D4.2 Development of a MPI based on MARIO Robot CGA-Module

Project Acronym:	**MARIO**
Project Title:	**Managing active and healthy aging with use of caring service robots**
Project Number:	**643808**
Call:	**H2020-PHC-2014-single-stage**
Topic:	**PHC-19-2014**
Type of Action:	**RIA**

D4.2

Work Package:	*WP4*	
Due Date:		M18
Submission Date:		27/07/2016
Start Date of Project:		01/02/2015
Duration of Project:		36 Months
Organisation Responsible of Deliverable:		IRCCS Casa Sollievo della Sofferenza
Version:		2.1
Status:		Final Version
Author name(s):	**Daniele Sancarlo** Grazia D'Onofrio Antonio Greco Francesco Giuliani Francesco Ricciardi Mary Kagkoura Thomas Messervey Misael Mongiovì Lazaros Penteridis Keith Cortis	IRCCS IRCCS IRCCS IRCCS IRCCS R2M Solution R2M Solution CNR ORTELIO NUIG
Reviewer(s):	Alessandro Leone David Bisset	CNR Lecce RU Robots
Nature:	☒ R – Report ☐ P – Prototype ☐ D – Demonstrator ☐ O - Other	
Dissemination level:	☒ PU - Public ☐ CO - Confidential, only for members of the consortium (including the Commission) ☐ RE - Restricted to a group specified by the consortium (including the Commission Services)	
Project co-funded by the European Commission within the Horizon 2020 Programme (2014-2020)		

Revision history			
Version	**Date**	**Modified by**	**Comments**
1.0	10/04/2016	Daniele Sancarlo	Initial Table of Contents and Executive summary
1.1	23/06/2016	Daniele Sancarlo	Introduction
1.2	29/06/2016	Grazia D'Onofrio	Revision and completed Section 3.2
1.3	12/07/2016	Francesco Giuliani	First draft of section 2.1
1.4	14/07/2016	Francesco Ricciardi	Contribution to section 2.1 – Template updates
1.5	15/07/2016	Francesco Giuliani Francesco Ricciardi	Completed Section 3.3
1.6	16/07/2016	Grazia D'Onofrio Daniele Sancarlo	Revised Executive Summary and completed Conclusion section
1.7	19/07/2016	Keith Cortis	Statistical methods updated
1.8	25/07/2016	Francesco Giuliani	General revision
1.9	26/07/2016	Francesco Giuliani	Final version including reviewers' comments
2.0	29/11/2016	Francesco Giuliani	Adding new sections upon request of the reviewers
2.1	5/12/2016	Daniele Sancarlo	Updated with internal reviewers comment

Executive Summary

This deliverable focuses on the development of a statistical model to determine the health status of subjects based on the CGA (Comprehensive Geriatrics Assessment) approach through the MARIO Robot. The main points are shown below:

- Literature search on the Mortality risk factors and prediction tools for the elderly in the Community-Dwelling, Nursing Home and Hospital settings.
- Determination of the key variables to be included in the model and choice of the sensors and apps to be implemented in MARIO to assess and rehab health, cognitive and functional status of patients with dementia.
- Development of the statistical model to determine health status to be implemented in MARIO

This document constitutes the input to choose the most appropriate ways to develop functionalities to make MARIO innovative and relevant for the assistance of patients with dementia from a healthcare point of view.

Table of Contents

Index of Tables

1. Introduction

Information and Communication Technologies (ICT) have the potential to improve the access, quality, safety and efficiency of patient care and their use may be particularly helpful in the care of patients with dementia (PwD). A multidimensional evaluation could be very useful in these patients in order to determine the best therapeutic strategy but it could be relatively time consuming.

The comprehensive geriatric assessment (CGA) is a clinical multidimensional-multidisciplinary approach methodology, widely validated and diffused around the world, which gives a framework for the delivery of interventions addressing relevant and appropriate issues related to frail elderly subjects. CGA helps to characterize the health status of an elderly person's considering medical, psychosocial, functional, and environmental domains in order to build a tailored and effective intervention plan. In particular we have chosen to include in a standardized evaluation the following domains: basal and instrumental activities of daily living (ADL, IADL) [1,2]; Cognitive status [3]; Comorbidity [4]; Nutritional status [5]; Pressure sores risk [6]; Medication use; Social aspects; Vital parameters. Currently, on the market, one can only find ICT solutions capable to assist carers in performing CGA but there are not ICT platforms able to autonomously perform CGA.

The ambitious objective of this deliverable is to present a model built upon relevant health domains to build a health index based on the CGA approach that can be integrated in a robot solution. This solution will enable to perform autonomously the CGA evaluation and its accuracy will improve thanks to the possibility of a continuous monitoring of the many variables included in the model. This will eventually result in improving the quality of care, reducing time spent by healthcare professional in data acquisition. From a research point of view, this work could also ease the search for new mathematical models to represent health changes.

1.1. Work Package 4 Objectives

WP4 objectives are:

- Determine the mathematical model of a Multidimensional Prognostic Index capable dynamically to detect the health changes occurring in a subject living at home or in hospital.
- Develop the CGA-module for the robot platform through a multidisciplinary interaction using an end user centered design approach.
- Select which technologies could be more adequately fitted on the MARIO platform starting from the specific user needs of the elderly.
- Customization of the comprehensive geriatric assessment (CGA) approach to the service robot context.
- Create a shared health database useful for the integration with other health and social services to improve the monitoring of the health status and multidisciplinary cooperation.

1.2. Purpose and Target Group of the Deliverable

Task 4.2 is titled “Development of a MPI based on MARIO Robot CGA-Module” and its objective is to develop a statistical model for MPI from the domains acquired from the robotic CGA module. The focuses of this task are the followings:

1) to compare the performance of the model, possibly using proxy values obtainable through the mining of large database of elderly patients resident at home or hospitalized with complete CGA;

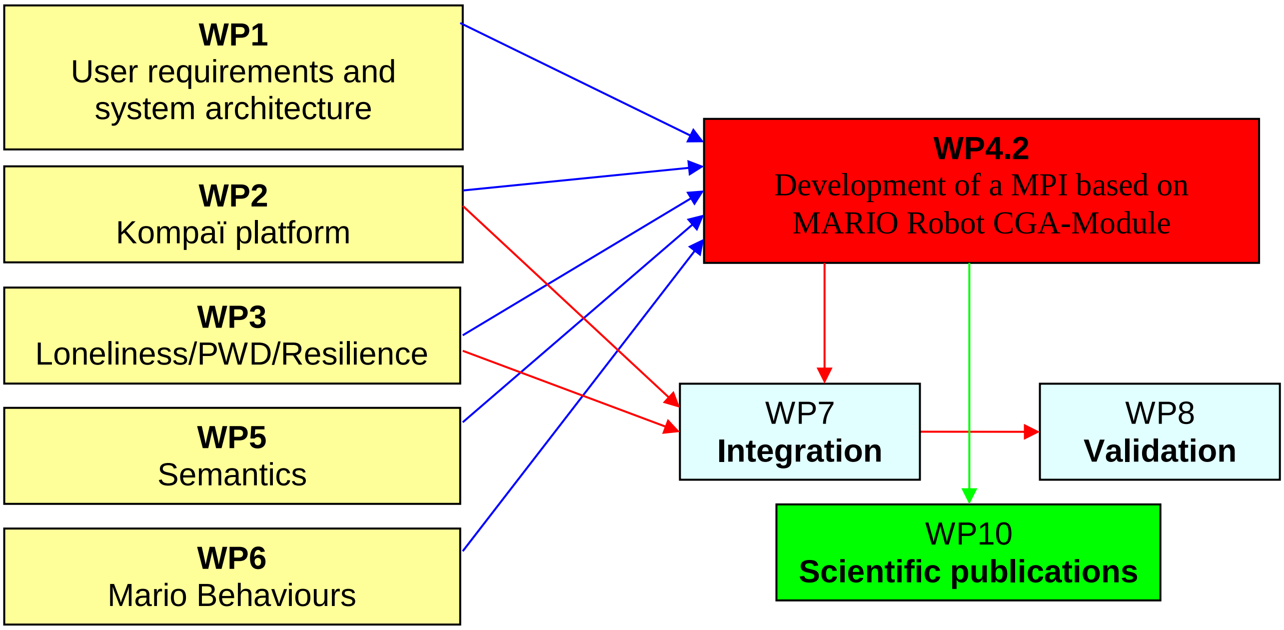
2) to redefine autonomously the weight of the different domains used to compute the MPI, through iteratively re-calculating it over time;

3) to re-determine the scoring rules improving accuracy and therefore the estimate of the health status.

