Evaluation of Two Mobile Nutrition Tracking Applications for Chronically Ill Populations with Low Literacy Skills

Katie A. Siek¹, Kay H. Connelly², Beenish Chaudry², Desiree Lambert³, and Janet L. Welch⁴

¹University of Colorado at Boulder, Department of Computer Science
430 UCB, Boulder, CO 80309 USA
ksiek@cs.colorado.edu

²Indiana University, Department of Computer Science
Lindley Hall 301E; 150 S. Woodlawn Ave.; Bloomington, IN 47405 USA
connelly,bchaudry@cs.indiana.edu

³Indiana University Hospital, Adult Renal Services
Indianapolis, IN 46202 USA
desireelambert@hotmail.com

⁴Indiana University, School of Nursing
Indianapolis, IN 46202 USA
jwelch@iupui.edu

Abstract In this chapter, we discuss two case studies that compared and contrasted the use of barcode scanning, voice recording, and patient self reporting as a means to monitor the nutritional intake of a chronically ill population. In the first study, we found that participants preferred unstructured voice recordings rather than barcode scanning. Since unstructured voice recordings require costly transcription and analysis, we conducted a second case study where participants used barcode scanning or an Integrated Voice Response system to record nutritional intake. We found that although the latter input method provided participants with a faster method to input food items, participants had difficulty using the system despite training.

Keywords: Assistive Technologies, Human Computer Interaction, Chronic Kidney Disease, Health Informatics, Integrated Voice Response, Consumer Health Informatics, Self Management, Hemodialysis
1. Introduction

Chronic diseases, such as Chronic Kidney Disease (CKD) and heart disease, are among the leading causes of death and disability in the world. At least half of the chronic disease related deaths could be prevented by adopting a healthy lifestyle, such as good nutrition, increased physical activity, and cessation of tobacco use. Researchers believe that the world must put a higher priority on interventions to help prevent and successfully manage chronic illness (Preventing Chronic Diseases: A Vital Investment, 2005).

Current interventions to help chronically ill populations improve their nutritional health and self-manage therapeutic diets include paper-based food diaries, 24 hour recalls, and food frequency questionnaires (Dwyer, Picciano, & Raiten, 2003; Resnicow et al., 2000) Patients who use these interventions must have high literacy and memory recall skills. Unfortunately, over a quarter of the United States population do not have the necessary literacy or numeracy skills needed to successfully self-monitor themselves (Kirsch et al., 1993). If people cannot self-monitor themselves, they cannot manage their chronic conditions (HRSA Literacy) and may lead them to worse health outcomes (Schillinger et al., 2002). In addition, to administer current interventions medical professionals must spend a significant amount of time evaluating the data from paper-based forms.

We are currently developing a mobile handheld application to assist CKD patients on Hemodialysis monitor and maintain their nutritional intake. Initially, we thought a personal digital assistant (PDA) would be the best solution for health professionals and patients (Connelly, Faber, Rogers, Siek, & Toscos, 2006). Participants could scan barcodes on food items for their primary input or select items from an interface as a secondary input. These input mechanisms are idea for low literacy populations because there is no reading required – participants only have to identify a barcode or select a picture. Health professionals could easily administer the intervention and evaluate data without intermediate steps of electronic transcription. The low literacy chronically ill participants benefit from
using the application because they can use the application anytime they consumed a food item, receive immediate visual feedback on their nutritional intake, and make decisions on a prospective basis. In addition, the interface and content could be customized for populations with varying literacy and computing skills.

In this chapter, as part of a larger study, we will compare and contrast the use of barcode scanning, Integrated Voice Response System (IVRS), and patient self reporting as a means to monitor the nutritional intake relative to their dietary prescription of CKD patients. In the first case study we found that participants preferred unstructured voice recordings rather than barcode scanning. Unstructured voice recordings are difficult to automatically parse and require transcription. We had to find out if patients would use a menu-based structured voice input system, such as IVRSs for automated recognition. In the second case study, we explored participant use of an IVRS and found although the system provided participants with a quicker way to input food items, participants had difficulty using the system and some could not use the system despite training. We will discuss the methodology and findings from these two case studies. We will conclude the chapter with lessons learned during the user study and provide considerations for future areas of research.

2. Related Work

PDAs with scanner input and mobile phones used for IVRS input gather information in many domains. PDAs and scanners have been used to show clinicians videos about specific unit appliances (Brandt, Björgvinsson, Hillgren, Bergqvist, & Emilson, 2002), save and search for information about food products, music, and books (Bernheim, Combs, Smith, & Gupta, 2005), and obtain information about an environment from embedded barcodes (Fitzmaurice, Khan, Buxton, Kurtenback, & Balakrishnan, 2003). Mobile phones used for IVRSs have been used for patient counseling to enhance time spent with health professionals (Glasgow, Bull, Piette, & Steiner, 2004) and assess patient status with chronic illnesses such as depression, cancer, heart failure, and diabetes (Piette, 2000). In this section, we discuss specifically how PDAs and
mobile phones have been used for interventions and nutritional monitoring.

### 2.1. PDA Nutrition Monitoring Interventions

Currently, there are many PDA applications that can assist with the self-monitoring of nutritional intake. The United States Department of Agriculture (USDA) has a PDA nutrient database that provides people with a mechanism for looking up the nutritional information of foods. Users must correctly type the first few letters of a food item they are looking for into a search box and then click through a series of menus to find the appropriate food item based on portion size and preparation ("USDA Palm OS Search," 2008).

DietMatePro ("DietMatePro") and BalanceLog ("BalanceLog") use the USDA database along with other fast food nutritional information to create a PDA program that provides users with a way to save consumption information for a set of specific nutrients. CalorieKing ("CalorieKing") uses its own nutritional database and provides users the ability to save consumption information. In addition, it has a nutritional tracking application specific to diabetic populations. The applications are similar to the USDA database in that users must be able to spell the first few letters of food items. Unlike the USDA database, users must type in portion size. Food items are also broken up into three subsections - breakfast, lunch, and dinner. The nutritional analysis is given on a separate screen.

