Delft University of Technology

A Feasibility Study of a Social Robot Collecting Patient Reported Outcome Measurements from Older Adults

Boumans, Roel; van Meulen, Fokke; Hindriks, Koen; Neerincx, Mark; Olde Rikkert, Marcel

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

ORIGINAL RESEARCH

A Feasibility Study of a Social Robot Collecting Patient Reported Outcome Measurements from Older Adults

Roel Boumans[1,2] · Fokke van Meulen[1] · Koen Hindriks[2] · Mark Neerincx[2] · Marcel Olde Rikkert[1,3]

Abstract

Patient reported outcome measures (PROMs) are an essential means for collecting information on the effectiveness of hospital care as perceived by the patients themselves. Especially older adult patients often require help from nursing staff to successfully complete PROMs, but this staff already has a high work load. Therefore, a social robot is introduced to perform the PROM questioning and recording task. The study objective was to design a multimodal dialogue for a social robot to acquire PROMs for older patients. The primary outcomes were the effectiveness, the efficiency, and the subjective usability as perceived by older adults of acquiring PROMs by a social robot. The robot dialogue design included a personalized welcome, PROM questions, confirmation requests, affective statements, use of a support screen on the robot displaying the answer options, and accompanying robot gestures. The design was tested in a crossover study with 31 community-dwelling persons aged 70 years or above. Answers obtained with the robot were compared with those obtained by a questionnaire taken by humans. First results indicated that PROM data collection in older persons may be carried out effectively and efficiently by a social robot. The robot's subjective usability was on average scored as 80.1 ($\pm$11.6) on a scale from 0 to 100. The recorded data reliability was 99.6%. A first relevant step has been made on the design trajectory for a robot to obtain PROMs from older adults. Practice variation in subjective usability scores still asks for technical dialogue improvements.

Keywords Social robot · Patient reported outcome measures · Humanoid · Multi-modal dialogue · Older adults

1 Introduction

Patient report outcome measurements (PROMs) are questionnaires to record a patient's opinion on the status of their health condition, health behavior, or their evaluation of received healthcare. PROM data are obtained from the patient without interpretation of the patient's response by

✉ Roel Boumans
 roel.boumans@radboudumc.nl

1 Department of Geriatric Medicine, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

2 Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

3 Radboud Alzheimer Centre, Radboud University Medical Center, Nijmegen, The Netherlands

a clinician or anyone else [1]. Several health organizations advise that patients should be routinely asked for these "patient reported outcomes" [1–3]. They consider information from the patient's perspective essential to support a patient-centered approach to care. A survey of nearly 100,000 clinical trials published between 2007 and 2013 found that a PROM was used in 27% of these trials [4]. For older adults, PROMs are specifically important because they may also be a means to express the patients' actual and desired quality of well-being. In this respect, most older adults consider their quality of well-being recovery by a hospital intervention more important than increased longevity [5, 6].

The process of obtaining PROMs may require help from clinical staff; this is often necessary but time consuming. Furthermore, time may be needed to enter the data in an electronic health record. However, the administrative workload for nurses for writing nursing reports and nursing handovers is already high [7, 8]. Therefore, a relevant aim is to decrease the time spent by nurses on administration. This allows more time left for providing the fundamentals of care, such as sharing fear and sorrow, securing appropriate nutrition,

hydration, personal hygiene, sleep, rest, and interpersonal communication [9].

Electronic PROM tools (ePROs) are applications on a computer, tablet, or smartphone, in which people can enter their responses to questions [10, 11]. Advantages above pen-and-paper solutions are the automatic storage of the patient's responses in their personal health record, the automated calculation of scores, and the ease to present processed results in a brief report to medical staff. However, many patients have difficulty using computers, tablets, or smartphones because of their lack of digital literacy [12]. Also physical or cognitive problems, disabilities, or chronic diseases can make this technology difficult to use [13–15]. Other specific problems with exchanging tablets between patients are privacy threats and risks of spreading infections [10].

A speaking social humanoid robot may be an alternative for using paper forms, tablets, or computers, if it is capable to conduct a dialogue on the status of the patient's health. In that scenario, the patient needs only to answer the questions by voice. State-of-the-art social robots can include advanced dialogues that incorporate additional introductions, explanations, and background information. The robot can use affective statements such as "I am sorry to hear that" where appropriate. Moreover, the social robot could spend more time on the PROM interaction than nurses may have available. Obviously, the social robot shares many of the advantages identified for ePROs, such as electronic data storage, data processing, and reporting.

