DR 1.5: Design, ontology and evaluation of PAL prototype supporting shared child-caregiver responsibility

Olivier Blanson Henkemans, Elettra Oleari, Francesca Sacchitelli, Clara Pozzi, Francesco Sardu, Daniele Baranzini, Willeke van der Vught, Mark Neerincx

TNO, Soesterberg, The Netherlands
Fondazione Centro San Raffaele, Milan, Italy
Ospedale San Raffaele, Milan, Italy
‹Olivier.blansonhenkemans@tno.nl›

Project, project Id: EU H2020 PAL / PHC-643783

Project start date: March 2015 (48 months)

Due date of deliverable: March 31, 2019

Actual submission date: March 31, 2019

Lead partner: TNO

Revision: final

Dissemination level: PU

In the third design-test cycle of PAL, we refined the prototype based on the outcomes of the previous cycles and additional user feedback sessions. The new protype was improved on its overall *usability*, *goal setting & achievements* framework, personalised *child-agent interactions*, *educational content*, *gamification* and the *dashboards for parents* (PAL Inform). The underpinning knowledge and rationale was recorded in the ontologies and Socio-Cognitive Engineering Tool (SCET; see PAL deliverable D6.13). In an experiment, care with PAL was compared with care as usual for the diabetes self-management of children with T1DM (age 7-14), focusing on (1) determinants of behaviour (knowledge, awareness and skills), (2) the actual self-management behaviour (i.e., performing self-management activities and adhering to personalised diabetes goals), and (3) the treatment outcomes (i.e., glycaemic control). PAL showed to improve children's knowledge, where time of PAL usage correlated positively with this knowledge. In addition, PAL proved to increase the self-care score (particularly for the younger children), and enhanced the diabetes related quality of life score of children. Compared to this PAL prototype version (3.0), children spent their time more efficiently with an improved version (3.5).

Acronyms

T1DM	Type 1 Diabetes Mellitus
WP1	Work Package 1
HCPs	Health Care Professionals
y.o.	years old
NA	not available (or not necessary)

1 Executive Summary

The overall objective of WP1 is (I) to ensure that user needs, human factors knowledge and technical opportunities are addressed well throughout the development process, so that the PAL support functions fit with stakeholders' need for situated care (in the Netherlands and Italy) and (II) to develop and maintain a reusable, evidence-based, design knowledge-base on these functions and their effect on self-management. During the third and final cycle, WP1 consisted of the last cycle of the user-centred evaluations for evidence-based blended care solution, covering a personalised robot and its avatar and various educative mHealth applications for children with T1DM, their parents and HCP. During the final experiment, behavioural changes (e.g. regime adherence and responsibility) and clinical health outcomes (e.g. Quality of Life and glycaemic control) were measured. The main objective of this year was to capture the difference in behavioural and clinical outcomes between participant that used the PAL system and participants that received care as usual. Besides that, technological factors and usability of the PAL system were measured throughout the experiments which enable WP1 to compare the usability of two versions of the PAL application. This document discusses the research and development activities in of the third and final cycle. The focus was on the evaluation of the effect of the PAL personalized educational framework on specific T1DM determinants, especially focusing on the acquisition of a shared child-caregiver responsibility, a better awareness on the correct diabetes self-management behaviours and glycaemic control. To this extent a final long-term evaluation involving children in Italy and the Netherlands, their parents and reference HCPs, was set-up.

The first step was generation and refinement of the PAL design specification, for PAL 3.0, to achieve an ongoing and impactful, use of the system by children, in regard to T1Dm self-management. To achieve this aim, we introduced or improved the following aspects in the PAL system, derived from results of previous cycles:

- *General usability of the MyPAL app* - in order to improve children's engagement with the system and guarantee a more intuitive and smooth interaction with it;
- *Goal setting & achievements* – the PAL educational framework based on achievements/goals and tasks has been thoroughly revised, enriched in contents and linked to real life situations which children with T1DM could face;
- *Enriched interaction* – the child-agents (both embodied and virtual) interaction have been improved, in order to provide a more personalised (via episodic memory), sensitive (e.g. through affective explanations), contextualised (environment-related) feedbacks of the robotic pal, so as to strengthen the bond with the child and maximise the engagement;
  - *Additional educational material* – the MyPAL app games have been enriched in contents and levels, as well as a new video-viewer and two

new games (i.e. the Memory game and the Dance) has been added to its suite;

- *A gamification approach* – the MyPAL app has been equipped with an activity-based reward system which enables them to unlock additional features by which personalising their MyPAL app. This strategy has been undertaken in order to increase MyPAL adherence for long term usage;
- *Monitoring functionality for parents* (PAL Inform) – the last component of the PAL System, a monitoring dashboard dedicated to parents through which they could follow their children's educational progress.

Subsequently, we evaluated the improved PAL system. The aim was to evaluate the effect of the use of PAL version 3 on diabetes self-management in children and preadolescents with T1DM. These effects were measured at three primary levels, in children with PAL 3.0 and care as usual (CAU) respectively: (1) determinants of behaviour (knowledge, awareness and skills), (2) the actual self-management behaviour (i.e., performing self-management activities and adhering to personalised diabetes goals), and (3) the treatment outcomes (i.e., glycaemic control). In regard to diabetes knowledge, children in the intervention group showed a stronger increase after three months, than children in the wait-list group (p=.05). Moreover, we found a correlation between time paying with the MyPAL app and children's knowledge. Also, children in the intervention groups had a stronger increase in self-care score (p=.01). Moreover, younger children in the intervention group showed a stronger increase in score, in comparison with their older peers in the intervention group. We did not find an effect of PAL on parental stress and glucose regulation (including HbA1c and % in healthy range). However, we did find an effect on diabetes related quality of life in children (p=.02).