Gaining insight from the aforementioned items and thanks to the modularity of MARIO, more domains will be added to the scoring system that will be used to compute the health status of PwD. This deliverable will help organizations in making decisions on how to best use MARIO in the care of PwD.

1.3. Relations to other Activities in the Project

This WP will receive as its input the user requirements and the system architecture from WP1. From WP2, this WP will receive specifications of the Kompaï platform which will later be shared with WP7 (Integration) and used for the validation. Throughout the iterative development process in the project, the results from WP3 (Loneliness/PWD/Resilience), WP5 (Semantics), and WP6 (Mario Behaviours) will benefit also from the work of WP4. It is envisioned that this WP will also produce noteworthy scientific publications (WP10).



Relationships between Work Packages

1.4. Document Outline

In Section 2 the materials and methods used for the reviewing process will be presented. Section 3 will present the statistical method developed to detect health changes in elderly patients. The conclusions of this document will be presented in Section 4.

1.5. About MARIO

MARIO addresses the difficult challenges of loneliness, isolation and dementia in older persons through innovative and multi-faceted interventions delivered by service robots. The effects of these conditions are severe and life-limiting. They burden individuals and societal support systems. Human intervention is costly but the severity can be prevented and/or mitigated by simple changes in self-perception and brain stimulation mediated by robots.

From this unique combination, clear advances are made in the use of semantic data analytics, personal interaction, and unique applications tailored to better connect older persons to their care providers, community, own social circles and also to their personal interests. Each project objective is developed with a focus on loneliness, isolation and dementia. The impact centres on deep progress toward EU scientific and market leadership in service robots and a user driven solution for this major societal challenge. The competitive advantage is the ability to treat tough challenges appropriately. In addition, a clear path has been developed on how to bring MARIO solutions to the end users through market deployment.

2. Materials and methods

2.1. Literature Search

The search strategy and analysis was informed by: the study's aims, previous systematic reviews using qualitative data, and best practice recommendations in the research literature [7, 8].

Literature searches were conducted in the MEDLINE, PUBMED, SCOPUS databases, using the OVID search interface. The search queries included 1) mortality, 2) Elderly OR older, 3) Prognostic factor, 4) Prognostic Indexes. Only English language articles were included, due to lack of resources for translation. Reference lists of included articles and relevant review articles were examined to identify any studies which the electronic search strategy may have missed.

For this manuscript, a single reviewer examined abstracts retrieved by the electronic search to identify articles meriting a full review. Full length articles were then reviewed before data were extracted from relevant papers. The inclusion/exclusion criteria used for review protocol were the following. Inclusion criteria: 1) age ≥ 60 years; Exclusion criteria was no English editing (as we lacked resources for translation).

2.2. Robotic MPI

For the methods used to compute the robot based MPI you can refer to:

- The method to determine the MPI in its standard formulation (using eight domains: Activities of Daily Living, Instrumental Activities of Daily Living, Short Portable Mental Status Questionnaire, Comorbidity Index, Mini Nutritional Assessment, Exton Smith Scale, Number of Medications, Social Support Network) obtained by IRCSS and presented in paragraph 3.1
- The definition of a statistical model to determine the health status and thus a robot based MPI, presented in paragraph 3.3 of the present document

3. Results

The results are defined in three main paragraphs: the first two assessing the elements determining health status and the third determining the mathematical model to calculate a health index continuously combining different kind of data and using a holistic approach.

3.1. Mortality risk prediction tools for the elderly

The following paragraphs present the state of art on the field of computing indexes able to estimate mortality risks in the elderly.

Paragraph 3.1.4 will present in depth the experience made by IRCSS in the development of the Multidimensional Prognostic Index (MPI).

3.1.1. Community-Dwelling Older Adults

We identified 6 indices for community-dwelling older adults. Indices estimated mortality risk from 1 year [9] to 5 years [10]. The highest-risk group from Schonberg et al. at 9-year follow-up predicted 92% mortality (95% CI, 86%-96%) [11]. Gagne et al. [9] developed a mortality risk score to predict 1-year mortality by combining conditions in the Romano et al. [12] implementation of the Charlson et al. index [13] and the van Walraven et al. [14] implementation of the Elixhauser et al system [15]. The model had good discrimination and was well calibrated.

Reclassification measures compared the model favorably against the Romano/Charlson and van Walraven/Elixhauser indices.The 15-month index by Mazzaglia et al. [16] is a 7-item questionnaire for primary care physicians that was developed in 2470 primary care patients who were 65 years and older residing in northwestern Florence, Italy, and validated in a sample of 2926 similar patients residing in southwestern Florence. The model was well calibrated and had good discrimination, but it predicted the narrowest range of mortality of any examined index (0%-10% risk). Carey et al. [17] developed a 2-year index for community-dwelling elderly individuals from a sample of 4516 adults 70 years and older from the eastern, western, and central United States who had been interviewed in the Asset and Health Dynamics Among the Oldest Old (AHEAD) study in 1993.

Carey et al. subsequently validated the index in 2877 similar interviewees from the southern United States. The index had good discrimination and was well calibrated across all 3 risk levels but predicted only a narrow range of mortality (5%-36% risk). The index by Carey et al. for 3-year mortality [18] was developed in functionally impaired, nursing home–eligible, community-dwelling adults who were 55 years and older in the years 1988 through 1996, living in the western United States (n=2232), and enrolled in the Program of All-Inclusive Care for the Elderly (PACE), a senior daycare program providing multidisciplinary services. Validation was conducted in PACE participants from the eastern and midwestern United States (n=1667). The index was well calibrated but showed only moderate discrimination. Accuracy was similar for 1-year mortality. Lee et al. [19] developed a 4-year mortality index in community-dwelling adults older than 50 years from the eastern, western, and central United States who were interviewed in the Health and Retirement Survey of 1998 (81% participation rate, n=11701). To test geographic transportability, the index was validated in interviewees from the southern United States (n=8009). The Lee et al. index was well calibrated and showed very good discrimination. The index by Schonberg et al. [10] to predict 5-year mortality was developed from a nationally representative sample of adults older than 65 years (n = 16 077) who responded to the 1997-2000 National Health Interview Survey (NHIS) (74% participation rate); it was well calibrated and had good discrimination in a random sample of n=8038 adults drawn from the same data source. Schonberg et al. [11] then further validated the index in respondents to the 2001-2004 NHIS (n=22057, 25% aged > 80 years, 57% female, 12% dependent in at least 1 instrumental activity of daily living, 18% with diabetes, 15% with cancer) and found no change in discrimination (*C* statistic, 0.75). The Kaplan-Meier method demonstrated widening separation between risk groups out to 9 years.

3.1.2. Nursing Home Residents

Two indices were developed for the nursing home, both using the Minimum Data Set (MDS), a clinical and administrative data set that is federally required of all US nursing homes. The MDS Mortality Rating Index by Porock et al. [20] to estimate 6-month mortality in nursing home patients was developed using data from all Missouri long-term care residents in 1999. Study authors later created a simplified version of this model using the same data set [21]. The revised Flacker and Kiely [22, 23] long-stay index for 1-year mortality was developed and validated from the MDS using a split sample of nursing home residents who were 65 years and older and residing longer than 1 year in Medicare- certified nursing homes within New York (n=63077). Both indices demonstrated very good discrimination and were well calibrated across a wide range of mortality risk levels, except the revised Flacker and Kiely for the highest risk group (20% difference). Kruse et al. [24] prospectively validated indices by Porock et al and Flacker and Kiely in a small, prospective, single nursing home study in 2007 (n=130, mean age 83 years, 61% female, 24% dementia, 23% congestive heart failure). For the Porock et al index, discriminatory ability was lower in the validation study by Kruse et al (*C* statistic, 0.59; 95% CI, 0.46-0.72) than in the original derivation study by Porock et al (*C* statistic, 0.75) or using the simplified score (*C* statistic, 0.76). For the revised Flacker and Kiely index, discriminatory ability was the same in both the original derivation study by Flacker and Kiely (*C* statistic, 0.71) and the external validation by Kruse et al (*C* statistic, 0.72; 95% CI, 0.62-0.81).