There is already some evidence that social robots are useful for answering health-related questions. Experiments have been done where participants answered health questions posed by a robot using data entry on a touch screen attached to the robot [16, 17]. In another experiment, health-related questions were posed to a participant in a so called "Wizard-of-Oz" setup [18], which means that human operators remotely enter the statements to be said by the robot and the participant is actually interacting with a human operator instead of an autonomous robot system [19]. To our knowledge, social robots have not yet been used for autonomous PROM questioning.

Based on the aforementioned research and the fact that the pen-and-paper interview with a nurse is still the most common option for conducting PROM questionnaires among older persons, the decision was made to focus on the comparison between a social robot and a nurse in this proof-of-concept study. A multimodal dialogue was designed for a social robot to obtain a valid patient reported outcome. The research question was defined as: what is the effectiveness, efficiency, and subjective usability of the robot-taken PROM questionnaire (RP), when compared with a human-taken PROM questionnaire (HP)?

This paper describes the design of a multimodal dialogue for a social robot to acquire PROMs (patient reported outcome measures) for older patients. The robot is able to pose PROM questions and record their answers. The main contribution of this paper is that it reports a first research on evaluating robot-mediated data acquisition on PROMs in older participants.

2 Design of the PROM Interaction

Following the situated cognitive engineering method [20, 21], the design process started from a reference scenario in which the social robot is located in a room and the patient is brought to the robot by a nurse for an interview. The patient would sit in front of the robot and initiate the dialogue. Then the robot would start asking a range of questions and would react almost immediately (within 0.5 s) to the answers given. When all questions were answered, the robot would thank the participant. The modes of behavior for the robot toward the patient were decided to be aiming at cheerfulness, politeness, responsibility, intellect, logic, helpfulness, personalization, trust, and convenience [22–24].

In the next step, a dialogue representative for most PROMs was made. A range of typical questions with varying answer sets such as dichotomous and polytomous items, linear scales, visual analogue scales, and questions for numbers or dates was selected. PROM questionnaires currently in use at the Geriatrics department were reviewed: the Personal Wellbeing Index [25], the Malnutrition Universal Screening Tool [26], pain assessment using a Visual Analogue Scale [27, 28], the Pittsburgh Sleep Quality Index [29], the Barthel index [30], and The Older Persons and Informal Caregivers Study questionnaire (TOPICS) [31]. Fifteen questions were selected on well-being, malnutrition, pain, sleep, and ability to perform certain activities of daily living from these PROMs (see supplementary material for the questions used).

The Pepper robot from Softbank Robotics (Tokyo, Japan) was selected as robot platform 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 older persons (Fig. 1). Pepper is a humanoid robot 1.21 m tall and 26 kg in weight. It has a 10.1" screen on its chest. The screen was used to display the question-and-answer (Q&A) options for all questions except those on birth date and nationality. The Dutch speech recognition and speech functionality was made by Nuance (Burlington, MA, USA). For Pepper's arm and body motions during the interaction, the robot's ALSpeakingMovement and ALListeningMovement modules were used, which launched random arm and body animations typical for a neutral communication. In both modules, the robot eyes followed the human head to keep eye contact.
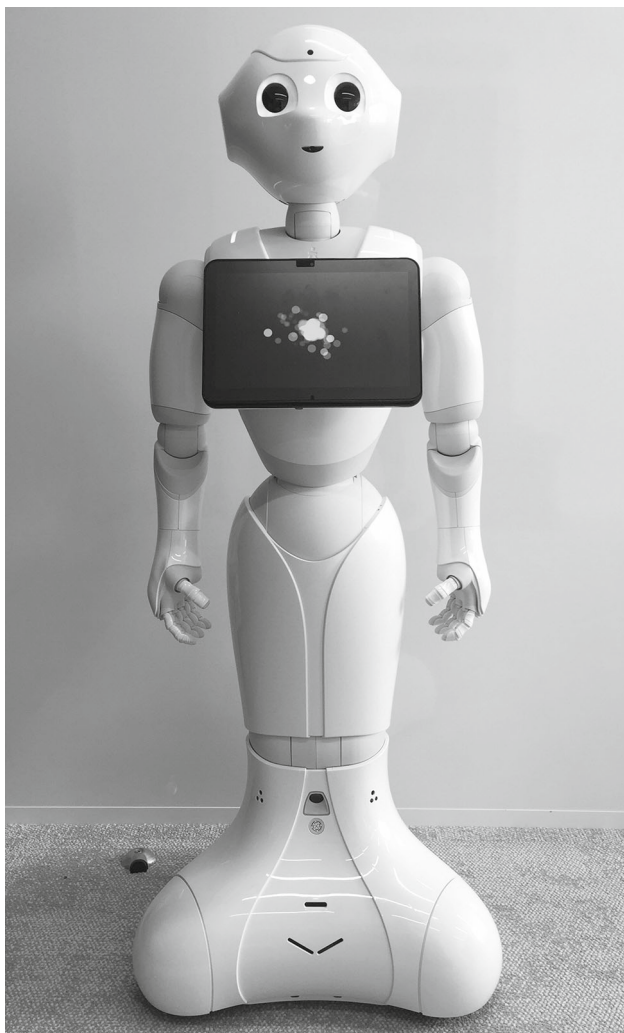
Fig. 1 The Pepper robot

