Summative evaluation of a sensor-based cognitive assistive technology: Impact on quality of life and perceived utility

Annica Kristoffersson PhD[a,c,*]

Ella Kolkowska PhD[b]

Amy Loutfi PhD[c]

[a]School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden; [b]School of Business, Örebro University, Örebro, Sweden; [c]School of Science and Technology, Örebro University, Örebro, Sweden; *Corresponding author: annica.kristoffersson@mdh.se

A. Kristoffersson, E. Kolkowska, A. Loutfi. Summative evaluation of a sensor-based cognitive assistive technology: Impact on quality of life and perceived utility. Gerontechonology 2019;18(2):59-69; https://doi.org/10.4017/gt.2019.18.2.001.00 **Background** Current research shows that perceived improvement of quality of life (QoL) is among the most important factors influencing acceptance of technology by older adults. **Research Aim/question(s)** The purpose of this study was to assess the utility of a sensor-based cognitive assistive technology (CAT) deployed in real homes and its impact on older adults' and their informal caregivers' perceived QoL. **Methods** During a period of five months, we conducted a summative evaluation by deploying a CAT in the home of couples in which one of the partners had a self-perceived memory decline. We applied a mixed method approach. Quantitative and qualitative data was collected via subjective forms, observations and 28 interviews conducted prior to the deployment, when the CAT was deployed and one month after the CAT removal. **Results** The study showed that the CAT had a psychosocial impact on older adults with a self-perceived memory decline and their informal caregivers but no impact on subjective forms measuring physiological and mental health, anxiety, depression or QoL. Additionally, we found that the CAT lacked a number of functionalities and that the test persons experienced several problems of different nature when using it in their homes. This affected the users' experienced impact of the CAT. **Conclusions** This study contributes to the literature on sensor networks' impact on QoL. While previous studies often use general forms intended to assess QoL, this summative evaluation indicates that using such instruments without considering contextual factors is not meaningful. The study also outlines a number of factors that sensor network providers might consider in order to increase their products impact on QoL and their perceived utility.

Keywords: summative evaluation, sensor networks, cognitive decline, quality of life, utility

Introduction

Modern information and communication technologies and related assistive devices are often foreseen as a solution to the problem of an increasing number of older adults and associated costs. One example of such a technology is sensor networks that are assumed to be ready to install in home environments upon purchase. Typically, they are composed by a number of sensors that communicate with each other via radio waves or Wi-Fi. There are several application areas, including smart homes and cognitive assistive technologies (CATs) that can support and thereby prolong the older adults' ability to live independently. The CATs offer functionalities such as alarms, reminders to turn off electronic devices, and to close doors, water taps et cetera. The smart homes target a broader public and offer functionalities such as the ability to check the status of and control electric devices from remote. Common for most sensor network providers is their claim to not only impact the quality of life (QoL) of individuals using them but also increase their safety and security (trygghet in Swedish). In literature, QoL is often related to safety and security (Nordgren, 2013). According to Swedish National Board of Health and Welfare (2012), safety and security are often subjectively defined as a compound of physical, psychological and existential aspects of well-being and a necessity for coping with daily activities. It is difficult to outline what makes us feel safe and secure but easier to outline what makes us feel unsafe and insecure. These feelings may oc-

cur when we worry about our own situation or the situation of others, but also be associated with fear, anxiety and/or panic over the situation (Swedish National Board of Health and Welfare, 2012). For that reason, safety and security is often estimated by measuring factors such as physiological and mental health, anxiety and depression as well as QoL- aspects of well-being. Yet, a number of reviews (e.g., Brandt, Samuelsson, Töytäri, & Salminen, 2011; Liu, Stroulia, Nikolaidis, Miguel-Cruz & Rincon, 2016; and Peetoom, Lexis, Joore, Dirksen, & De Witte, 2015) shows that studies of technology with respect to non-technical aspects such as QoL are very scarce. Recent reviews (e.g., Liu et al., 2016; and Siegel & Dorner; 2017) indicate that studies on the outcomes of using QoL technologies, i.e., sensor networks, in real settings are too diverse, few and small. Therefore, more studies dedicated to investigating the process of deploying sensor networks in real homes and their impact on the users' perceived QoL are needed (Liu et al., 2016; Siegel & Dorner, 2017).

The purpose of the summative evaluation (Schulz, Beach, Tabolt Matthews, Courtney, & De Vito Dabbs, 2012) presented in this article was to assess how a sensor-based CAT deployed in real homes for a period of 4-5 months contribute to the older adults' and their relatives' QoL, what psychosocial impact the CAT has, and which factors affect the perceived utility.

Brief literature review and research questions

QoL is a multidimensional construct covering a number of objective and subjective measures including the "physical health and functioning, emotional health, cognitive functioning, role performance, work productivity, and life satisfaction", p. 2397 (Schulz et al., 2012). This definition is in line with WHO's (World Health Organization, 1997) broad understanding of QoL, which says that QoL is affected by many factors such as the context of the culture, value systems, and "physical health, psychological state, level of independence, social relationship and their relationship to salient features of their environment", p. 1. Therefore, a multidimensional approach considering these different factors needs to be taken (Siegel & Dorner, 2017) to understand a sensor-based CAT's impact on the users' QoL.

