



OPEN ACCESS

Robot for health data acquisition among older adults: a pilot randomised controlled cross-over trial

Roel Boumans,^{1,2} Fokke van Meulen,¹ Koen Hindriks,² Mark Neerincx,² Marcel G M Olde Rikkert¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2018-008977>).

¹Department of Geriatrics, Radboud University Medical Center, Nijmegen, The Netherlands

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

Correspondence to

Roel Boumans, Department of Geriatrics, Radboudumc, Nijmegen, The Netherlands; roel.boumans@radboudumc.nl

Received 17 October 2018

Revised 20 February 2019

Accepted 24 February 2019

Published Online First

20 March 2019



► <http://dx.doi.org/10.1136/bmjqs-2019-009514>



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Boumans R, van Meulen F, Hindriks K, et al. *BMJ Qual Saf* 2019;**28**:793–799.

ABSTRACT

Background /Objectives Healthcare professionals (HCP) are confronted with an increased demand for assessments of important health status measures, such as patient-reported outcome measurements (PROM), and the time this requires. The aim of this study was to investigate the effectiveness and acceptability of using an HCP robot assistant, and to test the hypothesis that a robot can autonomously acquire PROM data from older adults.

Design A pilot randomised controlled cross-over study where a social robot and a nurse administered three PROM questionnaires with a total of 52 questions.

Setting A clinical outpatient setting with community-dwelling older adults.

Participants Forty-two community-dwelling older adults (mean age: 77.1 years, SD: 5.7 years, 45% female).

Measurements The primary outcome was the task time required for robot–patient and nurse–patient interactions. Secondary outcomes were the similarity of the data and the percentage of robot interactions completed autonomously. The questionnaires resulted in two values (robot and nurse) for three indexes of frailty, well-being and resilience. The data similarity was determined by comparing these index values using Bland-Altman plots, Cohen's kappa (κ) and intraclass correlation coefficients (ICC). Acceptability was assessed using questionnaires.

Results The mean robot interview duration was 16.57 min (SD=1.53 min), which was not significantly longer than the nurse interviews (14.92 min, SD=8.47 min; $p=0.19$). The three Bland-Altman plots showed moderate to substantial agreement between the frailty, well-being and resilience scores ($\kappa=0.61$, 0.50 and 0.45, and ICC=0.79, 0.86 and 0.66, respectively). The robot autonomously completed 39 of 42 interviews (92.8%).

Conclusion Social robots may effectively and acceptably assist HCPs by interviewing older adults.

INTRODUCTION

An important set of medical data consists of patient responses to medical questionnaires, such as patient-reported outcome measurements (PROM).^{1–3} PROM data

provide essential information about a patient's health status and the effectiveness of the delivered care.¹ A survey of nearly 100 000 clinical trials published between 2007 and 2013 found that PROMs were used in 27% of these trials⁴; however, interviewing an older patient for a PROM is a time-consuming administrative task for healthcare professionals (HCP), whose time is often very limited. This problem is further exacerbated by the increasing shortage of medical personnel⁵; therefore, patients are frequently asked to provide the data themselves using computers, tablets or smartphones.^{6,7} Many patients, in particular older patients, have difficulties using digital technology solutions because of their lack of digital literacy⁸ or their disabilities (eg, low vision).⁹ In cases where older patients are requested to complete forms via the internet, the non-response rate is high (74%) and increases with age.¹⁰

Social robots can be viewed as humanoid robots with which a person can interact like with another person.^{11,12} They are emerging as potential supporting technologies for HCPs, and their potential for involvement with patient data collection is currently under investigation.¹³ The use of social robots in the care of older patients has been widely investigated^{14–21}; however, to the best of our knowledge, their ability to independently conduct a health status questionnaire in a hospital setting has not yet been evaluated. Our study therefore adds to the scarce research on robot-assisted surveys.

