

Designing for Reflection: Chronic Obstructive Pulmonary Disease Patients as Self-Trackers in Telehealth

Peder Walz Pedersen

Aalborg University
Rendsburggade 14
9000 Aalborg, DK

pwpe08@student.aau.dk

Stephanie Githa Nadarajah

Aalborg University
Rendsburggade 14
9000 Aalborg, DK
snadar11@student.aau.dk

ABSTRACT

Telehealth (TH) systems for early detection and treatment of chronic conditions have seen increased use. But the effects on user needs and concerns when healthcare provider continuously monitor and patients provide subjective and objective data over time is poorly understood. Personal Informatics literature informed the analysis of interview data from six Chronic Obstructive Pulmonary Disease (COPD) patients to improve understanding of user needs and concerns in the use of a state of the art TH solution. While patients generally felt taken care of, the system in many ways did not meet user needs. It included no access to previously entered data leading to difficulties assessing reliable subjective measures and provided no support on reflection and follow-up action.

Author Keywords

self-tracking; personal informatics; reflection; telehealth; chronic obstructive pulmonary disease;

INTRODUCTION

In response to ageing societies, there is an increasing need for people to take an active role in their own health and well-being. Designing technologies that support self-reflection, awareness and self-management of chronic conditions have particularly been of interest to HCI researchers to support quality of life issues. Chronic Obstructive Pulmonary Disease (COPD) is a progressive lung disease in which the airways are damaged. People with chronic health conditions such as COPD occasionally experience exacerbations that, if not detected and treated early, result in an increased use of healthcare services and a decline in health-related quality of life. Previous studies have shown poor self-management among COPD patients, who do not respond to the early warning signs, because they have difficulties recognizing the onset of an exacerbation or its importance, resulting in delayed recognition and treatment of exacerbations.

Telehealth technologies where patients log and keep track of important variables and symptoms (e.g. self-reported shortness of breath or fluctuations in oxygen saturation measures) are widely used today by healthcare providers to remotely monitor patients and support early detection and initiation of treatment. We refer to the ongoing activity of logging data (objective or subjective) during concrete *episodes* over time

as *self-tracking*. Self-tracking is well-understood from literature on the Personal Informatics movement, where people voluntarily, self-initiate and successfully use technology to enhance self-awareness, self-reflection and thereby change behaviours, improve health or other aspects of life. While the health literature provides little information on, how patients are supported in their self-tracking efforts in the telehealth context, Personal Informatics literature investigates interaction design aspects of self-tracking important for the successful integration of technology in self-tracking interventions.

We synthesized findings from the disparate fields of telehealth and Personal Informatics and used it as an analytical lens for material from semi-structured interviews with COPD patients using a telehealth solution to understand, how we can support patients in their self-tracking efforts (Study 1). While patients generally felt safe, many struggled with subjective data entries that employed references to perceived baselines. The state of the art system they used provided little support on follow-up actions and reflection by excluding access to historical data.

We used findings from the literature and the study to propose design solutions through mockups, focusing on supporting self-reflection that could potentially support patients in the early detection and treatment of exacerbations. We conducted individual feedback sessions on design mockups with five COPD patients (Study 2) and found that not all patients were willing to engage in self-reflection and identified some major concerns related to the design proposals.

We redesigned and implemented a prototype to evaluate it in a real context during a two-weeks trial by six new COPD patients (Study 3). The trial informed, how four concrete design decisions affects reflection among COPD patients. This paper presents key findings on how telehealth systems in the future may be designed to support reflection, finish by drawing some key concerns for those intending to design and evaluate reflection.

BACKGROUND

In recent years, Personal Informatics or Quantified Self (QS) tools have received an increasing interest in the field of human-computer interaction with the introduction of low-cost mobile applications, wearables and advances in sensor technologies. Personal Informatics help people understand themselves through self-tracking of personally relevant in-

formation for the purpose of self-reflection and gaining self-knowledge [6].

Researchers have proposed different models for understanding, how self-trackers concretely use Personal Informatics tools over time. Li et al. proposed the cascading five-staged *Personal Informatics Model* describing, how self-trackers transition between: (1) *Preparation* (determining variables, tools and frequency of tracking), (2) *collection* (logging data), (3) *integration* (preparing data for reflection e.g. by aggregating and analysing data), (4) *reflection* (examining data to gain self-knowledge) and (5) *action* (deciding what to do with said knowledge) [6].

Little is known about self-tracking practices around telehealth systems in the health context, but we found that despite telehealth does not completely reflect the practices of Personal Informatics (because there are multiple stakeholders/users in telehealth), the above-mentioned stages still apply. We have illustrated the differences between stakeholders roles in Personal Informatics and Telehealth on Figure 1.