Schulz et al. (2012) developed a roadmap for designing, developing and evaluating QoL technologies which states that it is a must to administer QoL, and condition specific, health related measures to assess a technology's effect on end-users QoL. Based on the roadmap, evaluations can be formative, summative or both. While the formative evaluation focuses on collecting data to assess the acceptability or functionality of features in a system being developed, the summative evaluation focuses on collecting data to assess the usefulness of an already mature technology.

Recent literature reviews show that the level of technology readiness for smart homes and home health-monitoring technologies is low (Liu et al, 2016; Siegel & Dorner, 2017). Liu et al. (2016) found that the majority of studies are based on technologies in development and testing phases or technologies in demonstration or pilot phases. Siegel and Dorner (2017) argue that this fact can be a reason for the technical problems experienced during almost all reviewed studies. Despite this, negative effects of the systems are seldom reported. Both Siegel and Dorner (2017) and Molka-Danielsen and Moe (2013) discuss the possibility of bias caused by stakeholders et cetera. On the contrary, the study presented in this paper assesses the use of a fully operating and mature sensor-based CAT with focus on its impact on QoL and perceived utility. As stated earlier, the concepts of QoL and safety and security are difficult to assess. Thus, in line with previous literature (see Siegel & Dorner, 2017 who surveyed instrumental use) we assess these concepts using forms for physiological and mental health (SF-12), anxiety and depression (HADS) and general QoL (QOLS-S). Psychosocial impact was assessed using PIADS. Further information on the measures is provided in 'Procedures and measures'. In sum, this study answers the following research questions:

RQ1: What impact, if any, does a CAT have on physiological and mental health, anxiety, depression and general quality of life?
RQ2: What is the psychosocial impact, if any, of CAT use on older adults with a self-perceived memory decline and their informal caregivers?
RQ3: What factors affect the users' perceived utility of a CAT?

The informal caregivers could gain from technology aiming at helping older adults. Side effects of caregiving include; anxiety and depression (Lou et al., 2015), stress (Fonareva & Oken, 2014), a deterioration of social life (Medrano, López Rosario, Núñez Payano, & Reynoso Cappellán, 2014) and sleep quality (Liu et al., 2017; McHugh, Wherton, Prendergast, & Lawlor, 2012). Similarly to McCloskey, Jarret, Stewart, and Keeping-Burke (2015) and Molka-Danielsen and Moe (2013), we involved the stakeholder informal caregivers in the summative evaluation of the sensor-based CAT presented in this article.

Methods

Inspired by Schulz et al.'s (2012) roadmap and shortcomings in prior research, this article reports on the results of a summative evaluation of QoL. This section provides information about

Table 1. A summary of the CAT's nine functionalities (F1-F9) and their division into five categories

Functionality category	Functionality number	Functionality description
Calendar visualization	F1	Time and date display.
	F2	Calendar for structuring daily tasks.
Calendar reminders	F3	Reminders about upcoming events.
Communication	F4	Skype
Sensor-based reminders	F5	Reminder if stove on but no motion in kitchen for x minutes.
	F6	Reminder if an electronic device, e.g., coffee brewer, has been turned on for x minutes.
	F7	Reminder if the entrance door is opened while the stove is on.
	F8	Reminders specifically relating to the opening of the entrance door can be issued.
	F8a	A reminder to act, e.g., lock the door.
	F8t	A reminder issued only during a certain time frame, e.g., during the night which can be useful for people mixing up day and night time.
Sensor-based actuation	F9	A lamp is automatically lit when a sensor has detected motion beside the bed.

the CAT deployed, the participants, the mixed method approach and data analysis.

CAT description

The deployed CAT consists of a digital calendar and a number of environmental sensors. The calendar issues a reminder/alarm to the user when one of the pre-configured situations in which a sensor-based reminder should be triggered occurs. The reminder is issued in text along with an image and a pre-recorded voice instruction. A checkbox can be added to each activity entered into the calendar. When a calendar reminder is issued, the user is requested to confirm that action has been taken by ticking the box. *Table 1* shows the functionalities. Column one shows the five functionality categories offered by the CAT: calendar visualization, calendar reminders, communication, sensor-based reminders, and sensor-based actuation. Each of these categories includes one or more functionalities (F1-F9), presented in column two. These functionalities are described in the third column. It is recommended to position the calendar where the user spends a lot of time. The calendar is not accessible from phones, and sensor-based reminders are only issued via the digital calendar.

Participant characteristics

Summative evaluations typically start by testing a technology with a limited number of participants with similar background and extend it to cover larger, more diverse groups (Schulz et al., 2012). Following their recommendation, we focused on a limited number of test persons. This allowed us to get to know all participants.

The CAT was deployed for a duration of 4-5 months in seven households of men (Mean = 71.6 years old) who had a self-perceived memory decline of varying kind. Each of them lived with their wife (Mean = 65.7 years old). Hence, all informal caregivers involved in this study were women. Two of them were still working and balancing the work with caring for their husbands while the other five were retired. All participants provided their informed consent but test site 6 requested to have the CAT removed after a short period of use. The type of accommodation varied from three room apartments to real houses. Information on each test person, wife, and type of accommodation is provided in *Table 2*. Additional information on the test sites and CAT is provided in Kristoffersson, Kolkowska and Loutfi, 2014.