3.1.3. Hospitalized Older Adults

Eight indices were developed to estimate mortality risk for hospitalized older adults. Five were intended for use in the emergency department or on hospital admission [25-29] and 3 after hospital discharge [30-32]. The “Silver Code” by Di Bari et al. [27], a 1-year index for emergency triage of individuals aged 75 years and older, was developed and validated using administrative records of patients admitted to the hospital via the emergency department from Florence, Italy, in 2005 (n=10913). They achieved 91% linkage across 4 administrative data sets (demographics, hospitalizations, prescription medications, and mortality). Random split sample validation was conducted on half the cohort. The index was well calibrated and discriminatory ability was moderate. Fischer et al. [28] conducted a retrospective medical record review to develop a 1-year index for hospitalized elderly individuals using 4 pre-specified predictors called the CARING criteria, collected at admission. Their sample included patients admitted to the medical service of a US Department of Veterans Affairs hospital in a 4-month period in 1999 (n=873). Participants admitted in the first 2 months of the study period were included in the development cohort; the remainder were in the validation cohort. The model had very good discrimination and a reported error rate of 0.26 in the validation cohort. Youngwerth et al. [33] later prospectively tested the external validity of the CARING criteria in a younger, sex balanced sample from a university hospital in 2005 (n=427, average age 54 years, 50% female). No *C* statistic was reported for the external validation. The Burden of Illness Score for Elderly Persons by Inouye et al. [25] updated previous indices developed by the same group [34, 35] by adding functional and laboratory data to diagnoses from administrative data to estimate 1-year mortality. Participants were drawn from a prospective study of individuals aged 70 years and older who were hospitalized at Yale–New Haven Hospital from 1989 through 1991 (n=525). The study was validated in a sample of 1246 participants from 27 Connecticut hospitals who were 65 years and older with a principal discharge diagnosis of pneumonia from 1995 through 1996. The investigators demonstrated improvement in the *C* statistic with the addition of laboratory and functional and cognitive measures to administrative data (validation *C* statistics, administrative alone, 0.59; all measures, 0.77). The model was well calibrated at the extremes but was less accurate in middle risk groups (Table 4). Pilotto et al. [26] used information from the standardized Geriatrics Assessment, performed at admission, to develop a 1-year prognostic index for hospitalized individuals aged 65 years and older in a sample of 838 consecutively admitted patients to the geriatrics unit of an Italian hospital in 2004, validating in 857 participants from 2005. They subsequently tested the model's accuracy at 1 year and 1 month in participants from the same hospital from 2005 to 2007 (n=4088) [36]. The model was well calibrated and demonstrated good discrimination in the larger validation study (*C* statistic, 0.71; 95% CI, 0.70-0.74), and performance was similar at 1 month (*C* statistic, 0.76; 95% CI, 0.73-0.79). Teno et al. [29] developed a nomogram to predict 1- and 2-year mortality based on medicine and ICU patients aged 80 years and older who were enrolled in the Hospitalized Elder Longitudinal Project (HELP) from 5 different hospitals across the United States from 1993 to 1994 (n=1266). Teno et al tested the reproducibility of the index in 150 random samples from the original 1266 patients. The Teno et al nomogram is convenient in that it predicts multiple end points from a single score. The index includes the APACHE III scale, which requires arterial blood gas measurement. Levine et al. [32] developed a 1-year prognostic model for hospitalized elderly individuals after discharge using data from a cohort of patients admitted to hospitalist and non-hospital physicians at the University of Chicago Hospitals from July 1997 through June 1999 (development cohort, n=2739) and July 1999 through June 2001 (validation cohort, n=3643). The index had moderate discriminatory ability and was well calibrated. Walter et al. [30] developed a 1-year index for elderly individuals after hospital discharge using secondary data from a study of patients aged 70 years and older who were hospitalized between 1993 and 1997 at the University of Hospitals

Cleveland (development cohort, n=1495) and the Akron City Hospital (validation cohort, n=1427). The model demonstrated good discrimination and was well calibrated across risk groups. Rozzini et al. [37] subsequently externally validated the index's performance predicting 6-month mortality in a retrospective analysis of 840 consecutively admitted participants to a hospital in Italy and found monotonic increases in mortality for each predicted risk level (observed 4%, 10%, 25%, and 46% 6-month mortality).

Taking into account the indexes described so far, IRCCS has developed in the past years a widely validated Multidimensional Prognostic Index to be applied to hospitalized elderly patients. It will be the base on which we will build a Multidimensional Prognostic Index in a new version to encompass variables read by the Robot or sensor network as detailed in paragraph 3.3. In the following we describe the main principles that led to the development of MPI.

3.1.4. The Multidimensional Prognostic Index (MPI)

Since mortality in older subjects results from a combination of biological, functional, psychological, pathological, and environmental factors, tools that effectively identify patients with low life expectancy should take a multidimensional approach [38]. Previous attempts to develop a prognostic index for older patients based on their functional, biological, and environmental characteristics were performed on population-based [39, 40], community-dwelling [17, 19], or institutionalized subjects [41]. Two studies developed a prognostic index for mortality in hospitalized elderly patients from measures of physical and cognitive functions [42] or from demographic characteristics, functional disability, co-morbidity, length of hospital stay, and laboratory measures [30]. To our knowledge, no prognostic index for mortality in hospitalized elderly patients has been developed that fully utilizes the wide range of information provided by a standardized CGA, which is the most accurate and sensitive diagnostic tool for evaluating and monitoring elderly patients for clinical decision-making purposes [43, 44].

In IRCCS CGA was carried out using assessment instruments widely employed in geriatric practice. Functional status was evaluated by the Activities of Daily Living (ADL) index [1], which defines the level of dependence/independence of six daily personal care activities, including bathing, toileting, feeding, dressing, urine and bowel continence, and transferring (in and out of bed or chair), and by the Instrumental Activities of Daily Living (IADL) scale [2], which assesses independence in eight activities that are more cognitively and physically demanding than those in the ADL index, including managing finances, taking medications, using the telephone, shopping, using transportation, preparing meals, doing housework, and washing. Cognitive status was assessed by the Short Portable Mental Status Questionnaire (SPMSQ), a 10-item questionnaire that assesses orientation, memory, attention, calculation, and language [3]. Co-morbidity was examined using the Cumulative Illness Rating Scale (CIRS) [4]. The CIRS uses 5-point ordinal scales (score 1–5) to estimate the severity of pathology in each of 13 systems, including cardiac, vascular, respiratory, eye-ear-nose-throat, upper and lower gastroenteric diseases, hepatic, renal, genitourinal, musculoskeletal, skin disorders, nervous system, endocrine-metabolic, and psychiatric behavioral problems. Based on the ratings, the two following scores are derived: (1) the

Comorbidity Index (CIRS-CI), which reflects the number of concomitant diseases and is derived from the total number of categories in which moderate or severe levels (grades 3–5) of disease are quoted (ranging 0–13); and (2) the Severity Index (CIRS-SI), which reflects the overall severity of diseases and the average rating of 13 disease categories, excluding psychiatric behavioral problems (ranging 1–5). Nutritional status was explored with the Mini Nutritional Assessment (MNA) [5], which includes information on (1) anthropometric measures (body mass index [BMI]: body weight/height2, mid-arm circumference in cm [MAC], calf circumference in cm [CC], and weight loss); (2) lifestyle, medication, and mobility; (3) number of meals, food, and fluid intake and autonomy of feeding; and (4) self-perception of health and nutrition. The Exton-Smith Scale (ESS) was used to evaluate the risk of developing pressure sores. This five-item questionnaire determines physical condition, mental condition, activity, mobility, and incontinence. For each item, a score from 1 to 4 is assigned [6]. Medication use was defined according to the Anatomical Therapeutics Chemical Classification code system (ATC classification) [45], and the number of drugs used by patients at admission was recorded. Patients were defined as drug users if they took a medication of any drug included in the ATC classification code system at the time of admission. Social aspects included household composition, home services, and institutionalization.

In order to develop a MPI that correctly reflects the multidimensional impairment of the hospitalized geriatric patient, a cluster analysis on CGA data of the development cohort population was initially made for evaluating the independence of variables and identifying the most relevant domains of the CGA in predicting mortality outcome. The cluster analysis showed a correlation among ADL, IADL, SPMSQ, ESS, and MNA and evident independence among the previous variables and comorbidity (CIRS) and medication use, which were correlated with each other and social aspects. Thus we started to develop a MPI considering only three variables: ADL, medication use, and social aspects. This “three-domain” MPI, while in a Cox regression analysis produced an acceptable separation among the survival curves of the three groups of patients (low, moderate, and severe risk of death), resulted in an unsatisfactory prognostic index for 1-year mortality (OR, 0.635; 95%CI, 0.141–2.871). Following a step-wise method, other domains of the CGA, one at a time, were progressively included in the model, and relative Cox and logistic regression analyses were performed.

