Proof of Concept of a Social Robot for Patient Reported Outcome Measurements in Elderly Persons

Roel Boumans[1,2], Fokke van Meulen[2], Koen Hindriks[1], Mark Neerincx[1], Marcel Olde Rikkert[2]

[1] Delft University of Technology,
Delft, the Netherlands
{r.j.l.boumans, k.v.hindriks, m.a.neerincx}@{tudelft.nl}

[2] Radboud University Medical Center,
Nijmegen, the Netherlands
{fokke.vanmeulen, marcel.olderikkert}@{radboudumc.nl}

ABSTRACT

Medical staff uses Patient Reported Outcome Measurement (PROM) questionnaires as a means of collecting information on the effectiveness of care delivered to patients as perceived by the patients themselves. Especially for the older patient group, the PROM questioning poses an undesirable workload on the staff. This proof of concept paper investigates whether a social robot with a display can conduct such questioning in an acceptable and reliable way. A set of 15 typical questions was selected from existing PROM questionnaires. For the asking, answer-processing and responding, a multi-modal robot-dialogue was designed and implemented. In a within-subjects experiment, 31 community-dwelling older participants answered the 15 questions in two conditions: questioning by the robot, versus questioning by a human. The main part of the robot questioning provided reliable answers, but took somewhat more time compared to human questioning. The experiment demonstrated the feasibility of a social robot for an acceptable and reliable collection of PROM data from older persons.

ACM Reference Format

Roel Boumans, Fokke van Meulen, Koen Hindriks, Marc Neerincx, Marcel Olde Rikkert. 2018. Proof of concept of a Social Robot for Patient Reported Outcome Measurements in Elderly Persons. In *HRI'18 Companion: Conference on ACM/IEEE International Conference on Human-Robot Interaction, March 5-8, 2018, Chicago, IL, USA.* ACM, NY, NY, USA, 2 pages. DOI: https://doi.org/10.1145/3173386.3177013

1 INTRODUCTION

A Patient Reported Outcome Measure is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else [5]. PROMs are regularly acquired by healthcare professionals through administering questionnaires to patients. The necessity for hospitals to acquire PROMs is based on the need for providing evidence of performing value-based health care. Up till recently PROM data were mainly collected with paper-and-pencil methods, and since a few years e-health solutions such as apps on tablets or smartphones are used, where answer options can be selected by touch buttons.

However, elderly people often do not have the e-health literacy for using these devices [6], or find the technology difficult to use due to their disabilities and chronic diseases. A social robot which can conduct a verbal dialogue, supported by gestures and an answer display, would not require e-health literacy from the patient. There is evidence that such social robots are perceived as more supportive than e-health solutions such as an interactive tablet [2]. Our research question was therefore twofold: 1) how to design a dialogue, using a social robot with a screen, for acquiring PROMs by direct verbal and visual communication, and 2) what is the resulting usability in terms of user satisfaction, effectiveness and efficiency?

HRI '18 Companion, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-5615-2/18/03.
https://doi.org/10.1145/3173386.3177013

2 METHOD

Figure 1- Participant in chair viewing the robot

We designed the interaction using the situated Cognitive Engineering method [3]. We used relevant values for supporting patients in the hospital context (i.e., helpfulness, cheerfulness, politeness, responsibility, intellect and logic) to derive the interaction design requirements. Next we selected 15 questions on wellbeing, malnutrition, pain, sleep, and ability to perform certain activities of daily living from existing PROMs. The question types included dichotomous and polytomous items, linear scales, visual analogue scales, and asking for texts, numbers or dates. Robot arm motions were used that support a question and answering (Q&A) interaction, such as spreading your arms low with hands palms up. The robot's tablet was used as screen for Q&A options; all robot interaction was designed for speech. We programmed the interaction design in a robot Pepper v1.7 from Softbank Robotics (Tokyo, Japan) because of its user friendly programming environment, its ability to communicate in Dutch and its friendly human-like appearance, which was expected to appeal to elderly.

The experiment was designed as a non-blinded, non-randomized controlled trial. We compared the PROM interactions with the robot and with a human in a paired t-test within subjects with n=31, with alternating groups to minimize carry-over effects. The interview setup consisted of a room, in which the participant sat in a chair facing the robot from a distance of about 1.2 meter (Figure 1). The robot-taken PROM questionnaire (RP) procedure started by asking the participant to begin the PROM dialogue with the robot by saying "Hello Pepper" to the robot. We compared with a human-taken PROM questionnaire (HP). This HP procedure started with the researcher asking the same PROM questions using the same script, with the Q&A options shown on a laptop. The research approach has been reviewed and approved by the Ethical Commissions of our universities.