*Table 2. Description of participants. TPn = test person n, Wn = wife n, * denotes the two wives who were still working*

Test site	Accommodation	Internet	Age TPn	Age Wn
TS1	Five room apartment	Cable	77	68
TS2	Three room apartment	3G/4G	74	69
TS3	Three room apartment	Cable	82	82
TS4*	Two story detached house	Fiber	54	54
TS5	Two story summer residence	3G/4G	71	68
TS6	Four room row house	3G/4G	75	66
TS7*	Three room apartment	Fiber	68	53
Mean			71.6	65.7
SD			8.3	9.1

Procedures and measures

According to Schultz et al. (2012), there is no single, universally accepted instrument for measuring a technology's impact on the quality of life. The authors stress that at a general level, the technologies could be assessed using general QoL instruments such as WHO-QoL (The WHO-

Table 3. A summary of the data collection procedures and methods used at E1-E3

Time	Test person	Wife	Together
E1 (before CAT deployment)	Consent form Socio-demographics SF-12 HADS QOLS-S Second interview	Consent form Socio-demographics HADS QOLS-S Second interview	First interview
E2 (at CAT removal)	SF12 HADS QOLS-S PIADS	HADS QOLS-S PIADS	
E3 (one month after CAT removal)	SF-12 HADS QOLS-S	HADS QOLS-S	Third interview

QOL Group, 1994) or health-related QoL; SF-36 (McHorney, Ware, Lu, & Sherbourne, 1994) or the EQ-5D (EuroQol Group, 1990) while other instruments should be used to assess the specific domains targeted by a particular technology. Siegel and Dorner (2017) argue further that to be able to understand how technologies affect the different dimensions of the users' QoL, researchers cannot rely solely on outcomes from instruments. Other methods need to be used for studying how the technology affects the older adults and their caregivers in their context. As suggested by Schulz et al. (2012) and Siegel and Dorner (2017), this study applied a mixed method approach to assess the CAT's impact on the users' QoL, its psychosocial impact and perceived utility. Qualitative and quantitative data was collected through observations, semi-structured interviews and subjective forms. In line with earlier studies (see the survey by Siegel & Dorner 2017), we administered subjective forms assessing general QoL, anxiety, depression, physical health and mental health (RQ1) and psychosocial impact (RQ2). We administered subjective forms and conducted semi-structured interviews with the participants at three different occasions: (E1) prior to the deployment of the CAT, (E2) in conjunction with its removal, and (E3) one month after the removal. This approach provided a rich set of qualitative and quantitative information regarding the participants' experiences of using the CAT (*Table 3*).

The qualitative information was collected during a total of 28 interviews. The first two (sets of) interviews, that were conducted at E1 served to: (1) Gather information on how to deploy the CAT, and (2) identify expectations and needs based on the first step of the IPPA (Individualised Prioritised Problems Assessents) method (Wessels et al., 2002). The third semi-structured interview, that was conducted at E3, considered the way in which the study was conducted, the subjective forms used, how well the CAT met the participants' expectations, the CAT's utility and their QoL. During the test period, we regularly visited the participants to check how the CAT was used (i.e., which and how many events were planned in the calendar) and to collect the participants' reflections regarding the usage of the CAT. All in all, approximately 15 observations were made, and 15 unstructured interviews with the test persons and their wives were conducted during the test period. The observations made and comments provided during our visits to the participants were written down. After the system removal, we checked the calendar events for each of the test sites to see how the CAT was used during the test period.

Considering the rich set of data collected, each quantitative form assessing the CAT's impact on physiological and mental health, anxiety, depression and general QoL (i.e., RQ1) were chosen with consideration to the test persons' self-perceived memory decline. We chose forms that placed a minimum burden on the participants, i.e., short forms that were available in the participants' mother tongue, Swedish. In addition, the selected forms supported the analysis

*Table 4. Test persons physiological (PCS12) and mental (MCS12) health according to SF-12 at baseline E1, end of study E2 and follow-up E3. * two items unanswered and one answered in non-numeric characters. ** missed answering one question*

Test person	PCS12			MCS12		
	E1	E2	E3	E1	E2	E3
TP1	*	41.89	44.77	*	56.50	45.60
TP2	55.74	39.97	29.35	35.35	52.67	56.92
TP3	40.64	32.42	40.88	31.39	33.64	27.28
TP4	32.36	**	31.66	30.35	**	38.91
TP5	49.33	54.19	56.61	60.78	53.08	53.72
TP7	53.24	49.27	42.75	56.73	58.10	40.28
Mean	46.26	43.55	41.00	42.92	50.80	43.79
SD	8.63	7.55	8.98	13.10	8.82	9.86

*Table 5. Test persons level of anxiety and depression according to HADS at baseline E1, end of study E2 and follow-up E3. * missed filling out the form*

Test person	Anxiety			Depression		
	E1	E2	E3	E1	E2	E3
TP1	5	5	6	5	9	7
TP2	3	4	4	3	4	4
TP3	11	14	11	8	11	10
TP4	6	*	1	10	*	7
TP5	3	6	9	2	4	5
TP7	0	0	0	0	0	4
Mean	4.67	5.80	5.17	4.67	5.60	6.17
SD	3.40	4.58	3.98	3.45	3.93	2.11

of data over time. SF-12, a short-form health survey (Ware, Kosinski, & Keller, 1996) was used to assess the impact of the test persons' physiological and mental health. Other forms used were HADS, Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983), and QOLS-S, The quality of Life Scale (Burckhardt & Anderson, 2003). In addition, we used PIADS (Day & Jutai, 1996) which is a form developed for assessing the psychosocial impact of an assistive device to study RQ2. It has been shown that PIADS can be used to predict abandonment and retention of an assistive device (Day, Jutai, & Campbell, 2002). The majority of forms were administered to participants at E1, E2 and E3. *Table 3* provides a summary of the evaluation procedure and the qualitative and quantitative data collection methods used in this study.