Fig. 2 Schematic view of the interview setup

3 Evaluation

3.1 Experimental Setup

The experiment is designed as a non-blinded controlled crossover trial. Each participant had a RP interaction and a HP interaction. The RP and HP interactions were planned to take place with a 2-week wash-out period in between to minimize learning effects. The order of the RP and HP interactions for the participant was based on the time of signing in for the experiment. Community-dwelling older participants were recruited by advertisements in a local newspaper and through welfare organizations for older persons. Inclusion criteria were age above 70 years, Dutch speaking, and no cognitive impairments. Because of the lack of data for this type of human–robot interaction trials, the required sample sizecould not be calculated, and the aim was therefore pragmatically set to 30 participants.
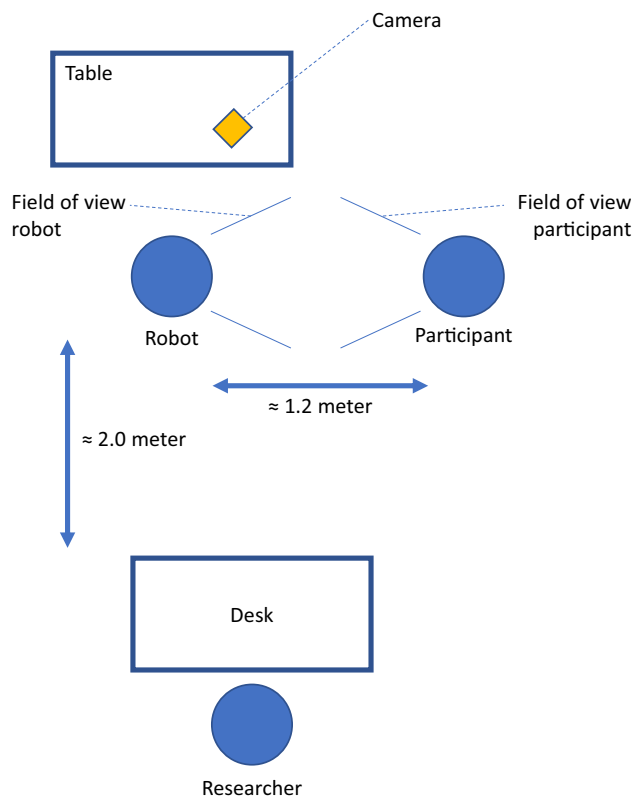
The interview setup consisted of a room in which the participant sat on a chair facing the robot at a distance of about 1.2 m. The heads of the robot and the participant were at the same height. The robot was the main object of view for the participant (Fig. 2).

The intention was that the participant was able to complete the interview without any help. Since this would probably be the first time that these older adults would interact with a social robot, during the RP interaction a researcher was present in the room for reassurance. For example, if the participant did not know how to proceed, the participant could ask the researcher what to say to the robot. During video analysis, such an event will be noted as an off-script event. All interactions were recorded with a Flip mino HD video camera (Cisco Systems, CA, USA).

The research plan has been reviewed by the Medical Ethical Review Board of the Radboud university medical center (dossier number 2017-3392); the board did not consider the study as a medical experiment, and therefore the research plan was not subject to national legislation for medical experiments in human beings. Informed consent was obtained from all individual participants included in the study. The study has been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

3.2 The Procedure for the Interactions

The RP interaction started with the researchers inviting the participant to sit in a chair opposite the robot. First, the participant completed a 3-min training dialogue with the robot under guidance of the researcher. If the participant was comfortable to proceed after completion of the training dialogue, they could initiate the PROM questionnaire by saying "Hello Pepper." If during RP interaction the robot would malfunction, the researcher could be asked for help.