Researchers at Indiana University studied how three people with CKD used DietMatePro to monitor nutritional consumption over a three-month period. They found participants had difficulty navigating standard PDA menu navigation and preferred using a large PDA screen with touch sensitive icons (Dowell & Welch, 2006). Sevick and colleagues evaluated how five CKD participants used BalanceLog over a four-month period. They found that participants improved their dietary intake using the electronic nutrition monitoring system (Sevick et al., 2005). Both applications evaluated in these studies required significant literacy and cognitive skills.

Stephen Intille et al. created a proof-of-concept PDA application that
provides users with a way to scan food items and obtain nutritional information to assist users in making healthy choices (Intille, Kukla, Farzanfar, & Bakr, 2003). The application did not have an extensive UPC/nutrition database because none are freely available. Although the application does not allow users to save intake information, the application shows that integration of scanners and nutrition information is possible given enough resources.

Researchers at Microsoft created a generic barcode look-up system that gave participants the opportunity to look up product information available online about specific food items. During their five-week study with twenty participants familiar with PDA technology, they found participants had mixed reactions to the system in terms of enjoyment and usefulness. Similar to a recent mobile phone study at Georgia Tech (Patel, Kientz, Hayes, Bhat, & Abowd, 2006), participants in the Microsoft study did not always bring the PDA with them despite being enthusiastic PDA owners (Bernheim et al., 2005).

In addition to PDA monitoring of nutrition, there have been great strides in mobile phone nutrition monitoring applications. Those who use the commercial application myFoodPhone take pictures of foods they are consuming with their mobile phone and post the pictures to an online food journal to receive feedback from a nutritionist ("myFoodPhone"). However, users must have access to a computer and be able to properly upload the information. Tsai and colleagues developed a mobile phone application where participants input food items via the keypad and immediately receive feedback on caloric balance on the phone screen. During the month-long feasibility study with 15 college-educated participants, they found participants preferred the mobile phone input system to traditional paper and pen journaling methods (Tsai et al., 2006). These applications use mobile phone input via pictures or key presses, but a more natural input interaction would be voice recognition software. In the next subsection, we discuss the use of IVRSs in health interventions.

2.2 Integrated Voice Response Systems in Interventions

IVRSs in healthcare have been used for reminders, surveys, screening and
assessments, and disease management (Lavigne, 1998). A review of IVRS feasibility studies in populations with chronic illnesses such as depression, cancer, heart failure, and diabetes led Piette to conclude that IVRSs are feasible for chronically ill populations, including populations that have mental health problems or low-income (Piette, Weinberger, & McPhee, 2000). According to Mundt et al. (2002), IVRSs benefit healthcare because they ensure procedural standardization, automatic data scoring, direct electronic storage, and remote accessibility from multiple locations.

Long-term alcoholism and coital studies have supported the feasibility of interventions (Aharonovich et al., 2006; Helzer, Badger, Searles, Rose, & Mongeon, 2006; Mundt et al., 2002; Hays, Irsula, McMullen, & Feldblum, 2001; Schroder et al., 2007), though the populations are well educated and technically savvy. Notably Aiemagno et al. (1996) assessed substance abuse treatment needs among 207 homeless adults, finding some evidence of greater disclosure of risky behaviors with IVRS.

Long-term IVRS usage has had mixed reporting rates and health-related quality of life benefits. A 91 day coital study by Schroder et al. (2007) found a significant decrease in self-reports over time, while a two-year study with daily reports of alcohol consumption by Helzer et al. (2006) had a 91.7% reporting rate, but compensated participants per call. Daily alcoholism reports among HIV patients found a decrease in drinking over time (Aharonovich et al., 2006). In contrast, an IVRS intervention with diabetes patients found no measurable effects on anxiety or health-related quality of life (Piette et al., 2000).

Disease management IVRSs that act as diaries have improved participant satisfaction over paper diaries (Hays et al., 2001). Two recent studies have challenged this result (Weiler, Christ, Woodworth, Weiler, & Weiler, 2004; Stuart, Laraia, Ornstein, & Nietert, 2003). Weiler et al. (2004) conducted a 3-week, 3-way, cross-over trial including 87 adults with allergic rhinitis recording daily through an IVRS or paper diary. A majority (85%) of the participants preferred the paper instrument, whereas only 4% preferred the IVRS. Stuart et al. (2003) conducted a year-long study with 642 patients to enhance antidepressant medication compliance. One of three different treatment strategies included a 12-week IVRS
component, yet no significant differences in patient compliance were found and 50% of the 232 patients assigned to the IVRS component either never used the system or stopped before the 12 weeks were completed.

IVRSs in healthcare typically limit response input to yes/no or numeric responses (Levin & Levin, 2006). Recent work exploring how IVRS vocabulary is expanded in a two week pain monitoring study by Levin et al. found that number of sessions per subject ranged from 1 to 20, accumulating 171 complete sessions and 2,437 dialogue turns. Only 2% of responses recorded were out-of-vocabulary. Though volunteers in the evaluation were not trained, the results suggested that training sessions could have significant value and that IVR-based data collection is not a replacement for existing data collection, but simply another option for healthcare providers and researchers.

Whereas the research discussed in this section primarily focuses on how well educated, technically savvy users interact with various technology interventions for monitoring in their everyday lives, our work deals with how non-technical users with varying literacy skills use two different types of input mechanisms. The IVRS literature especially shows how compliance is studied with this technology, but it does not research if participants could use the system and how the system can be improved. We are iteratively studying input mechanisms because our target population will depend on the application for their personal health and thus will have to find using the application efficient and enjoyable for long-term adoption. This chapter details two case studies that provided insight into finding the ideal input mechanism for nutrition monitoring.
3. Case Study 1: Barcode and Unstructured Voice Recording

In this section we present our initial formative study that examines what, when, and how CKD participants input food items into an electronic intake monitoring application. The study required that participants complete PDA application training exercises, meet with researchers during dialysis sessions three times per week, and use the Barcode Ed application during two study phases for a total of three weeks. Table 1 shows that there was a three week break between the two phases that allowed researchers to evaluate the data and decide on future directions for the application. All interactions with participants were done during dialysis treatment in an urban, hospital-based, outpatient dialysis unit. We documented how we conducted user studies in a dialysis ward in previous work (Siek & Connelly, 2006).

### 3.1. Methodology

In this section, we discuss why we selected the hardware and application used for this case study.