Data analysis

While quantitative data collected from a small sample of participants cannot be generalized to cover larger groups of people, our goal was to investigate whether trends could be revealed from the data. Hence, the quantitative data collected using SF-12, HADS, QOLS-S, and PIADS was analyzed according to the instructions for each form. The answers provided to SF-12 were summarized into two sub-scores, PCS12 (physiological health) and MCS12 (mental health). Blank items resulted in no score for PCS12 and MCS12. The answers provided to HADS were summarized into two scores, anxiety and depression. A blank item on a HADS form resulted in no score for either anxiety or depression. The QOLS-S score is the sum of all answers provided, i.e., a "total score" could be calculated even if there were blank items.

The answers provided to PIADS were summarized into three sub-scores (competence, adaptability and self-esteem) and a total score. Blank items were a problem also for PIADS. In such occasions, an indicative score was calculated by replacing blank items with 0 on a -3 to +3 scale. A table visualizing the psychosocial impact for each individual was built.

We subjectively analyzed the qualitative data collected during the 28 semi-structured interviews to derive information on the soundness of using the quantitative forms for assessing RQ1 and RQ2, and (RQ3) what factors influence CAT users' perceived utility.

The data collected during our visits to the participants (comments and observations) and analysis of the calendar events at the end of the test period allowed us to verify the verbal and written responses from the participants, since there may be large differences between actual behaviour and reported behaviour.

Results

A CAT's impact on physiological and mental health, anxiety, depression, and general QoL (RQ1)

The analysis of how the SF-12 (PCS12 and MCS12) and HADS (Anxiety and Depression) changed over time resulted in the same conclusion. Unequal variance t-tests between 1) E1 and E2, 2) E1 and E3, and 3) E2 and E3, yielded no significant difference for any of the four measures. An analysis on QOLS-S measures over time was not conducted since only two of the test persons and three wives answered all questions. While there were no significant differences in the scores, we wish to highlight that the CAT had no positive impact on physiological and mental health, anxiety, depression or general QoL for the majority of test per-

Table 6. The wives level of anxiety and depression according to HADS at baseline E1, end of study E2 and follow-up E3

Wife	Anxiety			Depression		
	E1	E2	E3	E1	E2	E3
W1	1	1	1	2	1	2
W2	3	9	5	4	9	5
W3	8	8	3	4	4	4
W4	7	13	8	9	14	13
W5	14	13	14	10	12	9
W7	7	7	14	4	2	8
Mean	6.67	8.50	7.50	5.50	7.00	6.83
SD	4.11	4.07	5.06	2.93	4.97	3.62

*Table 7. Test persons and wives perceived quality of life according to QOLS-S at baseline E1, end of study E2 and follow-up E3. N denotes the number of items answered, only n number of questions are included in the score. * TP3 did not fill out the form at E3*

Test Person	n	E1	E2	E3	Wife	n	E1	E2	E3
TP1	12	79	69	64	W1	15	82	77	76
TP2	16	95	85	85	W2	16	92	78	95
TP3	12	67	57	*	W3	15	80	80	86
TP4	14	52	45	32	W4	16	74	59	78
TP5	16	93	78	89	W5	16	78	84	85
TP7	13	77	73	71	W7	15	73	76	59

sons. As shown from *Tables 4* to *7*, the scores rather worsened over time.

The soundness of using the three quantitative forms SF-12, HADS, and QOLS-S in this context was discussed during the third semi-structured interview (because of the space limitation, only a few examples are provided). Two test persons had problems with the first item in the SF-12 form (*"In general, would you say your health is:"*). Rather than rating their health on a scale ranging from excellent to poor, they wanted to emphasize that their physiological health was good.

"This general health, what is meant with it? There is mental health and there is physiological health. You can be in a wheelchair but be awarded the Nobel Prize. [..] I think mental and physiological health are two different things." (TestPerson5)

Regarding QOLS-S, only two of the six test persons participating throughout the study answered all questions at all three occasions. The number of questions answered varied from 12-16. Among the wives, three answered all questions and three answered all questions but one. If this was a larger study with a corresponding spread in number of questions answered, the response rate to this questionnaire would have been low (33% for the test persons). Two test persons explained that some items were not directed towards older adults (in particular item 3, 4, 7, 8, and 11, see Burckhardt & Anderson, 2003). For example, item 4 assessed their current satisfaction level with having and raising children.

Comments from the informal caregivers

An important comment regarding QOLS-S that we consider worth reflecting upon when designing summative evaluations of technologies with older adults in which subjective forms are being used was:

"I think that I sometimes felt that it was a bit difficult to answer the questions. [..] It had nothing to do with the apparatus. [..] Sometimes, I felt that the answer is not wrong but it has more to do with his (TestPerson4's) illness and other factors. [..] For example, NN has not felt very good lately. It affects me." (Wife4)

Wife7 was concerned about what the response from TestPerson7 would tell us as she often disagreed with his response. Additionally, while HADS is designed to assess the current anxiety and depression level, Wife7 was confused when filling out the HADS form.