Thus the eight-domain MPI (i.e., a total of 63 items in eight domains of the CGA) resulted in the best index in predicting 1-year mortality in this population. For each domain, a tripartite hierarchy was used (0, no problems; 0.5, minor problems; and 1, major problems) based on conventional cutoff points derived from the literature for the SPMSQ [3], MNA [5], EES [6], and ADL/IADL [46] or by observing the frequency distribution of the patients at various levels to identify points of separation for co-morbidities and number of drugs. The specific threshold used to define the three categories are shown in Table 1. The sum of the calculated scores from the eight domains was divided by 8 to obtain a final MPI score from 0 to 1. For analytical purposes, absolute values of MPI were not considered; we preferred to express the MPI as low (MPI value≥0.33), moderate (MPI between 0.34 and 0.66), and severe risk (MPI≤0.66), according to previous rule-based indices used for exploring multidimensional impairment in elderly subjects [47]. In order to determine the best MPI cutoff points, MPI values were classified into three categories: low, moderate and severe risk as follows: Then we fixed d = (B - A) value to 0.1, obtaining the following cut-off point sequence: Setting the d values to 0.2, 0.3, . . . , 0.8, for each cut-off point sequence, the degree of separation between the MPI curves for patients with low, moderate, and severe risk was computed: thus we found that the point values that produced the best separation among the curves corresponding to the different MPI grades were A = 0.33 and B = 0.66. In this model the term "separation among the curves" refers to the maximal value of R, that is, R = Δr1 + Δr2 where Δr1 = $\int_0^{t_{max}}$ [r_L (t) - r_M (t)]dt and Δr2 = $\int_0^{t_{max}}$ [r_M (t) - r_S (t)]dt. In the integral, r_L, r_M, and r_S refer to the survival curves for a low, moderate, and severe patient risk. Martingale residuals were used to explore the functional form of the relationship between the MPI and mortality and to verify whether the thresholds for the definition of the MPI group were appropriate. The proportionality assumption was verified by plotting Ln[-ln(survival function)] with time.

To test the hypothesis that the prognostic value of the aggregated MPI was superior to the prognostic value of its single components considered individually, a logistic model was carried out on the individual parameters. Age, ADL, IADL, SPMSQ, CIRS, MNA, EES, and number of drugs were evaluated as continuous variables, while social support network and MPI were evaluated as ordinal variables, based on the assumption of an equidistance between single unit values. Sex was analyzed as a dichotomous variable.

We assessed the predictive accuracy of the final model by looking at the two components of accuracy: calibration and discrimination. Calibration of the model was assessed by comparing the predicted mortality with the actual mortality in the development and validation cohorts. The discrimination of the model was assessed by calculating the receiver operating characteristic (ROC) curves for the development and validation cohorts. A *p* value of 0.05 was considered for statistical significance.

3.2. New parameters in the evaluation of health status

Currently, technologies exist that hold great promise to expand the capabilities to measure health parameters. Due to their relatively low cost they seem to be ready to spread to the wide community. They are also expected to improve diagnostics and monitoring and to maximize independence and participation of individuals to their care processes. We describe in the following sections the ones that have been so far selected for MARIO implementation:

- Beddit Sleep Monitor

- FITBIT

- ZephyrLIFE™

They, together with the results obtained through the following MARIO apps:

- CGA app

- Music and Flash game apps

- News app

- Hobby app

will form the base dataset upon which to build the model described in paragraph 3.3.

3.2.1. Beddit Sleep Monitor

Beddit Sleep Monitor [48] measures sleep with unobtrusive force sensors. It measures the forces caused by the body on the bed with a flexible film sensor that is placed below the bed sheet. The measurement methodology poses scientific challenges because physiological information (heart rate, respiration, etc.) that are vital for analyzing sleep cannot be readily extracted from the sensor's signal, but requires sophisticated signal analysis methods.

The measurement of mechanical cardiac activity from the platform supporting the body is called ballistocardiography (BCG). Each time the heart beats, the acceleration of blood generates a mechanical impulse that can be measured with a proper force sensor, such as the Beddit Sleep Monitor.

Measuring the heart rate from BCG or similar mechanical signals is much more complex than measuring the heart rate using electrocardiogram (ECG), the most commonplace cardiac measurement method. Individual heartbeats can be detected in an ECG signal relatively easily, by locating a clear spike (called the QRS complex, from the consecutive named spikes Q, R, S of the ECG heartbeat) that accompanies each heartbeat. However, with BCG, the cardiac impulses are less pronounced and more variable than the salient shape of the QRS complex.

Beddit Sleep Monitor is also capable of measuring respiratory activity as respiration causes the chest to move measurably. There are three main motivations for measuring respiration unobtrusively during sleep. First, respiration conveys information about the general condition of the patient, so the deterioration of health can be detected with respiration monitoring. Second, sleep-related breathing disorders (SRBD) such as sleep apnea represent a major share of sleeping problems. Third, the structure of sleep can be analyzed based on respiration, because sleep stages have differing effects on respiration. The respiration measurement is a 4-step process where first the parts of the signal that contain gross movements are discarded. Then the respiration signal is low-pass filtered on 4 distinct f Hz frequencies with potentially disturbing phenomenons taken into account at around 2x f Hz. Therefore, at least one of the filters will result in an output signal that has the respiration frequency intact but the disturbance removed. Then the respiration cycles are detected from each filtered signal. A respiration cycle begins at a local maximum and ends at the next local maximum in the signal. In addition, the amplitude of each respiration cycle is calculated by taking the difference between the signal value of the local maximum that starts the cycle and the minimal signal value in the cycle. Lastly, the final sequence of respiration cycle lengths is compiled from the four signals based on the stability of respiration cycle amplitudes in each signal. The correct signal is typically selected, because the signal that contains frequencies up to the respiratory frequency is more stable in its amplitude than a signal that also contains higher-frequency disturbing phenomena.

As sleep correlates with a low level of motility, circadian rhythmicity can be estimated with a method called actigraphy. An accelerometer sensor is worn on the wrist 24 hours a day, which allows estimating the daily alternation between sleep and wakefulness.

Due to its limited accuracy, actigraphy is typically used for the overall characterization of sleeping patters over a period of at least a week. The Beddit Sleep Monitor detects the gross movement of the person sleeping. Even if Beddit is not a medical device and shouldn't be used for any kind of medical diagnoses, the excessive movement during the night could potentially be a sign of, for example, periodic limb movement disorder. In such a case, the user should be in contact with an appropriate doctor. The movement information is analyzed by detecting discrete events of movement from the BCG signal. That is done by dividing the high-pass filtered (cut-off frequency 5 Hz) signal into three-second windows. Each window is detected as movement if the difference between signal minimum and maximum in the segment is above a fixed threshold.

The measurement data provided by the Beddit Sleep Monitor has been tested and validated [49-52].

3.2.2. FITBIT

Fitbit Charge HR has the potential to use heart rate-derived algorithms to contribute to estimates of energy expenditure based on activity intensity [53, 54]. Recent evidence suggests this method has acceptable validity, however there is inherent variability, demonstrating that the accuracy of this device is dependent on the device used, the type and intensity of activity [55]. Given the rapid consumer uptake of this device, it is critical to determine its accuracy to measure these variables across a variety of modes and intensities given its potential to have a major influence on lifestyle behavior and weight management.

3.2.3. ZephyrLIFE™

ZephyrLIFE™ remote patient monitoring (RPM) [56] is a complete solution designed to monitor individual or multiple patients seamlessly while delivering periodic updates and event driven physiological alerts to clinical care teams, all without entangling wires or cumbersome sensor systems.

ZephyrLIFE RPM monitors patients throughout the care continuum in the hospital, in transition, and at home.

While in the hospital, ZephyrLIFE RPM is used with the general care population to assist in avoiding never events, optimizing care, and assisting in prioritizing treatment.

During transition and in the home, ZephyrLIFE RPM enhances the level of medical support while promoting patient independence and dignity.

