



GUARDIAN

The social robot companion to support homecare nurses

D4.1 Evaluation protocols / instrument report

Project acronym:	GUARDIAN
AAL JP project number:	aal-2019-6-120-CP
Deliverable Id :	D4.1
Deliverable Name :	D4.1 Evaluation protocols
Status :	V1.0 (will be iteratively updated throughout project phases)
Dissemination Level :	PUBLIC
Due date of deliverable :	M16 (April 2021)
Actual submission date :	April 30, 2021
Author(s):	Jessica Rochat (HUG), Alexandra Villaverde (HUG)
Lead partner for this deliverable :	HUG
Contributing partners :	VIL, CCARE, INRCA, UNIVPM, UNIGE, TU/e



VERSION HISTORY

Version	Authors	Date	Description
0.1	HUG (Jessica Rochat, Alexandra Villaverde)	08-12-2020	First draft
0.2	HUG (Jessica Rochat, Alexandra Villaverde)	20-01-2021	Second draft
0.3	HUG (Jessica Rochat, Alexandra Villaverde)	28-01-2021	Third draft
0.4	HUG (Alexandra Villaverde)	22-02-2021	Recruitment information and planned measurements
0.5	HUG (Alexandra Villaverde)	31-03-2021	Detailed evaluations points
0.6	HUG (Alexandra Villaverde)	22-04-2021	Detailed measurements
0.7	HUG (Alexandra Villaverde)	26-04-2021	General review, update due dates
0.8	CCARE (Judith de Koning)	29-04-2021	Partner's review
0.9	CCARE (Janna Alberts); VILANS (Dirk Lukkien)	30-04-2021	Partner's review
1.0	HUG (Alexandra Villaverde)	30-04-2021	Final version (will be iteratively updated for pre-alpha, alpha & beta)



TABLE OF CONTENTS

1	Executive summary	6
2	Introduction	8
3	Recruitment of end-users	8
	<i>3.1 End-user's identification</i>	<i>9</i>
	3.1.1 Primary	9
	3.1.2 Secondary	9
	3.1.3 Tertiary	9
	<i>3.2 Participation criteria</i>	<i>9</i>
	3.2.1 Formal caregiver inclusion	10
	3.2.2 Informal caregiver inclusion	10
	3.2.3 Senior inclusion	10
	3.2.4 Senior exclusion	10
	<i>3.3 Channels of recruitment</i>	<i>10</i>
	3.3.1 Switzerland (HUG)	10
	3.3.2 The Netherlands (VILANS; ZNWV)	10
	3.3.3 Italy (INRCA)	11
	<i>3.4 Informed consent and financial compensation</i>	<i>11</i>
	3.4.1 Switzerland (HUG)	11
	3.4.2 The Netherlands (VILANS)	11
	3.4.3 Italy (INRCA)	11
4	Testing phases overall	12
	<i>4.1 User involvement in WP4</i>	<i>12</i>
	<i>4.2 Global work plan</i>	<i>13</i>
5	Heuristic evaluation	14
	<i>5.1 Work plan and sample size</i>	<i>14</i>
	<i>5.2 Installation and material</i>	<i>14</i>
	<i>5.3 Heuristic evaluation methodology</i>	<i>14</i>
	5.3.1 Bastien & Scapin's heuristics	15
	5.3.2 Nielsen's Gravity scale	15
6	Usability testing	16
	<i>6.1 Work plan and sample size</i>	<i>16</i>
	<i>6.2 Installation and material</i>	<i>16</i>
	<i>6.3 Usability test methodology</i>	<i>17</i>
	6.3.1. Scenarios	17
	6.3.2 Debriefing and questionnaires	17
	6.3.3 Data analysis	17
	6.3.4 Testing environment	17
	6.3.5 Qualitative interviews	18
	<i>6.4 Usability test measurements</i>	<i>18</i>



6.4.1	Accessibility	18
6.4.1.1	IBM usability questionnaire	18
6.4.2	Desirability	18
6.4.2.1	Online questionnaire	19
6.4.3	Factors influencing user acceptance	19
6.4.3.1	POBA	19
6.4.4	General usability	19
6.4.4.1	SUS	19
6.4.5	Personalization	19
6.4.5.1	Online questionnaire	20
6.4.6	System and log data	20
6.4.6.1	Guardian Eco-system questionnaire	20
7	Alpha & Beta pilot test	20
7.1	<i>Work plan and sample size</i>	20
7.1.1	Pre-alpha pilot test	20
7.1.2	Alpha pilot test	21
7.1.3	Beta pilot test	22
7.2	<i>Installation and material</i>	23
7.2.1	Pre-alpha pilot test	23
7.2.2	Alpha pilot test	23
7.2.3	Beta pilot test	23
7.3	<i>Pilot test methodology</i>	24
7.3.1	Pre-alpha and alpha pilot test	24
7.3.1.1	Scenarios	24
7.3.1.2	Online questionnaires	24
7.3.1.3	Data analysis	24
7.3.1.4	Testing environment	24
7.3.1.5	Qualitative interviews	24
7.3.2	Beta pilot test	24
7.3.2.1	Responsible Innovation questionnaire	24
7.4	<i>Pilot test measurements</i>	24
7.4.1	Pre-alpha pilot test	25
7.4.2	Alpha and beta measurements	25
7.4.2.1	Care burden	25
7.4.2.1.1	EDIZ questionnaire	25
7.4.2.2	Ethical design	26
7.4.2.2.1	Online questionnaire	26
7.4.2.3	Impact of the technology	26
7.4.2.3.1	Nurses' questionnaire	26
7.4.2.4	Perceived health	26
7.4.2.4.1	SF12	26
7.4.2.5	Perceived stress/anxiety	26
7.4.2.5.1	GAD-7 questionnaire	26
7.4.2.6	Quality of life	26
7.4.2.6.1	EQ5D questionnaire	26
7.4.2.7	Self-efficacy	27
7.4.2.7.1	UTAUT questionnaire	27
7.4.2.8	Social connectedness	27
7.4.2.8.1	SCS questionnaire	27
7.4.2.9	Technology acceptance	27
7.4.2.9.1	TAM questionnaire	27



7.4.2.10	Willingness to pay	27
7.4.2.10.1	Online questionnaire	27
7.4.2.11	Workload	27
7.4.2.11.1	Reduced planned care hours	27
8	Conclusion	28
9	References	28



1 Executive summary

International projects such as GUARDIAN require extensive and accurate evaluations on the use and usability of the social robot and its prototypes, before any home implementation. Expert evaluations and usability tests will thus be conducted to improve all related interfaces by ensuring that the features and the design match the audience's needs and can, therefore, finally be tested at home. The aim is to maximize the robot's benefits and market fit as a useful companion for homecare and as an improving tool regarding the social connectedness and independency of frail seniors.

Then, to appraise the strength of adhesion, but also to rank the robot's weak and strong points, different pilot tests will be set up to present and introduce Misty II in the home of seniors needing home cares. Alpha and beta pilot tests will allow us to check back if the ideas, concepts, clickable mock-ups and prototypes are satisfactorily designed and if there are still elements left to be improved to satisfy the client. The goal of these evaluations is to take the insights of the end-users into account, corroborating each perspective of them to the whole set of prototypes. Alpha and beta pilot tests offer the best real conditions to analyse the interaction between man and machine. The natural environment of those tests offers a qualitative indicator to determine if Misty II has chances of being successful.