The HP interaction started with the researcher asking the same PROM questions. The dialogue script as programmed in the robot was used to make both interviews comparable. The answering options were shown on a laptop to mimic the robot's screen.

3.3 Data Analysis

Three empirical sources to evaluate this proof of concept of the RP interactions were selected: data recordings of the robot itself, analysis of the interaction video's, and questionnaires for the participants on both interactions. The interview by the robot was defined to be efficiently conducted if it was completed within a reasonable time [32], compared with duration of the HP interaction (ratio HP/RP duration > 0.5). It was expected that, during the actual RP dialogues, some off-script events might occur. Off-script events are events that do not follow the preprogrammed script of questioning and answering. Two off-script event types were anticipated. The first off-script event type was one raised by the participant, when for instance, the robot did not respond to the answer given by the participant, and therefore the participant had to repeat the answer. In the second off-script event type, the participant asked for help from the researcher. Both RP and HP interaction videos were reviewed, and observed events were written down on forms and categorized. The effectiveness of RP interaction was determined by counting the number of off-script events that occurred during the interactions [33].

A question/answering interaction "set" was defined as one participant completing one Q&A set including confirmation and optional repeats or clarifications. With X participants completing Y Q&A-sets with the robot, X*Y interaction sets were obtained. The number of off-script events can be related to the number of interaction sets. It is possible that more than one off-script event occurs during one interaction set. Because this is a feasibility study, validity issues between questions were not studied.

An evaluation questionnaire was used to ask participants to score their subjective usability after both the RP interaction and the HP interaction. The questionnaire consisted of 11 statements to be scored on a 7-point Likert scale (totally disagree—disagree—slight disagree—neutral—slight agree—agree—fully agree, equivalent to scores 1–7). The statements were based on the Almere model for assessing acceptance for assistive social agent technology [34] and selected and adapted to conducting PROM questionnaires (Table 1). All statements are formulated positively since negations complicate wanted understanding of the questions [35]. The overall usability score was determined using the method for the System Usability Scale [35]:

$$T_s = \frac{100}{N \times L} \sum_{i=1}^{N}(s_i - 1) \tag{1}$$

Here, T_s is the total score on a 0–100 scale, i the item number, N the total number of items, L the Likert range minus one, and s_i the score per item. A usability score is called "high" if the mean score is over 80. To compare the usability scores of RP and HP, the evaluation questions 1–8 and 10 of Table 1 were used, with replacing the word "robot" in the question by "nurse" for the HP interaction. Questions 9 and 11 were not fitting HP interaction and were not included in the comparison.

All participants were asked to compare both experiences by asking them to score the statements "Do you find a difference in answering the questions by human or robot?," "Would you mind if these questions are asked by a robot instead of a human?," "Would you feel more at ease with the human?," and "Did you consider answering the questions from the robot more difficult?" on a 7-point Likert scale. This could indicate the preference for one of the methods, RP or HP. The recorded data reliability was assessed by comparing the answers recorded electronically by the robot with the answers stated by the participants as heard in the video. The correlation between the answers on the questions on life in general, health in general, weight, and activities of daily living, as given to the robot and the nurse, were determined, because these were not likely to change over the period between the RP and HP interactions.

The Castor research data management system (Castor EDC, Amsterdam, the Netherlands) was used to record study data. SPSS version 22 (IBM, USA) and Microsoft Excel 2007 (Microsoft, USA) were used for statistical analysis. The resulting outcome scores for effectiveness, usability, and the times measured for efficiency were compared in paired one sample t tests between subjects for normally distributed data, and by the Mann–Whitney U test for non-normal distributions. Standard deviations are presented between parentheses. Correlations between continuous variables in the TOPICS answers were analyzed with Spearman's ρ_s. The datasets generated during and/or analyzed during the current study are available from the corresponding author on request.