Our hypothesis is that the social robot task time for autonomously conducting lengthy PROM surveys in older adults does not differ significantly from the task

time if an HCP conducts the survey (the current practice). We already showed proof of concept regarding the acceptability and effectiveness of social robots interacting with a group of older volunteers, but did not compare this with regular care.²² In this study, we aimed to test our hypothesis with community-dwelling older adults using a specifically designed robot-participant interaction programme on the Pepper robot.²³ This social robot has a friendly engaging appearance and a height of 1.2 m as preferred by older adults.²⁴ The voice recognition capability of Pepper is based on matching with a preprogrammed set of words. The robot further combines the recognition of a face in its camera image and the direction of the voice sound signal to turn its head to the person talking. We measured the task completion percentage without HCP intervention, the agreement between data obtained via the HCP and robot-conducted surveys, and compared the task duration and acceptability of these methods of data collection.

METHODS

Trial design

The experiment was designed as a non-blinded randomised controlled cross-over trial to compare data acquisition via robot (robot-participant: RP) and nurse (nurse-participant: NP) interactions with older participants. Each participant answered three questionnaires administered by the robot in one session and the same three by a nurse in another session. This within-subject cross-over design was selected to minimise variance not related to the signal of change and better detect differences in appreciation of the HCP and the robot. A 2-week washout period was used between sessions to minimise the learning effects. The 2-week period is a compromise between a longer washout, by which carry-over effects could be further reduced, and the increasing probability for intercurrent morbidity in these older subjects, which would limit comparability. Participants were randomly assigned by the researcher to two study groups using their sign-up dates, with one group encountering the nurse in the first session and the robot in the second session, and the other group encountering the reverse order of interviewers. This counterbalancing was applied to avoid learning and boredom effects.

Participants

Participants were recruited by newspaper advertisements or through local older adult organisations in the period from November 2017 to January 2018. The inclusion criteria were as follows: aged over 70, Dutch speaking, living independently and no cognitive disabilities.

Interaction design

The interaction design was focused on the patient's self-assessment of their current frailty, well-being

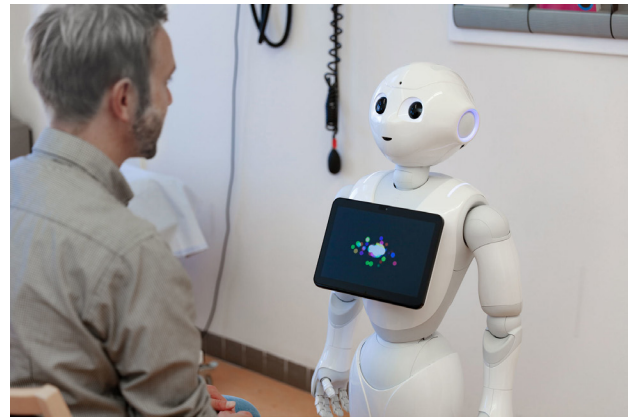


Figure 1 Person being interviewed by the Pepper robot.

and resilience in coping with illness. These assessments were performed using the TOPICS short form (TOPICS-SF) questionnaire,²⁵ the Personal Wellbeing Index (PWI)²⁶ and the Resilience Scale,²⁷ respectively.

Experimental procedure

During the RP session, the nurse welcomed the participant and accompanied them to the examination room with the robot (figure 1). The nurse and the participant sat opposite the robot, and the nurse explained that she had a new robotic assistant to help in her administrative tasks by verbally administering questionnaires. The participant received an instruction card explaining his dialogue options, which were also displayed with a large font size for easy readability on the robot's screen (online supplementary figure 1).²⁸ This allows the participant to think about the options independent of memory function, and select the most appropriate answer. After a short training dialogue for the participant, the nurse instructed them how to command the robot to start the actual RP interaction, and then left the room, leaving the participant alone with the robot. On the participant's start command, the robot began the interview with the questionnaires, asking for confirmation of each answer it registered. On interview completion the robot thanked the participant. This procedure is further detailed in online supplementary appendix 1.

The NP interaction procedure was comparable to the RP procedure, except that the nurse interviewed the participant, showed the questions and answer options on a paper form and noted the given answers.