Evaluations' outcomes will conduct, step by step, to the final product development, taking significant social and ethical issues into account. The iterative evaluation process ensures that most of the issues have been solutioned and is supported by some principles of the responsible innovation. All evaluations are therefore endowed with ethical values and principles that are indispensable for the smooth running and success of the project.

Acronyms used in this deliverable

VIL	Vilans
CCARE	ConnectedCare Services B.V
SRS	Smartrobot.solutions
JEF	JEF S.r.l.
TU/e	Eindhoven University of Technology
UNIGE	University of Geneva
HUG	University hospitals of Geneva
UNIVPM	Università Politecnica della Marche
INCRA	National Institute of Health and Science on Aging
ZNWV	Zorggroep Noordwest-Veluwe
IT	Information technology
RI	Responsible Innovation
P1	Prototype 1
P2	Prototype 2
P3	Prototype 3



2 Introduction

GUARDIAN's social robot co-design has been based on the requirements identified in the previous interviews of seniors, informal and formal caregivers. The designs, interfaces, services and screens made in order to respond to end-users' needs and requirement, will be submitted to end-users' evaluations.

A whole set of different evaluations have then been considerate, to progressively check if the produced elements of Misty II meet the global expectations of the end users. Assessments' iterations on the prototypes created are then necessary to know whether the improvements gradually made are indeed responding to the needs and requirements identified earlier.

The focus of this deliverable is to explain in broad terms all the required evaluations such as heuristic tests, usability testing, pre-alpha testing and alpha and beta pilot tests, and to provide information about participants recruitment, evaluation deadlines, methodology and measurements for each kind of test.

The whole approach attempts to implement certain aspects of responsible innovation (RI), which aim to ensure good ethical evaluation proceedings. This basis will also be an added-value useful to support us in the prototype's improvement. RI enables to create responsible evaluations in order to progressively produce a final social robot that could be socially acceptable and not subject to the release of ethical detriment on the end-users' use.

3 Recruitment of end-users

This section will gather and summarize all information needed regarding the involvement of end-users taking part into GUARDIAN's quality and impact evaluations.

First, let's briefly recall all the steps that require participant recruitment to better situate and understand at which step, the heuristic evaluation, the usability testing, the pre-alpha testing and the different pilot tests, alpha and beta, will unroll in the project.

As partly shown in the table below, the GUARDIAN project follows from the start a unique iterative design, research and development methodology with three streams, focusing on: co-design & personalization, ethical & value-sensitive design and business modelling & cost-effectiveness.

Even though the table below shows co-design evaluations, this deliverable only specifies and will be focus on expert evaluation, usability testing, pre-alpha, alpha and beta pilot tests.

	Wireframe evaluation	Scenario & mock-up evaluation	Expert evaluation	2 rounds of ethical design & business modelling	Usability testing	Pre-alpha pilot test	Alpha pilot test	Beta pilot test
CH	9 end-users	6 end-users	1-2 end-users	9 end-users	10 end-users	3 end-users	12 end-users	30 end-users
IT	6 end-users	5 end-users	1-2 end-users	9 end-users	10 end-users	3 end-users	12 end-users	30 end-users



NL	25 end-users	10 end-users	1-2 end-users	9 end-users	10 end-users	3 end-users	12 end-users	30 end-users
Total	38 end-users	21 end-users	5 experts	27 end-users	30 end-users	9 end-users	36 end-users	90 end-users

Table 1. Overview of the number of end-users involved all along the project

3.1 End-user's identification

The GUARDIAN project aims to offer direct benefits related to assistive technologies, for three groups of end-users: the homecare nurses also called formal caregivers, the informal caregivers and the seniors. The whole target audience will be requisitioned and invited to take part in the planned tests at their own location which will therefore take place in Switzerland (HUG), Netherlands (ZNVV) and Italy (INRCA).

3.1.1 Primary

Primary refers to formal caregivers as doctors, nurses, medical members related to the frail senior's home care for whom the GUARDIAN social robot will be the *"eyes, ears and communication channel"*. Formal caregivers will, by this ideology, use the social robot to support individual treatments of frail clients thereby reducing their workload and increasing their access to important information in a timelier manner. GUARDIAN must then allow formal caregivers to spend less of their professional activity on time-consuming routine tasks/check-ups and provide them enhancement and work pleasure by reducing stress.

3.1.2 Secondary

Secondary refers to informal caregivers as relatives, often children, friends, neighbors, anyone who is daily supporting and carrying out tasks and responsibilities regarding to a frail senior's home care. Informal caregivers experience high levels of stress and mental-physical fatigue as they worry about the frail senior's health problems and have difficulties finding a work-life balance due to the additional care tasks. GUARDIAN will allow them to have a form of relay given that the social robot will keep them remotely informed on the frail senior's state of wellbeing. The GUARDIAN social robot can then be seen by informal caregivers as the *"eyes, ears and communication channel"* which relays important information when they are not around the senior.

3.1.3 Tertiary

Tertiary refers to frail seniors as men and women over 65 years old, with fragile health, requiring medical treatment support to remain independent whether on a preventive or curative way.

Frail seniors frequently experience loneliness as formal and informal caregivers cannot be present 24/7. GUARDIAN will aim to provide them a social companion that can take over tasks from caregivers like reminding them to take their medication or to have some lunch. The robot that we want to implement would then be a life support tool that should lengthen the independence of the senior.

3.2 Participation criteria

The inclusion criteria maintained so far are still valid. They are important to respect because they characterize and represent participants who fully correspond to the potential end-users of Misty II. To deviate from these inclusion and exclusion criteria is to run the risk of not meeting the needs analyzed.



3.2.1 Formal caregiver inclusion

- who provides home care
- living in [country of the study] or cross-border workers
- at least one-year experience
- at least 18 years old
- good written and oral comprehension of the language of the interview

3.2.2 Informal caregiver inclusion

- relatives or close friends of a senior receiving home care.
- living in [country of the study] or cross-border workers, supporting senior's life
- at least 18 years old
- good written and oral comprehension of the local language

3.2.3 Senior inclusion

- 65 years old or older
- receiving home care
- living in [country of the study] or cross-border
- good written and oral comprehension of the local language

3.2.4 Senior exclusion

- diagnosed mild cognitive impairment or dementia.

3.3 Channels of recruitment

To ensure that the right participants are included in the evaluations, each partner must have its own recruitment technique, and ensure that the inclusion criteria mentioned above are met. Several channels are available for this purpose and everyone is free to proceed in the easiest perceived way. Nevertheless, considering the pandemic situation, it is recommended to proceed with digital recruitment channels.

3.3.1 Switzerland (HUG)

Swiss participants will be recruited among several social networks such as *Facebook*, *Twitter* or *LinkedIn*. Among them, the official HUG website, which is one of the most consulted in Geneva, is requested because of the good visibility it offers to recruit end-users. Although this communication chain is normally effective, during the pandemic situation, priority was given to announcements related to Covid-19, thus reducing the visibility of the GUARDIAN project.