The advantages of ZephyrLIFE RPM are the followings:

- Ambulatory/wireless monitoring throughout the whole care continuum
- Monitors changes in heart rate and respiration rate
- Monitors changes in activity level and position

ZephyrLIFE RPM consists of the following components:

- BioPatch wireless device
- Wireless communications (ECHO radio system in the hospital, Wireless/cellular at home)
- Monitoring Interface (Central monitoring stations in the hospital, Cloud-based interface for remote access)
- BioPatch wireless device - It is a rechargeable vital signs monitor that uses standard, off-the-shelf ECG electrodes. The BioPatch wireless device is designed for continuous monitoring.

The BioModule sensor is lightweight and provides ECG, and Respiration Rate, with 3-D Accelerometry to measure posture, position, and patient activity.

- Echo radio – Easy to install stand-alone radio. It provides complete ward coverage for ambulatory patients.
- Cellular – Smartphone/tablet, based in a charging cradle or taken outside the home for mobile monitoring.
- Monitoring interface.
- Central monitoring station for hospital monitoring managed through a touch-screen terminal.
- Cloud-based interface provides remote access and can monitor multiple patients.
- Alert triggers set globally or on a per-patient basis.
- Interoperable with EHR/PHR and service providers.
- Zephyr’s API will enable channel partner HL7 / portal / web services.
- Zephyr supports EHR/ADT integration.

The measurement data provided by the ZephyrLIFE™ has been tested and validated [56-58].

3.2.4. CGA app

The usefulness of the CGA in the planning and personalization of care for elderly people is known [59]. To our knowledge, MARIO will make available for the first time a CGA app.

The carers and patient will be identified through RFID sensors. MARIO will approach and face the user. The CGA App will need to be used in a situation where the user can access the touch screen as the variation in responses and the critical nature of the data means that a higher word recognition confidence is needed with respect to spoken responses. This may mean the touch screen will be used more than in other Apps because the verbal responses may be too complex to analyse.

The CGA app in this project consists of two modules: the first is the classical module including variables acquired through questionnaires and the second through an automated continue monitoring of vital signs and other domains such as cognitive, functional, social and emotional aspects. For the first module, two options to trigger the app are going to be developed: one is manual (vocal command, useful at the admission) and the second based on a specific algorithm strictly related to final configuration of the robot. One example could be to trigger the event using as variables: date of precedent evaluation, significant variation in vital signs, and variation in the autonomous evaluations of defined parameters of interest.

MARIO will speak and present questions to the user and gather user responses. This data will be stored in an anonymised form in MARIO's store. The data will be encrypted such that only staff members who own a digital key can unencrypt the data. Staff will not be able to identify the patient from the data stored.

Care Staff will also capture responses from the patient while the CGA App is operating. In order to create parallel data that can be used to assess the capture of responses by MARIO a staff member will also record the answers to the questions. At the admission of the patient, the carers trigger the CGA and the results are recorded in parallel. The staff members will not observe the responses the user gives on the screen nor will they see what MARIO thought the answer was. This allows aggregated data from the MARIO collection process and the manual process to be compared.

All the data will be treated only by healthcare professionals and by the principal investigator of the study avoiding problems related to privacy and ethical aspects. The success of the MARIO interface and the understanding any differences between the capture of data by MARIO and by care staff will be evaluated. Please refer to the following 3.3 paragraph to gain insight on how this could play an important role in the statistical model definition.

The text of the questions on the CGA form is not in a format that can be read out by MARIO. The questions are written using medical terminology and they are designed to be used as guidance in questioning by a human with experience of both the purpose of the CGA and the type of language used by the patient when describing their condition.

It has been necessary to transform the question text of the CGA into a more acceptable form that can be understood by the patients, displayed on MARIO's touch screen, and spoken to the user. As part of this simplification process, it has been necessary to break down the CGA questions into sub-questions so that the level of confusion is reduced.

With clear answers to these two questions the clinical practitioner can fill in the answer with reasonable accuracy. However, verbal responses are likely to be much more conversational and it may take several alternative questions before the clinical practitioner has enough information to assess the category of answer.

This questioning was converted into a format so that MARIO's interface is as effective as possible in both delivering the questions and capturing the answers. This can be achieved in a number of different ways:

a) MARIO only puts questions on its screen along with the three answers and expects the user to press the screen buttons representing one of the three answers.

b) MARIO speaks a question and puts on the screen the expected answers. This is based on questioning as above, where each question is broken down into sub-questions or it could directly ask the main question.

c) MARIO speaks a question, also putting it on screen, and speaks out the three possible answers one after the other pausing to see if the user responds verbally or by touching the answers that are displayed on the screen.

We will choose the approach that gives best results in terms of correspondences between the CGA results obtained by MARIO and the ones obtained by the healthcare professionals.

Although MARIO will identify users with RFID tags, since MARIO only stores the answers from the CGA questions anonymously, there will not be direct need to correlate the patients with their answers.

3.2.5. Music and Flash game apps

In a recent study, the experiences and individual characteristics were evaluated, during a 7-week computer activity program, for 27 persons with dementia recruited from three nursing homes in an American southern country [60]. This study has shown that listening to music was the most favorite activity and playing games was able to generate strong positive feelings.

About the Music app under development in MARIO, the starting point was the definition of the keywords to identify musical genres. This was due to personalize the app according to user preferences. The patients with dementia may be suffering from irritability, depression, and frustration: MARIO Music app will stimulate brain activity allowing them to relax, have some fun, or even experience a heightened sense of awareness as they enjoy listening to the music they bonded with while growing up. Listening to music beyond being a pleasant activity for the patients to enjoy with relatives can also create a context favorable to express their emotions. MARIO Music app will awaken synaptic brain activity in patients and allow for some moments of lucidity and recognition of their surroundings and those they love.

Regarding the Flash game app, several brain stimulation games were screened to be implemented and of these only four will be integrated. Nevertheless other games could be implemented in the future.

When people with dementia keep their minds active, their thinking skills are less likely to decline, medical research shows. Games, puzzles, and other types of brain training may help slow memory loss and other mental problems.

By monitoring the app usage, more information can be gathered on the state of health of PwD. Once the app will be available, we will select which parameters could be useful from a clinical point of view (e.g. time of use, length of each session of use, frequency, variation of activity, etc.).

3.2.6. News app

The MARIO News app is a service based on the contents delivered by the most important newspapers and news agencies around the globe to adjourn the patients about important topics and discover the news relevant for the geographical location of the patient.

Drawing on thousands of sources, News app will present verified stories and send an alert on the MARIO screen. The PwD will agree or not to let MARIO read the news.

3.2.7. Hobby app

This app is going to be developed based on the focus groups performed with patients, formal and informal caregivers at the beginning of the project. In the Hobby app the following activities are going to be implemented for the patients with dementia:

- Playing cards
- Photo & Scrapbooking Activities
    - Sort photos by topic, subject, type or date. Mix them up after you finish so they can be sorted in a different way next time.
    - Assemble a photo collage.
    - Make a scrapbook, pasting photos onto the pages and writing notes about the memory beside the photo. One can also use a photo album with plastic sleeves.
    - Label old family photos so to have that information later on.
    - Reminisce about the focus of the photo.
- Reading Activities
    - Picture books
    - The Bible or Bible stories
    - Short story collections
- Activities involving humor
    - Watch or listen to comedy TV shows, movies and old radio
    - Start a humorous notebook or scrapbook
    - Laugh over funny family memories

3.3. Definition of the statistical model to determine health status

Probabilistic and machine learning techniques are now an essential part of building robots (or embedded systems) designed to operate in the real world. These systems must contend with uncertainty and adapt to changes in the environment by learning from experience. Uncertainty arises from many sources, such as the limitations inherent in modelling a complex environment, noise and perceptual limitations in sensor measurements, and the approximate nature of algorithmic solutions. Building intelligent machines also requires that they adapt to their environment. Few things are more frustrating than machines that repeat the same mistake over and over again [67].

Based on the aforementioned data discussed in the previous sections, we have chosen to define the model using one of the following statistical approaches that are usually used in robotics:

- Fundamentals of Uncertainty
  - Cox Axioms
  - Maximum Entropy
  - Online algorithms, regret minimisation
- Probability Theory
  - Probability is the branch of mathematics that studies the possible outcomes of given events together with the outcomes' relative likelihoods and distributions. In common usage, the word "probability" is used to mean the chance that a particular event (or set of events) will occur expressed on a linear scale from 0 (impossibility) to 1 (certainty), also expressed as a percentage between 0 and 100%. The analysis of events governed by probability is called statistics [62]
- Bayesian Filters
  - Bayesian analysis is a statistical procedure which endeavours to estimate parameters of an underlying distribution based on the observed distribution. Begin with a "prior distribution" which may be based on anything, including an assessment of the relative likelihoods of parameters or the results of non-Bayesian observations. In practice, it is common to assume a uniform distribution over the appropriate range of values for the prior distribution [62].
- Monte Carlo
  - Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties. The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically. It was named by S. Ulam, who in 1946 became the first mathematician to dignify this approach with a name, in honour of a relative having a

propensity to gamble (Hoffman 1998, p. 239). Nicolas Metropolis also made important contributions to the development of such methods [62].