"Sometimes, one had to remind oneself that they are talking about before and after. One is afraid to lose focus on that." (Wife7)

To summarize results regarding RQ1, we can say that it is difficult to measure the CAT's impact on physiological and mental health, anxiety, depression and general QoL, using the commonly

*Table 8. Test persons (left) and wives (right) perceived psychosocial impact of using the CAT according to PIADS at system removal, E2. * TP1 answered with two alternatives on three questions. W7 answered with two alternatives on two questions. For both cases, the alternative closest to 0 was used in the calculations. ** TP3 did not answer four questions. Blank answers were encoded as 0*

Test person	Competence	Adaptability	Self-Esteem	Total	Wife	Competence	Adaptability	Self-Esteem	Total
TP1*	0.08	0	-0.13	0.04	W1	0	0	0	0
TP2	1.25	0.50	0.75	0.96	W2	0.58	1.17	0.38	0.65
TP3**	0.42	0.67	0.13	0.38	W3	0.08	0.67	0.13	0.23
TP4	1.33	0.67	0.88	1.04	W4	0.17	0.00	0.13	0.12
TP5	1.50	1.50	1.63	1.54	W5	0.75	0.50	0.13	0.50
TP7	2.75	3.00	2.13	2.62	W7*	2.75	3.00	2.63	2.77
Mean	1.06	0.91	0.79	0.95	μ	0.62	0.76	0.49	0.61
SD	0.89	0.98	0.78	0.85	σ	0.91	1.00	0.88	0.91

recommended subjective forms. First, the questions are not always adjusted to the targeted group (older adults). Second, some questions are confusing. Third, it is difficult to answer the questions solely focusing on the CAT because other factors than the deployed technology such as life situation, or sickness of a child may influence physiological and mental health, anxiety, depression and general QoL.

A CAT's psychosocial impact (RQ2)

As shown in *Table 8*, the psychosocial impact varied from none or minimal for a few participants, e.g., TestPerson1 and Wife1, to an obvious effect for others. The table indicates that the CAT had a larger psychosocial impact on the test persons than on the wives. However, unequal variance t-tests show that the difference is not significant for any of the sub-scores or the total score.

The scores were presented to the participants during the third interview. Interestingly, a consistent comment was that the results were correct, i.e., the scores were low among the participants who felt that the CAT affected them to a little extent. We could also confirm the results based on the observations during our visits to the participant and based on the analysis of the calendar events at the end of the test period. Looking closely at the scores for TestSite5 and TestSite7 who desired to keep the calendar after the end of the summative evaluation, the PIADS scores, which maximums are 7, also indicate a clear psychosocial impact for TestPerson5, TestPerson7 and Wife7. The results support Day et al. (2002).

Factors influencing the perceived utility of a CAT (RQ3)

The meaning of QoL interconnects with the CAT's functionalities. A number of comments from the wives indicate that two important aspects of QoL are their test persons' possibility to check the calendar themselves and to get reminders about taking their medicine. The calendar and its checkbox-functionality worked as a means for ascertaining some of the wives that the test persons had taken their medicine and/or eaten.

"It was an assistive device because when you (TestPerson1) got the alert about taking the medicine, you immediately took it and ticked the checkbox in the calendar." (Wife1)

"The daily life has worked. I mean, I know what is in the refrigerator and previously it was common that he (TestPerson7) *forgot to have lunch. Then, when I came home and asked him to take out the garbage, he could get dizzy because of having forgotten to eat and/or take his medicine."* (Wife7)

In fact, the two wives who were still working found the calendar and calendar reminders very useful as they made their test persons more independent.

"Now, he (TestPerson7) *could checkout the calendar to see what was going to happen when I was away."* (Wife7)

"It was good that it issued reminders when I was not at home. [..] It increased the chance of activities being conducted." (Wife4)

Several comments indicate that the CAT lacked functionalities that could have increased the QoL. The wives who worked or occasionally spent one or two days away desired the ability to access the CAT from their phones.

"When I went away for a couple of days, I would have liked being able to access the calendar so that I could see what activities had been checked. [..] Also, it would have been good if I could have checked if the door was closed and that you (TestPerson1) had gone to bed." (Wife1)

"One could register when he (TestPerson4) *went out and when he came home so that I could know if he has been away for too many hours. That I would like to have. [..] It could have been good if one could have seen that you had actually heated the food in the microwave oven, if I could have checked that you had eaten while I was working."* (Wife4)

Several participants informed that it would be useful if the CAT could provide information related to risk of fire, burglary, and flooding when about to exit the home. This includes information regarding the status of different electric devices such as the iron, stove or TV, but also the status of doors, windows and water taps.

Five main factors affecting the perceived utility of the CAT could be identified when analyzing the qualitative data collected during the third interview; unrealized expectations on the CAT, lack of understanding, misconfiguration, form factors and aesthetics.

The analysis of qualitative data collected during the first two interviews and the comparison with observations made while the CATs were deployed, has shown that there was quite a number of deviations between the expected deployments and the actual deployments. There were several reasons for these deviations. For instance, F5 and F8 needed a stove sensor which could not be installed for physical or technical reasons at three test sites. The sensor was too big to fit in the drawer underneath the cooktop (TestSite3 and TestSite4) and the stove was connected without electrical socket at TestSite5. While TestSite3 and TestSite4 initially did not worry about the fact that the stove sensor could not be deployed as part of the CAT because of being too big, both of them forgot the stove on while the CAT was deployed. The unrealized expectations decreased the perceived utility of the CAT.