In addition, we conducted an implementation study in which we compared PAL 3.0 with an improved version 3.5. The latter contained functions aiming at further contributing to PAL's usability, key working elements regarding diabetes self-management goal-achievement and finally ongoing use. When we look at the difference between PAL 3.0 and 3.5, we found that children spent their time on the system more efficiently working toward their learning goals in version 3.5 of the PAL system. Also, children in the wait-list group used the MyPAL app more than children in the PAL group, indicating a higher level of motivation to use with PAL 3.5. In regard to participation in the study, children positively evaluated PAL and we did only had one child who dropped out. Moreover, the majority of children participating in the intervention group with PAL in phase 1 also continued to participate in phase 2. Parents were also likely to engage in the MyPAL app with their child, such as playing an educational game together. Children and parents provided also some suggestions on how to further improve PAL, such as increasing the variety of games and increased personalization of the games' difficulty level.
Based on these results, we can conclude that a blended care solution such as PAL, including personal robot and diabetes education app with avatar, and dashboards for HCP and parents, can have a positive effect on diabetes self-management determinants and behaviour in children with T1DM. Also, through application of gamification principles, goal setting, and enriched CRI, children are more positive about the system and motivated to use it. Although we did not find an effect on

health outcomes within three months, it may be that with a prolonged use the PAL system, will have positive effects on glucose regulation.
These findings are valuable for the engineering community and for the medical and patient community, consisting of children with chronic illness such as T1DM and their families. For the engineering community, our results show that an iterative (i.e., cyclic) human-centred approach, such as Socio-Cognitive Engineering, can aid to develop usable and useful systems through strong involvement of end users and stakeholders, with different background (e.g., Dutch and Italian) and characteristics (e.g., age, gender, number of years with illness, and type of medical device).. For the medical and patient community, results show that a blended care solution such as PAL, can have a positive impact not only self-management determinants, such as knowledge, but also on self-management behaviour and patients' quality of life. This provides important guidance for future development and implementation of such solutions on a larger scale (e.g., in other European countries) and in in other domains (e.g., other chronic illnesses).

2 Embedding in project

Through the entire project lifetime, the domain experts and end-users of the PAL system have been heavily involved in the design (e.g., refinement of the user requirements) and evaluation activities (e.g., determination of the outcome measures and participant in the experiment). In an iterative way, we called on their expertise, experiences, opinions, preferences, feedbacks and behavioural responses, to derive new and refine "old" functional requirements. This was during formal evaluation Cycles 1 and 2 (spring/summer of 2016 and 2017) and the complementary needs assessment and feasibility study activities during the diabetes camps, both in the Netherlands and Italy (summer/autumn of 2015, 2016 and 2017). Results of these activities lead us to the implementation of the final Pal System version 3, evaluated in spring-autumn 2018 (see Section 5, 6 and 7).

2.1 Role of the WP

WP1 contributes to the development of the PAL System by the application of a human-centred development approach, in which the domain and stakeholder's needs and expertise, as well as the relevant theories and human factors knowledge, are systematically addressed to design and evaluate the system. Expected effects are related to the implemented functions and specific instruments are used to measure these. Main novel functionalities are related to amongst others gamification principals, goal setting & achievement, enriched child-robot interaction and parental monitoring. The impact of these functionalities were mainly assessed through blood glucose measurements, surveys, and log-data from the system.

3 Tasks, objectives, results

3.1 Planned work

The overall goals for WP1 are (1) to ensure that user needs, human factors knowledge and technological opportunities are addressed well throughout the development process, so that the PAL support functions fit with stakeholders needs for situated care (in the Netherlands and Italy), and (2) to develop and maintain a reusable, evidence-based, design knowledge-based on these functions and their effect on self-management.

In this final period (September 2017 to February 2019), we entered the third and final design and evaluation cycle of the PAL System. This time the focus of our research was to evaluate the effect of the PAL personalized educational framework on specific T1DM determinants, especially focusing on the acquisition of a shared child-caregiver responsibility, a better awareness on the correct diabetes self-management behaviours and glycaemic control. To this extent a final long-term evaluation involving children in Italy and the Netherlands, their parents and reference HCPs, was envisioned.

3.2 Deviations from workplan and necessary steps taken

The principal deviation from the workplan consisted of an extension of the Cycle 3 experiment. Based on our power calculation at the onset of the study, our aim was to include 60 children. To achieve this number we recruited children the hospitals involved in the study, the Netherlands (Meander MC and Ziekenhuis Gelderse Vallei) and one hospital in Italy (Ospedale San Raffaele). Moreover, we agreed with another Dutch hospital VUmc to recruit approximately 20 children amongst their population. Unfortunately, the location of the diabetes clinic for children was closed, due to financial issues, before we could start the study. As a result, we recruited only 42 children, of which one dropped out at the start of the study. Therefore, we started a second recruitment campaign in the Netherlands and this generated another 8 participants. Due to project planning constraints, they only participated in the first phase of the experiment, the one devoted to the study of the effects on T1DM determinants.

4 Actual work performed

The current PAL activities concerned requirements analysis, ontology engineering and evaluations. These evaluations took place, both in Italy and in the Netherlands, in real life settings: in hospitals, at home and at the diabetes camps, as described in the following.

4.1 Method

The PAL project applies a situated Cognitive Engineering (sCE) methodology to design and test: (1) an Embodied Conversational Agent (ECA) for children, (2) several (educative) child-ECA activities, and (3) dashboards for caregivers. This project entails multi-disciplinary research & development of a blended care system with the involvement of a large diversity of stakeholders in all phases of the R&D cycle. An ontology construction helps to identify (interrelated) key concepts that should be univocally addressed in the design (e.g., requirements), implementation (e.g., dialogues) and evaluations (e.g., goal attainment). Furthermore, it enforces the systematic integration of relevant theories on social, cognitive and an effective processes into the support system (e.g., on bonding, goal-driven learning and emotion). Both formative and summative evaluations are conducted to improve the design in an incremental, human-centered, design process.

4.2 Evaluation

To this extent, we firstly performed feasibility studies regarding the last components/features to be implemented for the third PAL System release (3.0), both with children (at summer and autumn diabetes camps in Italy and the Netherlands) and with expert sessions with health care professionals (HCP) from participating Dutch and Italian hospitals. The results of these complementary investigations have been thoroughly described in *Deliverable 1.4 "Design, ontology and evaluation of PAL prototype supporting diabetes regimen adherence*" [M36], and we will only be briefly recall them in the present document (see Section 5). Then, we did a formal evaluation of the last PAL System release (version 3 – experimental Cycle 3) involving children with T1DM aged 7-14 y.o., in hospital and home settings on a 9 months duration trial both in the Netherlands and Italy (for more details please see Section 7).

For the sake of clarity, in Section 5 we will recall the key results and insights from Cycle 2 experimental activities and from the feasibility studies, to provide an overview of the bases upon which we developed the last mile of our work. In Section 6 we will describe the improvements and the current final status of the PAL System, which was used for the Cycle 3 investigation, which will be summarised in Section 7.