#### 3.1.1. Hardware

We chose an off-the-shelf Palm OS Tungsten T3 PDA for our study. The

<table>
<thead>
<tr>
<th>Study Phase #</th>
<th>Length of Phase</th>
<th>Motivating Research Question(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>1 week</td>
<td>1. Can participants find, identify, and successfully scan barcodes on food items?</td>
</tr>
<tr>
<td>Break</td>
<td>3 weeks</td>
<td></td>
</tr>
</tbody>
</table>
| Phase 2       | 2 weeks         | 1. Will participants remember how to use this application after a 3 week break?  
|               |                 | 2. Will participants actively participate without meeting with researchers every other day? |

Table 1: Overview of Case Study 1
Tungsten T3 has an expandable screen, large buttons, voice recorder, SDIO slot, 52 MB of memory, and Bluetooth. We chose an off-the-shelf PDA so the results could be useful to the consumer health informatics community for future studies.

The Socket In-Hand SDIO card scanner (Socket Scanner) was chosen as the barcode scanner because it was small, easy to use, and gave visual and audio feedback to users. Participants must press the predefined scanning button, line up the scanning light perpendicular to the barcode, and hold the PDA and object steady. The PDA beeps and shows appropriate feedback when participants have successfully scanned a barcode. Previous studies have shown that CKD patients can use the Tungsten T3 and Socket Scanner (Moor, Connelly, & Rogers, 2004)

3.1.2. Application Design

We created a simple application, Barcode Ed, because we wanted to isolate participants' ability to scan and yet have an alternative input mechanism (e.g., voice input) to record all food items consumed. In initial interviews, half of the CKD patients said they did not eat any foods with barcodes. However, once they were prompted, we found they primarily ate frozen, canned, and prepared foods. Thus, for participants to use an easy input mechanism like scanning, they would have to learn how to identify barcodes and use the scanner. We only used scanning and voice recording in this study because we did not want to overburden novice computer users with a complex interface because they may have decreased cognitive function during treatment (Martin-Lester, 1997).
Figure 1: Screen shots from Barcode Ed. (a) Home Screen; (b-c) Voice recording and playback screens; (d-e) Barcode Scanning feedback screens
Figure 1. Since our user group had low literacy skills, we relied on icons 11mm large with some text for navigation. We found these CKD patients could view icons 10mm or larger (Moor et al., 2004). When participants turned on the PDA, they would view the Home screen. Participants could choose to voice record by pressing the Voice button or scan a barcode by pressing the Scan button. As soon as participants pressed the Voice button, the application would begin voice recording and show participants how many minutes and seconds they recorded on the Voice recording screen. When participants were finished recording, they could press the Stop button and play back their recording on the Voice recording play back screen. When participants were satisfied with their recording, they could return to the Home screen. When participants pressed the Scan button, participants could see a red laser line emitted by the scanner. Participants lined the scanner line perpendicularly across the barcode they were attempting to scan. If the food item was successfully scanned, a green check mark would appear on the Barcode scanning success screen. If the food item was not successfully scanned, a red “X” would appear on the Barcode scanning unsuccessful page and participants could decide whether to scan again or return to the home screen and voice record the item instead.

The application recorded the time the participant first pressed a Scan or Voice button, the barcode number or voice recording, and the time the recording was saved. We also recorded how many times participants played back their voice recordings. We did not record how many failed barcode scans were attempted because it was difficult to differentiate when a participant was scanning the same object or gave up and attempted to scan a new object during the same period of time. Also, participants sometimes did not use the scan button on the Barcode scanning unsuccessful page - instead they went to the Home screen and then pressed the scan button again. The times recorded assisted us in determining when participants recorded what they consumed. Recording the number of voice recording playbacks gave us insight into how participants used the application.
3.2. Participants

Participants were asked to participate in the study during their dialysis session. They had to be (1) over 21 years of age, (2) able to make their own food or have the ability to go out and purchase food, (3) willing to meet with researchers during each dialysis session during the week, and (4) willing to carry the PDA and scanner with them and input food items consumed. Ten participants volunteered for the study. During the first phase, one participant could not participate anymore because of a medical emergency and another participant dropped out because he did not want to record what he was eating ($n = 8$). We lost two participants during phase two for similar reasons ($n = 6$).

The average age of participants was 52 years old (s.d. = 16.28). Half of the participants were male; all of the participants were black. One participant completed an associate degree, four participants graduated from high school, and one participant completed 10th grade. Participants had been receiving dialysis treatments on average of five years (s.d. = 3.5 years).

Only four participants reported using a computer. Usage frequency ranged from every couple of months to once a week for a half hour. Participants primarily played games and surfed the Internet. Only two of the participants owned a mobile phone that they used for emergencies only.

The participants were equally divided about how many food items they consumed had barcodes - some thought all and some did not think any food items had barcodes. Five patients said they did not have to monitor any nutrients or fluid. However, by the end of the first phase, the researcher had established a trusting relationship with the participants and found that all of them had to monitor fluid and nutrients such as sodium, potassium, phosphorus, and protein. None of the patients recorded their fluid or nutrient consumption prior to the study.
3.3. Design and Procedure

We met with participants during dialysis sessions four times during each phase of the study for approximately 30 minutes. During the first session, we collected background information and taught participants how to turn the PDA on, insert the scanner, and use the application. Participants practiced scanning various food items and voice recording messages. Researchers met with participants during the study sessions to discuss any problems participants may have had with the PDA, retrain participants how to do certain tasks (e.g., barcode scanning), and collect recordings and barcodes from the PDAs via Bluetooth. The researchers played back the voice recordings to ensure the correct information was transcribed and informed participants if they voice recorded a food item that could have had a barcode. Participants returned the PDAs at the end of each phase of the study, talked to researchers about their experience, and verbally completed a modified Questionnaire for User Interface Satisfaction (QUIS) (Chin, Diehl, & Norman, 1988) survey. Participants received ten dollars (U.S.) for every time they met with researchers for a total of thirty dollars during phase 1. For phase 2, participants received five dollars each time they met with the researcher for a total of fifteen dollars.