Outcomes

The primary outcome measure was the time required for completion of the questionnaires in the RP and NP interactions. The secondary outcome measures were the data similarity, and the percentage of RP interactions completed autonomously (without HCP intervention). We also evaluated the opinion of the participants on the acceptability of using the robot technology for clinical interviews.

Sample size

Using G*power for a dependent t-test (two tailed) within subjects,²⁹ a sample size of 36 people was calculated to be required to detect a 0.5 effect on the efficiency (time) of PROM completion when the power was set at 0.90 and using an alpha of 0.10.

Data analysis

The RP answers were recorded electronically by the robot, while the NP answers were recorded on paper. All data were stored in the Castor data management system.³⁰ The data were analysed using SPSS statistical software (V.22; IBM) and Microsoft Excel (Office 365; Microsoft, Redmond, WA, USA).

The autonomous completion percentage was determined as the number of RP interactions without interrupting events, as a percentage of the total number of RP interactions. An interrupting event is defined as any HCP intervention necessary for further continuation of the interview by the robot, for example, because of a robot failure. The task duration of each interview was calculated as the difference in time between the first and last answers. In the RP interactions, the time was recorded electronically by the robot, and for the NP interaction the time was calculated from the interview recording.

The data similarity was calculated for three indexes. The Frailty Index (FI) was calculated from the 18-question TOPICS-SF,³¹ excluding any missing values. FI is used for the phenotypic categorisation of participants as Frail, Prefrail or Robust, where Prefrail was equivalent to two to five deficits reported in the TOPICS-SF questionnaire (giving an FI of 0.1–0.25).³² The overall PWI was calculated from the average scores on questions 2–8 of the PWI, which was converted to a value on a scale from 0 to 100, with higher scores reflecting higher well-being.²⁶ The PWI was categorised into three categories: low, medium and high, with cut-offs pragmatically defined as the overall mean \pm SD of all index values from PWI_{nurse} and PWI_{robot} . The Resilience Scale resulted in a Resilience Index (RI) between 25 and 100, where higher scores reflected higher resilience.²⁷ The RIs were converted using gender-specific norm values, then categorised into low, medium or high-resilience categories.

Each participant thus obtained six indexes: FI_{nurse} , PWI_{nurse} , RI_{nurse} , FI_{robot} , PWI_{robot} and RI_{robot} . Following the method of Bland and Altman,^{33 34} the agreement between the two assays for each index was analysed using scatter plots of the samples, where $S(x,y) = ((Index_{nurse} + Index_{robot})/2, Index_{nurse} - Index_{robot})$. The intraclass correlation coefficients (ICC) for the continuous measures (the indexes) were determined in SPSS using a two-way mixed model that analysed the absolute agreement between the robot and nurse measurements. Additionally, Cohen's kappa (κ) was calculated for the ordinal measures to analyse the inter-rater agreement between RP and NP. A $\kappa < 0$ was

characterised as no agreement, 0.00–0.20 as slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement and 0.81–1.00 as almost perfect agreement.³⁵ Indices were first measured as a discrete integer value within a finite interval, and next—for the purpose of classifying patients into groups—were converted into a categorical value, and therefore both ICC and κ were calculated.³⁶ Carry-over effects were determined by comparing the sequential results of both study groups.

Finally, the participants were asked to score the acceptability of the robot using Almere questionnaires.³⁷ These questionnaires assessed distinct properties of robot usability: attitude towards the robot, facilitating conditions, anxiety, perceived sociability, social influence, perceived ease of use, social presence, perceived enjoyment, trust and perceived usefulness (online supplementary table 1). Each construct was judged using a seven-point Likert scale. The answers were converted into a value on a 0–10 scale, with higher scores reflecting higher acceptability. Constructs consisting of more items were averaged, and negatively formulated items were reversed in the summation.

RESULTS

Participants

Forty-two people (45% female) participated in this study, all of whom lived independently in a Dutch city, were native Dutch speakers, were an average of 77.1 (SD=5.7) years old and on average completed their secondary education. Hearing aids were used by four participants (10%), and spectacles were used by 34 of them (83%). Forty-five per cent of our subjects had recent health service contacts (29% no contact, 26% unknown). Online supplementary figure 2 shows the participant flow diagram while online supplementary table 2 provides the participant demographics. All participants completed the allocated treatment, that is, interview sessions; therefore, the intention-to-treat and per-protocol analyses were identical.