The majority of participants informed that they did not fully understand the capabilities of the CAT at E1 or while it was deployed despite having received information at two different occasions (E1 and when the CAT was deployed).

"I think I did not realize all the things I could use the CAT for. I used it for the activities and it is a bit sad because I could have tried these other things. [..] I am mainly thinking about the door functionality and the fact that we let the reminder be "the door is open". We could have recorded a message reminding us about locking the door. That thing, and reminders about buying medicine." (Wife1)

"When there is an activity in the calendar, it clearly shows the time and duration of the activity but then it is crossed over, why not just remove it? [..] I think it can be confusing for those who need to remember taking the medicine. "Have I taken it or not?"" (TestPerson2 never discovered the checkbox-functionality.)

The participants suggested several ways in which the understanding of CATs can be improved. TestSite2-4 and TestSite7 would have preferred receiving information on how to use the CAT both at the time of deployment and shortly after. One suggestion was to start by focusing on how to add activities and then continue with information on how to log on to the CAT from work and extra functionalities during the next visit. In addition, several participants would have appreciated user groups where they could have shared their experiences and tips on how the CAT could be used.

The majority of participants chose not to inform the researchers about problems with the CAT, i.e., misconfigured sensors-based reminders. This choice affected the perceived utility of the CAT. An example of such a misconfiguration is that TestPerson7 pointed out that he had a problem remembering the keys on the outside of the entrance door while preparing the deployment. While a suitable sensor-based reminder would have been *"Bring the keys inside"*, the CAT issued the annoying reminder *"The door is open"*.

"It is a bit tedious to hear that the door is open every time I open it. Then, I say "I know" and shut it." (TestPerson7)

Regarding form factors, three test sites spending a lot of time in their summer houses, complained on the calendar's weight and size. While having aimed to bring the calendar with them when going to the summer houses, it was perceived as being too big and heavy for doing so. Therefore, they recommended to replace the current calendar with a tablet-sized calendar.

The calendar never entered a screen saver mode. While this problem was partly overcome during the study by automatically adjusting the screen's brightness during evenings/nights, several test sites had negative feelings about the screen's bright light. On the other hand, TestPerson3 informed us during the third interview that he missed the extra light coming from the screen as it supported him when going to the toilet during the night. This finding corroborates with the findings of van Hoof, Kort, Rutten and Duijnstee's (2011) whose participants also commented on aesthetics and disturbing light.

Finally, while the deployed CAT was relatively simple, most participants could point out several features that could be useful in a CAT when the study ended, something they were unable to do before gaining the experience of using a CAT. Additional factors that could have increased the perceived utility include: the possibility to print out notes (e.g., shopping lists and activities), medication lists, sensor-based reminders sent to the participants' phones and the ability to access the status of sensors from the phone, the ability to communicate with home care, and automatic system monitoring so that a warning is issued when there is a system failure.

Discussion and conclusions

This paper has presented a summative evaluation of a Sensor-Based Cognitive Assistive Technology (CAT). The CAT was deployed in the home of seven older men with a self-perceived memory decline during approximately five months. All of them had a similar socio-demographic situation in that they lived together with their informal caregivers (spouses). Data of quantitative and qualitative nature was collected at several occasions from both the men and their wives. The study results are summarized as follows:

First, we can conclude that using widely recommended quantitative subjective forms to assess a sensor-based CAT's impact on factors associated with QoL is questionable because: (1) the questions included in the forms are not always adjusted to the targeted group (older adults), (2) some questions are confusing, (3) contextual and environmental factors may influence the CAT's impact on physiological and mental health, anxiety, depression and general QoL.

Our results are based on a limited number of participants. However, they confirm findings in previous studies. Regarding QOLS-S, only a minority of the participants provided an answer to all items. Regarding SF-12, two out of six test persons did not answer the questionnaire correctly at E1 and E2 respectively. There were two different errors, blank answers or answers in text instead of numbers. Iglesias, Birks, and Torgerson (2001) administered the SF-12 questionnaire (stem and

leaf layout on 9 out of 12 questions) and found that 26.6% of the SF-12 questionnaires returned had one or more missing responses. The authors recommend an alternative layout (York SF-12, in which the response alternatives are provided for every question) that had a better response rate in their study, i.e., only 8.5% of the York SF-12 questionnaires returned had one or more missing responses. In another study, de Souza and Corrente (2013) found that using smileys instead of verbal answers may improve the response rate in studies with older adults.

In our study, we asked our test persons and wives to reason upon the use of the four different forms. This allowed us to understand more exactly what older adults experienced as difficult or unsuitable. Such information is not collected in studies that just administer questionnaires.