Competency skills tests were administered at the end of the second and fourth meeting of the first phase and during the first and last meeting of the second phase to test basic Barcode Ed skills - turning the PDA on; inserting the scanner; scanning three to five objects with different physical qualities; voice recording with playback; and do a combined barcode scanning and voice recording sequence. The items participants had to scan ranged from a cardboard soup mix box that is easy to scan because of the material; a can of chips that is somewhat difficult to scan because of material and barcode orientation; and a bag of candy that is difficult to scan because it is amorphous and made of shiny material. Researchers measured how many times it took participants to successfully complete each task. We measure the time it took to complete each competency skill with the Barcode Ed application.
Participants were instructed to scan or voice record food items when they consumed the items. Participants were instructed to scan the barcodes on food items first and voice recording items only if they could not scan the barcode or if a food item did not have a barcode. When participants mastered scanning and voice recording, researchers encouraged participants to note via voice recording how much they were consuming and the portion size. Each participant was given a phone number of a researcher to contact if they had any questions during the study. Participants were given a visual state diagram of the application to assist them with any questions regarding use of the application that had images Figure 1.

3.4. Findings

The key findings of our study were:

- Participants preferred voice recording once they mastered the application
- Participants with low literacy skills needed extra instruction on how to sufficiently describe food items for voice recordings
- Participants reported more individual food items with the Barcode Ed application than what they thought they consumed
- Electronic monitoring provides researchers with ways to identify participant compliance

In this section, we present the results in more detail.

3.4.1. Barcode Scanning and Voice Recording Frequency

One of the motivating factors for the first phase of the Barcode Education study was to teach participants how to identify and scan barcodes. In Figure 2, we see that there was a learning curve associated with identifying and scanning barcodes during the first study phase. Participants voice recorded more individual food items during the first few days of the study because they were either unsure of where the barcode was located on the food item or were unable to scan the barcode. Gradually during the week, we noticed an increase of barcode scans up until the last day of the first
A goal of the second study phase was to see if this trend of increased barcode scans would persist and if participants would continue actively participating in the study without meeting with researchers every other day. The first two days of the second study phase were promising because participants were scanning everything they consumed and only voice recorded items without barcodes (e.g., fresh produce). However, after the second day, participants realized *everything* had barcodes and were overwhelmed with the amount of time it took to scan each individual food item. Thus, during the third and fourth day of the study, participants began voice recording food items they had previously scanned to save time.

The lack of items input at the end of phase one shown in Figure 2 can be attributed to not seeing a study researcher to encourage them to participate at the end of the week. Indeed, three participants acknowledged that they had forgotten to input foods on more than one
occasion because they had not been visited by a researcher. Participants were more likely to forget to input foods on weekends (days six, seven, thirteen, and fourteen).

During the second week of the second study phase, participants rarely scanned barcodes and typically voice recorded what they consumed. The voice recordings listed multiple food items in an unstructured manner. For example, one participant recorded, “I ate a small apple, a lunch meat sandwich, and a boost for lunch. I ate … eggs, and bacon for breakfast. Tonight for dinner I am planning on eating…”

When we asked participants why they scanned more on the 13th day of the study, they told us that they had remembered they would see a researcher on the following day to finish the study. Of course, the researchers called the participants to remind them to bring the PDAs to the last day of the study.

3.4.2. Voice Recording Food Items

We thought voice recording food items was an easy alternative input method when participants could not scan. However, participants with low literacy skills were initially unable to give sufficient identifying information in their voice recordings. Since the participants were unable to read the name on the food item, they were not able to say what they were eating (e.g., Lucky Charms cereal). Instead, participants said, “I had cereal for breakfast.” When we met with participants and played the recordings for transcription, we were able to suggest ways to be more descriptive (e.g., describe what is on the box) to help us identify the food items. After two to three sessions, the low literacy participants recorded more descriptive input (e.g., I ate the cereal with the leprechaun and rainbow on the box) and it was easier to identify what they were eating. However, even with descriptive input, we were unable to identify three of the items mentioned in the 195 recordings.
3.4.3. Barcode Ed versus Self Reported Food Items

In pre-study interviews, participants told us they had good and bad days that affected how much they consumed and discussed how many meals they typically consumed on each of these days. The participants usually had a good and bad day fairly recently and could easily describe to us the exact number of items they consumed. We asked participants if they had a good or bad day each time we met during the first study phase. We then compared how many items they electronically input to how many items they said they would consume, including the type of day they were having in the calculation. Participants ate more than they estimated for an average of three days (s.d. = 2.875) during the seven day period. When participants did consume more than they estimated, they typically consumed on average 3.5 more items than estimated – nearly doubling their normally recorded intake of 4.4 items (s.d. = 3.27)¹.

¹ The standard deviation is large because it depends if participants were having a good or bad day in terms of consumption and physical health.
3.4.4. Participant Compliance

For this study, we loosely defined compliance as inputting at least one food item a day. Similar to traditional monitoring methods, participants could back fill and modify their compliance record. However, unlike traditional methods, with electronic nutrition monitoring, researchers can identify this behavior more quickly. For example, a participant back filled entries in Figure 4 (green circle) by recording what he had consumed for the last two days since he had not actively participated. Another indicator of back filling is the number of times a participant recorded a food item that could be scanned during a short time interval since participants cannot scan items that have been consumed and discarded.

Participants were unaware that we were recording the date and time of inputs and thus assumed if they said, “Today, on February 11, I ate…” the researcher would not know that it was recorded on February 12. When we showed participants similar graphs as shown here, participants attempted to decrease backfilling or were more truthful in disclosing lack of participation. In addition to backfilling, we see in Figure 4 an example of End-Of-Study compliance where the participant realizes the end of the study is near and increases participation in hopes the researcher will not notice.

![Figure 4: Example of voice recordings, barcode scans, and voice recordings that should have been barcode scans (wrong record) a participant made during the first phase. The participant did back filling as shown by the green circle and increased input during the end of the study. The dotted lines denote the next day. Faces denote when researchers met with participants.](image-url)
We discussed in Section 3.4.1. that once participants realized *everything* had a barcode on it, participants began to voice record more. We see this behavior in Figure 4 – the participant starts to scan items, but then starts to *hoard* consumption information in one voice recording a day. The participant told us in a post-study interview that reporting everything he ate in one voice recording was more time efficient.