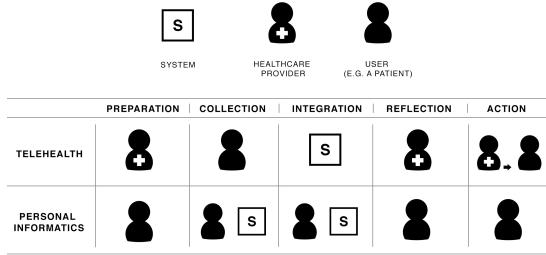


Figure 1. Stakeholders model bla bla

In telehealth, healthcare providers mandate and predefine what symptoms, how often and with what tool patients should track (*preparation*). Patients collect both objective numerical (e.g. oxygen saturation measures) and subjective binary data (e.g. yes/no answers to whether dyspnea has increased more than usual) (*collection*). The system integrates the data and based on predefined individual normal ranges flags data for follow-ups (*integration*). Trained nurses or physicians review the self-tracked data (*reflection*) in telehealth and if needed contact and advise the patient on potential initiation of treatment (*action*) [13, 11].

Epstein et al. found that *collection*, *integration* and *reflection* are ongoing processes that can occur simultaneously and categorize these activities as *tracking and acting*. They further divide the *preparation* stage into *deciding* (deciding to self-track personally relevant data) and *selecting* (selecting a tool to track with) and integrate *lapsing* (e.g. due to an oversight or holidays) and *resuming* into their *Lived Informatics Model* [4, 12].

While Li et al.s and Epstein et al.s work describes the different stages self-trackers transition between, Rivera-Pelayo et al. discuss Personal Informatics in light of reflective learning theory (or learning by reflection) [?]. Based on the work of Boud et al., they propose a framework consisting of three

dimensions in which technology can be integrated to support reflection: (1) *Tracking*, (2) *triggering* and (3) *recalling and revisiting*. Rivera-Pelayo et al. describe *tracking* as logging data that serves as the basis for the reflective process. The data can both be experiences (e.g. feelings or physiological data), but also outcomes (e.g. gained insight or changes in behaviour). *Triggers* have the purpose of raising awareness and detect discrepancies. These are related to the initiation of reflection based on the logged data. Finally, enrichment and data presentation facilitates *recalling and revisiting* past experiences.

We use the *Lived Informatics Model* [4] and Rivera-Pelayos framework as an analytical lens to identify user needs and barriers for self-tracking in the following. We further aim to provide an understanding of, how technology can be used to support people in their self-tracking efforts.

Preparation

The *preparation* stage includes includes people getting motivated to track (e.g. because of goal they have in mind), which guides the decision on, what data to track and selecting the tool to track with.

Motivation and Goals

As described by Epstein et al., people self-track for various reasons, but not all people self-track with a concrete goal in mind [4]. One example is people, who track out of natural curiosity about what their data might reveal about themselves [6, 4]. We refer to this as self-tracking for *life experience*. From the literature, we classified four main drivers for sustained tracking among people with a health-related condition: (1) *documentation* (e.g. to create records for their healthcare providers [1]), (2) *communication* (e.g. to communicate their condition to family members [8]), (3) *self-knowledge and advice* (e.g. to get a sense of the current state of their condition or to get advice from their healthcare-provider [8, 1]) and (4) *self-improvement* (e.g. to change or maintain a behaviour in order to improve well-being or lifestyle [8, 1, ?]).

Barriers to motivation include strong emotional adversity to reflection on data, because it reminds people of negative aspects of their illness [6, 1], tracking the wrong data or not tracking well enough to gain benefits [3], effort [2, 10], reliability and relevance [9, 13, 4] and mismatches between subjective feeling and objective measures [1].

Selection of Data and Tool

Previous studies show that, sometimes healthcare providers give little support on what symptoms to track and how to track (e.g. frequency of tracking) [10]. This is an issue in health conditions that involve many symptoms that arise unexpectedly, where patients have to decide on additional symptom tracking [10, ?]. Trackers use tools, such as notebooks, health diaries and specific applications to do additional tracking or sometimes develop tools themselves that are cumbersome and incomplete [10]. This later affects the *integration* and *reflection* stages, where healthcare providers struggle with interpreting data sets with additional and not always relevant information [3, ?].

Little is known about what motivates to self-track in tele-health. As previously mentioned (See Figure 1), the decision on what to data track and which tool to use, is in telehealth context decided by a healthcare provider. Rivera-Pelayo et al. mention that selection of what data to track has an effect on the acceptance from the user and the efficiency of learning by reflection [?].

Collection

The *collection* stage deals with logging of the data. Logging too many things can be challenging and lead to tracking fatigue [2, 10], but since logging data is the basis for triggering and supporting *reflection*, it is of importance to choose as unobtrusive a method as possible, which is sufficiently reliable [?]. We previously described self-tracking as logging data (either objective or subjective) during concrete episodes over time. We now distinguish between data being either automatically logged (e.g. physiological sensors that log heart rate) and manually logged (e.g. self-reporting feelings or ideas).

Manual Logging

Manual logging requires responsibility and motivation that has to be kept over time from trackers and can therefore be burdensome compared to automatic logging [6, ?]. Some self-trackers find manual logging time consuming and requiring effort [1], hampering incorporation of tracking into daily routines [14, 1]. One example is, when objective data has to be manually logged, it can require preparation time, e.g., resting before taking physiological measures, which increases time and effort [14]. Data granularity impede logging, when trackers overthink, while trying to rate mood on a scale from 0 to 10 [9] and lack of baseline to compare with cause difficulties for trackers with chronic conditions, when self-reporting on severity of symptoms, because they are constantly symptomatic [13]. Another downside of manual logging is that sometimes self-trackers postpone manual logging [8], which induces recall bias that that affects the data reliability.