Table 1 Evaluation questions, variables, and scores for the RP interaction with older participants

	Evaluation questions	Variable	Mean score	SD
1.	I approve of this robot asking me questions on my health	Attitude towards technology	87	17
2.	I find it easy to answer the questions of the robot	Perceived ease of use	87	16
3.	I think the robot understands my answers correctly	Trust	76	23
4.	I find the talking speed of the robot satisfactory	Perceived ease of use	85	19
5.	I find the robot reacting pleasantly to my answers	Perceived enjoyment	88	13
6.	I think the robot's body motions were appropriate	Perceived sociability	67	25
7.	I find the robot made a friendly impression	Perceived enjoyment	88	11
8.	I would trust the robot if it would give me advice	Trust	66	21
9.	I find the robot funny	Perceived sociability	85*	17
10.	I find the answering of questions from the robot a pleasant way to provide data on my health	Perceived usability	78	21
11.	I think my family and friends would approve of using a robot to provide data on my health	Social influence	68*	22

*Not included in subjective usability evaluation of the HP interaction

4 Results

Thirty-one community-dwelling older participants (45% female) with an average age of 76.2 (2.0) years completed both sessions. No participants were observed as frail or ill at the moment of the interviews. No clear anxiety among participants was observed during the RP interaction. No participants had speech impediments. Their education level was high: 57% of the male participants and 82% of the female participants had a college or university degree. Of all participants, 35% had visited an outpatient clinic in 2017 for treatment, for checkups, or for emergency care.

All 31 participants completed all 15 Q&As; therefore, 465 interaction sets were obtained. The participants showed 100% adherence to both interactions. Two participants have not answered the (repeated) e-mails on the HP interaction evaluation; therefore, only 29 evaluations for the paired comparisons on usability could be used.

The average number of off-script events caused by the participant for the RP interaction was 5.2 ($\pm$3.2) per participant, and in total 162. In 83 events, these concerned barge-in errors where a participant answered "yes" or "no" too quickly on one of the seven confirmation questions or the four ADL questions. In 26 events, the participants had to repeat their answer a bit louder for the robot to understand. In 11 events, the participants used an answer not in the answer list, realized this because the robot did not react, and corrected themselves.

Other off-script events were: the participant making a funny remark which the robot did not understand, the participant gave a wrong answer at first and corrected themselves, the participant did not understand the question, and the participant did not hear the robot.

The average number of off-script events where the researcher was asked to help for continuation with the interview was 2.0 ($\pm$1.2) per participant, and in total 55. The researcher explained how to give the answer, and after the participant did so, the robot continued with the interview. In 22 sets, help was needed with fluently stating their birth day. In 15 sets, the participant answered yes or no too quickly. In 11 sets, the participant used an answer not in the answer list. The following events occurred only for a maximum of three times: the participant did not understand the question; the participant did not hear the question; or the participant stated the answers too softly. In all cases the interview was completed.

The number of off-script events in HP interactions was on average 3.3 ($\pm$2.7) per participant, and in total 106. A qualitative analysis of the HP dialogue videos showed that these off-script events can be categorized as follows:

- Participant gave a long explanatory answer (51 sets);
- Participant gave a short answer that is not in the answer list (32 sets);
- Participant gave a premature answer (19 sets);

- Participant posed a clarification question to the human (three sets);
- Participant answered with a joke (one set).

The nature of the off-script events differed between HP and RP interactions. For the RP interaction, they were due to volume issues or "jumping the gun," whereas during HP interaction, participants clarified their answers.

The RP task efficiency ratio was 5 min 32 s/7 min 11 s = 0.77, where the average duration for the RP interaction (without training time) was 7 min 11 s (0 min 42 s; range: 6 min 00 s–9 min 36 s) and for the HP interaction 5 min 32 s (0 min 49 s; range: 4 min 19 s–8 min 57 s). The shortest RP interview duration with all answers given immediately correct was 5 min 57 s. The recorded data reliability analysis shows only two sets (0.43%) in which a wrong answer had been registered by the robot. Thus, the recorded data reliability is 99.6%.

The participants scored the overall subjective usability score as 80.1 ($\pm$11.6) for the robot and 84.0 ($\pm$10.7) for the human; these scores are not significantly different (Mann–Whitney U = 528.5, n1 = 31, n2 = 29, $p <$0.05, two-tailed). The participants' opinions on the interaction are provided in Table 1. The first group (n = 17), who were first interviewed by the robot and then by the human, scored the subjective usability for the robot on average as 78.2 ($\pm$12.9) and for the human as 82.0 ($\pm$10.5). The second group (n = 12), who were interviewed in reverse order, scored the subjective usability for the robot as 82.4 ($\pm$9.8) and for the human as 86.7 ($\pm$11.0). Thus, carry-over effects changing the difference in RP vs HP evaluation were absent ($p <$0.001). The mean time between both interviews was 15.7 days.