Task durations and autonomy

Both task times were positively skewed, and a paired t-test was allowed.³⁸ Participants completed their RP interaction with a mean task duration of 16.57 min (SD=1.53 min), while the mean NP task duration was not significantly shorter (14.92 min, SD=8.47 min, $n=42$, $t=1.33$, $p=0.19$). Three of the 42 RP interactions required an interruption event by the nurse, once because of a technical failure and twice because of a participant start-up command failure. The autonomous completion percentage was therefore 92.8%. The Human-Participant (HP) interactions showed a 100% completion.

Frailty Index

The FI data agreement is shown by the Bland-Altman plot in figure 2. The FI differences showed a normal

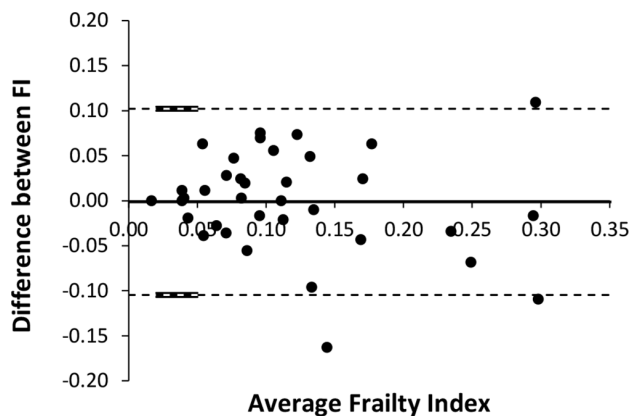


Figure 2 Bland-Altman (BA) plot of the differences in the Frailty Index (FI) values determined from the nurse–participant and robot–participant data. The solid line in this BA plot shows the mean difference between the two methods. The dashed lines are the limits within which 95% of the values of the difference per individual fall.

distribution. The mean FI difference between the data acquired during the RP and NP interactions was 0.001 (SD=0.053), and the lower and upper limits of agreement (LOA) were -0.105 and 0.102 , respectively (95% CI -0.107 to -0.102 , and 0.100 to 0.104 , respectively). A systematic error of 0.13% was observed in the FI differences, but no systematic trend was detected. No significant difference was found between the RP and NP FIs ($t=-0.16$, $df=41$, $p=0.87$), and the κ value was 0.61, indicating substantial agreement. The ICC was 0.79 (95% CI 0.65 to 0.88; $F(41,41)=8.49$, $p<0.001$). The carry-over effects were symmetric. The participants gave an average of 4.1 ± 2.4 different answers to the robot or nurse, where 85% differed by one grade on the item scale, 12% differed by two grades and 3% differed by three or more grades.

Personal Wellbeing Index

The PWI differences were normally distributed (figure 3). The PWI mean difference was 0.44

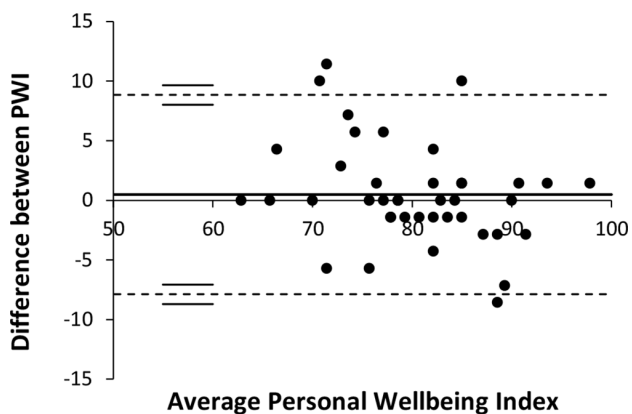


Figure 3 Bland-Altman (BA) plot of the differences in the Personal Wellbeing Index (PWI) values determined from the nurse–participant and robot–participant data. The solid line in this BA plot shows the mean difference between the two methods. The dashed lines are the limits within which 95% of the values of the difference per individual fall.