Surprisingly, we found that the obtained scores from using the forms often worsened rather than improved when using the CAT. Using the interviews, we could understand that it often depended on other events and situations in the participants' lives. This finding is in line with previous studies arguing that an intervention, such as the deployment of a CAT, is just one piece in the complex puzzle of daily living. Including both environmental contextual factors and QoL factors is recommended by Molka-Danielsen and Moe (2013). Similarly, van Bronswijk and Kearns (2009) discussed a number of other environmental factors that may affect the independence of older people. Hence, using a mixed method approach combining the results with qualitative information from interviews is recommended when studying the impact of a technology (Siegel & Dorner, 2017). Therefore, based on our result and the previous research (i.e., Molka-Danielsen & Moe, 2013), we argue that contextual factors cannot be ignored when assessing a CAT's utility since the contextual factors may have a larger impact on its users than the system itself. Consideration of notes from observations made and interviews with the participants were needed while analyzing the results of data collected using the subjective forms. Hence, a mixed method approach is recommended when assessing the impact of, e.g., a sensor network.

Second, discussing the psychosocial impact as measured using the PIADS form with the older test persons and their wives, a consistent comment was that the results were correct. I.e., the CAT had a larger psychosocial impact on the test persons than on the wives and the impact was clearly higher among those wanting to keep the calendar after the end of the summative evaluation.

Third, regarding RQ3, we identified a number of factors which affect the perceived utility of the CAT. Factors affecting the perceived utility include: unrealized expectation on the CAT, lack of understanding of the system's capabilities, misconfiguration, form factors and aesthetics. The practical implication of these findings is that sensor network providers may minimize the effect of these factors by offering appropriate education and support groups in which users can discuss how the CAT can be used as well as ensuring that the CAT is configured according to the user requests. It is also important to create realistic expectations regarding the technology being deployed to avoid disappointment among the users. Our findings are also interesting for the research community. For example, Siegel and Dorner (2017) found in their literature review that most studies in the field reported positive outcomes even though many studies are based on technologies in development or pilot phases (Liu et al., 2016). Our findings support Molka-Danielsen and Moe's (2013) and Siegel and Dorner's (2017) discussion on the possibility of bias in the presentation of results in some of the studies since we found that five of our seven test sites actually did not find the CAT usable in their situation and chose not to keep it. We also identified several factors that decrease the perceived utility of the deployed technology.

This study contributes to a better understanding of how sensor-based CATs impact QoL and utility. In general, this study confirms Zwijsen, Niemeijer, and Hertogh's (2011) findings about the importance of including external factors in the evaluation such as (1) previous experience of technology, (2) environmental factors (i.e., the home layout), and (3) other events occurring in life. We could see that these factors influenced the perceived QoL and the utility of the CAT. On a similar note, privacy issues in relation to using a CAT was studied in a parallel study (Kolkowska & Kristoffersson, 2016) which resulted in a number of privacy by design principles. In short, the results corroborate with Pol et al.'s (2016) finding that health care professionals' ability to access sensor data outweigh privacy concerns although the CAT tested in this study did not offer such a functionality.

Acknowledgements
The authors would like to express their gratitude to the test persons and wives involved in this study and to the developer of the sensor-based cognitive assistive technology deployed. This work was supported by the Länsförsäkringar Research Foundation.

References

Brandt, Å., Samuelsson, K., Töytäri, O., & Salminen, A. (2011). Activity and participation, quality of life and

user satisfaction outcomes of environmental control systems and smart home technology: a systematic review. Disability and Rehabilitation: Assistive Technology, 6(3), 189-206. https://doi.org/10.3109/17483107.2010.532286

Burckhardt, C. S., & Anderson, K. L. (2003). The Quality of Life Scale (QOLS): Reliability, Validity, and Utilization. Health and Quality of Life Outcomes, 2, 7 pages. https://doi.org/10.1186/1477-7525-1-60

Day, H., & Jutai, J. (1996). Measuring the psychosocial impact of assistive devices: The PIADS. Canadian Journal of Rehabilitation, 9, 159–168.

Day, H., Jutai, J., & Campbell, K.A. (2002). Development of a scale to measure the psychosocial impact of assistive devices: lessons learned ant the road ahead. Disability and rehabilitation, 24, 31-37.

de Souza, L.B., & Corrente, J.E. (2013). Evaluation of the Flanagan Quality of Life Scale for older adults. International journal of Science Commerce and Humanities, 1(4), 10-20.

EuroQol Group (1990). EuroQol–a new facility for the measurement of health-related quality of life. Health Policy, 16, 199–208.

Fonareva, I., & Oken, B. S. (2014). Physiological and functional consequences of caregiving for relatives with dementia. International psychogeriatrics/IPA, 26, 727-747. https://doi.org/10.1017/S1041610214000039.

Iglesias, C.P., Birks Y.F., & Torgerson, D.J. (2001). Improving the measurement of quality of life in older people: The York SF-12. Q J Med, 94, 695-698.

Kolkowska, E., & Kristoffersson, A. (2016). Privacy by design principles in design of new generation cognitive assistive technologies. Proceedings of the 31st IFIP TC 11 International Conference, SEC 2016, 384-397. https://doi.org/10.1007/978-3-319-33630-5_26

Kristoffersson, A., Kolkowska, E., & Loutfi, A. (2014). Assessment of Expectations and Needs of a Sensor Network to Promote Elderly's Sense of Safety and Security. Proceedings of The Seventh International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, Centric 2014, 22-28.

