cessation: Using Placebo Tailoring to Examine the Influence of Expectancies and Personalization. Health Psychology, 24, 2, (2005) 179-188.

76. Woolford, S., Clark, S., Strecher, V., and Resnicow, K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare, 16, 8, (2010) 458-461.

APPENDICES

Appendix A Unsigned Consent Form

Northeastern University
Human Subject Research Protection

960 Renaissance Park
Northeastern University
Boston, MA 02115-5000
Tel: 617.373.7570; Fax: 617.373.4595

Northeastern University, College of Computer and Information Science
Name of Investigator(s): ***Timothy Bickmore, Amaura Kemmerer, Laura Pfeifer***
Title of Project: Methods for Conducting Web-Based Health Interviews

Request to Participate in Research
We would like to invite you to participate in a web-based online survey. The survey is part of a research study whose purpose is to examine how self-reported health data can be acquired over a long period of time. This survey will ask you questions about alcohol consumption, and should take about 5-10 minutes to complete per session, for up to 16 weekly sessions. We are asking you to participate in this study because you are a student at Northeastern University. **You must be at least 18 years old to take this survey.**

The decision to participate in this research project is voluntary. You do not have to participate and you can refuse to answer any question. Even if you begin the web-based online survey, you can stop at any time.

The possible risks or discomforts of the study are minimal. Answering the personal survey questions may bring up emotional feelings.

There are no direct benefits to you from participating in this study. However, your responses may help us learn more about health interviews that take place online.

As a token of our appreciation for completing the survey, you will be entered into a weekly drawing for $25. If you win, you will receive your prize via email.

Your part in this study is anonymous to the researchers. However, because of the nature of web based surveys, it is possible that respondents could be identified by the IP address or other electronic record associated with the response. Neither the researcher nor anyone involved with this survey will be capturing those data. Any reports or publications based on this research will use only group data and will not identify you or any individual as being affiliated with this project.
If you have any questions regarding electronic privacy, please feel free to contact Glenn C. Hill, Director of Information Security and Identity Services, via phone at 617-373-7718, or via email at privacy@neu.edu.

If you have any questions about this study, please feel free to contact Laura Pfeifer, phone: 617-373-4605, email: laurap@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

If you have any questions regarding your rights as a research participant, please contact Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University, Boston, MA 02115. Tel: 617.373.7570, Email: irb@neu.edu. You may call anonymously if you wish.

By clicking on the survey link below you are indicating that you consent to participate in this study. Please print out a copy of this consent form for your records.

☐ I am at least 18 years old, I am a student at Northeastern University and I agree to participate in this study.

Continue   No Thanks

Appendix B Study Advertisements

Got 10 Minutes?

Complete a short survey about alcohol and you could win $25.

APPROVED
NU IRB# 10-08-04
VALID 6/22/11
THROUGH 6/22/12

You can participate every week!

Participants must be over 18 and a student at Northeastern University.

http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey
http://wonder.ccs.neu.edu/survey

APPROVED
NU IRB# 10-08-04
VALID: 1/9/12
THROUGH: 6/22/12

Got 10 Minutes?

Complete a short survey about alcohol and you could win $25.

You can participate every week!

Participants must be over 18 and a student at Northeastern University.

http://wonder.ccs.neu.edu/survey

APPROVED
NU IRB# 10-08-04
VALID 1/9/12
THROUGH: 6/22/12

http://wonder.ccs.neu.edu/survey

Got 10 Minutes?

Complete a short survey about alcohol and you could win $25. You can participate every week!

Participants must be over 18 and a student at Northeastern University.

The DRINKS Survey team,
http://wonder.ccs.neu.edu/survey

Appendix C Agent Survey

Age ____

○ Female

○ Male

Please indicate how comfortable you would be talking to each character about your weekly alcohol consumption.

Not at all Comfortable		Neutral		Very Comfortable
1	2	3	4	5

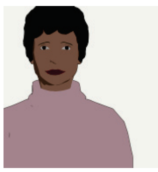
Not at all Comfortable		Neutral		Very Comfortable
1	2	3	4	5

Not at all Comfortable		Neutral		Very Comfortable
1	2	3	4	5

Not at all Comfortable		Neutral		Very Comfortable
1	2	3	4	5

Please place the characters in rank order (1-4) indicating which character you would prefer to talk to about your weekly alcohol consumption. (1 = most preferred, 4 = least preferred).

____



____

____

____

Appendix D Weekly Self-Report Question Set

We'd like to ask you about your experiences with alcohol. It's important to clarify what one drink really means. When asked how much you drink, please use the following definitions:

For all questions, one drink equals:
- 4oz. wine
- 10oz. wine cooler
- 12oz. beer (8 oz. of Canadian, Malt Liquor, or Ice Beers, or 10 oz. of Microbrew)
- 1 Cocktail with 1 oz. of 100 proof liquor or 1 1/2 oz. of 80 proof liquor.