The participant's answers given to the robot and given to the human were compared. The answers have a strong correlation for the questions on satisfaction with life in general (Spearman's ρ_s = 0.900, same answers = 85%), health in general (ρ_s = 0.913, 74%) and on weight (ρ_s = 0.980, 77%). The dichotomous answers to the questions on the ability to travel independently were equal for 97% of the participants. The same levels were reached for the questions on shopping (97%), meal preparation (97%), and doing household tasks (90%).

5 Discussion

5.1 Global Evaluation

This experiment on the interaction of a robot with older participants in asking structured data showed that the robot interview was perceived as an acceptable way to provide PROM data. This is consistent with the results among a group of patients with Parkinson's disease [18]. The subjective usability of the robot was rated high. The system usability scale rating did not differ significantly between the PROM acquisition by the robot vs the human. Moreover, the robot–PROM interaction was highly reliable in registering the PROM answers as communicated. The design of the multimodal dialogue proved usable, although there certainly are some lessons learned and possible improvements identified by the off-script interactions.

The task efficiency in terms of completion time was moderate: the robot interaction took more time if you consider the time between first and last question. This may be different when analyzing the complete time from meeting the patient to saying goodbye, but this is more difficult to compare objectively. It is expected that efficiency can be improved by tailoring the multimodal interaction sets to specific questions. The effectiveness goal of the interaction for routine PROM acquisition in care pathways should be to obtain data without off-script events requiring external intervention. In future, no staff should need to be present during the interview. Some off-script events by the participant might not necessarily be a problem, e.g. stating an answer twice if the robot does not initially react, as long as it does not annoy the participant. Participants also caused off-script events during their interaction with the human, and this is considered normal. However, to improve the effectiveness, including answer screens for also the more obvious answer sets will be considered. Also the use of a timer function that enables the robot to take action if, after some time, it has not been able to understand the participant's statement will be studied. The observed correlations between the participant's answers on the same questions to the robot and to the human also gave confidence that use of the robot may result in valid PROM questioning.

These results may point out the potential usefulness of social robots for other patient groups. For example, for symptom reporting by children in pediatric oncology [36]. However, it may also be useful for some older adults who are reported to have problems with the Geriatric Depression Scale and the Patient-Reported Outcome Measurement Information System [15].

5.2 Strengths and Weaknesses

The strength of this study is the design, implementation, and evaluation of conducting PROM questionnaires by a robot and the comparison with data acquired by a nurse. A strict protocol was used for both interactions and based the investigation on validated usability scoring methods [34, 35], tailored for this study. As far as we know, this is the first study that evaluated robot-mediated data acquisition on healthcare outcomes (PROMs) in older participants.

A limitation of the study is that it had a non-blinded design, which is however unavoidable. It also included a small sample of highly educated participants, which may have inflated

acceptability. Frailty or illness were not measured objectively. Moreover, although participants in this study may have been representative for the older persons first seen in an outpatient setting, they are not representative of frail older patients admitted to hospital. For these patients, a separate usability trial with a more representative group of participants is needed.

6 Conclusions and Future Work

The conclusion of this study is that a first relevant step has been made on the design trajectory for a robot to effectively and efficiently obtain PROMs from older adults. The robot is able to pose PROM questions and record their answers. The subjective usability was judged positively by the older participants who favorably accepted and appreciated the interaction with the social robot. However, several interaction elements were observed that still require improvements to obtain a higher effectiveness and efficiency.

This first positive proof of concept warrants further innovation, implementation, and evaluation of social robot interaction with older patients. Next steps should consist of further development of the quality of the interaction in co-creation with healthy older and more frail older participants. This will include exploration of direct PROM feedback to professionals, as well as application of this social robot technology in integrated care pathways [37] to have both patients and professionals benefit from an improved quality of care. Future opportunities might also include gathering patient reported outcomes in the patient's native language.

Compliance with Ethical Standards

Conflict of interest Roel Boumans is part-time employed with mijnrobots.nl, but this company has no financial interest in, or conflict with, the subject matter or materials discussed in this manuscript. Koen Hindriks is part-time CEO of Interactive Robotics, but this company has no financial interest in, or conflict with, the subject matter or materials discussed in this manuscript. Mark Neerincx, Fokke van Meulen and Marcel Olde Rikkert declare that they have no conflict of interest.