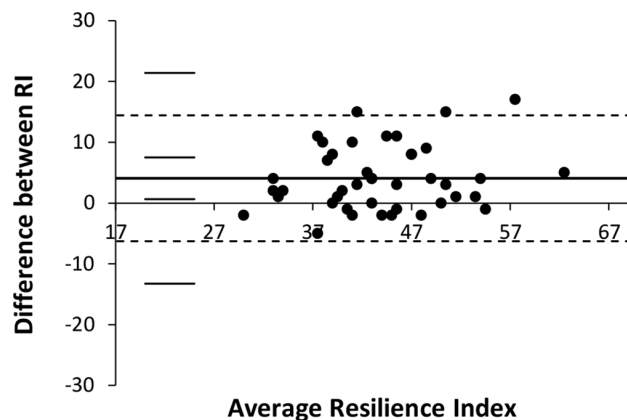


Figure 4 Bland-Altman (BA) plot of the differences in the Resilience Index (RI) values determined from the nurse–participant and robot–participant data. The solid line in this BA plot shows the mean difference between the two methods. The dashed lines are the limits within which 95% of the values of the difference per individual fall.

(SD=4.27), with lower and upper LOAs of -7.94 (95% CI -8.69 to -7.18) and 8.81 (95% CI 8.06 to 9.58), respectively. A systematic error $<0.5\%$ was observed, with no trend detected in the PWI differences. No significant difference was detected ($t=0.67$, $df=41$, $p=0.51$), and the κ value was 0.50, which indicated moderate agreement. The ICC was 0.86 (95% CI 0.76 to 0.92; $F(41,41)=13.53$, $p<0.001$). The carry-over effects were symmetric. The participants gave an average of 3.2 ± 1.9 different answers to the robot and nurse, 81% of which differed by one grade on the item scale, 12% differed by two grades and 7% differed by three or more grades.

Resilience Index

The RI Bland-Altman plot is presented in figure 4. The mean difference was 4.07 (SD=5.29) with lower and upper LOAs of -6.30 (95% CI -13.27 to 0.66) and 14.45 (95% CI 7.48 to 21.41), respectively. No trend was observed, but we did note a significant systematic difference ($t=4.99$, $df=41$, $p<0.01$): participant RIs scored by the nurse were higher than those scored by the robot. The κ value was 0.45, indicating moderate agreement between the NP and RP values. The ICC was 0.66 (95% CI 0.23 to 0.84; $F(41,41)=7.04$, $p<0.001$). The carry-over effects were asymmetric: for the nurse-then-robot group, RI_{nurse} was 44.0 and RI_{robot} was 41.4, showing a significant decrease in RI ($p=0.014$), while for the robot-then-nurse group, RI_{robot} was 42.3 and RI_{nurse} was 47.7, indicating a significant increase in RI ($p<0.001$). The participants gave an average of 7.9 ± 4.0 different answers to the robot and nurse, where 97% differed by one grade on the item scale, 2% differed by two grades and 1% differed by three grades.

Acceptability

The acceptability of using the robot was scored from the Almere model variables (online supplementary

table 3). The participants had generally positive feelings about the robot (mean 7.4, SD=1.7) and found it easy to use (mean 7.7, SD=1.0). The robot did not invoke anxiety among participants (mean 1.3, SD=1.4), instead causing joyful feelings when used (mean 7.3, SD=1.7).

Participants were invited to provide additional comments on the RP interaction. Similar comments made by four or more participants (10% of the study population or more) are summarised here. Eleven participants said that at some time during the interview the robot did not immediately understand them, but the available dialogue options enabled them to solve the problem. Five participants found the robot speech unclear. Four participants found certain questions to be inappropriate for interviews conducted by a robot. Four participants said they wanted to be able to further elucidate their answers.