### 4. Case Study 2: Barcode and IVR

In this section we present our follow-up study that examines what, when, and how CKD participants input food items into an electronic intake monitoring application and an IVRS with a borrowed mobile phone. Similar to the first case study, participants complete PDA application and mobile phone training exercises, meet with researchers during dialysis sessions, and use either the PDA barcode monitoring application or the mobile phone IVRS over a two week period. Participants were recruited and trained as the same dialysis unit from the first case study.

#### 4.1. Methodology

In this section, we discuss the hardware selected for the study and design of the applications used for capturing participant input.

**4.1.1. Hardware**

We designed an application to run on a PDA with an attached barcode scanner to test participants’ ability to scan barcodes of food items. For the PDA, we chose an off-the-shelf Pocket PC from Hewlett Packard: the iPAQ hx2495b. We decided to use an iPAQ for the second case study because the Windows CE operating system provides a better rapid prototyping environment with Visual Studio .NET CF. The iPAQ hardware includes a large, color, touch screen, stylus and large buttons. We used the same SDIO In-Hand Scan Cards (SDSC Series 3E).
We provided participants with a Nokia 6682 mobile phone to provide participants the ability to record food at any time. The phone has a high-resolution color screen and large buttons. As with the PDAs, we provided soft leather cases with belt clips to the participants. We programmed the phone so that pressing any button would dial the number for recording their food items.

**4.1.2. Application Design**

The scanning application was similar to the Barcode Ed application used in the first case study. The only difference in the application was that participants did not have the ability to record unstructured voice recordings. If the food item did not have a barcode, the participant could not record the food item.

We implemented an IVRS that could be accessed with any phone to test participants’ ability to use structured voice input. As Figure 5 shows, we implemented the IVRS by transferring a call through a Session Initiation Protocol (SIP) gateway to Voxeo, an IVRS platform provider. The caller identifier was then submitted to our web server where a CGI script selected...
participant grammar files (Nuance GSL Grammar Format), returning a VoiceXML form to collect items.

The initial grammar included 152 food items and 2 command operators, ‘done’ and ‘wrong.’ The same grammar was available at every prompt. ‘Done’ submitted the results and terminated the call. ‘Wrong’ incremented a counter, such that if said twice without an intervening positive recognition, the participant was prompted to voice record the item for addition to the grammar. With food items, 45 were single words (e.g., bagel), 12 were compound words (e.g., fish sticks), 27 used optional phrase operators where a portion need not be uttered (e.g., French fries; French is conditional) and 50 optional phrase operators initially existed. There were 4 subset uses of the disjunction operator [] (e.g., ([green baked] beans) is valid for ‘green beans’ or ‘baked beans’).

We updated the grammar throughout the study based on participant interviews and the items voice recorded through IVRS interaction. The Voxeo platform also provided detailed logs of each call, identifying the caller and the interaction sequence between the participant and VoiceXML prompts. The interaction sequence logs included timeouts, grammar recognition errors labeled No Match, prompts, and recognitions.

With a completed call, two lists of items and counter variables were submitted to a MySQL Database—a list for food items misinterpreted by the IVRS when identified as wrong by the participant and a list of identified food items. When a participant recorded an item for addition to their grammar, the WAV file was submitted to our web server, written to disk, and a VoiceXML file returned to continue prompting for additional food items.

4.2. Participants

We used the same criteria for selecting participants as we described in case study one. Nine people volunteered for the study, but three dropped out before completion. Two people dropped out after the second day due to
lack of interest and one person was forced to drop out at the end of the first week because she had to undergo emergency surgery and remained in the hospital during the second week of the study. This high dropout rate is consistent with our previous studies and is a result of working with this type of chronically ill population. Here, we report on the six participants who completed the study ($n=6$).

The participants’ average age was 55 years, with a standard deviation of 10.9 years. The youngest participant was 36 and the oldest was 65. Four of the participants were female. Five participants identified themselves as Black or African American, and one as White. One participant had a ninth grade education, two had completed high school and three had some community college.

One participant had undergone dialysis for 23 years. The remaining participants ranged from 2-5 years of dialysis treatment. Two participants said they did not try to keep track of their nutrient or fluid consumption. Two participants did not keep track of nutrients, but attempted to limit their fluid intake by either not drinking liquids over the weekend or “staying conscious” of how much they are drinking. Two participants claimed to keep track of both nutrients and fluid. One used a journal and was conscious of portion sizes; the other could not describe their method of monitoring but said they carefully monitored sodium and potassium intake. We have found in previous studies that participants in this population often tell researchers what they think they want to hear in regards to their nutrient and fluid consumption, regardless of the reality.

Two participants were very familiar with computers. One took surveys on the Internet, while the other used his laptop daily, including bringing it to the dialysis sessions. One participant had some familiarity with computers. This participant had a computer at home, but did not use it very often. The final three participants said they were not familiar with computers, although one had three years of typing experience and said she could use a keyboard. Three participants owned mobile phones.

### 4.3. Design and Procedure

For most participants, the study lasted a total of two weeks. However
some participants had extra time with one of the applications because bad weather caused them to miss the dialysis session in which they were supposed to change technology. For these participants, we extended the total length of the study to ensure they had a minimum of one week with each technology.

We primarily used the same methods described in the first study. In this section, we describe additions we made to the methods. For the phone application, we taught participants how to turn the phone on and off, how to dial the number to record their meals and how to record food items with the voice recognition application, making sure to speak one food item at a time very clearly.

During each session, the researcher asked participants about any problems they were having with the application, if there were any food items they did not record, why they did not record a food item, when and how they used the application and their general opinions about its usefulness. In addition, we asked participants to list the foods they had eaten in the last 24 hours so that we could compare their recall with what they recorded with the applications.

Similar to the first study, competency tests were given to participants during all but the final day of the study. For the mobile phone, participants were asked to record their last meal, which required them to turn the mobile phone on, dial the number, and follow the prompts to record the meal. We recorded the number of times participants attempted to complete each task and noted any difficulties they were having. If necessary, we retrained and retested the participant.