References

1. National Quality Forum (2013) Patient reported outcomes (PROs) in performance measurement. National Quality Forum, Washington
2. Canadian Institute for Health Information (2015) PROMs background document. https://www.cihi.ca/sites/default/files/document/proms_background_may21_en-web.pdf
3. NHS Digital (2017) Patient reported outcome measures (PROMs) in England—a guide to PROMs methodology. NHS Digital, Leeds
4. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL (2015) Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007–2013). Contemp Clin Trials 43:1–9
5. Hofman CS, Makai P, Boter H, Buurman BM, de Craen AJ, Olde Rikkert MG et al (2015) The influence of age on health valuations: the older olds prefer functional independence while the younger olds prefer less morbidity. Clin Interv Aging 10:1131–1139
6. Bakker FC, Persoon A, Bredie SJ, van Haren-Willems J, Leferink VJ, Noyez L et al (2014) The CareWell in Hospital program to improve the quality of care for frail elderly inpatients: results of a before–after study with focus on surgical patients. Am J Surg 208(5):735–746
7. Hendrich A, Chow MP, Skierczynski BA, Lu Z (2008) A 36-hospital time and motion study: how do medical-surgical nurses spend their time? Perm J 12(3):25
8. Van Veenendaal H, Wardenaar J, Trappenburg M (2008) Tijd voor administratie of voor de patiënt? Een onderzoek naar de administratieve belasting van verpleegkundigen. (Time for administration or for the patient? An investigation into the administrative burden for nurses). In: Kwaliteit in Zorg—Tijdschrift over kwaliteit en veiligheid in de zorg. Vakmedianet, Alphen aan de Rijn, Netherlands, vol 2, pp 24–26 (in Dutch)
9. Kitson A, Conroy T, Kuluski K, Locock L, Lyons R (2013) Reclaiming and redefining the fundamentals of care: nursing's response to meeting patients' basic human needs. University of Adelaide, Adelaide, SA, Australia, School of Nursing
10. Schick-Makaroff K, Molzahn A (2015) Strategies to use tablet computers for collection of electronic patient-reported outcomes. Health Qual Life Outcomes 13(1):2
11. Shah K, Hofmann M, Schwarzkopf R, Pourmand D, Bhatia N, Rafijah G et al (2016) Patient-reported outcome measures: How do digital tablets stack up to paper forms? A randomized, controlled study. Am J Orthop (Belle Mead NJ) 45(7):E451–E457
12. Watkins I, Xie B (2014) eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 16(11):e225
13. Smith A (2014) Older adults and technology use. Pew Research Center. http://assets.pewresearch.org/wp-content/uploads/sites/14/2014/04/PIP_Seniors-and-Tech-Use_040314.pdf. Accessed 15 May 2017
14. US Department of Health and Human Services (2009) Guidance for industry—patient-reported outcome measures: use in medical product development to support labelling claims. US Department of Health and Human Services, Washington
15. Paz SH, Jones L, Calderón JL, Hays RD (2017) Readability and comprehension of the Geriatric Depression Scale and PROMIS® physical function items in older African Americans and Latinos. Patient 10(1):117–131
16. Mann JA, MacDonald BA, Kuo IH, Li X, Broadbent E (2015) People respond better to robots than computer tablets delivering healthcare instructions. Comput Hum Behav 43:112–117
17. Kidd CD, Breazeal C (2008) Robots at home: understanding long-term human–robot interaction. In: Proceedings of the 2008 IROS

IEEE/RSJ international conference on intelligent robots and systems, Nice, France, pp 3230–3235

18. Briggs P, Scheutz M, Tickle-Degnen L (2015) Are robots ready for administering health status surveys: first results from an HRI study with subjects with Parkinson's disease. In: Proceedings of the 10th annual ACM/IEEE international conference on human–robot interaction. ACM, Portland, OR, USA, pp 327–334

19. Kelley JF (1984) An iterative design methodology for user-friendly natural language office information applications. ACM Trans Inf Syst TOIS 2(1):26–41

20. Neerincx MA, Lindenberg J (2008) Situated cognitive engineering for complex task environments. Ashgate Publishing, Aldershot