3 EVALUATION

We scored the results with dedicated usability questionnaires and variables derived from the Almere model [1]. The user satisfaction was determined in two steps. First participants were asked to score usability statements on the domains Perceived Ease of Use, Perceived Enjoyment, Perceived Sociability, Perceived Usability, Trust and Social Influence. The scoring range was a 7-point Likert scale (totally disagree – disagree – slightly disagree – neutral – slightly agree – agree – fully agree, equivalent to scores 1-7). Secondly, these scores were used to calculate the satisfaction with the method for the System Usability Scale [4] to arrive at a value between 0 and 100. The effectiveness number was defined as the number of off-script events per participant during the interaction. These events could be anything from not waiting until the robot finished speaking (barge-in), not giving a predefined answer or speaking unclearly. The task efficiency was defined as the RP task completion time, compared to HP completion time. The technical reliability was determined by comparing the answers recorded digitally with the answers stated by the participants as heard on a video of the interaction.

4 RESULTS

Thirty-one participants (45% female) with an average age of 76.2 (sd=5.2) completed both sessions. The completion of all 15 questions is equivalent to 465 participant/robot interaction cases. The average duration of the time elapsed between RP and HP interaction was 15 days. Due to two unanswered evaluation questionnaires we could use only 29 evaluations for the paired comparisons.

The effectiveness number for the RP interactions was (M=7.5, sd=3.4) and for the HP interactions (M=3.3, sd=2.7) ($t(30)=5.5$, $p<.001$). The task duration for the RP interactions was (M= 7'11" , sd=0'42") and for the HP interaction (M=5'32", sd=0'49") ($t(30)=13.6$, $p<.001$). The RP task efficiency was 5'32"/7'11" = 0.77. The participants scored the subjective usability as (M=80.1, sd=11.6, n=31) for the robot and (M=84.0, sd=10.7, n=29) for the human, which is not significantly different (Mann-Whitney U = 528.5, $n_1=31$, $n_2=29$, $p<0.05$, two-tailed). The usability parameter Trust scored lower (M=66) than the usability average ($p<.001$); two participants explained: "It depends on who programs the robot.". Also the parameter Social Influence scored lower (M=68) ($p<.003$); participants reported sceptic remarks from family and friends on the use of social robots in health care.

The carry-over effects were symmetric. The technical reliability analysis showed 2 cases (0.4%) in which the robot registered a wrong answer, thus the correctness of the data is 99.6%. Participants stated that they felt hardly any difference between answering the questions from the robot versus the human, they felt at ease with the robot, and did not mind answering this type of questions from a robot.

5 DISCUSSION

This proof of concept showed that the robot interview was perceived as an acceptable way to provide PROM data. The subjective usability indicates that older persons value the interaction with the social robot as positive. However the effectiveness number was too high. Analysis showed that it was mainly caused by two events: a barge-in effect, causing an unnecessary waiting time, and non-fluent pronunciation of a certain phrase. Both can be solved by new dialogue elements. A limitation of the study is its small population sample with highly educated and motivated elderly participants. The strength of this study is a demonstration of a real-life PROM questionnaire interaction with elderly people by a robot.

6 CONCLUSIONS

We conclude that it could be feasible for this social robot to effectively and efficiently obtain Patient Reported Outcome Measures from elderly persons. Answers to questions were correctly recorded. The effectiveness can be further improved.

Acknowledgement

The 4TU research centre Humans & Technology and the Radboud university medical center have funded this research.

7 REFERENCES

[1] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: the Almere model. *Int. J. Soc. Robot.* 2, 4 (2010), 361–375. DOI:https://doi.org/10.1007/s12369-010-0068-5

[2] J.A. Mann, B.A. MacDonald, I.H. Kuo, X. Li, and E. Broadbent,. 2015. People respond better to robots than computer tablets delivering healthcare instructions. *Comput. Hum. Behav.* 43, (2015), 112–117. DOI:https://doi.org/10.1016/j.chb.2014.10.029

[3] M.A. Neerincx and J. Lindenberg. 2008. *Situated cognitive engineering for complex task environments.* Ashgate Publishing Limited Aldershot.

[4] J. Sauro and J.R. Lewis. 2016. *Quantifying the User Experience - Practical Statistics for User Research* (2nd edition ed.). Morgan Kaufman.

[5] US Department of Health and Human Services. 2009. Guidance for Industry - Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims.

[6] I. Watkins and B. Xie. 2014. eHealth Literacy Interventions for Older Adults: A Systematic Review of the Literature. *J. Med. Internet Res.* 16, 11 (November 2014), e225. DOI:https://doi.org/10.4017/gt.2014.13.02.206.00 10.2196/jmir.3318