Q1 How many days of the week did you drink alcohol during the past 7 days?
I did not drink at all.*[Code = 1]*
Once a day or less*[Code = 2]*
About once a month*[Code = 3]*
Two to three times a month*[Code = 4]*
Once or twice a week*[Code = 5]*
Three to four times a week*[Code = 6]*
Nearly every day*[Code = 7]*

Q2 Please indicate the average number of drinks you consume on one occasion at parties or when socializing
None, I don't drink.*[Code = 1]*
1*[Code = 2]*
2*[Code = 3]*
3*[Code = 4]*
4*[Code = 5]*
5*[Code = 6]*
6*[Code = 7]*
7*[Code = 8]*
8*[Code = 9]*
9*[Code = 10]*
10*[Code = 11]*
11*[Code = 12]*
12*[Code = 13]*
13*[Code = 14]*
14*[Code = 15]*

15 or more*[Code = 16]*

Q3 Think of the occasion you drank the most this past 7 days. How <u>much</u> did you drink?
0 drinks*[Code = 1]*
1 drink*[Code = 2]*
2 drinks*[Code = 3]*
3 drinks*[Code = 4]*
4 drinks*[Code = 5]*
5 drinks*[Code = 6]*
6 drinks*[Code = 7]*
7 drinks*[Code = 8]*
8 drinks*[Code = 9]*
9 drinks*[Code = 10]*
10 drinks*[Code = 11]*
11 drinks*[Code = 12]*
12 drinks*[Code = 13]*
13 drinks*[Code = 14]*
14 drinks*[Code = 15]*
15 drinks*[Code = 16]*
16 drinks*[Code = 17]*
17 drinks*[Code = 18]*
18 drinks*[Code = 19]*
19 drinks*[Code = 20]*
20 drinks*[Code = 21]*
21 drinks*[Code = 22]*
22 drinks*[Code = 23]*
23 drinks*[Code = 24]*
24 drinks*[Code = 25]*
25+ drinks*[Code = 26]*

Q4 Think of the occasion you drank the most this past 7 days. How <u>many HOURS</u> did you spend drinking on that occasion?
I did not drink in the past month.*[Code = 1]*
Less than one hour*[Code = 2]*
1 - 2 hours*[Code = 3]*
2 - 3 hours*[Code = 4]*
3 - 4 hours*[Code = 5]*

4 - 5 hours*[Code = 6]*
5 - 6 hours*[Code = 7]*
6 - 7 hours*[Code = 8]*
7 - 8 hours*[Code = 9]*
8 - 9 hours*[Code = 10]*
9 - 10 hours*[Code = 11]*
More than 10 hours*[Code = 12]*

Q5 How many drinks per occasion do you consider moderate (not excessive) for yourself?
None, I don't drink.*[Code = 1]*
One*[Code = 2]*
Two *[Code = 3]*
Three*[Code = 4]*
Four*[Code = 5]*
Five*[Code = 6]*
Six or more*[Code = 7]*

Q6 When you choose not to drink excessively, what are some of the important reasons why not? (Check all that apply)
It interferes with my school work.*[Code = 1]*
It interferes with my job.*[Code = 2]*
I am worried about the negative effect on my health.*[Code = 3]*
I do not want to lose control.*[Code = 4]*
It interferes with my athletic activities.*[Code = 5]*
I do not like the way I act when I drink that much.*[Code = 6]*
Other (please specify)*[Code = 7] [TextBox]*

Different things happen to people as a result of their alcohol use. Some of these things are listed below. Please indicate how many times each has happened to you during the **past 7 days** while you were drinking alcohol:

Q7 Had a hangover
Not applicable/Don't drink*[Code = 1]*
Didn't happen*[Code = 2]*
1 time*[Code = 3]*
2 times*[Code = 4]*

3 times*[Code = 5]*
4 times*[Code = 6]*
5 or more times*[Code = 7]*

Q8 Attended class hung over or been unfocused in class
Not applicable/Don't drink*[Code = 1]*
Didn't happen*[Code = 2]*
1 time*[Code = 3]*
2 times*[Code = 4]*
3 times*[Code = 5]*
4 times*[Code = 6]*
5 or more times*[Code = 7]*

Q9 Missed a class because of drinking
Not applicable/Don't drink*[Code = 1]*
Didn't happen*[Code = 2]*
1 time*[Code = 3]*
2 times*[Code = 4]*
3 times*[Code = 5]*
4 times*[Code = 6]*
5 or more times*[Code = 7]*