DISCUSSION

Our findings suggest that social robots can autonomously and acceptably interview older adults and collect valid PROM data. The primary outcome showed that 93% of the RP interactions were autonomously completed, and provided reliable FI, PWI and RI data. This demonstrated that a social robot may assist HCPs in collecting PROM data from older adults, which may free more of the HCP's time for basic and specialised healthcare tasks.^{13 39} The questionnaires were completed an average of 1.63 min (SD=0.02 min) faster when performed by a nurse rather than the robot, with no significant difference. However, the large SD in the task time by the nurse condition was unexpected. The interview duration for five participants was more than 28 min, whereas for another five participants the interview duration was less than 9 min. This is caused by some participants elucidating on their answers (increasing interview time), or by the nurse handling some questions with some participants much quicker and not asking for confirmation (decreasing interview time).

The overall agreement between the outcomes of the frailty, resilience and well-being assessments was fair to good; for example, only one of the six subjects categorised as frail by the nurse was considered prefrail by the robot, and none were considered robust. It should be noted that inter-rater and repeated-measurement differences in the PROM data can appear due to (A) 'real' changes of person's frailty, personal well-being and resilience, and (B) the overall reliability of the measurement instrument. Based on an earlier study on the responsiveness and stability of the longer term (ie, 3–12 months) PROM ratings in a community-based older adult population, similar to the one we have tested, we presumed that in a 2-week period the PROM outcomes would not relevantly change.⁴⁰ However, we have not formally assessed this stability over 2 weeks, which is a limitation of our study. Only for resilience,

a small but significant difference appeared between the nurse and the robot 'rater', that is, the nurse acquired a higher resilience score than the robot. A possible reason might be that the type of questions on resilience nudged the participants to give a more positive answer to the nurse than to the robot, because patients often want to make a favourable impression to nursing staff due to the power imbalance in the nurse–patient relationships.⁴¹ Also, several participants judged the resilience scale too undifferentiated, which may have contributed to this behaviour. The reason for score difference is not likely to be caused only because of the technical nurse versus robot difference, as then we also expect to find a significant difference for the frailty and the well-being indexes. Future research is needed to investigate if this is a persistent effect and, if yes, how to interpret it.

The participant scores of Attitude towards the Robot, Perceived Ease of Use, and Perceived Enjoyment were in line with similar studies, such as the results reported by Briggs *et al* for their Parkinson's Disease Questionnaire-39.⁴² They teleoperated a Nao robot to conduct a health status survey with people with Parkinson's disease, with the researcher typing the questions subsequently spoken by the robot. The participant's verbal response was heard by the teleoperator, who directed the reaction of the robot. Briggs *et al* found that participants reacted positively to the robot overall; however, the robot served more as an intermediary for a human and less as the independent entity used in our study. Broadbent *et al*⁴³ reported that 75% of the participants in their study responded positively to the use of a robot in an interview situation. They used an iRobi robot to monitor patients with chronic obstructive pulmonary disease (COPD) at their rural homes over a 4-month period. The robot completed weekly clinical COPD questionnaires with the patient by verbally asking questions, answerable via a touch screen on the robot. This touch screen interaction is quite different from the voice-controlled interaction used in the present study.⁴⁴ During method definition, we considered including an additional comparison with tablet-based surveys but decided against this because of the significantly lower user acceptability in a similar population.^{12 45}

The results may be generalisable to older outpatients or patients visiting general practice, as 45% of our subjects had recent health service contacts. The results cannot yet be translated to older inpatient groups, who in general have more severe functional or cognitive limitations.

The strength of our study is that it is the first to use a social robot for PROM data collection among older adults in a clinical outpatient setting. We administered three well-validated and frequently used questionnaires in a novel way, as compared with the paper, computer, laptop or smartphone-based data collection methods commonly used. The voice-controlled interaction

facilitated a natural manner of communicating, and the design included enough dialogue options to enable the dialogue to be completed without intervention in the majority of cases.

Our study also had some limitations. First, there was some selection bias by our voluntary response recruitment, which resulted in relatively highly educated participants. Moreover, our respondents were not actual patients. Rerunning the study as a real-world application with frail older patients of different socio-economic status will be important for a generalisation of our study results. Second, the robot's audio sensing and language processing functions still have room for improvement. Third, patients could not be blinded to group allocation. The final limitation is that we did not study the evaluation of the robot-patient interaction as perceived by nurses and other HCPs. This should also be carried out before the wide-scale use of robot healthcare assistants.