5 Background Analyses

To refine (and include possible new ones) the PAL system specification in view of the final release, we conducted, through the application of the sCE method, a requirements analysis, ontology engineering and evaluations. This was based on the result of previous test cycles and most predominantly the results of the previous experimental Cycle 2 (exerting the PAL system version 2.0) and user needs analysis, design and evaluation (i.e., feasibility studies on PAL 3.0 pilot versions). In this section we discuss the requirements analysis in regard to generate and refine design specification.

5.1 Summary of cycle 2 findings

The experiments at the hospitals in the Netherlands and Italy of the PAL 2.0 version elicited the following main insights (for more details, please see *Deliverable 1.4*).

1. Regarding T1DM-related outcomes, we did not perceive a clear adherence to the self-monitoring self-management activities through the MyPAL Timeline. Nor did we find an effect of the use of the PAL system on the children's quality of life;
2. Children were positive about PAL 2.0 version and bonded with its robotic actor, perceiving the PAL robot and its avatar as similar and as buddies who were supporting and making them happy;
3. However, despite an increased usability was detected since Cycle 1 (PAL version 1.0) by children, a considerable number of participants dropped-out early. We have considered the following possible causes of attrition:

    - Actually the PAL 2.0 back-end was greatly improved back-end, but the System lacked of novelty on the front end;
    - The majority of children already interacted with Embodied Conversational Agents (ECA);
    - The experimental period covered ~5 months, which included months around the summer vacations.

4. Moreover, parents of the participating children were critical, due to practical shortcomings of the use MyPAL app, including technical issues, at home in the daily routine and reported a lack of motivation to use MyPAL in their children as time passed.

5.2 Summary of camps feasibility studies

During the feasibility studies held at the camps in Italy and the Netherlands, we aimed at *(1)* collecting parents and children perceptions and needs, for the development of the Pal Inform dashboard (summer camp Italy); *(2)* refinement of the Objective functionality for increased goal-oriented use of PAL; *(3)* test strategies for goal-oriented use of PAL and increased MyPAL adherence, during long term usage (including, gaming principles and enriched Child-Robot Interaction, CRI) (camp Netherlands).

5.2.1 PAL inform investigation

For the development of *the Pal Inform dashboard*, an user-centered investigation, involving 39 children/pre-adolescents and 45 parents, has been put in place in order to derive coherent and value-sensitive functional specifications. A set of co-design methods has been outlined to create a proper and engaging research environment, including: structured questionnaires and a focus group session for the participating parents, role-play sessions for children (for more details on the experimental evaluations please see *Deliverable 1.4 Annex 2*).

Results showed that (parent's perspective):

- The majority of the parents would like to monitor, through PAL Inform: the glycaemic values (100%) and corresponding timing (83%), as well as the insulin doses (83%) and the amount of CHO intake per meal (100%);
- The duration of eventual sport activities is another important aspect to be possibly monitored in relation to the corresponding glycaemic trends;
- Knowing the emotional status of their children (input into MyPAL) is perceived as an added value to the daily therapy management;
- Parents would like to be aware of the progresses in the T1DM educational path of their children, as long as the fulfilment of their personalized goals, in order to help them (if needed), praise and reward them whenever they accomplish a new goal and understand their actual improvements towards autonomy.

From the children's point of view, instead:

- They agreed on the importance of sharing information regarding glycaemic values, CHO counting and insulin doses;
- An aspect that did not find a great interest was the possibility to share the number and detail of meals away from home;
- With respect to additional information on sports, they were willing to share any relative aspect;

For the emotion-related part, they were prone to share only an overview of their feelings, like an emoticon but not any personal information written the MyPAL diary section.

5.2.2 Enhancing a goal-oriented PAL usage

In regard to the refinement of the Objective functionality for increased goal-oriented use of PAL, two separate bunch of activities were organized with Italian and Dutch HCPs (i.e. paediatric diabetologists and nutritionists) to create Goals which were more meaningful and relevant for children, by relating them to challenges the children face in their lives with diabetes. Firstly, multiple brainstorming sessions were held, during the Italian summer camp stay, where HCPs were presented the current version of the PAL system Goal and Task tree (visualized through the PAL Control dashboard) and, together or individually depending on the cases, they went through each content and corresponding difficulty (knowledge) level, in order to improve the existing goals and find out new ones (for more details please look at *Deliverable 1.4 Annex 2)*. The second part was during a PAL project meeting in Italy, with various Dutch and Italian diabetes doctors and nurses in involved in the project. The session moderated by PAL researches, who extensively were involved in the previous research cycles and had expertise on children's experience and needs in the context of PAL. The aims was to fit the goals with achievements of children in their daily life with T1DM. For example, T1Dm goals and tasks are to know the amount of carbs in food and to practice counting carbs daily. An achievement for children is to have a sleep over with a friend, for which the goal knowing know the amount of carbs is a prerequisite. Following this structure, the HCP allocated goals and tasks to achievements for children. Finally, the tasks to help fulfil goals were linked to activities performed with PAL, such as playing quiz, Break&Sort and Memory, using the Timeline, and watching education video's. As such, these sessions resulted in a new Objective function for PAL, which on the one hand is appealing for children, as it engaged on things they want to achieve in their daily life with T1DM and at the same time aids them to fulfil T1DM self-management goals.

On this topic, a third session was organized with children during the Dutch and Italian camp, to improve the usability of interface of Objective functionality in MyPAL. It consisted of a usability test of the MyPAL Objective functionality as offered in the PAL 2.0 version. Children were asked to go through the Objective pages and perform and comment on different tasks. Various issues were addressed, including inconsistencies, unreliability of difficulty level, lack visibility of system state, and lack of documentation (from more details, please see *Deliverable 1.4 Annex 2).*

At last, in regard to test strategies for goal-oriented use of PAL and increased MyPAL adherence, during long term usage, 22 children, aged 8-12 y.o., at the Dutch camp (October 2017) were presented with various strategies concerning 1) applying gamification principles and 2) stimulating bonding with PAL actor, through repeated interaction with embodied and avatar version of the PAL robot. The main results were:

- Gamification principles were evaluated positively and children could specify which principle they preferred. The most preferred was receiving coins, buying objects such as new backgrounds for the app;
- The robot was more positively evaluated than the avatar, but the robot and MyPAL app with avatar could teach them equally about T1DM. This is relevant, as children can take the MyPAL app home, but not the robot (for the moment).