Other solutions have been sorted out to have a broadly and better impact on recruitments: use our personal contacts, call on participants from other projects, develop collaborations with associations or institutions related to the seniors' home care.

3.3.2 The Netherlands (VILANS; ZNWV)

In the Netherlands, the recruitment aspect will be handled by ZNWV and partly supported by Vilans, particularly in the flyer designing for the whole-set of evaluations. To facilitate the



recruitment process, an online awareness session will be held at ZNWV, which should bring the request for participation to the attention of a wider public.

To ensure the required number of participants, additional recruitment resources will be also deployed, the use of the newsletter being part of it. As the different evaluations take place at different times, the newsletter will allow the announcement to be disseminated in a timely manner.

3.3.3 Italy (INRCA)

Regarding the recruitment channel, Italian partners will use the contact list of their neurology unit to communicate with potential participants of previous projects and initiatives, as well as with other people who might be interested in contributing to the project evaluations.

This method makes possible to be in contact with clients met before and with whom a relationship of trust has been already established. It simplifies the generally difficult recruitment channels in a pandemic period as flawed as the current one.

3.4 Informed consent and financial compensation

A general consent form has been initially created and shared with all partners to allow each of them to update and translate it, in any case of need. The consent form contains general information about the project, the stage in which end-users' participant will be requested, with a quick description of the tasks to be completed. At the end of the document, the rights of the participants as well as the criteria for participation are specified. By dating and signing the document's last page participants give their consent to take part into the evaluation process.

Financial compensation will be offered to participants to thank them for their involvement. The amount of the financial compensation can vary regarding the time required and the country.

3.4.1 Switzerland (HUG)

For ease and to avoid contact during this Covid-19 situation, the consent form will be sent by email to participants. Before the planned evaluation, the consent form should be signed and returned to HUG by scan or simple photo, if they want to be involved in the study process.

For their contribution on the prototypes' evaluations each participant will receive money as a financial compensation. In the days following their intervention, they will receive a financial sheet by email that must be returned signed, to finally proceed to the payment due.

3.4.2 The Netherlands (VILANS)

Given the pandemic circumstances, the informed consent will have to be signed digitally. If the person is not able to provide a signature on the document, an email with their consent will be sufficient. The most important is to have written proof of their willingness to participate in evaluations, regardless of the form.

To thank the involved participants, the Netherlands offers a choice between two different financial compensations. They are free to choose between a voucher to use on internet or a voucher to purchase books (bol.com).

3.4.3 Italy (INRCA)

In view of the pandemic situation, the consent form will be sent digitally and returned the same way. Except for the description of the study's purpose and the expectation regarding participants, the consent form will remain significantly the same.



Regarding to compensation's aspects, partners from INRCA cannot give money to any participants involved in the prototype testing taking part in Italy. As a compensation, they will then offer free medical exams to participants who free their time to take part in the project.

4 Testing phases overall

4.1 User involvement in WP4

Step	Deadline	Netherlands	Switzerland	Italy	Total
Heuristic Evaluation (HE)	M17 (May 2021)	1-2 experts	1-2 experts	1-2 experts	5 experts

Table 2. User involvement for heuristic evaluation (HE)

Step	Deadline	Netherlands	Switzerland	Italy	Total
1 st usability test	M21 (September 2021)	3-4 formal carers	3-4 formal carers	3-4 formal carers	10 formal carers
		3-4 informal carers	3-4 informal carers	3-4 informal carers	10 informal carers
		3-4 seniors	3-4 seniors	3-4 seniors	10 seniors
Total		10 participants	10 participants	10 participants	30 participants

Table 3. User involvement for usability testing

Step	Deadline	Netherlands	Switzerland	Italy	Total
Pre-alpha pilot test	M23 (November 2022)	1 formal carer	1 formal carer	1 formal carer	3 formal carers
		1 informal carer	1 informal carer	1 informal carer	3 informal carers
		1 senior	1 senior	1 senior	3 seniors
Total		3 participants	3 participants	3 participants	9 participants

Table 4. User involvement for pre-alpha testing

Step	Deadline	Netherlands	Switzerland	Italy	Total
Alpha pilot test	M28 (April 2022)	4 formal carers	4 formal carers	4 formal carers	12 formal carers
		4 informal carers	4 informal carers	4 informal carers	12 informal carers
		4 seniors	4 seniors	4 seniors	12 seniors
Total		12 participants	12 participants	12 participants	36 participants

Table 5. User involvement for alpha pilot test

Step	Deadline	Netherlands	Switzerland	Italy	Total
Beta pilot test	M36 (December 2022)	10 formal carers	10 formal carers	10 formal carers	30 formal carers
		10 informal carers	10 informal carers	10 informal carers	30 informal carers

		10 seniors	10 seniors	10 seniors	30 seniors
Total		30 participants	30 participants	30 participants	90 participants

Table 6. User involvement for beta pilot test

4.2 Global work plan

Step	Sub step	Deadline	Responsible
Heuristic Evaluation (HE)	1 st Mock-up available	M15 (march 2021)	CCARE
	HE Protocol	M16 (April 2021)	HUG - VILANS
	Running of HE	M17 (May 2021)	HUG – ZNWV - VILANS - INRCA
	Sharing results	M17 (May 2021)	HUG – VILANS - INRCA
	Feedback report	M17 (May 2021)	VILANS
1 st usability test	1 st prototype integration	M17 (May 2021)	CCARE
	Protocol and template	M17 (May 2021)	VILANS
	Participant's recruitment	M17 (May 2021)	HUG – ZNWV- INRCA
	Running of usability test	M18-M19 (June and July 2021)	HUG – ZNWV - VILANS - INRCA
	Sharing results	M20 (August2021)	HUG – VILANS- INRCA
	Feedback report	M21 (September 2021)	VILANS
Pre-alpha pilot test	Protocol for the pre-alpha pilot test + update of D4.1	M21 (September 2021)	HUG
	Participant's recruitment	M21 (September 2021)	HUG – ZNWV- INRCA
	Running of the pre-alpha	M22 (October 2021)	HUG – ZNWV - VILANS- INRCA
	Sharing results	M23 (November 2021)	HUG – VILANS- INRCA
	Feedback report	M23 (November 2021)	HUG
Alpha pilot test	Protocol for the alpha pilot test + update of D4.1	M24 (December 2021)	HUG
	Ethics committee approvals (if necessary)	M24 (December 2021)	HUG
	Recruitment of participants	M24 (December 2021)	HUG – ZNWV – VILANS - INRCA
	Running of the alpha pilot test (3 months)	M27 (March 2022)	HUG – ZNWV - VILANS - INRCA
	Analysis and sharing of the results	M27 (March 2022)	HUG – ZNWV - VILANS - INRCA
	D4.3 Alpha pilot test report	M28 (April 2022)	HUG
Beta pilot test	Protocol for the beta pilot test + update of D4.1	M29 (May 2022)	HUG



	Ethics committee approvals (if necessary)	M29 (May 2022)	HUG
	Recruitment of participants	M29 (May 2022)	HUG – ZNWV – VILANS - INRCA
	Running of the alpha pilot test (6 months)	M35 (November 2022)	HUG – ZNWV - VILANS - INRCA
	Analysis and sharing of the results	M35 (November 2022)	HUG – VILANS - INRCA
	D4.4 Beta pilot test report	M36 (December 2022)	HUG

Table 7. User involvement overview

5 Heuristic evaluation

Heuristic evaluation refers to the “Task 2.4 Expert and end-user mock-up evaluation” which could not be carried out in time and is therefore dealt with and merged in this evaluation deliverable. This task will be performed from M16 (April 2021) to M17 (May 2021) before moving on to usability testing. The task leader is Vilans and the participants are: HUG, UNIGE, INRCA, TU/e.