Participants were paid ten dollars (US) at the end of each week of the study, for a total of twenty dollars. Payment did not depend on the number of times they recorded food items.
4.4. Findings

The key findings of our second case study were:

- Participants spent less time recording input with the IVRS
- Participants performed better with the scanner application on non-dialysis days and better with the IVRS on dialysis days
- Participants can record more items consumed with the IVRS, but the scanner application is more usable for a larger audience
- Input mechanism preference is not always linked with the participants’ performance with the technology

4.4.1. Barcode Scanning and IVRS Frequency of Use

Despite participants using each technology for at least seven days, we found that in reality participants used the PDA to scan items on average only five days (s.d. = 1.4 days) and the mobile phone to input items with the IVRS on average of 4.5 days (s.d. = 2.95 days). We found that participants who used the technologies on most of the study days did so because they enjoyed using the application systems and wanted to tinker with the technology to identify breaking points. In addition, participants mentioned a desire to help advance medical research to help themselves and their peers. Participants also mentioned the compensation rewards, although the compensation was not dependent on frequency of use. Participants who did not use the technologies regularly in the study sometimes forgot the PDA in their homes and expressed a reluctance to integrate technologies into their daily routines. We found no correlations between personal computer and mobile phone usage outside of the study and their willingness to incorporate the technology into their lives.
We examined usage patterns more closely by looking at participant input sessions. We defined an input session for the PDA scanner application as events that occurred within 10 minutes of each other because we found participants took longer to scan items in realistic situations (e.g., cooking meals). We defined an input session for the IVRS as any time a participant called into the system.

When we analyzed usage of each technology on a per input session basis, we found participants overall had more input sessions with the IVRS than with the PDA (13.67 input sessions versus 11.67 input sessions), but they had similar amount of input sessions when averaged over the week (1.95 input sessions versus 1.67 input sessions). In Table 2, we show the total and average number of sessions each participant had with each device, and the total and average time spent in each session. Participants 1-3 had the

<table>
<thead>
<tr>
<th>PDA</th>
<th>#sessions (avg.)</th>
<th>length (avg.)</th>
<th>#sessions (avg.)</th>
<th>length (avg.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18 (2.57)</td>
<td>72:23 (4:01)</td>
<td>10 (1.43)</td>
<td>24:10 (2:25)</td>
</tr>
<tr>
<td>2</td>
<td>16 (2.29)</td>
<td>29:07 (1:49)</td>
<td>25 (3.57)</td>
<td>28:19 (1:08)</td>
</tr>
<tr>
<td>3</td>
<td>4 (0.57)</td>
<td>5:27 (1:22)</td>
<td>4 (0.57)</td>
<td>0:04 (0:01)</td>
</tr>
<tr>
<td>CP</td>
<td>4</td>
<td>19 (2.71)</td>
<td>48:48 (2:34)</td>
<td>22 (3.14)</td>
</tr>
<tr>
<td>5</td>
<td>6 (0.86)</td>
<td>9:17 (1:33)</td>
<td>13 (1.86)</td>
<td>17:41 (1:28)</td>
</tr>
<tr>
<td>6</td>
<td>7 (1.00)</td>
<td>16:14 (2:19)</td>
<td>8 (1.14)</td>
<td>0:52 (0:07)</td>
</tr>
</tbody>
</table>

Table 2: Number and length of time (minutes:seconds) of sessions for each device. Averages are calculated per week.²

We examined usage patterns more closely by looking at participant input sessions. We defined an input session for the PDA scanner application as events that occurred within 10 minutes of each other because we found participants took longer to scan items in realistic situations (e.g., cooking meals). We defined an input session for the IVRS as any time a participant called into the system.

When we analyzed usage of each technology on a per input session basis, we found participants overall had more input sessions with the IVRS than with the PDA (13.67 input sessions versus 11.67 input sessions), but they had similar amount of input sessions when averaged over the week (1.95 input sessions versus 1.67 input sessions). In Table 2, we show the total and average number of sessions each participant had with each device, and the total and average time spent in each session. Participants 1-3 had the

² On day three of the study, participant 5 showed the mobile phone application to a friend, and had a 39 minute phone where she recorded 31 food items. We removed this outlier from the data analysis.
PDA the first week of the study, while participants 4-6 had the mobile phone.

Looking at the time participants spent on input gives us insight into how realistic it would be to use these systems in their everyday lives. If technology is going to take too much time, then individuals will not be willing to use it. We see in Table 2 that participants spent less time on input sessions when using the IVRS in comparison to the PDA scanning application. Scanning took more time because (1) occasionally the scanner popped out of the SDIO card holder and had to be replaced multiple times and (2) participants were multitasking during scanning sessions and input food items as they were doing an activity (e.g., cooking a meal) instead of input all at once later on (e.g., right after eating). Participants’ who multi-tasked with the PDA application showed that they are willing to integrate the technology into their lives. However, it also shows that raw input times may not be the best measure of efficient usage of the PDA application.

4.4.2. Performance

Besides the actual usage of the technologies in this study, we wanted to study the participant performance with each input mechanism. For this study, we defined performance as the ratio of unsuccessful to successful attempts at recording food items. We observed that performance was not consistent on all days. The ratio of unsuccessful to successful barcode scans on dialysis days was two times higher than on non-dialysis days (2.43 to 1.11). Conversely, we found participants performed better with voice recording on dialysis days – they had better performance on three out of the four non-dialysis days. Thus, on non-dialysis days participants performed better with the scanner application and on dialysis days, participants performed better using the IVRS.

We also studied how participants interacted with the IVRS. Unlike the first study, participants would have to say items one at a time and use command operators to record food items. We found on average that 53% of the time participants did not use command operators correctly during IVRS sessions. Participants did not say, “Wrong,” when items were not
recognized by the IVRS for 27% of the total calls. Participants did not say, “Done,” when they finished their calls 26% of the total calls. These errors effect how the IVRS interprets the input and thus could affect giving participants feedback on their food consumption in future implementations.

4.4.3. Electronic Input versus Self Reported Food Items

We asked participants to recall all of the food they ate in the last 24 hours each time we met with them. We then compared their 24 hour recall to the foods they electronically input into either the scanning program or IVRS with Venn diagrams shown in Figures 5 and 6. The relative ratios between these three numbers provide us insight into how participants used the electronic application.