6 Ongoing development of PAL

6.1 From Cycle 2 to the generation and refinement of PAL 3 design specification

The aim for the generation and refinement of the PAL design specification, for PAL 3.0 was to achieve on ongoing and impactful, use of the system in children, in regard to T1Dm self-management. To achieve this aim, we focused on the following aspects:

- *General usability of the MyPAL app* - in order to improve children's engagement with the system and guarantee a more intuitive and smooth interaction with it;
- *Goal setting & achievements* – the PAL educational framework based on achievements/goals and tasks has been thoroughly revised, enriched in contents and linked to real life situations which children with T1DM could face;
- *Enriched interaction* – the child-agents (both embodied and virtual) interaction have been improved, in order to provide a more personalised (via episodic memory), sensitive through a sentiment-mining approach, contextualised (environment-related) feedbacks of the robotic pal, so as to strengthen the bond with the child and maximise the engagement;
- *Additional educational material* – the MyPAL app games have been enriched in contents and levels, as well as a new video-viewer and two new games (i.e. the Memory game and the Dance) has been added to its suite;
- *A gamification approach* – the MyPAL app has been equipped with an activity-based reward system which enables them to unlock additional features by which personalising their MyPAL app. This strategy has been undertaken in order to increase MyPAL adherence for long term usage;
- *Monitoring functionality for parents* (PAL Inform) – the last component of the PAL System, a monitoring dashboard dedicated to parents through which they could follow their children's educational progresses.

To achieve this aim, we performed the following requirement analysis activities. First, after analysing the data of the experiment in the Dutch and Italian hospital of PAL 2. in cycle 2, we updated the requirements baseline for PAL third release. The PAL researchers went through list of requirements and *(1)* refined existing requirements to ensure improvement of PAL adherence and contribute to goal-oriented self-management; *(2)* added requirements, based on new insights, in regard to usability, interaction, motivation, content and monitoring possibilities. During requirements refinement and augmentation, in addition lessons learned of previous PAL research activities, we also incorporated theory. This included literature on strategies engaging on goal achievement, motivation in children (e.g., gamification), child-robot interaction (CRI) and the role of the family in T1Dm self-management. This generated a baseline for PAL 3 with 46 requirements, which contributed to PAL adherence (e.g., and/or goal-oriented self-management), see overview in Appendix 6.

Second, with a PAL project team of researchers, developers and HCPs, we prioritized the requirements list. As it was not possible to fulfil all requirements before the start of the final PAL evaluation study, for implementation's efforts causes, it was decided to split the release of the third version into two subsequent ones: *PAL 3.0* covered 21 requirements and *PAL 3.5* covered an additional 14. Therefore, we specified two phases in the final experiment (see Section 7), in which we focused for PHASE 1 on a PAL 3.0 which is motivating and contributes to T1DM self-management in children aged 7-14y over a prolonged period of time. The corresponding features implemented were therefore focusing on providing a more rich and engaging T1DM educational support, e.g. the video viewer was added to see informative short videos on T1DM topics, a personalised goal-based selection of the Quiz question was implemented, the new Goal dashboard was put in place, educational feedbacks by the robotic pal were added. And for PHASE 2, we focused on a PAL 3.5 which in addition contains functions contributes to the further implementation of PAL in the health care and home setting, maximising the engagement and the usability of the system. Examples of the corresponding features are: the insertion of the "tip of the day" by the pal agent, the dance game was added, the episodic memory maximised as well as the sentiment mining module. As such, we could develop a PAL 3.5 that covers all the requirements for assisting the child, health professional and parent to advance the self-management of children with type 1 diabetes aged 7-14y.

Finally, with the requirements core functionalities were specified, with use cases and claims. These were used to iteratively develop PAL 3.0 and 3.5. Including expert review and evaluation of mock-ups and prototypes, during diabetes camps in the Netherlands and Italy and expert sessions, pilot of PAL 3.0 in the period before the cycle 3 experiment and the final experiment in Dutch and Italian hospitals.

A total of 11 requirements were not integrated in the PAL 3.0 nor in PAL 3.5, which can be clustered in two groups (see appendix 6): 1) child self-monitoring in MyPAL and 2) interactive data representation in PAL Inform. For these two clusters, we propose the following, in regard to future testing. In regard to self-monitoring in children of data, such as insulin dose, carbs, medicine, finger prick areas, in MyPAL, it is important to find a balance between the benefit of collecting this data and the burden it puts on the child. We found throughout the different cycles of the PAL project, that children disliked self-monitoring, e.g., keeping a diary, and they often do not adhere to the advice of the HCP to do so. In the Netherlands even more than in Italy. This was also the primary argument not to integrate this requirement cluster in MyPAL 3.0/3.5. A suggestion for future development, is to make as much as possible use of existing data, already collected by the child through the meter, sensor or other mobile eHealth-apps. However, this is only possible if companies providing these services do not experience the same non-adherence, and are prepared to share their (raw) data for the PAL system to process and use. This is an important starting point for future research and development, which is maybe not so much a technical issue but a strategic one.

In regard to interactive data representation in PAL Inform, it is important to establish in co-creation with parents, what information they would like to receive, and how it should be presented for them to make the best assessment of the diabetes management of their child and what action to take (with the child together). We anticipate that that between parents there are different preferences, in regard to content (e.g., some parent would like more data and some less) and representation (e.g., some parent understand figures better and some text), and it should be adaptable by the parent. We propose that in future research multiple interfaces of PAL Inform are tested with various parents, on utility, usability and privacy (see section 6.2).

When looking at the results of the final cycle (see Chapter 7), we anticipate that, when succeeding in fulfilling these two clusters of requirements, the PAL system could have a positive effect on ‘shared responsibility child & parents’ (self-management behavior). With PAL 3.0, we did not find a significant effect on this outcome measure. However, if the child would perform more self-monitoring activities, and the parent would have a better view of these activities, it could aid parents and the child to better align their management actions and in time agree how to divide the responsibility (and shift it towards the child over time).

6.2 PAL Inform implementation

On the basis of the user-requirements elicitation described in paragraph 5.2.1, the monitoring dashboard for parents has been included into the PAL System 3.5 release version. It is a software module that enables parents to monitor the learning progress and a customizable aggregated view of the data entered by the children in their MyPAL interactive timeline. Importantly, accordingly to the requirements investigation, no private data of the child is presented in this overview, such as pictures or texts and diary entries. This choice should guarantee a balance between medically and system-usage relevant data and the child’s privacy. For more details, please see Deliverable 2.3 “Parental Monitoring Interface”.

6.3 Goal-oriented PAL

All the outcomes of the usability study have been taken into account to refine the design of the Objective functionality interface for PAL System 3 version (2018 – see Figure 1).