Heuristic evaluations are often required to anticipate use problems and to improve the prototype that will be presented in the following steps. Furthermore, the experts’ vision and opinion on the project, either in general or on some specific points, is a plus for the progress of GUARDIAN.

5.1 Work plan and sample size

Step	Deadline	Responsible
Set up protocol	M16 (April 2021)	HUG - VILANS
Running of HE	M17 (May 2021)	HUG – ZNWV - VILANS - INRCA
Analysis and sharing of the results	M17 (May 2021)	HUG – VILANS - INRCA
D4.2 Guardian usability report	M17 (May 2021)	VILANS

Table 8. Workplan for heuristic evaluation

5.2 Installation and material

As this kind of evaluation comes into consideration rather quickly in the development of the project, we can only apply it to the first prototype (P1) and on the interfaces that are currently available. The heuristic evaluation will then go only through two services, *the Guardian Caregiver Application* and *the Senior’s tablet application*, each with several screens to analyze.

5.3 Heuristic evaluation methodology

The experts requested will have to navigate through the different screens of the services, by following a specific order of tasks to be done. The tasks will be relatively easy, such as creating an account, login and navigate on the website, with some more specific requests as: creating medication reminders, adding a senior profile, change an appointment, etc.

Carrying out those tasks, experts will progressively detect usability problems that might arise to the end-users and will classify them in the table created specifically for the heuristic evaluation. This table aims to rank all use-related problems encountered in the manoeuvre of both services. Bastien &



Scapin's heuristics will quantify the problems in terms of numbers. An indication of quality has always some benefits, that's why, to determine and rate the importance of the problems underlined by all experts, the Nielsen's gravity scale will be also set up on the heuristics table.

It is therefore through these two means that experts will suggest improvement solutions for the problems identified.

5.3.1 Bastien & Scapin's heuristics

The 8 heuristic principles from Bastien & Scapin, listed below, are essential when making user interfaces UI evaluations. All interfaces and services need to be tested on *the guidance, the workload, the explicit control, the adaptability, the error management, the consistency, the significance of codes* and on *the compatibility*; to determine the satisfaction's level of the product built so far.

These criteria will allow a first scan of use related problems. To offer a wider appreciation of the criteria of Bastien and Scapin, here are some brief descriptions:

1. Guidance evaluates the means deployed to orient and inform users throughout their interactions with the interface. It can involve prompting, grouping, immediate feedback and legibility.
2. Workload concerns elements of the interface reducing the perceptual or cognitive load of users, increasing then the dialogue efficiency. It can involve information density and brevity regarding to concision and minimal actions.
3. Explicit control must provide references on the system processing of explicit user actions, and type of control that users have on the processing of their actions by the system.
4. Adaptability is the capacity to behave accordingly to the users' needs and preferences by analyzing the flexibility and the user experience.
5. Error management refers to the means available to prevent, reduce and recover from errors or invalid data entry. It takes account of the error protection, the quality of error messages and the error correction.
6. Consistency is about the interface design choices that have been implemented and their homogeneity. The design of the information must remain consistent from one channel to another.
7. Significance of codes is about the adequacy between the object or information displayed or entered, and its referent. Codes are significant to the users when there is a strong semantic relationship between the codes and the items they refer to.
8. Compatibility concerns the coherence between environments and applications. It refers to the match between users' characteristics and task characteristics on one hand, and the organization of dialogue for a given application, on the other hand.

5.3.2 Nielsen's Gravity scale

The Nielsen's Gravity scale will be useful for experts' review because it will simplify the ranking of all the use related problems detected that will be classified into priorities. If the score is 0, there is no usability problem; if the score is 1, there is just a cosmetic problem easily to fix; score 2 means that a minor usability problem has been detected which fixing is on a low priority; score 3 means that a major usability problem has emerged and has a high fixing priority. If the score



reaches 4, there is a usability catastrophe that will be needed to be fixed before the releasing otherwise the use of the product is highly questionable.

6 Usability testing

Usability testing refers to the “Task 4.2 Usability testing of Guardian”. This task will be performed from M16 (April 2021) to M21 (June 2021). The task leader is Vilans and the participants are: HUG, UNIGE, INRCA, TU/e.

For the usability testing of Guardian, 30 end-users from the end-user organizations (UNIGE/HUG, INRCA & ZNWV/Vilans) will participate in usability testing to find out whether the interaction with, and mediation of the system is satisfactory. These evaluations will take place in controlled environments and aim at identifying any usability problems, collect qualitative and quantitative data, and determine the participant's satisfaction with the product.

Special attention will be given to test the Human-Robot Interaction with older people, and we will also pilot the robot and platform for one day in the homes of the seniors and let them keep a diary to gain insight into the experiences in respect to acceptance and usability. The usability testing will be repeated until the tool reaches a sufficient maturity.

6.1 Work plan and sample size

Step	Deadline	Responsible
Set up protocol for the 1 st usability test	M17 (May 2021)	VILANS
Running of the 1 st usability test	M18-19 (June-July 2021)	HUG – ZNWV- VILANS - INRCA
Analysis and sharing of the results	M20 (August 2021)	HUG – VILANS- INRCA
D4.2 Guardian usability report	M21 (September 2021)	VILANS

Table 9. Workplan for usability testing

Step	Deadline	Netherlands	Switzerland	Italy	Total
1 st usability test	M21 (September 2021)	10 participants	10 participants	10 participants	30 participants

Table 10. Sample size for usability testing

6.2 Installation and material

Usability testing will take over the two services (Caregiver Application and Senior Application) and related interfaces and screens of P1 that have been improved based on the Heuristic Evaluation. Participants will have tasks to complete via the interfaces in order to go through the different possible functionalities and will be finally submitted to questionnaires.

The two screenshots below show how the issue of medical treatment appears and is deployed on the different services. As the pictures show, the caregivers can set a medicine to be taken, while the seniors will report it.