Q10 Been a passenger in a car with a driver who had been drinking
Not applicable/Don't drink*[Code = 1]*
Didn't happen*[Code = 2]*
1 time*[Code = 3]*
2 times*[Code = 4]*
3 times*[Code = 5]*
4 times*[Code = 6]*
5 or more times*[Code = 7]*

Q11 Forgotten what happened when you were drinking (blacked out)
Not applicable/Don't drink*[Code = 1]*
Didn't happen*[Code = 2]*
1 time*[Code = 3]*
2 times*[Code = 4]*
3 times*[Code = 5]*
4 times*[Code = 6]*

5 or more times*[Code = 7]*

Q12 Urinated or vomited in a public setting

Not applicable/Don't drink*[Code = 1]*

Didn't happen*[Code = 2]*

1 time*[Code = 3]*

2 times*[Code = 4]*

3 times*[Code = 5]*

4 times*[Code = 6]*

5 or more times*[Code = 7]*

Q13 Injured yourself as a result of drinking

Not applicable/Don't drink*[Code = 1]*

Didn't happen*[Code = 2]*

1 time*[Code = 3]*

2 times*[Code = 4]*

3 times*[Code = 5]*

4 times*[Code = 6]*

5 or more times*[Code = 7]*

Q14 Gotten into a physical fight as a result of drinking

Not applicable/Don't drink*[Code = 1]*

Didn't happen*[Code = 2]*

1 time*[Code = 3]*

2 times*[Code = 4]*

3 times*[Code = 5]*

4 times*[Code = 6]*

5 or more times*[Code = 7]*

Q15 Had intercourse when you ordinarily would not

Not applicable/Don't drink*[Code = 1]*

Didn't happen*[Code = 2]*

1 time*[Code = 3]*

2 times*[Code = 4]*

3 times*[Code = 5]*

4 times*[Code = 6]*

5 or more times*[Code = 7]*

Q16 Failed to use safe sex practices when you ordinarily would have

Not applicable/Don't drink*[Code = 1]*
Didn't happen*[Code = 2]*
1 time*[Code = 3]*
2 times*[Code = 4]*
3 times*[Code = 5]*
4 times*[Code = 6]*
5 or more times*[Code = 7]*

This next section asks about your behaviors while drinking. Please indicate how often you do the following based on the corresponding choices.

During the past 7 days, if you "partied"/socialized, how often did you . . .?

Q17 Switch between alcoholic and non-alcoholic beverages
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*
Never*[Code = 1]*

Q18 Determine, in advance, not to exceed a set number of drinks
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*
Never*[Code = 1]*

Q19 Choose not to drink alcohol
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*
Never*[Code = 1]*

Q20 Use a designated driver
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*
Never*[Code = 1]*

Q21 Eat before and/or during drinking
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*
Never*[Code = 1]*

Q22 Have a friend let you know when you've had enough
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*
Never*[Code = 1]*

Q23 Keep track of how many drinks you were having
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*
Never*[Code = 1]*

Q24 Pace your drinks to 1 or fewer per hour
Not applicable/Don't drink*[Code = 0]*
Always*[Code = 5]*
Usually*[Code = 4]*
Sometimes*[Code = 3]*
Rarely *[Code = 2]*

Never*[Code = 1]*

Q25 Avoid drinking games

Not applicable/Don't drink*[Code = 0]*

Always*[Code = 5]*

Usually*[Code = 4]*

Sometimes*[Code = 3]*

Rarely *[Code = 2]*

Never*[Code = 1]*

Q26 Drink an alcohol look-alike such as: non-alcoholic beer, punch, juice, or water

Not applicable/Don't drink*[Code = 0]*

Always*[Code = 5]*

Usually*[Code = 4]*

Sometimes*[Code = 3]*

Rarely *[Code = 2]*

Never*[Code = 1]*

Q27 Think about your BAC (Blood Alcohol Concentration)