Figure 1. The new Layout of the Achievement ("Traguardi") and related Goal and Tasks visualization

In year 3, we redeveloped the objective section to ensure that the objectives were more personalized and understandable and as such to motivate the children to work towards these objectives with the PAL system. Development of the objectives section was done via multiple structured brainstorm sessions with health care professionals from Meander and Gelderse Vallei that have contact with children with diabetes in the age range of 7 until 14 almost every working day. Figure 2 shows the templates that are filled out during the development sessions. The content was validated by 2 HCP. During the experiments of year 3, we used the new objective section for the first time. As to evaluate how HCP would use such a system, at the onset of the study, we only instructed the HCP how the objectives were structured and how to use the tool, but did not direct them on how many objectives they should assign to a child. We saw that HCP were eager to select rather more objectives than less so that the child could choose which one it would like to do and so that the child was not running out of tasks during the testing period.

While evaluating the assigned and achieved objectives after the experiments of year 3, we found out that children had many tasks assigned and that they were unable to reach them all. This is not surprising since the HCP were rather generous with assigning objectives. However, based on this, we could argue that for future use, it would be better to select less objectives at a time to enable the child to achieve most of them (but be aware to select enough objectives so that the child could keep learning). Besides that, we could enrich the task set so that the tasks is more valuable for a goal and with that, decrease the amount of tasks that need to be done to achieve a goal. Furthermore, if the system is more flexible towards different users, it might enable a functionality were a child could choose a type of task it likes most and with that task, achieve the goal (e.g. one child might like doing a quiz while another child likes the memory game more, both tasks can be appropriate to achieve a goal and enabling the child to choose might be more effective than making the child do both). Additional details are described in Appendix 5.

Goal: Inject insulin

ID (not required)	II2.1, II2.2, II2.3, II3.1, II3.2
Description	The child first has to be able to prepare stuff for the injections with parents. Furthermore it needs to be able to inject his/her self with help and after that without. I can inject insulin with help - I can inject insulin myself
Skill	(see table below)
Achievement(s)	Sleepover, Birthday party, Dinner in restaurant, School camp
Topic(s)	
Dependencies	(see table below)

Skill:

Development level	Skill level	Tasks
2	I know	**QUIZ (II2)** I know when and how to inject insulin.
2	I can do with help	**REAL WORLD TASK** *injection with help*
2	I can do myself	**REAL WORLD TASK** *prepare stuff injection* timeline task (not yet in application): *Declare the body area used* Declare who performed the insulin injection *description: Alternate the areas for the insulin injection* *Perform insulin injection autonomously*
3	I know	**QUIZ (II3)** I know when and how to inject insulin and what to take into account.
3	I can do myself	**REAL WORLD TASK** *do an injection without help,*

Dependencies:

Device	Nationality
☑ Pen ☐ Pump	☑ NL ☑ IT

Goal - Measure blood glucose value

ID (not required)	MBG1.1, MBG1.2, MBG2.1, MBG3.1, MBG3.2
Description	The child needs to learn how to measure his/her blood glucose value within four levels. Currently, this goal is specified within three skill levels (2, 3 and 4) namely, it needs to understand when to measure, how to measure with help and after that how to measure without help. Besides that, within the timeline they can indicate that they know to change fingertips (knowledge level 1) and have performed the test autonomously (knowledge level 4?)
Goal levels + corresponding tasks	(see table below)
Achievement(s)	Swimming, Sleepover, Sports training, Sports game, Play outside, Birthday party, Lazy weekend, Dinner in restaurant, Sleepover, Extra candies , Roller coaster Go on Holiday
Dependencies	no dependencies (both for pen/pump and NL/IT)

Goal levels + corresponding tasks:

Development level	Skill level	Tasks
1	I know	**QUIZ (MB1)** **VIDEO (Watch measure bg VMBG)**
1	I can do with help	**REAL WORLD TASK** *measure with help (strip in meter, lancet in lancing device)*
2	I can do myself	**Timeline - insert BG x times a day for 2 days (italy 6 and netherlands 3 times)** ***(not sure if it will be in 3.0)*** **Timeline - glycemic test before (and in Italy after) a meal.** **Insert the glycemic value in correspondence of each meal** **VIDEO (Watch measure bg VMBG)** **REAL WORLD TASK** *description: without help do a measurement + Measure the glycaemia before every meals*
3	I know	**QUIZ (MB3)** *self-knowing when to measure*
3	I can do myself	***TimeLine - Glycemic test before and after a sports event (not sure if it will be in 3.0)*** Insert a glycemic value before any sport activity *description: Measure the glycaemia before sport*

Figure 2 – Templates of two goals linked to achievements and tasks, content was provided by HCP's. Each goal level is personalized for children with a specific development level (ranging from 1 until 4) and includes tasks that are appropriate for that specific group.

6.4 Ontology engineering

The development and implementation of the PAL ontology has been a continuous activity. The current PAL ontology consists of *eight* sub-ontologies, *seven* of which are "truly" independent and capture "really" different knowledge. *One* further ontology brings them together through the use of hand-written interface axioms, employing axiom constructors such as *rdfs:subClassOf* and *owl:equivalentProperty*, or by posing domain and range restrictions on certain underspecified properties. The TBox (in short, the concepts) and RBox (in short, the hierarchy of the concepts) of the PAL domain stays constant, i.e., will *not* change over time. Only relation instances from the ABox (in short, the value of a concept) might undergo a temporal change, e.g., the weight of a child at certain times, but, e.g., *not* the birthdate.

This independence of the subontology has as advantage that it provides clean sub-ontologies which can be reusable in other projects and/or domains. Besides that, this structure has a practical advantage that different project partners can work on the ontologies simultaneously, without interfering with each other.