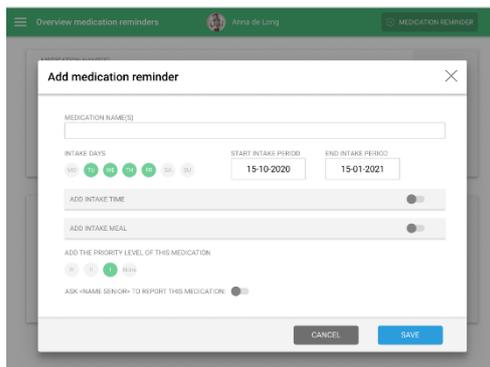


Image 1. Caregiver Application.
Add medication reminder

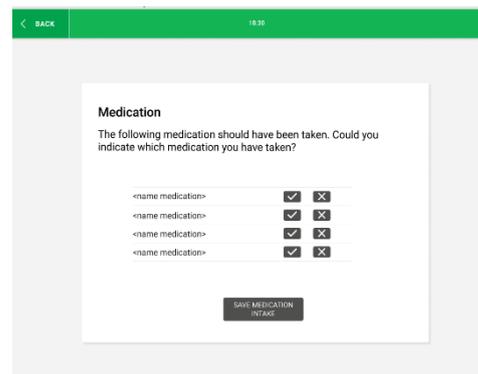


Image 2. Senior Application.
Add medication reminder

As the equipment is in its early stages and the health measures do not guarantee that all the project partners will be able to deploy the test in person, the user tests will be done remotely.

6.3 Usability test methodology

This chapter will define the methodology deployed on the formative usability test which will be performed to have a continuously and repeatedly evaluation on the first prototype.

6.3.1. Scenarios

Scenarios are a classic step in usability testing as they can give a visual idea of the project's value to our participants and allow to awaken deeper reflections that are appropriate to develop projects such as this one. GUARDIAN's scenarios allow to illustrate the benefits of using the social robot, but also gave the opportunity to show several ranges of possibilities and functionalities intended to make it useful. Scenarios will thus be walked through by the participants to bring them a more qualitative access to the robot possibilities.

The scenario's protocol will be created by CCARE and will be sent to all project's partner having to pass those tests, in order to ensure that the same methodology is used.

6.3.2 Debriefing and questionnaires

Methodologically speaking, the scenarios will not be enough. They will be necessarily followed by a debriefing session and some questionnaires. The aim is to collect information on the use of the interfaces and get an idea of the overall satisfaction of the users. Questionnaires will allow end-users to put into perspective their user experience. Through questionnaires we expect better and specific information on the use issues, to improve P1, and prevent other prototypes from mistakes that can already be corrected.

6.3.3 Data analysis

The debriefing and the questionnaire's result will be used as a basis to define main categories, where data will be clustered. All the information will be recorded on screen videos and transcript as data to analyze. The data analysis will also each country the opportunity to make a qualitative analysis on the problems encountered by the end-users involved in the assessment. The main conclusions will finally lead to tam improvement of prototype 1.

6.3.4 Testing environment

To have a proper and relevant usability test, participants should take part and be assessed in a controlled environment, e.g., a testing room or a dedicated space with all infrastructures



and details like real life's context. In our case, usability testing will be done at the end-user's home, on a remote way. Doing the test at home ensures that the real environment is respected.

6.3.5 Qualitative interviews

As GUARDIAN's development is in his early stages, interviews will be set up after improvement of the prototype 1, in order to check if the user experience problems identified during usability testing have been resolved or not. The aim is then to get more rich data and qualitative feedbacks, to know what end-users like and dislike, for which reasons and what would they like to improve. These interviews are necessary to ensure the smooth running of the project, as they allow design improvement and an additional insurance to ensure compliance with Responsible Innovation.

The methodology involved considers measurements such as: accessibility, desirability, factors influencing user acceptance, general usability, personalization and system & log data. All these elements are presented in the next section:

6.4 Usability test measurements

Measurement	Evaluation's tool
Accessibility	IBM usability questionnaire
Desirability	Online questionnaires
Factors influencing user acceptance	POBA's
General usability	The System Usability Scale (SUS)
Personalization	Online questionnaires
System & log data	GUARDIAN eco-system

Table 11. Usability testing's measurements and questionnaires related overview

6.4.1 Accessibility

Guardian has to remain accessible regarding end-users who must be able to do the actions wanted in due time and to benefit from a tool that grant them to meet and satisfy their social or professional needs. Evaluating accessibility means ensuring a system which corresponds to ethical values. Indeed, working on good accessibility is essential to any project, as it reflects the fact that each end-user has a fair access to a good user experience.

6.4.1.1 IBM usability questionnaire

Accessibility will be assessed through the IBM usability satisfaction questionnaire which is used to understand the perceived usability of GUARDIAN's social robot. This questionnaire uses a 7-point response scale, which is divided between the feeling of agreeing or disagreeing strongly. Each scale point has an assigned value, it is therefore possible to settle a numerical value that assess accessibility based on the fact that the lower the score, the better the usability.

6.4.2 Desirability

Desirability is important to ensure that a social robot, such as Misty II, represents a current expectation and is relevant and/or requested in the market place. It is not only the usefulness of the robot that is



evaluated by this measure, but also its design appearance and the functionalities that might appeal to users and create or increase a desirability phenomenon.

6.4.2.1 Online questionnaire

To gather desirability appreciation, an online questionnaire will be setup to have access to the end-users' subjective perceptions on the willingness to interact with the first prototype, and potentially the other two prototypes. Online questions will then be based on the emotional design aspect of GUARDIAN to record data on the end-users' feelings regarding the whole design and usefulness. Knowing whether or not the robot is desirable to end-users, allows us to assess the likelihood of Misty's adoption.

6.4.3 Factors influencing user acceptance

Performance and effort expectancy, social influence and facilitating conditions are the four factors that might influence user acceptance. Regarding this measurement, GUARDIAN's social robot has to offer higher performances to improve the quality of the senior's home treatment follow-up. It must be easy to set up and used by providing a less-effort and simple system. Social influence is also a factor to be analysed in order to know if social and general opinion can have a weight on the decision to use or not a social such as ours. In summary, environment's enablers or barriers must be highlighted to understand the perception of ease (or difficulty) to perform tasks.

6.4.3.1 POBA

POBA is the acronym of Piloting an Outcome Based Approach. It will be used to analyze the providing environment where P1 is involved. POBAs' aims then to gather and summarize the environment's outcomes discerned by end-users, on the three geographical locations. The results of this outcome-based approach will inform us on practical barriers, issues and opportunities related to the user behavior.

6.4.4 General usability

It is important to assess whether the final participants are globally satisfied, able to use easily the new device in order to improve their daily and work lives. This measurement looks quite similar to the accessibility but has a particular interest in others usability aspects regarding designs created so far, such as effectiveness, efficiency, satisfaction, engagement, error tolerance and ease of learning of Misty II.

6.4.4.1 SUS

SUS is the acronym of System Usability Scale, the most used instrument to assess general usability of a product like GUARDIAN's social robot. It spreads through 10 questions, including the possibility of responding with a Likert Scale of 5 rank points going between 1, *strongly disagree* to 5, *strongly agree*. The calculation of the points awarded will give a numerical value which will evaluate if the usability's quality is sufficient or not.

6.4.5 Personalization

Allowing a system like Misty II to be customized and personalized ensures, in an ethical way, that everyone can adapt it to their own usage needs. This aspect is really important in projects where seniors with special needs are involved. The possibility of having a personalized system, undoubtedly shows the tool as being much more attractive to end users.