- Inverse sensors model
- Binary Bayes Filters
- Kalman Filtering
    - An algorithm in control theory introduced by Kalman (1960) and refined by Kalman and Bucy (1961). It is an algorithm which makes optimal use of imprecise data on a linear (or nearly linear) system with Gaussian errors to continuously update the best estimate of the system's current state [62]
- Undirected Graphical models
    - Also called Markov Networks [66].
- Hammersley-Clifford Theorem
    - The Hammersley–Clifford theorem is a result in probability theory, mathematical statistics and statistical mechanics that gives necessary and sufficient conditions under which a positive probability distribution can be represented as a Markov network (also known as a Markov random field). It is the fundamental theorem of random fields [63].
- Gaussian Processes
    - In probability theory and statistics, Gaussian processes are a family of statistical distributions (not necessarily stochastic processes in which time plays a role). In a Gaussian process, every point in some input space is associated with a normally distributed random variable. Moreover, every finite collection of those random variables has a multivariate normal distribution.
    - A Gaussian process is a statistical distribution X_t, $t \in T$, for which any finite linear combination of samples has a joint Gaussian distribution [64].
- Kernel Embeddings, kernel Bayes rules
    - In machine learning, the kernel embedding of distributions (also called the kernel mean or mean map) comprises a class of nonparametric methods in which a probability distribution is represented as an element of a reproducing kernel Hilbert space (RKHS) [65].
- Fast Approximate Kernel methods applications

The choice of statistical technique that will be used is based on numerous experiments that will consist of:

1. Use the complete set of features defined in Table 1 below (derived from Section 3.2) on all the different techniques mentioned.

2. The best X (e.g., 3) statistical techniques are chosen depending on the results obtained.
3. A feature selection process then takes place on the statistical techniques outlined in Step 2, where this is repeated on a smaller number of feature sets until a final set of features are selected. Such feature reduction will produce the final set of features which consist of the most relevant and useful ones for the purposes of our task in determining the health status (e.g., top 10).

We are aware that MARIO may not generate enough data, especially for some of the cited learning methods. Given this assumption and given that we need to achieve a consistent result within the time frame of the project, the statistical method to be implemented could be the same that led IRCCS to the computation of MPI [26] by enriching the model from elements resulting from human-robot interaction and the possibility of obtaining additional parameters from the sensors connected to MARIO.

In its classical formulation MPI can be described by the formula:

$$\sum_{i=1}^{N} k_i T_i$$

where T_i are the results obtained making a simple average of the answers of patients to the questions pertaining each single domain (ADL, iADL, etc.). It can have values ranging from 0 to 1.

A first extension of this model could be expressed by the following formula

$$\sum_{i=1}^{N} (k_i \bar{T}_i \pm \varepsilon_i)$$

In this case $\bar{T}_i$ is different from T_i because MARIO could interpret not every answer in the correct way. MARIO will in fact give a confidence level to the answers of the patient but if this is below a given threshold the answers will be discarded as not understood by the robot. If this is the case T_i will be computed as the average value taken on a smaller number of items in respect with the T_i computed in the case of the "traditional" eight-domain MPI computation.

This model includes also ε_i, a sort of misclassification error, i.e. MARIO gives a given answer a confidence level above the threshold (so for MARIO the answer is classified as one of the possible answers) but MARIO misunderstands the real meaning of the answer. This error can be computed considering an experimental setting in which the patient gives its answers to MARIO, the answers are recorded and then double checked by a healthcare professional to assess the differences between his MPI calculation and the one computed by MARIO. In all cases we consider in the model the MPI calculated by a healthcare professional as a sort of golden standard, i.e. in a "perfect" scenario with all ε_i equaling zero and the formulas above coincide.

The linear model to calculate the health index, the MARIO Robot based Multidimensional Prognostic Index, can be represented as:

$$MPI_R = \sum_{i=1}^{N} (k_i \bar{T}_i \pm \varepsilon_i) + \sum_{j=1}^{M} k_j R_j$$

Where R_j is, as an example:

R_1, Heart Rate

R_2, Respiratory Rate

R_3, Blood Pressure

R_4, App performance (for examples Flash Games performance)

Basing on conventional cut-off points derived from the literature, present guidelines and medical recommendations, we will consider three possible levels for each R_j (0, no problems, 0.5 minor problems, 1, major problems).

For R_j we'll consider present guidelines and medical recommendations for the elderly.

For R_4-like variables, i.e. the ones derived from patient's performances at completing some task, following the same approach, we will calculate the cut off points considering the mean values of an elderly subject without cognitive decline, all the performances obtained by the patients during the trial periods and dividing them into determined classes thanks to healthcare professionals supervision.

This model will be extended to encompass the real R-like variables to be determined on what MARIO will offer in terms of apps relevant from a medical point of view and in terms of data gathered by the external sensors.

Considering the technologies reported in Section 3.2 we present in the following table the complete set of data we will include in the model and their mapping into existing scales/domains validated, to our knowledge, in the literature in order to determine their efficacy in the corresponding domain and consequently their impact on the life of the subjects and their caregivers.

Sensor	**Variable**	**Format**	**Existing Test/Scale/Domain**
Beddit	Total sleep time	time	Neuropsychiatric Inventory, Hamilton Rating Scale for Depression
	Time to fall asleep	time	Neuropsychiatric Inventory, Hamilton Rating Scale for Depression
	Sleep efficiency	percentage	None

	Awake time	time	Neuropsychiatric Inventory, Hamilton Rating Scale for Depression
	Snoring time	time	Physical health status
FitBit	Number of steps	integer	Exton-Smith Scale, Tinetti Balance Scale
	Heart rate	frequency	Physical health status
	Burned calories	energy consumption	Mini Nutritional Assessment
	Walking distance	lenght	Exton-Smith Scale, Tinetti Balance Scale
	Time of rest	time	Tinetti Balance Scale
ZephyrLIFE	Respiratory rate	frequency	Physical health status
	R-R interval	time	Physical health status

The vital parameters, integrated with the other data, will be essential to determine the frequency and deepness of the subject evaluation performed by the robot and to improve the accuracy, in term of sensibility and specificity, of the health index and consequently the kind of action to be performed. A great effort will be spent to reduce at minimum not required interactions in order to improve the platform acceptability. As with traditional MPI calculation, the values of MPI_R will be expressed as an absolute value ranging from 0 to 1 and to simplify the data interpretation it will be classified in three risk classes: low (MPI_R value ≤0.33), moderate (MPI_R between 0.34 and 0.66) and severe (MPI_R >0.66). This is according to the rule-based indices developed in the past for the evaluation of multidimensional impairment in elderly subjects [26].

In particular, to determine the cut off points for the aforementioned risk classes an analysis will be performed to set them as to best separate the survival curves for patients with low, moderate and severe MPI. Given our formulation of MPI_R it has to be demonstrated that an “equal” distribution of cut off points (i.e. low risk with MPI_R value ≤0.33, moderate risk with MPI_R between 0.34 and 0.66 and severe with MPI_R >0.66) is the best one.

The relationship between MPI score group and time to death will be analyzed with a Cox proportional hazard regression model. A Martingale residual analysis will be used to explore the functional form of the relationship between MPI classes and mortality.

A particular focus will be given to formally assessing the prognostic value of MPI_R with respect to its prognostic value if compared to the prognostic value of each of its components, through a logistic model. A calibration and discrimination of the model will be assessed comparing the predicted mortality with the actual one and computing the ROC curves for each cohort (development and validation).

MPI_R can be used to reach various goals. In clinical settings, it can be used to target personalized interventions for patients belonging to different groups of MPI_R value. Once the MARIO robot functional capabilities will be known, it could be also used to select which patients could benefit most by the use of the robot. This can be done only after a data intensive validation phase of the model in different cohort of patients before it can deliver results back to clinical staff in a meaningful way such that they can be used in decision making or long term assessment of methods.