- The *PAL Objective Model (POM)* ontology for *Diabetes Mellitus EDUcation (DMEDU)* is describing concepts and their relations related to learning objectives. It describes objectives consisting of achievement(s), learning goal(s) and learning task(s).
- In the *Domain* ontology, concepts are described that are related to the PAL domain, without any particular epistemological background. It describes characteristics of diabetes, the PAL system, the users and locations. In addition, information necessary for specific dialogue modelling in the project have been put into this ontology. These classes and properties constitute the information state of the dialogue components.
- The *Episodic Memory (EpiMem)* ontology is based on the affection-based episode ontology of Lim et al. (2013), which combines the Ontology-based unified robot knowledge (OUR-K; Lim et al., 2011) with temporal episode ontology in order to model event episode knowledge. As the temporal episode ontology, this EpiMem ontology is built around the 5W1H (When, Where, Who, Why, What and How) principle of Han et al. (2013). Using a model that can understand Natural Language and can decompose an input text to match to the concepts of 5W1H, information of an episode can be stored in the ontology.
- The *Agreement* ontology defines policies that enable the generation, activation and closing of agreements between agents, for example, about what information is shared amongst others and what information is not (permissions, prohibitions and obligations). It is inspired by Kayel et al. (2018) and Mioch et al. (2018). An agreement consists of an antecedent that is triggered by a condition (e.g. the debtor is at a location within a certain time frame). If there is an agreement that is triggered by this condition, the consequent will happen (e.g. the debtor has to share/notshare information with a creditor).
- The *Semantics* ontology is tailored to the specific needs of games, played in PAL (quiz & sorting game). For practical purposes, we deviate from

the FrameNet Frame semantics and developed a simple frame semantics, oriented along thematic roles. Among other things, this ontology was meant to be a reference point for natural language generation and interpretation, to specify which frames are used in the project. This should also support multilinguality, linking concrete realizations in the different languages to the abstract concept as, e.g., Multilingual WordNet does.

- The *Affect* ontology is based on the Emotional State Model for Affective Semantics (ESMAS), which was designed especially for the PAL project. ESMAS takes several existing models (James Russell's Circumplex Model of Emotions, the Schachter-Singer theory of emotion and Joseph Forgas' Affect Infusion Model) as a basis and integrates these to describe how Mood and Emotion continuously influence each other. One's emotional state is the direct result of these influences and can be approximated with the ESMAS. The ontology serves three purposes. First, it represents the emotional state of the child and formulate an appropriate response when questioned. Second, it supports the communication of the child's emotions, using common/understandable labels for emotional states. Third, it enables to infer valence and intensity of the child's emotion.
- The *smalltalk* ontology started as background knowledge for small conversations initiated by the robotic actor of the PAL project. The purpose of these interactions is to start *off-activity-talk* (OAT, aka. social talk) with the user. The data for these conversations consists of Starters, Prompts, Disclosures and Closure parts to conduct such a dialogue. In addition, concepts necessary for conducting less scripted dialogues have been added, like parameters that play a role during OAT (intimacy level, topic, or valence), and concepts such as *smalltalk:Liking*, to represent preferences uttered by a user during a conversation, but also preferences for the robotic actor, to equip it with a unique personality. The purpose of this ontology is to provide the data structure specifications for all kinds of OAT related dialogue, which then can be used in later sessions as reference items.
- The *Task* ontology provides a basis for a task world model that can be operationalized in the PAL software. It provides a structured model for task decomposition (tasks and goals), task world modeling (roles, groups and agents) and objects involved in the task (object). The ontology is an adapted version of the ontology for Task World Models of Welie et al., (1998).

This ontology engineering approach proved to be helpful for the incremental development of the PAL knowledge base and PAL's dialogues and user interfaces. First, the creation of the ontology made tacit knowledge of the health-care professionals explicit in a formal (logically correct) model that is interpretable by the relevant human stakeholders *and* the PAL system. It resulted in an extendable set of self-management objectives (focusing on learning), with a coherent and concise structure. From this structure, new content was generated, such as quiz questions, break&sort tasks, new memory "card set", educational videos and timeline tasks, which are all tied together per child consistently (i.e., supporting transparent *personalization*). Second, via the ontologies, social and cognitive theories have been integrated into the PAL system in a transparent and verifiable way (e.g., affect, memory, agreement and explanation). Third, the content with its structure of the dashboards and dialogues are based on the ontology. Finally, it should be noted that ontology engineering is an iterative process. Draft ontologies for *Feedback* and *Explanations* have been developed for PAL, which are being completed and validated in current research.

7 Evaluation of integrated prototype

7.1 Experiment

Objective: The primary objective of the cycle 3 experiment was to evaluate the effect of the PAL-system on diabetes self-management determinants and behaviour and as a result on the health of children with T1DM. The PAL-System consisted of a robot and a MyPAL-app, containing an avatar of the robot. Children could play educational activities with the robot in the hospital and with the avatar at home and have talks about their day and their diabetes. Also, children could work on personal diabetes objectives, set together with their health care professional (HCP) and parents. Parents and HCP could monitor to what extend children achieve their objectives, through dedicated, secured web pages, part of the PAL-System. HCP could tailor through their webpage the content of the activities to offer children additional support to achieve their objectives. The secondary objective was to evaluate the usability and use of the system, both from children's perspective and their parents and healthcare professionals.
Study population: The study population consisted of children with T1DM, aged 7-14, with an T1DM onset of more than 6 months at the study time, good command of the Dutch or Italian language, and without major comorbidities that may affect study participation, followed by Healthcare Professionals (HPCs) in two hospitals both in the Netherlands (Meander MC and Ziekenhuis Gelderse Vallei, ZGV) and one hospital in Italy (Ospedale San Raffaele). A third Dutch hospital was targeted to recruit additional children, but this hospital was unfortunately closed down before the start of the study, due to financial issues. A second recruitment campaign was launched in the tow Dutch hospitals Meander and ZGV.
Study design: The study design was a randomized controlled trial with a wait-list control group (Figure 2). The children were randomly assigned either to an intervention group (group A) or a control-wait list group (group B). The intervention group were accustomed through the use of a first version of PAL (V.3.0) for 3 months (phase 1). After 3 months the effectiveness, in terms of T1DM-related indicators, of PAL 3.0 were measured. The study was envisioned to have a break during the summer holydays period, in order to avoid the inclusion of some confounding factors in the results: (i) the scarce usage of the PAL system during holydays, (ii) the possible seasonality influence on T1DM clinical indicators, (iii) the possible influence, for the Italian sample, of the educational summer camps organised by San Raffaele Hospital. After the summer break, both groups (A & B) were asked to play with a second version of PAL (V.3.5) for another period of 3 months (phase 2). At the end of this second phase, the added value of several extended features will be evaluated in regard to their impact on usability, use and implementation potentiality of the PAL-system.