Not applicable/Don't drink*[Code = 0]*

Always*[Code = 5]*

Usually*[Code = 4]*

Sometimes*[Code = 3]*

Rarely *[Code = 2]*

Never*[Code = 1]*

Q28 Deliberately pick a drink that would affect you more slowly

Not applicable/Don't drink*[Code = 0]*

Always*[Code = 5]*

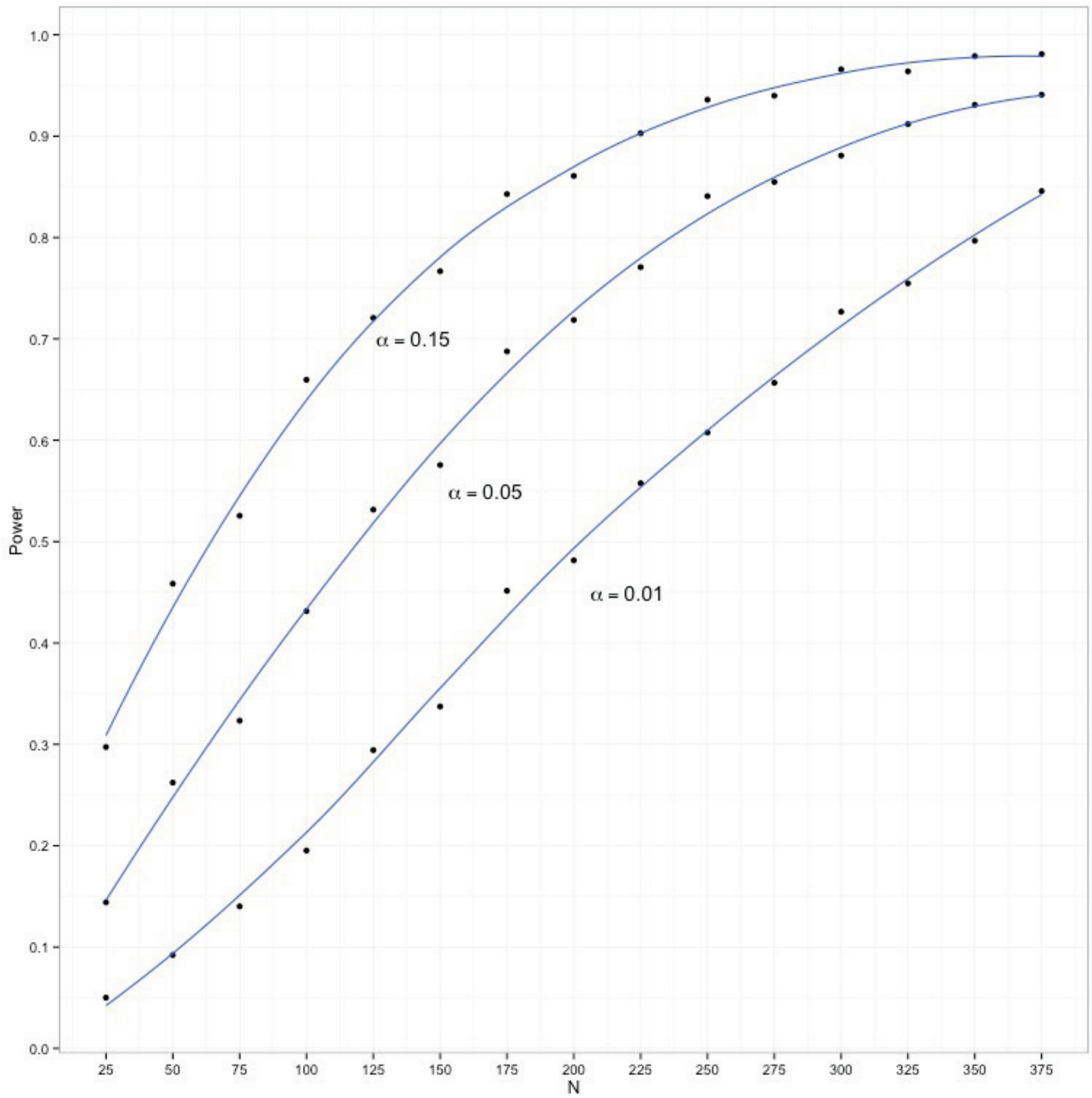
Usually*[Code = 4]*

Sometimes*[Code = 3]*

Rarely *[Code = 2]*

Never*[Code = 1]*

Appendix E Power Curves for three α levels



Appendix F Experience Questionnaire

Participants completed this form once, online, after the entire study was over.

Did you enjoy filling out the weekly surveys?

Not at all So –So Very Much

Overall, how useful were the surveys to you?

Not at all So –So Very Much

How burdensome was filling out the weekly surveys?

Not at all So –So Very Much

How time-consuming was it to complete the weekly surveys?

Not at all So –So Very Much

Thinking back, how honest were your answers to the weekly survey questions?

Not at all So –So Completely Honest

[For FEEDBACK participants only]

How useful was the weekly feedback you received at the end of each weekly survey?

Not at all So –So Very Much

How much did the weekly feedback influence your decision to fill out the survey each week?

Not at all So –So Very Much

What motivated you more when deciding whether or not to fill out the survey?

- The personal feedback
- The entry into the drawing for $25

[for AGENT participants only]

How satisfied were you with Tanya?

Not at all So –So Very Much

How much did you trust Tanya?

Not at all So –So Very Much

You filled out the survey [only x times, why?] [x times. What made you come back and fill it out several times?]

Any other thoughts?