The Venn diagrams for voice and scanning show that participants did not record everything they ate. Indeed, participants were somewhat limited with their ability to electronically record because the scanning application
required all recorded items to have barcodes and the IVRS required the items be in the database to be recognized. We found that sometimes participants electronically recorded items they did not eat. One participant in particular recorded non-food items. Overall, it appears that participants can capture more items they consume with the IVRS. However, more participants with varying abilities can capture items they consume with the scanning applications as shown by only one participant not using the scanning application as opposed to two participants not using the IVRS successfully. We also see that providing alternative input mechanisms, scanning or IVRS, did not motivate participants 4, 5, and 6 to input a majority of the food items they consumed during the study.

### 4.4.4. Electronic Input Preference

At the end of the study, we asked participants which device they preferred. Overall, two participants preferred scanning and four preferred voice. Once we identified their preferred device, we looked at their performance with

![Venn diagram of food items in 24 hour recall and items reported to IVRS.](image_url)

*Figure 7: Venn diagram of food items in 24 hour recall and items reported to IVRS.*
each input mechanism as described in previous sections and pictorially compared preference with performance as shown in Figure 8. We found that performance influenced preference in only 3 participants. Participants 4 and 6 chose the IVRS input despite not being able to successfully use the system. They told us that they still preferred the mobile phone despite moderate success with the scanning application because they were comfortable with using phones and with practice, could improve using the IVRS.

<table>
<thead>
<tr>
<th>ID/Preferred Device</th>
<th>Scanning</th>
<th>Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image" alt="Scanning" /></td>
<td><img src="image" alt="Voice" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image" alt="Scanning" /></td>
<td><img src="image" alt="Voice" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Scanning" /></td>
<td><img src="image" alt="Voice" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image" alt="Scanning" /></td>
<td><img src="image" alt="Voice" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image" alt="Scanning" /></td>
<td><img src="image" alt="Voice" /></td>
</tr>
<tr>
<td>6</td>
<td><img src="image" alt="Scanning" /></td>
<td><img src="image" alt="Voice" /></td>
</tr>
</tbody>
</table>

*Figure 8: Participant preference of electronic input mechanism and overall performance with each input mechanism.*
5. Discussion

Even though barcode scanning is a quick method for inputting individual food items, our results show that it may not be usable over an extended period of time when participants do not receive immediate feedback about their nutritional intake. Participants were overwhelmed with the amount of work associated with scanning every food item they consumed. However, participants did think that this application would be helpful for CKD patients who have recently been diagnosed with the chronic illness to assist them in learning about the restrictive diet. Participants thought CKD patients in their first year of dialysis treatment would be more likely to spend extra time scanning barcodes if it meant clinicians could give them better feedback about their diet and health. Another possibility for an electronic self monitoring application would be to have people use it periodically (e.g., quarterly when dietitians are conducting nutritional assessments with patients) to raise awareness and help them stabilize their diet.

We did not anticipate the amount of training participants needed to create descriptive voice recordings. In retrospect, it made sense that people with low literacy skills would not be able to gather enough data from the food item to identify it. Transcribing the data was time consuming, but was easier as the study continued because the participants typically consumed the same food items. Researchers need a better understanding of their user group so they can accurately identify food items that may be culturally or economically influenced. Since our user group has a restrictive diet, not being able to identify food items is unacceptable since it can have such a drastic change in participants’ overall health.

Participants’ underestimation of what they thought they would consume in comparison with what they actually consumed has been documented by other nutrition researchers (Dwyer et al., 2003; Resnicow et al., 2000). However, electronic self monitoring gives more detailed information (e.g., date, time, food item) than 24 hour recalls and food frequency questionnaires as had been used in the previous studies. Indeed, the
standard deviation for days participants ate more than they estimated is large for our small sample. This is significant because of the participants’ restrictive diet – overconsumption of the restricted nutrients is dangerous and can result in death.

Backfilling and hoarding are subject to retrospective biases and may not completely be accurate. In addition, researchers have shown that memory recall is undependable – thus participants may not be able to accurately describe what they had consumed during the past days even if they are attempting to be accurate (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2003). The end of study compliance we discussed is similar to Rand’s parking lot compliance where participants attempt to be compliant by complying with the study procedure in the parking lot of the research facility (Brandt et al., 2002). Since it is difficult to scan food items once they are consumed (or disposed of), participants increased participation before the end of each study phase with voice recording or wrong records. It is difficult to determine if patients were increasing participation before dialysis sessions where they met with researchers because participants may have been having a bad day (e.g., not feeling well due to dialysis session recovery).

We occasionally had difficulties with participants forgetting the devices, especially the PDAs, at home when we met with them. Since the participants were not use to having these devices in their lives, it is not surprising that they forgot them sometimes. In a recent study with “enthusiastic” PDA owners, three out of the eight participants forgot their PDA during a scheduled observation time (Bernheim et al., 2005). In addition, we had a number of participants who had to stop the study early because of medical concerns or a lack of motivation to complete the study. Losing participants is not localized to chronically ill populations - in the enthusiastic PDA owner study, out of 20 total participants, only 5 participants finished all the tasks in the study (Bernheim et al., 2005)!

In case study two, the IVRS we used had difficulties recognizing inputs when the vocabulary became too large. Participants became increasingly frustrated when the system could not recognize even the simplest word,
such as “egg” during the second week of the study. Participants voiced their frustrations by attributing human traits to the hardware (e.g., “It [IVRS] was dragging last night.”). IVRSs that use large vocabularies must be robust and able to handle slight variations between words.

Despite updating the vocabulary each night and thereby increasing it by 30% by the end of the study, participants continued to voice record new food items not in the vocabulary showing that although participants typically eat the same foods, there is some variation that must be considered when designing a nutrition monitoring system. Another difficulty we encountered with the IVRS was that participants did not use the command operators correctly. It would be difficult to create an IVRS without some command operators to provide the system information about correctness and when to store the information. We attempted to use a minimal amount of command operators, but participants did not use them half of the time. We would encourage researchers to conduct more thorough training sessions with a speaker phone so they can hear participants’ utterings and the system response.

One weakness that all monitoring methods have is that we are not sure if participants are truthfully recording what they consume. Without subjecting participants to costly blood work or requiring participants to wear an invasive device that could detect what a person is eating, we can only assume participants are being truthful. As we discuss above, electronic self monitoring can help researchers identify noncompliant, untruthful trends more quickly and discuss non-compliance with participants, but this is not a fault proof method.