Many people are concerned about robots taking over human jobs. For this reason, we introduced the robot as an assistant to the HCP, not as a replacement. The HCP remains in control of patient care, but can ask the participant to be interviewed by the robot assistant. For the HCP, having a robot assistant is a new but helpful experience, and it is important that they are involved in its development.¹³ Our findings should stimulate the further study of the interaction modes between patients, HCPs and social robots. Our study indicates that autonomously acquiring PROM data with a social robot among older adults is an acceptable procedure for this group. These insights are needed for future studies in which an integrated care pathway solution including a social robot can be realised, where patients are admitted to a clinic and interviewed and guided along their treatment path by robot healthcare assistants.

In conclusion, we have shown that the use of a social robot for conducting PROMs may be a valuable tool for HCPs and an acceptable interviewer for older patients. We recommend the further study of multi-modal PROM questioning and the handling of open dialogues by social robots in the interactions between older patients and their HCPs. The next step is to study the implementation of social robot assistants in clinical practice.

Funding This study was funded by the Radboud Universitair Medisch Centrum (Grant No: 4TU Human&Technology).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by the Institutional Review Board of the Radboudumc.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data are available upon request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others

to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- 1 National Quality Forum. *Patient reported outcomes (pros) in performance measurement*, 2013.
- 2 NHS Digital. *Patient Reported Outcome Measures (PROMs) in England - A guide to PROMs methodology*, 2017.
- 3 Canadian Institute for Health Information. *PROMs background document*, 2015.
- 4 Vodicka E, Kim K, Devine EB, *et al*. Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007-2013). *Contemp Clin Trials* 2015;43:1-9.
- 5 UWV (Employee Insurance Agency). *Health care Factsheet*, 2018.
- 6 Schick-Makaroff K, Molzahn A. Strategies to use tablet computers for collection of electronic Patient-reported outcomes. *Health Qual Life Outcomes* 2015;13.
- 7 Shah KN, Hofmann MR, Schwarzkopf R, *et al*. Patient-reported outcome measures: how do digital tablets stack up to paper forms? A randomized, controlled study. *Am J Orthop* 2016;45:E451-E457.
- 8 Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. *J Med Internet Res* 2014;16.
- 9 Smith A. *Older adults and technology use*. Pew Research Center, 2014.
- 10 Ebert JF, Huibers L, Christensen B, *et al*. Paper- or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. *J Med Internet Res* 2018;20.
- 11 Breazeal C. Toward sociable robots. *Rob Auton Syst* 2003;42:167-75.
- 12 Hegel F, Muhl C, Wrede B, *et al*. *Understanding Social Robots. In: 2009 Second International Conferences on Advances in Computer-Human Interactions*, 2009: 169-74.
- 13 Archibald MM, Barnard A. Futurism in nursing: technology, robotics and the fundamentals of care. *J Clin Nurs* 2018;27:2473-80.
- 14 Shishehgar M, Kerr D, Blake J. A systematic review of research into how robotic technology can help older people. *Smart Health* 2018;7-8:1-18.
- 15 Bemelmans R, Gelderblom GJ, Jonker P, *et al*. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *J Am Med Dir Assoc* 2012;13:114-20.
- 16 Broekens J, Heerink M, Rosendal H, *et al*. Assistive social robots in elderly care: a review. *Gerontechnology* 2009;8:94-103.
- 17 Dautenhahn K, Campbell A, Syrdal DS. Does anyone want to talk to me?: reflections on the use of assistance and companion robots in care homes. In: *Procs 4th Int Symposium on New Frontiers in Human-Robot Interaction. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*. AISB, 2015.
- 18 Kachouie R, Sedighadeli S, Khosla R, *et al*. Socially assistive robots in elderly care: a Mixed-Method systematic literature review. *Int J Hum Comput Interact* 2014;30:369-93.