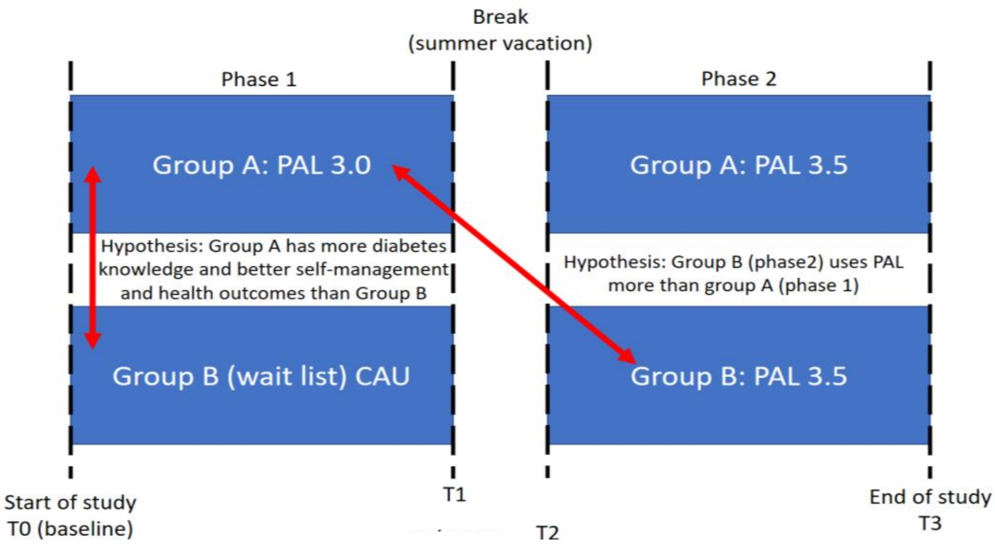

Figure 2. Design of the final experiment of PAL 3.0 and 3.5

Intervention: The children met the robot in the hospital and played an educational activity. Also, they set with their HCP and parents personal diabetes self-management objectives, to work on at home through their MyPAL-app. For example, they want to sleep over at a friend, which means they need to know how to count carbs and recognize when they have high and low glucose values. At home, to work on their objectives, they played educational activities with the avatar (including, quiz, memory, and sorting game) and conversations about their day (e.g., did they do sports, how did they feel) and their diabetes (e.g., their most recent glucose value). After approximately 3 months, they returned to the hospital, they play another activity with the robot and review the objectives with their HCP.

Main study outcome: As listed in Figure 3, the hypothesized main study outcomes were (1) an improvement in the intervention group in Type 1 Diabetes health outcomes: HbA1c values (measured with a finger prick) and glucose values (collected by a glucometer, as part of the standard diabetes management regime) of PAL 3.0 during phase 1. And (2) an improvement of diabetes knowledge (questionnaire).

The hypothesized secondary outcomes during phase 1 were (1) improvements in Diabetes-related Quality of Life (measured with the PedsQoL-DM questionnaire by both child and parent(s)), (2) self-management determinants and behaviour (measured with the Self-Care Inventory, SC, questionnaire and self-management goal achievement in the PAL-system), and (3) system usability (SUS, only intervention group).

The expected outcome in phase 2 was a further improvement of the usability and use of the PAL-System V3.5, with additional features, in comparison to V3.0.

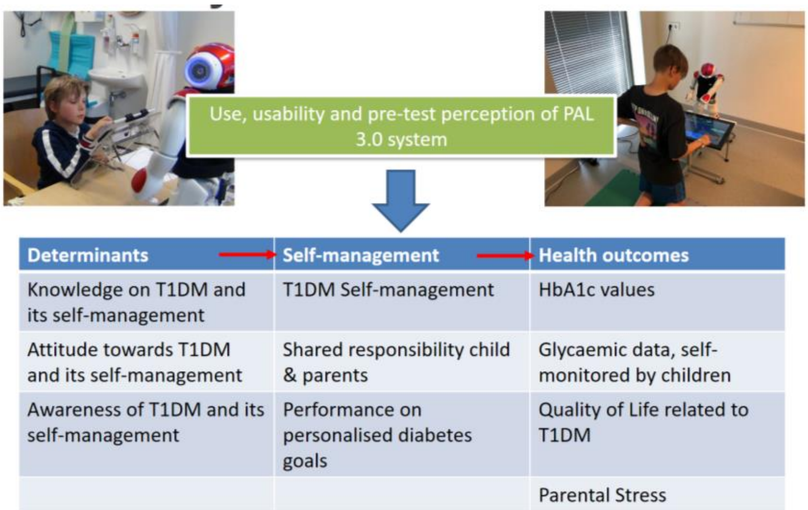

Figure 3. Overview of outcome measurers of the evaluation of PAL 3.0

Results: Children, their parents and HCP from hospitals in the Netherlands and Italy participated in the study. In the first recruitment campaign 21 children and their parents were recruited through the Dutch hospitals and 20 children and their parents were recruited through the Italian hospital. They were invited to participate in phase 1 and phase 2 of the study and were randomly assigned to the intervention or wait-list group. As a result of an extra recruitment campaign, an additional 8 children were recruited, who fulfilled solely phase 1 of the study, making a total of 49 families. Results of phase 1, with 49 children are as follows. In regard to diabetes knowledge, children in the intervention group showed a stronger increase after three months, than children in the wait-list group (p=.05). Moreover, we found a correlation between time paying with the MyPAL app and children's knowledge. Also, children in the intervention groups had a stronger increase in self-care score (p=.01). Moreover, younger children in the intervention group showed a stronger increase in score, in comparison with their older peers in the intervention group. We did not find an effect of PAL on parental stress and glucose regulation (including HbA1c and % in healthy range). However, we did find an effect on diabetes related quality of life in children (p=.02). When we look at the difference between PAL 3.0 and 3.5, we found that children spent their time on the system more efficiently working toward their learning goals in version 3.5 of the PAL system. Also, children in the wait-list group had a higher usage of the MyPAL app than children in the PAL group, which can indicate that the MyPAL 3.5 version resulted in a higher usability then the MyPAL 3.0 version. In regard to participation in the study, children positively evaluated PAL and we did only had one child who dropped out. Moreover, the majority of children participating in the intervention group with PAL in phase 1 also continued to participate in phase 2. Parents were also likely to engage in the MyPAL app with their child, such as playing an educational game together. Children and parents provided also some suggestions on how to further to improve PAL, such as increasing the variety of games and increased personalisation of the games' difficulty level.

Table 2. Characteristics of participants of the final cycle of evaluation PAL

	PAL group (N=24)	Wait list group (N=25)
Variables	Frequencies	Frequencies
Country of residence		
Dutch	14	14
Italian	10	11
Gender		
Female	12	10
Male	12	15
Age groups		
7-10yrs	10	11
11-14yrs	13	7
Diabetes onset age		
T1DM therapy - MDI (pen)		
T1DM therapy - CSII (pump)		
T1DM therapy - Sensor		
Previous PAL interaction		
Yes	19	9
No	5	16

Conclusion: A blended care solution such as PAL, including personal robot and diabetes education app with avatar, and dashboards for HCP and parents, can have a positive effect on diabetes self-management determinants and behaviour in children with T1DM. Also, through application of gamification principles, goal setting, and enriched CRI, children are more positive about the system and motivated to use it. Although we did not find an effect on health outcomes within three months, it may be that with a prolonged use the PAL system, will have positive effects on glucose regulation.