Liu, L., Stroulia, E., Nikolaidis, I., Miguel-Cruz, A., Rincon, A.R. (2016). Smart homes and home health monitoring technologies for older adults: A systematic review. International Journal of Medical Informatics, 91, 44-59. https://doi.org/10.1016/j.ijmedinf.2016.04.007

Liu, S., Li, C., Shi, Z., Wang, X., Zhou, Y., Liu, S., Liu, J., Yu, T., Ji, Y. (2017). Caregiver burden and prevalence of depression, anxiety and sleep disturbances in Alzheimer's disease caregivers in China. Journal of Clinical Nursing, 26(9-10), 1291-1300. https://doi.org/10.1111/jocn.13601

Lou, Q., Liu, S., Huo, Y. R., Liu, M., Liu, S. & Yong, J. (2015). Comprehensive analysis of patient and caregiver predictors for caregiver burden, anxiety and depression in Alzheimer's disease. Journal of Clinical Nursing, 24, 2668-2678. https://doi.org/10.1111/jocn.12870

McCloskey, R., Jarret, P., Stewart, C., & Keeping-Burke, L. (2015). Recruitment and Retention Challenges in a Technology-Based Study with Older Adults Discharged from a Geriatric Rehabilitation Unit. Rehabilitation Nursing, 40, 249-259. https://doi.org/10.1002/rnj.149

Medrano, M., López Rosario, R., Núñez Payano, A., & Reynoso Cappellán, N. (2014). Burden, anxiety and depression in caregivers of Alzheimer Patients in the Dominican Republic. Dementia & Neuropsychologia, 8(4), 384-388. http://dx.doi.org/10.1590/S1980-57642014DN84000013

McHorney, C. A., Ware, J. E., Lu, J. F. & Sherbourne, C. D. (1994). The MOS 36-item short form health survey (SF-36): III. Test of data quality, scaling assumptions and reliability across diverse patient groups. Medical Care, 32, 40-66.

McHugh, J. E., Wherton, J. P., Prendergast, D. K., Lawlor, B.A. (2012). Identifying opportunities for supporting caregivers of persons with dementia through information and communication technology. Gerontechnology, 10(4), 220-230. https://doi.org/10.4017/gt.2012.10.4.003.00

Molka-Danielsen, J. & Moe, C. E. (2013). Pilot design of a survey instrument for assessment of assistive ICT initiatives, Trondheim:Akademika forlag, 36.

Nordgren, A. (2013) Personal health monitoring: Ethical considerations for stakeholders. Journal of Information, Communication and Ethics in Society, 11, 156-173.

Peetoom, K. K. B., Lexis, M. A. S., Joore, M., Dirksen, C. D., & De Witte, L. P. (2015) Literature review on monitoring technologies and their outcomes in independently living elderly people. Disability and Rehabilitation: Assistive Technology, 10(4), 271-294. https://doi.org/10.3109/17483107.2014.961179

Pol, M., van Nes, F., van Hartingsveldt, M., Buurman, B., de Rooij, S., & Kröse, B. (2016). Older People's Perspectives Regarding the Use of Sensor Monitoring in Their Home. The Gerontologist, 56(3), 485-493. https://doi.org/10.1093/geront/gnu104

Schulz, R., Beach, S. R., Tabolt Matthews, J., Courtney, K. L., & De Vito Dabbs, A. J. (2012) Designing and Evaluating Quality of Life Technologies: An Interdisciplinary Approach. Proceedings of the IEEE, 100(8), 2397-2409. https://doi.org/10.1109/JPROC.2012.2200558

Siegel, C., & Dorner, T. E. (2017). Information technologies for active and assisted living—Influences to the quality of life of an ageing society, International Journal of Medical Informatics. 100, 32-45. https://doi.org/10.1016/j.ijmedinf.2017.01.012

Swedish National Board of Health and Welfare (2012) Äldreomsorgens nationella värdegrund – ett vägledningsmaterial, Ch Dialog Trygghet.

The WHOQOL Group (1994). The Development of the World Health Organization Quality of Life Assessment Instrument (the WHOQOL). In: Orley J., Kuyken W. (eds) Quality of Life Assessment: International Perspectives. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-79123-9_4

van Bronswijk, J. E. M. H. V., & Kearns, W.D. (2009). Preventive health engineering in earlier and later life. Gerontechnology, 8(2), 76-81. https://doi.org/10.4017/gt.2009.08.02.005.00

van Hoof, J., Kort, H. S. M., Rutten, P. G. S, & Duijnstee M. S. H. (2011). Ageing-in-place with the use of ambient intelligence technology: Perspectives

of older users. International Journal of Medical Informatics, 80, 310-331. https://doi.org/10.1016/j.ijmedinf.2011.02.010

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey Construction of Scales and Preliminary Tests of Reliability and Validity. Medical Care, 32, 220-233.

Wessels, R., Persson, J., Lorentsen, O., Andrich, R., Ferrario, M., Oortwijn, W., VanBeekum, T., Brodin, H., & de Witte, L. (2002). IPPA: Individually Prioritised Problem Assessment. Technology and Disability, 14, 141–145.

World Health Organization. Division of Mental Health and Prevention of Substance Abuse (1997). WHO-QOL: measuring quality of life. Geneva: World Health Organization. Available from: http://apps.who.int/iris/handle/10665/63482

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica, 67, 361-370. Zwijsen, S. A., Niemeijer, A. R., & Hertogh, C. M. (2011). Ethics of using assistive technology in the care for community-dwelling elderly people: An overview of the literature. Aging & Mental Health, 15(4) 419-427. https://doi.org/10.1080/13607863.2010.543662