6.4.5.1 Online questionnaire

A small questionnaire will bring out end-users' general perception on the preset customization system possibilities whether they are rather useful and/or sufficient. Questions will go on the satisfaction regarding the functionalities that can be personalized, evaluating if they are suitable, either quantitatively or qualitatively.

6.4.6 System and log data

Data on the log and the system must be directional and therefore able to guide the users, or even simplify their connection or browsing. The aim here is to ensure that everyone can log in intuitively and easily and that they can browse all the interfaces by finding exactly and quickly what they want.

6.4.6.1 Guardian Eco-system questionnaire

The purpose of this questionnaire is to ensure that the guardian ecosystem is nurtured by the social robot. It's then matter to assess if all end users are unified by the use of a robot-based system offering simplified information flows or if some technical improvements have to be done in order to support the end-users' relationship.

7 Alpha & Beta pilot test

The methodology for the research and piloting consists of iterative co-design and in-situ Alpha and Beta testing.

7.1 Work plan and sample size

7.1.1 Pre-alpha pilot test

Pre-alpha pilot test refers to the "Task 4.3 Alpha pilot test". This task will be performed from M21 (September 2021) to M23 (November 2022). The task leader is HUG and the participants are: Vilans, INRCA, TU/e. This test-step was added to ensure that P1 is now, following the modifications resulting from usability testing, functional, satisfactory and therefore ready for home use in a long term.

This test is therefore above all a guarantee of safety, enabling good starting-process for the following test, which is deployed over three months. In this pilot test we will enroll 9 end-users and it will focus on –among others – usability, acceptability and desirability of GUARDIAN. Pre-alpha pilots will run triads of senior clients and their formal and informal caregivers. These pilots will run in sequence at the University Hospital of Geneva in Switzerland, Italy at INRCA and in the Netherlands at ZNWV and Vilans, with n = 9 for each pilot site (meaning 1 triad with senior client and their formal caregiver and informal caregiver).

Activities related to this pilot test include the pilot test protocol creation, the recruitment of participants, and the installation of experimental setting; follow up of the pilot test. The measurements will run for at least 1 month and for efficiency reasons – next to interviews – mainly questionnaires will be used and automated.

Step	Deadline	Responsible
Protocol for the pre-alpha beta test + update of D4.1	M21 (September 2021)	HUG
Participant's recruitment	M21 (September 2021)	HUG - ZNWV - INRCA
Running of the pre-alpha	M22 (October 2021)	HUG - ZNWV - VILANS - INRCA
Sharing results	M23 (November 2021)	HUG – VILANS - INRCA



Feedback report	M23 (November 2021)	HUG
-----------------	---------------------	-----

Table 12. Workplan for pre-alpha pilot test

Each country partner has to recruit 3 participants to take part into the pre-alpha pilot test, which includes 1 senior, 1 informal caregiver and 1 formal caregiver. The Netherlands (ZNVV- Vilans), Italy (INRCA) and Switzerland (HUG) should, each of them, receive 1 robot in order to have one robot per triad. If we combine the participants from all partners, we get a total of 9 participants for the whole alpha pilot test and 3 robots to provide.

Step	Deadline	Netherlands (ZNVV and Vilans)	Switzerland (HUG)	Italy (INRCA)	Total
Pre-alpha pilot test	M23 (November 2021)	1 triad with 1 senior client, his formal caregiver and his informal caregiver	1 triad with 1 senior client, his formal caregiver and his informal caregiver	1 triad with 1 senior client, his formal caregiver and his informal caregiver	
Total		3 end-users	3 end-users	3 end-users	9 end-users

Table 13. Sample size for pre-alpha pilot test

7.1.2 Alpha pilot test

Alpha pilot test refers to the “Task 4.3 Alpha pilot test”. This task will be performed from M24 (December 2021) to M28 (April 2022). The task leader is HUG and the participants are: Vilans, INRCA, TU/e.

Once the product is sufficiently mature, formative evaluations will be organized in Switzerland, Italy, and the Netherlands. In this pilot test we will enroll 36 end-users and it will focus on – among others - usability, acceptability and desirability of GUARDIAN. Alpha pilots will run with triads of senior clients and their formal and informal caregiver. These pilots will run in sequence at the University Hospital of Geneva in Switzerland, Italy at INRCA and in the Netherlands at ZNVV and Vilans, with n = 12 for each pilot site (meaning 4 triads with senior client and their formal caregiver and informal caregiver).

Activities related to this pilot test include the pilot test protocol creation, the recruitment of participants, and the installation of experimental setting; follow up of the pilot test. The measurements will run for at least 3 months and for efficiency reasons – next to interviews – mainly questionnaires will be used and automated.

Step	Deadline	Responsible
Protocol for the alpha pilot test + update of D4.1	M24 (December 2021)	HUG
Ethics committee approvals (if necessary)	M24 (December 2021)	HUG
Recruitment of participants	M24 (December 2021)	HUG – ZNVV – VILANS - INRCA
Running of the alpha pilot test (3 months)	M27 (March 2022)	HUG – ZNVV - VILANS - INRCA



Analysis and sharing of the results	M27 (March 2022)	HUG – VILANS - INRCA
D4.3 Alpha pilot test report	M28 (April 2022)	HUG

Table 14. Workplan for alpha pilot test

Each country partner has to recruit 12 participants to take part into the alpha pilot test, which includes 4 seniors, 4 informal caregivers and 4 formal caregivers. The Netherlands (Vilans), Italy (INRCA) and Switzerland (HUG) should, each of them, receive 4 robots in order to have one robot per triad. If we combine the participants from all partners, we get a total of 36 participants for the whole alpha pilot test and 12 robots to provide.

Step	Deadline	Netherlands (ZNVV and Vilans)	Switzerland (HUG)	Italy (INRCA)	Total
Alpha pilot test	M28 (April 2022)	4 triads with 1 senior client, his formal caregiver and his informal caregiver	4 triads with 1 senior client, his formal caregiver and his informal caregiver	4 triads with 1 senior client, his formal caregiver and his informal caregiver	
Total		12 end-users	12 end-users	12 end-users	36 end-users

Table 15. Sample size for alpha pilot test

7.1.3 Beta pilot test

Beta pilot test refers to the “Task 4.4 Beta pilot test”. This task will be performed from M29 (Mai 2022) to M36 (December 2022). The task leader is HUG and the participants are Vilans and INRCA.

In the final evaluation stage, the platform will be evaluated (summative evaluation) with the triad of senior client and formal caregiver and informal caregiver at three locations: (1) at the University Hospital of Geneva in Switzerland, (2) in Italy, and (3) in the Netherlands. For each of these three pilot sites, n = 30, meaning 10 triads with senior client and their formal caregiver and informal caregiver will be involved per country. We will evaluate the GUARDIAN platform in the context of healthcare professional taking care of several clients at home, with the possible involvement of the informal caregiver, and with an additional focus on ethical design and cost-effectiveness. Activities related to this task include recruitment of participants, installation of experimental setting; follow up of the pilot. The pilot test will run for 6 months with a focus on quantitative measurements, and again questionnaires will be automated for efficiency reasons, i.e., online (T0, T1, T2, and T3). Questionnaires will measure - among others - quality of life (EQ-5D), self-efficacy, workload (see, 2.1).