Further details of the cycle 3 experiment in the Netherlands and Italy can be found in the Attachment 2, Effects of the use of the final PAL-System release on Type 1 Diabetes Mellitus self-management determinants in children aged 7-14.

7.2 Analysis of the sentiment

In WP 3, the sentiment mining module was developed to identify whether a text, provided by a child expresses a negative, neutral or positive sentiment. The motivation behind the sentiment mining module is that the sentiment can provide insight into the child's well-being at a certain moment. In the PAL context, a (prolonged) negative sentiment can indicate bottlenecks and can be communicated to, for example, a care giver. The sentiment can also be a measure of progress (e.g. increased sentiment regarding injections/sugar levels might indicate an that it is going better) or can be used to guide the dialogue with the avatar (e.g. the avatar might adjust its behaviour on the sentiment).

During the Y3 PAL experiment, a sentiment mining module was implemented (see PAL deliverable D3.4). For the Dutch participants, the model was trained on the Dutch Basilex Corpus. For the Italian participants, the model was trained on the

Italian SemEval dataset. The outcome of the sentiment mining module is discussed in this section.

While analysing the data, we found that the classification of the sentiment of the Dutch content was fairly accurate. Based on this output, we identified four main types of interaction:

Interaction with the robot in the hospital

1. In the hospital the robot asked various open questions as ‘what is your name’, ‘what is your age’ or ‘what is your hobby’. Answers to these questions like ’11 years old’, ‘basketball’, ‘purple’ often have a neutral sentiment.

Interaction with the robot avatar in the app

2. On login, or sometimes spontaneously, the avatar asks open questions (e.g. “how are you doing?” or “do you also think this is strange?”). These answers often have a positive or negative sentiment.
3. When children used the timeline they could enter free text to elaborate on their nutrition or their activities. Text about nutrition often has a neutral sentiment, while text about activities often has a positive or negative sentiment.
4. Closed questions or information provided by the avatar was often guided via buttons namely; ‘OK’, ‘YES’ and/or ‘NO’. These buttons are classified as neutral.

Table 1 shows the total of positive, neutral and negative detected sentiments per experimental phase and country. Analysing the Dutch dataset, we detected that most interactions with the Nao were of interaction type 1 and 4 and therefore often resulted in a neutral input. Because the timeline was not used that often, we only saw a limited set of interactions type 3. Type 2 interactions often occurred right after login which often resulted in a positive input. The total number of interactions did not differ significantly over the phases with exception of phase 2 in Italy. This increase can be attributed to the increased number of participants in that phase.

While analysing the data of the Italian log files, however, we found inaccurate outputs of the sentiment mining model. As can be seen in Table 1, there is a significant difference between the output of the Dutch model and the output of the Italian model. Analysing this difference, we found that many neutral words as ‘OK’, ‘ciao’, ‘9’, names and even non-words as ‘khbkijgvyjf’ that in the Dutch dataset were classified as neutral, were classified as positive in the Italian dataset. Besides that, also the negative sentiment was often classified incorrectly (e.g. “Hi nao, how are you?" and 'Gymanstics in general" were classified as having a negative sentiment). The inaccurate model explains the difference shown in Table 1. The inaccuracy of the Italian sentiment mining model can be explained by the fact that the Italian dataset did not contain data that was adjusted to children's vocabulary and therefore might not be suitable for the input we gather via the PAL system. Therefore, we conclude that the model of the Italian sentiment mining was unreliable and therefore unusable for this research. To improve the reliability of the Italian, but also to optimize the Dutch sentiment mining model, more suitable training sets should be provided to train the model.

Table 2. Number of positive, neutral and negative detected sentiments of the participants while communicating with PAL, per experimental phase and country.

	Positive	**Neutral**	**Negative**	**Total**
NL phase 1	43	413	5	461
NL phase 2	54	342	5	401
IT phase 1	389	4	24	417
IT phase 2	570	6	36	612

8 Conclusion

To our knowledge, this study is the first RCT of a multi-application blended care solution, encompassing a personal robot buddy, mHealth application and dashboards, for children, parents and HCP, in a clinical and home setting. Moreover, the PAL system as intervention, is based on multiple cycles of development and –formative and summative– evaluations with intensive involvement of end users, including children aged 7-14, their parents and HCP. As such, we managed to develop a system which has proven to have a positive impact on T1DM self-management in children, in collaboration with their informal and formal care givers. In addition, children, HCP and parents were positive about the PAL system. Most children indicated that they wanted to continue the MyPAL application and were even more enthusiastic about the PAL robot.

These findings are valuable for the engineering community and for the medical and patient community, consisting of children with chronic illness such as T1DM and their families. For the engineering community, our results show that an iterative (i.e., cyclic) user-centred approach, such as Socio-Cognitive Engineering, can aid to develop usable and useful systems through strong involvement of end users and stakeholders, with different background (e.g., Dutch and Italian) and characteristics (e.g., age, gender, number of years with illness, and type of medical device).. For the medical and patient community, results show that a blended care solution such as PAL, can have a positive impact not only self-management determinants, such as knowledge, but also on self-management behaviour and patients' quality of life. This provides important guidance for future development and implementation of such solutions on a larger scale (e.g., in other European countries) and in in other domains (e.g., other chronic illnesses).

References

Gi Hyun Lim, Suh Il Hong, Hyowon Suh, Semantic Robot Knowledge Framework for Delivery Service Tasks, The Abstracts of the international conference on advanced mechatronics : toward evolutionary fusion of IT and mechatronics : ICAM, 2010, Volume 2010.5, 2010.5, Pages 591-596.

Han, S., Lee, K., Lee, D., & Lee, G. G. (2013). Counseling dialog system with 5w1h extraction. In Proceedings of the SIGDIAL 2013 Conference (pp. 349-353).

Mioch, T, Peeters, MMM, Neerincx, MA. (2018). Improving Adaptive Human-Robot Cooperation through Work Agreements. 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).

Van Welie, M., van der Veer, G. C., & Eliens, A. (1998). An ontology for task models. In Design, specification and verification of interaction systems '98 (pp. 81-94). Vienna, Austria.