We recognize that the case studies presented in this chapter are relatively slow. Although researchers have shown that conducting usability studies with 4-6 participants will sufficiently provide enough data to determine the effectiveness and usability of a system (Nielsen, 2002; Virzi, 1992), we are currently conducting a larger scale study with a fully functioning version of the system. We recommend researchers who work with chronically ill populations conduct smaller studies to better understand their target user group better before conducting a larger study.
6. Future Work

The research discussed in this chapter provides many avenues for future research projects. Interface designers must find a way to visually display portion sizes that low literacy populations understand. Visualizing portion sizes is fairly complicated because the type of visualization must be customized based on the type of food. For example, water would have a different portion visualization than bread or meat. In addition, the portion size visualization has to be informed by current methods dietitians use to educate CKD patients about portion sizes.

We must find a way to verify consumption to ensure self monitoring assistive applications can provide participants with accurate measures of their dietary intake. Consumption verification could provide participants with reminders to record what they consume instead of estimate time reminders not based on actual context. Indeed, there is already work being done in this area (Amft & Troster, 2006), but we need to continue development to make less obtrusive or invasive devices for everyday use.

In the second case study, one team member spent a significant amount of time updating the IVRS vocabulary each night. We could decrease the update time and distribute the work load if we better utilize all research members’ time to help listen to and decipher unrecognized phrases throughout the day. This idea builds on the human solver attack for Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). Websites, such as web mail and blogs, make users who want an account or post a comment identify the wavy characters in a picture – this challenge-response is known as a CAPTCHA. In the CAPTCHA human solver attach, a computer script would automatically fill out an online form, identify a CAPTCHA, and then pass the CAPTCHA to a high traffic website and promise the Internet surfer something in return (e.g., free porn) for identifying the characters in the CAPTCHA. The computer script would then take this response, enter it in the form, and create a malicious third party account (Doctorow, 2004). In an IVRS update setting, team members would be prompted throughout the
day to identify phrases. Depending on ethics board approval, this method could be distributed among a broader Internet community for faster turnaround time.

7. Conclusion

In this chapter, we highlighted results from two case studies that compared and contrasted the use of barcode scanning, IVRS, and patient self reporting as a means to monitor the nutritional intake relative to their dietary prescription of CKD patients. When we found that participants preferred unstructured voice recordings rather than barcode scanning in the first case study, we decided to study structured voice recording in a follow-up study. We found in the second case study that although the system provided participants with a faster method to input food items, participants had difficulty using the system despite training. We are continuing to study if patients will increase their usage of nutrition monitoring systems if they receive immediate feedback.

Acknowledgments. Katie A. Siek was supported in part by the National Physical Science Consortium and Sandia National Laboratories/CA during case study 1. This work was supported by NSF grant EIA-0202048, a grant from the Lilly Endowment, and Grant R21 EB007083 from the National Institute of Biomedical Imaging and Bioengineering to J.L. Welch.

List of Main Acronyms
PDA – Personal Digital Assistant
CKD – Chronic Kidney Disease
IVRS – Integrated Voice Recognition System
CAPTCHA - Completely Automated Public Turing test to tell Computers and Humans Apart
About the Authors

BEENISH CHAUDRY is a doctorate student in the Computer Science Department at Indiana University. She is working on developing a Dietary Input Monitoring Application on PDA for low literacy dialysis population under the guidance of Prof. Kay Connelly.

KAY H. CONNELLY is an Assistant Professor in the Computer Science Department at Indiana University where she leads the Security for Ubiquitous Resources Group (SURG). She is the Associate Director of IU’s Center for Applied Cybersecurity Research and a member of the Pervasive Technology Labs. Dr. Connelly's research focuses on user acceptance of ubiquitous and mobile computing technologies where there is a delicate balance between such factors as convenience, control and privacy.

DESIREE LAMBERT is a Registered Dietitian who has worked with dialysis patients at the Indiana University Medical Center in Indianapolis. She is currently President of the Indiana Council on Renal Nutrition, which organizes quarterly educational meetings for Renal Dietitians. She is also an active member of the National Kidney Foundation of Indiana.

KATIE A. SIEK is an Assistant Professor in Computer Science at the University of Colorado at Boulder. She is interested in integrating pervasive technologies in health and wellness environments to study how technology affects interventions. Her research interests include human computer interaction, ubiquitous computing, health informatics, and performance support systems.

JANET L. WELCH is an Associate Professor and Chair of Adult Health at the Indiana University School of Nursing. Her program of research focuses on self-management of diet and fluid intake among adults receiving hemodialysis.
References


Monitoring for Food and Fluid Intake. *Nephrology Nursing

What We Eat in America-NHANES. *J. Nutr., 133*(2), 609S-623.

Fitzmaurice, G., Khan, A., Buxton, W., Kurtenback, G., & Balakrishnan,

Glasgow, R., Bull, S., Piette, J., & Steiner, J. (2004). Interactive behavior
change technology: A partial solution to the competing demands of
primary care. *American Journal of Preventive Medicine, 27*(2,
Supplement 1), 80-87.

of three daily coital diary designs and a phone-in regimen.
*Contraception, 63*(3), 159-166.

Health Resources and Services Administration. Health literacy. Available
at www.hrsa.gov/healthliteracy/

and Alcohol Consumption in Heavily Drinking Men: 2 Years of
Daily Data Using Interactive Voice Response. *Alcoholism Clinical
and Experimental Research, 30*(5), 802-811.

technology to encourage incremental.

America: A first look at the results of the National Adult Literacy
Survey. Washington DC: National Center for Education Statistics,
United States Department of Education.

Lavigne, M. (1998). Interactive Voice Response in Disease Management:
Providing Patient Outreach and Improving Outcomes. 1-16.


of Elderly, Younger, and Chronically Ill Novice PDA Users (No.
TR595).

myFoodPhone. myFoodPhone Nutrition, Inc. http://www.myfoodphone.com/

Nielsen, J. (2002). Why you only need to test with 5 users. from http://www.useit.com/alertbox/20000319.html


http://www.ars.usda.gov/Services/docs.htm?docid=5720