4. Conclusion

This document showed the development of the statistical method to compute MPI to be implemented in the MARIO Robot through the CGA-App. The statistical method developed represents a starting point to define a validated model representing the health status in different settings: Nursing Home, Hospital, and in a Community-Dwelling. The sets of variables, sensors and the apps selected represent one of the possible choices to compute the index. Eventually, based on the real development of the robot, results in literature, software and sensors it will be possible to update the methods using MARIO's "modular" structure. Moreover, the validation process will give significant information on measurement error and on the limits of the robot in term of acceptability and behavioural capability. An interesting aspect of this module is its strict relationship with the other modules apparently far from each other, due to its multidimensional nature. So for example, also monitoring results related to weather, reading news, listen music could be integrated autonomously in the model considering them as a part of the health status. The capability of this system to monitor a large number of parameters autonomously over time opens new prospects in terms of intervention and action to slow down cognitive impairment and ageing trying to improve functional independence.

5. References

1. Katz S, Downs TD, Cash HR, Grotz RC (1970) Progress in the development of an index of ADL. *Gerontologist* **10**, 20–30.

2. Lawton MP, Brody EM (1969) Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* **9**, 179-186.

3. Pfeiffer E (1975) A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc* **23**, 433-441.

4. Parmelee PA, Thuras PD, Katz IR, Lawton MP (1995) Validation of the Cumulative illness rating scale in a ge6riatric residential population. *J Am Geriatr Soc* **43**, 130-137.

5. Vellas B, Guigoz Y, Garry PJ, Nourhashemi F, Bennahum D, Lauque S, Albarede JL (1999) The Mini Nutritional Assessment (MNA) and its use in grading the nutritional state of elderly patients. *Nutrition* **15**, 116-122.

6. Bliss MR, McLaren R, Exton-Smith AN (1966) Mattresses for preventing pressure sores in geriatric patients. *Mon Bull Minist Health Public Health Lab Serv* **25**, 238-268.

7. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy* 2005; **10**: 45–53.

8. Hannes K, Macaitis K. A move to more systematic and transparent approaches in qualitative evidence synthesis: update on a review of published papers. *Qual Res* 2012; **12**: 402–42.

[9] Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S (2011) A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol* **64**(7), 749-759.

[10] Schonberg MA, Davis RB, McCarthy EP, Marcantonio ER (2009) Index to predict 5-year mortality of community-dwelling adults aged 65 and older using data from the National Health Interview Survey. *J Gen Intern Med* **24**(10), 1115-1122.

[11] Schonberg MA, Davis RB, McCarthy EP, Marcantonio ER (2011) External validation of an index to predict up to 9-year mortality of community-dwelling adults aged 65 and older. *J Am Geriatr Soc* **59**(8), 1444-1451.

[12] Romano PS, Roos LL, Jollis JG (1993) Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol* **46**(10), 1075-1079.

[13] Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* **40**(5), 373-383.

[14] van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ (2009) A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care* **47**(6), 626-633.

[15] Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. *Med Care* **36**(1), 8-27.

[16] Mazzaglia G, Roti L, Corsini G, Colombini A, Maciocco G, Marchionni N, Buratti E, Ferrucci L, Di Bari M (2007) Screening of older community-dwelling people at risk for death and hospitalization: the Assistenza Socio-Sanitaria in Italia project. *J Am Geriatr Soc* **55**(12), 1955-1960.

[17] Carey EC, Walter LC, Lindquist K, Covinsky KE (2004) Development and validation of a functional morbidity index to predict mortality in community-dwelling elders. *J Gen Intern Med* **19**(10), 1027-1033.

[18] Carey EC, Covinsky KE, Lui LY, Eng C, Sands LP, Walter LC (2008) Prediction of mortality in community living frail elderly people with long-term care needs. *J Am Geriatr Soc* **56**(1), 68-75.

[19] Lee SJ, Lindquist K, Segal MR, Covinsky KE (2006) Development and validation of a prognostic index for 4-year mortality in older adults. *JAMA* **295**(7), 801-808.

[20] Porock D, Oliver DP, Zweig S, Rantz M, Mehr D, Madsen R, Petroski G (2005) Predicting death in the nursing home: development and validation of the 6-month Minimum Data Set mortality risk index. *J Gerontol A Biol Sci Med Sci* **60**(4), 491-498.

[21] Porock D, Parker-Oliver D, Petroski GF, Rantz M (2010) The MDS Mortality Risk Index: the evolution of a method for predicting 6-month mortality in nursing home residents. *BMC Res Notes* **3**, 200.

[22] Flacker JM, Kiely DK (2003) Mortality-related factors and 1-year survival in nursing home residents. *J AmGeriatr Soc* **51**(2), 213-221.

[23] Flacker JM, Kiely DK (1998) A practical approach to identifying mortality-related factors in established longterm care residents. *J Am Geriatr Soc* **46**(8), 1012-1015.

[24] Kruse RL, Parker Oliver D, Mehr DR, Petroski GF, Swenson DL, Zweig SC (2010) Using mortality risk scores for long-term prognosis of nursing home residents: caution is recommended. *J Gerontol A Biol Sci Med Sci* **65**(11), 1235-1241.

[25] Inouye SK, Bogardus ST Jr, Vitagliano G, Desai MM, Williams CS, Grady JN, Scinto JD (2003) Burden of illness score for elderly persons: risk adjustment incorporating the cumulative impact of diseases, physiologic abnormalities, and functional impairments. *Med Care* **41**(1), 70-83.

[26] Pilotto A, Ferrucci L, Franceschi M, D'Ambrosio LP, Scarcelli C, Cascavilla L, Paris F, Placentino G, Seripa D, Dallapiccola B, Leandro G (2008) Development and validation of a multidimensional prognostic index for one-year mortality from comprehensive geriatric assessment in hospitalized older patients. *Rejuvenation Res* **11**(1), 151-161.

[27] Di Bari M, Balzi D, Roberts AT, Barchielli A, Fumagalli S, Ungar A, Bandinelli S, De Alfieri W, Gabbani L, Marchionni N (2010) Prognostic stratification of older persons based on simple administrative data: development and validation of the “Silver Code,” to be used in emergency department triage. *J Gerontol A Biol Sci Med Sci* **65**(2):159-164.

[28] Fischer SM, Gozansky WS, Sauaia A, Min SJ, Kutner JS, Kramer A (2006) A practical tool to identify patients who may benefit from a palliative approach: the CARING criteria. *J Pain Symptom Manage* **31**(4), 285-292.

[29] Teno JM, Harrell FE Jr, Knaus W, Phillips RS, Wu AW, Connors A Jr, Wenger NS, Wagner D, Galanos A, Desbiens NA, Lynn J (2000) Prediction of survival for older hospitalized patients: the HELP survival model: Hospitalized Elderly Longitudinal Project. *J Am Geriatr Soc* **48**(5)(suppl), S16-S24.

[30] Walter LC, Brand RJ, Counsell SR, Brand RJ, Counsell SR, Palmer RM, Landefeld CS, Fortinsky RH, Covinsky KE (2001) Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. *JAMA* **285**(23), 2987-2994.

[31] Dramé M, Novella JL, Lang PO, Somme D, Jovenin N, Lanièce I, Couturier P, Heitz D, Gauvain JB, Voisin T, De Wazières B, Gonthier R, Ankri J, Jeandel C, Saint-Jean O, Blanchard F, Jolly D (2008) Derivation and validation of a mortality-risk index from a cohort of frail elderly patients hospitalised in medical wards via emergencies: the SAFES study. *Eur J Epidemiol* **23**(12), 783-791.

[32] Levine SK, Sachs GA, Jin L, Meltzer D (2007) A prognostic model for 1-year mortality in older adults after hospital discharge. *Am J Med* **120**(5):455-460.

[33] Youngwerth J, Fischer SM, Min S-J, Kutner JS (2011) Who would benefit from a palliative care consult in the hospital? CARING About Prognosis (CAP) (313-C). *J Pain Symptom Manage* **41**(1), 192-193.

[34] Desai MM, Bogardus ST Jr, Williams CS, Vitagliano G, Inouye SK (2002) Development and validation of a risk adjustment index for older patients: the high-risk diagnoses for the elderly scale. *J Am Geriatr Soc* **50**(3), 474-481.

[35] Inouye SK, Wagner DR, Acampora D, Horwitz RI, Cooney LM Jr, Hurst LD, Tinetti ME (1993) A predictive index for functional decline in hospitalized elderly medical patients. *J Gen Intern Med* **8**(12), 645-652.