Step	Deadline	Responsible
Protocol for the beta pilot test + update of D4.1	M29 (May 2022)	HUG
Ethics committee approvals (if necessary)	M29 (May 2022)	HUG
Recruitment of participants	M29 (May 2022)	HUG – ZNVV – VILANS - INRCA
Running of the alpha pilot test (6 months)	M35 (November 2022)	HUG – ZNVV - VILANS - INRCA
Analysis and sharing of the results	M35 (November 2022)	HUG – VILANS - INRCA



D4.4 Beta pilot test report	M36 (December 2022)	HUG
-----------------------------	---------------------	-----

Table 16. Workplan for beta pilot test

Each country partner has to recruit 30 participants to take part into the alpha pilot test, which includes 10 seniors, 10 informal caregivers and 10 formal caregivers. The Netherlands (Vilans), Italy (INRCA) and Switzerland (HUG) should, each of them, receive 10 robots in order to have one robot per triad. If we combine the participants from all partners, we get a total of 90 participants for the whole alpha pilot test and 30 robots to provide.

Step	Deadline	Netherlands (ZNVV and Vilans)	Switzerland (HUG)	Italy (INRCA)	Total
Beta pilot test	M36 (December 2022)	10 triads with 1 senior client, his formal caregiver and his informal caregiver	10 triads with 1 senior client, his formal caregiver and his informal caregiver	10 triads with 1 senior client, his formal caregiver and his informal caregiver	
Total		30 end-users	30 end-users	30 end-users	90 end-users

Table 17. Sample size for beta pilot test

7.2 Installation and material

7.2.1 Pre-alpha pilot test

The alpha pilot test will be introduced by a pre-test whose aim is to check the improvements made into the unidirectional prototype (P1), since the obtention of the usability testing results. The pre-alpha test will then strengthen the first prototype which was exploratory up to then.

The aim is not only to carry out an iterative evaluation but also to develop more targeted and relevant questionnaires for the following pilot tests and prototypes evaluations.

7.2.2 Alpha pilot test

The alpha pilot test will be more substantial than the pre-alpha, as it will take six months. To ensure that the installation at the senior's home is secure and relevant, it is necessary to have a mature product. We will therefore test the prototype 2 (P2), which no longer offers a unidirectional system but a bi-directional one.

P2 can be seen as an improved version of P1, as all useful data from previous evaluations will be reused on alpha pilot test. The difference lies in the wider range of options and functionalities, which it is intended to be discovered by end-users and evaluate too.

7.2.3 Beta pilot test

The beta pilot test will be significant as it will last about 6 months. The time required has increased, but so has the quality of the product which will be tested at home, known as the third prototype (P3). We are then referring to the final and complete social-robot, the one that will be marketed on completion of all tests.

P3 can be seen as an improved version of P2, as all useful data from previous evaluations will be reused on beta pilot test. The difference lies in a more tangible access to the end-users' willingness to pay, and the gathering of more data related to the attractiveness of the product.



7.3 Pilot test methodology

7.3.1 Pre-alpha and alpha pilot test

7.3.1.1 Scenarios

The scenarios created must reflect our ideal expectations regarding the use of Misty II and take into account the perspectives of the end-users regarding its functionality and empowerment ability. The aim is to illustrate scenarios where the social robot is used responsibly respecting moral and ethical standards. P2 and P3 have to be scenario-based.

7.3.1.2 Online questionnaires

The online questionnaires' methodology will be requested for the evaluation of ethical design, cost-effectiveness, quality of life (EQ-5D), self-efficacy, workload, desirability and the impact of the technology with online questionnaires. Pre-alpha, alpha and beta pilot test will be explored and measured with mainly questionnaires which will be filled out by the participants.

7.3.1.3 Data analysis

The data collected must be analyzed as accurately as possible, regarding RI and its ethical values. The most sensitive information should be anonymized, as in any research evaluation.

7.3.1.4 Testing environment

Pilot test are used to be performed in the natural environment of the end-users, i.e., in the senior's home. P2 and P3 being already more elaborate than P1, they can be tested, one by one, in a real context to collect and analyze use-related problems.

7.3.1.5 Qualitative interviews

Interviews will be set up after each evaluation of the prototypes (either P1, P2, P3), in order to check if the user experience problems identified have all been resolved or not. The aim is to continuously get more rich data and qualitative feedbacks, to know what end-users like and dislike, for which reasons and what would they like to improve. These interviews are necessary to ensure the smooth running of the project, as they allow design improvement and an additional insurance to ensure compliance with Responsible Innovation.

7.3.2 Beta pilot test

Methodology from alpha pilot test is also used into the beta pilot test, it will then include scenarios, online questionnaires, data analysis and testing environment. Nevertheless, beta pilot test will adopt a particular methodology particularly based on responsible innovation.

7.3.2.1 Responsible Innovation questionnaire

Set up a methodology based on responsible innovation is important when it comes to projects involving artificial intelligence and people privacy. RI will ensure that seniors and caregivers, either formal or informal, stay in control of the robot. Actions on the robot must be transparent, no information must be hidden from the main end-users: the seniors.

7.4 Pilot test measurements

All measurements requested for the usability test (accessibility, desirability, factors influencing user acceptance, general usability, personalization, system and log data) will be requested again in the whole set of pilot testing (pre-alpha, alpha, beta test). The iteration of the measuring instruments as



online questionnaires, allows us to grasp all the problems related to usability, acceptability and desirability on the project's continuum. The system & log data, the personalization and the factors influencing user acceptance will also be call for to collect additional data useful for the appreciation of the work established so far on the prototypes.

7.4.1 Pre-alpha pilot test

The pre-alpha test will focus on revalidating the dimensions analyzed and the questionnaires employed, both used during the previous usability testing of the first prototype. This iterative evaluation helps to solidify data and to anticipate potential usability problems that may be identified further by end-users.

7.4.2 Alpha and beta measurements

In addition, new measures will be added such as: acceptance, care burden, ethical design, impact of technology, perceived health, perceived stress/anxiety, quality of life, self-efficacy, social connectedness, technology acceptance, willingness to pay, workload. These added measurements are the same for alpha and beta testing, with the sole difference of the insertion of the cost effectiveness, as a measurement to consider in the beta pilot test.

Measurement	Evaluation's tool
Care burden	EDIZ questionnaires
Ethical design	Online questionnaires
Impact of technology	Nurses' questionnaires (T0,T1,T2,T3)
Perceived health	SF12
Perceived stress/anxiety	Generalized Anxiety Disorder (GAD-7)
Quality of life (QoL)	EQ5D-5L
Self-efficacy	UTAUT
Social connectedness	Social Connectedness Index (SCI)
Technology acceptance	TAM
Willingness to pay	Online questionnaire
Workload	Reduced planned care hours

Table 18. Measurements and tools for the whole pilot tests

7.4.2.1 Care burden

The aim is to check and evaluate whether the prototypes have been effective enough in reducing the level of work-related stress among informal caregivers. As the social robot must be *"the eyes and ears"* of relative caregivers, it is important that P2 and P3 can work properly in the information relay. If so, the care pressure on the seniors' relatives should be reduced.