21. Neerincx MA (2011) Situated cognitive engineering for crew support in space. Pers Ubiquitous Comput 15(5):445–456

22. Harbers M, Peeters MMM, Neerincx MA (2015). Perceived autonomy of robots: effects of appearance and context. In: Proceedings of the international conference on robot ethics, Lisbon, Portugal

23. Rokeach M (1973) The nature of human values. The Free Press, New York

24. Lee N, Kim J, Kim E, Kwon O (2017) The influence of politeness behavior on user compliance with social robots in a healthcare service setting. Int J Soc Robot 9(5):727–743

25. Van Beuningen J, de Jonge T (2011) The Personal Wellbeing Index: construct validity for the Netherlands. Centraal Bureau voor de Statistiek, Den Haag

26. Stratton RJ, Hackston A, Longmore D, Dixon R, Price S, Stroud M et al (2004) Malnutrition in hospital outpatients and inpatients: prevalence, concurrent validity and ease of use of the 'malnutrition universal screening tool'('MUST') for adults. Br J Nutr 92(5):799–808

27. Gould D, Kelly D, Goldstone L, Gammon J (2001) Examining the validity of pressure ulcer risk assessment scales: developing and using illustrated patient simulations to collect the data. J Clin Nurs 10:697–706

28. Woodforde J, Merskey H (1972) Some relationships between subjective measures of pain. J Psychosom Res 16(3):173–178

29. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ (1989) The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 28(2):193–213

30. Mahoney F (1965) Functional assessment: the Barthel index. Md Med J 14:61–65

31. Lutomski JE, Baars MA, Schalk BW, Boter H, Buurman BM, den Elzen WP et al (2013) The development of The older persons and informal caregivers survey minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. PLoS ONE 8(12):e81673

32. Murphy RR, Schreckenghost D (2013) Survey of metrics for human–robot interaction. In: Proceedings of the 8th ACM/IEEE international conference on human–robot interaction (HRI), Tokyo, Japan, pp 197–198

33. Steinfeld A, Fong T, Kaber D, Lewis M, Scholtz J, Schultz A et al (2006) Common metrics for human–robot interaction. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction, Salt Lake City, UT, USA, pp 33–40

34. Heerink M, Kröse B, Evers V, Wielinga B (2010) Assessing acceptance of assistive social agent technology by older adults: the Almere model. Int J Soc Robot 2(4):361–375

35. Sauro J, Lewis JR (2016) Quantifying the user experience—practical statistics for user research, 2nd edn. Morgan Kaufmann Publishers Inc., San Francisco

36. Leahy AB, Feudtner C, Basch E (2018) Symptom monitoring in pediatric oncology using patient-reported outcomes: why, how, and where next. Patient 11(2):147–153

37. Olde Rikkert MGM, van der Wees PJ, Schoon Y, Westert GP (2018) Using patient reported outcomes measures to promote integrated care. Int J Integr Care 18(2):8

Roel Boumans graduated in 1988 from Delft University of Technology, Faculty Aerospace Engineering on Space Robotics. After graduation he worked for almost ten years on the European Robot Arm at Fokker Space. Currently he is conducting research on social robots for health care for Delft University of Technology and Radboud University Medical Center. He also founded a startup company for social robot applications for long-term care, mijnrobots.nl.

Fokke van Meulen graduated in 2011 from the University of Twente in Technical Medicine. He obtained his PhD from the same university in 2017 on the topic of the evaluation of daily life movements of stroke survivors using wearable sensors. Dr. van Meulen is currently employed by the Integraal Kankercentrum Nederland as clinical informatician.

Koen Hindriks graduated from the Rijksuniversiteit Groningen in 1996 in Computer Science, and obtained his PhD from Utrecht University on agent programming languages in 2001. Prof. Hindriks is currently professor in Artifical Intelligence at the Vrije Universiteit Amsterdam. He is also launched the spinoff company Interactive Robotics in 2016.

Mark Neerincx graduated from University Leiden in Cognitive Psychology in 1987 and obtained his PhD on the same subject from Rijksuniverstiteit Groningen. Prof. Neerincx is currently professor at Delft University of Technology in human-centered computing. He is also employed by TNO as principal scientist.

Marcel Olde Rikkert graduated as Medical Doctor from the University of Groningen in 1988. He obtained his PhD on methodological and ethical aspects of scientific research in geriatrics from the Radboud University Nijmegen in 1998. Prof. Olde Rikkert is currently head of the Geriatrics Department of the Radboud University Medical Center. He is also chair of the Radboud Alzheimer Center.