[36] Sancarlo D, D'Onofrio G, Franceschi M, Scarcelli C, Niro V, Addante F, Copetti M, Ferrucci L, Fontana L, Pilotto A (2011) Validation of a Modified-Multidimensional Prognostic Index (m-MPI) including the Mini Nutritional Assessment Short-Form (MNA-SF) for the prediction of one-year mortality in hospitalized elderly patients. *J Nutr Health Aging* **15**(3), 169-173.

[37] Rozzini R, Sabatini T, Trabucchi M (2001) Prediction of 6-month mortality among older hospitalized adults. *JAMA* **286**(11), 1315-1316.

[38] Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G (2004) Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol* **59**, 255–263.

[39] Saliba D, Elliott M, Rubenstein LZ, Solomon DH, Young RT, Kamberg CJ, Roth C, MacLean CH, Shekelle PG, Sloss EM, Wenger NS (2001) The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. *J Am Geriatr Soc* **49**, 1691–1699.

[40] Fried LP, Kronmal RA, Newman AB, Bild DE, Mittelmark MB, Polak JF, Robbins JA, Gardin JM (1998) Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. *JAMA* **279**, 585–592.

[41] Fillenbaum GG, Pieper CF, Cohen HJ, Cornoni-Huntley JC, Guralnik JM (2000) Comorbidity of five chronic health conditions in elderly community residents: determinants and impact on mortality. *J Gerontol* **55**, M84–M89.

[42] Inouye SK, Peduzzi PN, Robison JT, Hughes JS, Horwitz RI, Concato J (1998) Importance of functional measures in predicting mortality among older hospitalized patients. *JAMA* **279**, 1187–1193.

[43] Consensus Development Panel (David Solomon, Chairman) (2003) National Institutes of Health Consensus Development Conference Statement: Geriatric assessment methods for clinical decision-making. *J Am Geriatr Soc* **51**, 1490–1494.

[44] Rubenstein LZ (2004). Comprehensive Geriatric Assessment: from miracle to reality. *J Gerontol Med Sci* **59**A, 473–477.

[45] L'informatore farmaceutico (2004) OEMF International, Milano 2002. Guidelines for ATC Classification, NLN publication no.16. Uppsala, Sweden, Nordic Council on Medicines, 1985.

[46] Thomas VS, Rockwood K, McDowell I (1998) Multidimensionality in instrumental and basic activities of daily living. *J Clin Epidemiol* **51**, 315–321.

[47] Rockwood K, Stadnyk K, MacKnight C, McDowell I, Hébert R, Hogan DB (1999) A brief clinical instrument to classify frailty in elderly people. *Lancet* **353**, 205–206.

[48] http://www.beddit.com/?nab=1&nabe=6005962488414208:1. Accessed June 30, 2016.

[49] Paalasmaa J, Leppäkorpi L, Partinen M (2011) Quantifying respiratory variation with force sensor measurements. *Conf Proc IEEE Eng Med Biol Soc* **2011**, 3812-3815.

[50] Paalasmaa J, Waris M, Toivonen H, Leppäkorpi L, Partinen M (2012) Unobtrusive online monitoring of sleep at home. *Conf Proc IEEE Eng Med Biol Soc* **2012**, 3784-3788.

[51] Perez-Macias JM, Jimison H, Korhonen I, Pavel M (2014) Comparative assessment of sleep quality estimates using home monitoring technology. *Conf Proc IEEE Eng Med Biol Soc* **2014**, 4979-4982.

[52] Paalasmaa, J, Toivonen, H, Partinen, M (2014) Adaptive Heartbeat Modeling for Beat-to-Beat Heart Rate Measurement in Ballistocardiograms. *IEEE Journal of Biomedical and Health Informatics* **19** (6), 1945-1952.

[53] Keytel LR, Goedecke JH, Noakes TD, Hiiloskorpi H, Laukkanen R, van der Merwe L, Lambert EV (2007) Prediction of energy expenditure from heart rate monitoring during submaximal exercise. *J Sport Sci* **23**, 289–297.

[54] Luke A, Maki KC, Barkey N, Cooper R, McGee D (1997) Simultaneous monitoring of heart rate and motion to assess energy expenditure. *Med Sci Sport Exerc* **29**, 144–148.

[55] Parak J, Korhonen I (2014) Evaluation of wearable consumer heart rate monitors based on photoplethysmography. *IEEE*, pages 3670–3673.

[56] https://www.zephyranywhere.com/media/pdf/ZLHM_BR_P_ZephyrLIFE-Brochure_20131108_9700.0213.v2.pdf, Accessed July 1, 2016

[56] Bianchi W, Dugas AF, Hsieh YH, Saheed M, Hill P, Lindauer C, Terzis A, Rothman RE (2013) Revitalizing a vital sign: improving detection of tachypnea at primary triage. *Ann Emerg Med* **61**(1), 37-43.

[57] Chen IM, Phee SJ, Luo Z, Lim CK (2011) Personalized biomedical devices & systems for healthcare applications. *Front Mech Eng* **6**(1), 3–12.

[58] Kokonozi A, Astaras A , Semertzidis P, Michail E, Filos D, Chouvarda J, Grossenbacher O, Koller JM, Leopoldo R, Porchet JA, Correvon M, Luprano J, Sipilä A, Zamboulis C, Maglaveras N (2010) Development and clinical evaluation of a physiological data acquisition device for monitoring and exercise guidance of heart failure and chronic heart disease patients. *Computing in Cardiology*, pages 1099 – 1102.

[59] Ellis G, Langhorne P (2004) Comprehensive geriatric assessment for older hospital patients. *Br Med Bull* **71**, 45-59.

[60] Tak SH, Zhang H, Patel H, Hong SH (2015) Computer Activities for Persons With Dementia. *Gerontologist* **55**, S40-S49.

[61] https://www.alzheimers.org.uk/dementiafriendlycommunities. Accessed July 1, 2016.

[62] Mathworld Wolfram - http://mathworld.wolfram.com

[63] Lafferty, John D.; Mccallum, Andrew (2001). "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data"

[64] Rasmussen, C. E. (2004). "Gaussian Processes in Machine Learning". Advanced Lectures on Machine Learning. Lecture Notes in Computer Science 3176. pp. 63–71. doi:10.1007/978-3-540-28650-9_4

[65] A. Smola, A. Gretton, L. Song, B. Schölkopf. (2007). A Hilbert Space Embedding for Distributions. Algorithmic Learning Theory: 18th International Conference. Springer: 13–31.

[66] http://www.cedar.buffalo.edu/~srihari/CSE574/Chap8/Ch8-PGM-Undirected/9.1-UndirectedGraphs.pdf

[67] CS8803: Statistical Techniques in Robotics http://www.cc.gatech.edu/~bboots3/STR-Spring2015/

Table 1. Variables included in the model

In the following table we present the variables that will be included in the model but, given the flexibility of the model and possible development of novel MARIO apps this sets could be expanded to comprise also other variables that will be judged valuable from a medical point of view.

Names	Value	Measurement type	Data Acquisition
ADL	Scale (0-6) or Binary if we consider the 6 items constituting the scale	Periodic	Data acquired by MARIO through speech to text module
IADL	Scale (0-8) or Binary if we consider the 8 items constituting the scale	Periodic	Data acquired by MARIO through speech to text module
MNA	Scale (0-30) or numeric if we consider the 18 items constituting it	Periodic	Data acquired by MARIO through speech to text module or direct input
CIRS	Numeric	Periodic	Data acquired by MARIO through speech to text module or direct input
ESM	Scale (5-20) or numeric if we consider the 5 items constituting it	Periodic	Data acquired by MARIO through speech to text module or direct input
# of drugs assumed	Numeric	Periodic	Data acquired by MARIO through speech to text module or direct input
MMSE	Scale (0-30) or numeric if we consider the 11 items constituting it	Periodic	Data acquired by MARIO through speech to text module or direct input
Co-habitation state	Qualitative (3 possible options)	Once	Data acquired by MARIO through speech to text module or direct input
Age	Numeric	Once	Data acquired by MARIO through speech to text module or direct input

Sex	Binary	Once	Data acquired by MARIO through speech to text module or direct input
MARIO sensor network acquired data	Numeric	Continuously/Periodically	Data acquired by sensors as specified in the table presented in paragraph 3.3
MARIO App usage/performance	Numeric (for example, patient performance on MARIO Flash game app)	Periodically	Data acquire by MARIO App modules