7.4.2.1.1 EDIZ questionnaire

The 15 item, EDIZ questionnaire will be set up to measure the seniors' informal caregivers care burden on a 0-9 scale (0-3 means a low care burden pressure, 4-6 moderate, 7-9 severe). Questions will take account on the time pressure, as they work alongside the care they provide, and questions on the consequences of cares' pression on their own health.



7.4.2.2 Ethical design

By ethical design, we measure whether the design of the product pleases in use and is in conformity with all the participants and their particular needs.

7.4.2.2.1 Online questionnaire

A short online questionnaire can be used to evaluate if the robot is enjoyable and pleasant to use, if so, how much on a scale, if not, it would be a must to know what should be improved to meet end-users' expectations.

7.4.2.3 Impact of the technology

The aim is to evaluate the impact that the use of a new technology can have on a life issue. In this case, it's related to the evaluation of the use of P2 and P3, in order to help seniors to be rigorous in their treatment.

7.4.2.3.1 Nurses' questionnaire

On everyday life will be explored and measured with mainly questionnaires which will be filled out monthly by the nurses (formal caregivers), informal caregivers and clients, before the start of the trial (T0), after the first month (T1), the second (T2), and third (T3).

7.4.2.4 Perceived health

The overall health status regarding subjective perception, is a good indicator on the perceived health. The aim is to understand the impact of their perceived health on their daily lives and to verify if this tool can have a good impact on it.

7.4.2.4.1 SF12

As the SF-12 correlates highly with the SF-36 on mental and physical aspects, we can keep the twelve item self-report. Participants will be asked to rate their health on a physical, mental and social level, by responding to 12 survey questions.

7.4.2.5 Perceived stress/anxiety

It is also interesting to evaluate the stress of the end-users. It can give us an indication on the robot facilities, or not, to reduce anxiety related to care, if they are caregivers, or related to health condition, if they are seniors. The robot should reduce anxiety and stress factors.

7.4.2.5.1 GAD-7 questionnaire

As the name suggests it, this tool is a 7-item questionnaire which answers spread into 4 different frequencies: from never to almost every day.

7.4.2.6 Quality of life

Quality of life will gather data on health issues and life quality. The aim of this measurement is to quantify the quality of life that could be enjoyed in each extra life year gained from the using of a tool improving the follow-up treatment, such as GUARDIAN.

7.4.2.6.1 EQ5D questionnaire

EQ5D is a generic questionnaire that measures health-related quality of life. The term "5D" means that 5 dimensions are involved in this measure: mobility, self-care, usual activities, pain and discomfort, and anxiety and depression.



7.4.2.7 *Self-efficacy*

It's important to have indications on the self-efficacy of end-users, even more on seniors' perception. We will evaluate the perceiver ability of each participant to organize and execute the courses of action required to manage prospective situations.

7.4.2.7.1 *UTAUT questionnaire*

UTAUT is the acronym of Unified Theory of Acceptance and Use of Technology, which is a model partly applicable to quantify if self-efficacy is estimated on a low or high level. This tool will allow us to assess people's perception of the proper use of Misty II, the social robot.

7.4.2.8 *Social connectedness*

This measure takes in interest two different levels: the individual and the overall level. The aim is to gather and assess all aspects, like qualitative or quantitative social appraisal. The intrinsic aim is to ensure that the insertion, and use above all, of a new technological tool does not lead to social alienation for the person using it.

7.4.2.8.1 *SCS questionnaire*

SCS is the abbreviation for Social Connectedness Scale, and as the name suggest it, it's a questionnaire used to evaluate how end-users define themselves in a society such as the one they live in.

7.4.2.9 *Technology acceptance*

It's crucial to evaluate the technology acceptance to know if Misty II has chances to be used daily by a triad of end-users. Seniors, as caregivers, either formal or informal, will be questioned depending on two factors: the perceived usefulness and the perceived ease-of-use.

7.4.2.9.1 *TAM questionnaire*

TAM is the abbreviation for Technology Acceptance Model. More detailed versions are available, such as TAM 2. This model will take into account the perceived usefulness and the perceived ease-of-use of the social robot.

7.4.2.10 *Willingness to pay*

Pilot test will gather some data or indications on the possible amount of money that a customer will accept to pay for the use of a social robot such as Misty II. Gathering those data is crucial to find out with which price it could enter in the fit market.

7.4.2.10.1 *Online questionnaire*

A very short questionnaire can be submitted to find out what prices they would award per prototype, and ask what the price mentioned depends on, or even check if the possibility of renting and paying a subscription could be a possibility, or even a desire, for them.

7.4.2.11 *Workload*

Formal caregivers encounter workload that pressure them on the home cares provided. GUARDIAN's social robot being the eyes and ears for formal caregivers too, it must provide them a form of relay and planning simplification.

7.4.2.11.1 *Reduced planned care hours*

To assess whether the workload is indeed reduced and solved by the inclusion of Misty II at home, we will use the indicator of reduced planned care hours. Data from this qualitative evaluation will bring



to the light if the system is efficient enough regarding workload or whether it still have to be modified, whether they expect to receive state support or insurance in order to own one.

8 Conclusion

The global overview offered by this deliverable puts the different evaluations' expectations into perspective until December 2022. This deliverable highlights all the steps related to the design evaluation in order to improve, between each test, whether for P1, P2 or P3, all the functionalities, screens and interfaces related to the use of Misty II the robot.

More than a simple description, the deliverable provides information on the measurements that will be assessed but also on the questionnaires picked up to analyze them. The questionnaires presented are those originally planned and some will maybe be merged or removed, in order to not overload the users. The distribution of measurements between each evaluation gives a progressive idea of the quality of the social robot produced so far.

To conclude this deliverable, it is necessary to underline the fact that the pandemic situation, limiting each of our partner countries in its own way, has hampered the evaluations' deadlines. The workplans presented in this deliverable are to be taken as guides but they may come to change. The deliverable 4.1 makes noticeable for each test how the overlap of the different tasks is expected to take place.

9 References

- [1] Bobillier-Chaumon, M. & Dubois, M. (2009). Technology Acceptance and Acceptability in Organizations. In "Le travail humain". Volume 72, Issue 4. 355-382.
- [2] Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., and Mack, R. L. (Eds.), Usability Inspection Methods, John Wiley & Sons, New York, 25-64.
- [3] Gravelle, H. & Smith, D. (2001). Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Journal of Health Economics*, 10: 587-599.
- [4] Venkatesh, V. & al. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly* 27 (3): 425-478.
- [5] Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396.
- [6] Bandura, A. & Wood, R. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision-making. *Journal of Personality and Social Psychology*, Vol 56 (5): 805-814
- [7] Lee, R. M., & Robbins, S. B. (1995). Measuring belongingness: The Social Connectedness and the Social Assurance scales. *Journal of Counseling Psychology*, 42(2), 232–241.
- [8] Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340.
- [9] Breidert, C. Hahsler, M. Reutterer, T. (2006). A review of methods for measuring willingness-to-pay. *Innovative Marketing*, Volume 2, Issue 4