- 19 Mordoch E, Osterreicher A, Guse L, *et al.* Use of social commitment robots in the care of elderly people with dementia: a literature review. *Maturitas* 2013;74:14–20.
- 20 Robinson H, MacDonald B, Broadbent E. The role of healthcare robots for older people at home: a review. *Int J Soc Robot* 2014;6:575–91.
- 21 Vandemeulebroucke T, de Casterlé BD, Gastmans C. How do older adults experience and perceive socially assistive robots in aged care: a systematic review of qualitative evidence. *Aging Ment Health* 2018;22:1–19.
- 22 Boumans R, van Meulen F, Hindriks K, *et al.* *Proof of concept of a social robot for patient reported outcome measurements in elderly persons*. Chicago, IL, USA: ACM, 2018.
- 23 Who Pepper. Softbank Robotics Pepper. Who is Pepper? Available: <https://www.softbankrobotics.com/emea/en/pepper> [Accessed 14 Feb 2019].
- 24 Broadbent E, Tamagawa R, Kerse N, *et al.* Retirement home staff and residents' preferences for healthcare robots. In: *Robot and human interactive communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on IEEE, 2009: 645–50*.
- 25 Lutomski JE, Baars MAE, Schalk BWB, *et al.* The development of the older persons and informal caregivers survey minimum dataset (TOPICS-MDS): a large-scale data sharing initiative. *PLoS One* 2013;8:e81673.
- 26 Van Beuningen J, de Jonge T. The personal wellbeing index: construct validity for the Netherlands. *Cent Bur Voor Stat Discuss Pap* 2011.
- 27 Wagnild GM, Young HM. Development and psychometric evaluation of the resilience scale. *J Nurs Meas* 1993;1.
- 28 Fisk AD, Rogers WA. *Handbook of Human Factors and the Older Adult*. San Diego, California: Academic Press, 1997.
- 29 Faul F, Erdfelder E, Buchner A, *et al.* Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 2009;41:1149–60.
- 30 Castor. Faster, smarter medical research - Castor EDC - eCRF, ePRO, CTMS, 2018. Available: <https://www.castoredc.com/> [Accessed 2 Jul 2018].
- 31 Searle SD, Mitnitski A, Gahbauer EA, *et al.* A standard procedure for creating a frailty index. *BMC Geriatr* 2008;8.
- 32 Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. *J Gerontol A Biol Sci Med Sci* 2007;62:738–43.
- 33 Giavarina D. Understanding Bland Altman analysis. *Biochem Med* 2015;25:141–51.
- 34 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60.
- 35 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- 36 Mandrekar JN. Measures of interrater agreement. *J Thorac Oncol* 2011;6:6–7.
- 37 Heerink M, Kröse B, Evers V, *et al.* Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. *Int J Soc Robot* 2010;2:361–75.
- 38 Sauro J, Lewis JR. *Quantifying the User Experience - Practical Statistics for User Research*. 2nd edition. Morgan Kaufman, 2016.
- 39 Kitson A, Conroy T, Kuluski K, *et al.* *Reclaiming and redefining the fundamentals of care: Nursing's response to meeting patients' basic human needs*, 2013.
- 40 Lutomski JE, Santoso AMM, Hofman CS, *et al.* Responsiveness of the full-length and short form of the older persons and informal caregivers survey. *J Am Med Dir Assoc* 2017;18:804–5.
- 41 Shattell M. Nurse-patient interaction: a review of the literature. *J Clin Nurs* 2004;13:714–22.
- 42 Briggs P, Scheutz M, Tickle-Degnen L. Are robots ready for administering health status surveys? First results from an hri study with subjects with parkinson's disease. In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015: 327–34.
- 43 Broadbent E, Garrett J, Jepsen N, *et al.* Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. *J Med Internet Res* 2018;20.
- 44 Burgan D. *Dialogue systems and dialogue management*. DST group Edinburgh Edinburgh SA Australia, 2016.
- 45 Mann JA, MacDonald BA, Kuo I-H, *et al.* People respond better to robots than computer tablets delivering healthcare instructions. *Comput Human Behav* 2015;43:112–7.