Automatic Logging

Automatic logging shifts the effort from the trackers to the technology. It allows for configuration of frequency and precision, but also implies challenges in terms of filtering and aggregating the large amounts [?]. One of the main downsides of automatic logging is that it might reduce awareness and self-reflection [2, 7].

Integration

The *integration* stage can be more or less apparent to the tracker depending on whether integration is automatically done by the system, requiring less effort from the self-tracker [6]. Self-trackers sometimes postpone data exploration when integration does not happen automatically, since it involves tedious tasks, such as cleaning up data, formatting and running statistical tests [2, 3, 6].

Systems that require manual integration expect the user to be able to analyse the data and ascertain the best way of creating a representation. This is of interest to curiosity-driven self-trackers that want to integrate data manually and explore the novel insights that data can offer them. In contrast, self-trackers with a goal know what they are looking for in the

data and strive at using automatic integration systems allowing them to concentrate on reflection. The manual integration process is an iterative process of moving back and forth between representation creation and reflection [?].

Reflection

In the *reflection* stage, self-trackers reflect on the collected and integrated data by looking, exploring or interacting with visualizations of it [6]. While literature points at many different definitions of reflection, Fleck & Fitzpatrick identified five different levels of reflection (R0-R4) that indicate what types of activities and behaviours can be associated with reflection [5]. Levels consist of (R0) describing or stating without being reflective (*description*), (R1) describing with explanations in a reportive or descriptive way (*reflective description*), (R2) seeing things from different perspectives and trying to identify relationships (*dialogic reflection*), (R3) changing original point of view due to gained knowledge (*transformative reflection*) and (R4) seeing the wider perspective beyond the immediate context (*critical reflection*). While higher level indicates being more reflective, lower levels are prerequisites for becoming more reflective.

Conditions for Reflection

One of the condition for reflection is creating and allowing for time to reflect [5]. Li et al. distinguish between short-term reflection, where the self-tracker reflects immediately after logging the data and long-term reflection that might occur several days or weeks after [6]. While short-term reflection makes the tracker aware of the current status, long-term reflection allows for higher levels of reflection (at least R2), since the tracker can compare logged data between different times and explore trends and patterns.

Several authors mention that the one reflecting should be open-minded and willing to reflect [?]. As mentioned by xx, in some literature there is an implicit assumption that *skills* (such as critical analysis and evaluation) are necessary to engage in reflection [?, ?]. Fleck & Fitzpatrick mention that reflection can be developed with time and with the right support [5].

The reflective process further needs a trigger. People often need a reason (e.g. a purpose) or at least an encouragement to reflect [5, ?]. In psychology, Festingers cognitive dissonance theory describes, how a mismatch (psychological discomfort or dissonance) between an individual's attitude and behaviour can lead to rethinking ones attitude and behaviour [?]. The dissonance triggers reflection can be actively triggered (system explicitly tries to catch the users attention by highlighting a certain mismatch) or passively triggered (system only presenting the data and relying on the user to detect something that starts a reflective process). Dissonance may occur due to comparison between current level and a recommended level or goal, but some might prefer such comparison in response to oneself. For example, some self-trackers prefer to interpret their data in light of their own personal and medical history and/or symptoms, rather than striving for provider-recommended normal ranges [1].

In the following we describe different ways to trigger and support reflection identified in our literature review.

Reflective Questions or Prompting

One way of supporting reflection is through the use of reflective questions or prompts. Simply asking people to provide justification or explanation for e.g. events or actions supports at least *reflective description* (R1) [5]. The presence of another person can encourage reflection, especially in a dialogue among two uneven partners (i.e. two people not sharing the same understanding or experience), where one takes the role of asking questions.

Systems can also take this leading role of asking questions, but opposed to the previously mentioned example, where reflection among two people can be dialogue driven, the systems often only pose an initial reflective question [?]. An intelligent system could further support reflection through follow-up questions. Another ways to foster reflection is by prompting questions triggered by automatically logged context data [5]. Presenting data that is not usually visible encourages people to see things from another perspective and can potentially lead to looking for relationships and patterns (R2).

Visualizations

Despite simply looking at data is not considered reflective according to Fleck & Fitzpatricks levels of reflection, creating representation of data is considered a prerequisite to support higher reflective levels. Visualizations of data can help people in exploring their information and gaining insights [6, 2].

When designing visualizations, it is important to consider under which conditions reflection is to take place (e.g. time and effort expected from the user, skills and purpose for reflection) [?, ?]. Li et al. identified six types of questions people ask about their self-tracked data [7]. These are, getting to know (1) status (what is current status?), (2) history (what has status been in the past?), (3) goals (what goal is appropriate to pursue?), (4) discrepancies (how does current status compare to goal?), (5) context (what affects current status?), (6) factors (how are different variables related?). Depending on the conditions, supporting both simple (e.g. status charts) and detailed visualizations (e.g. of time series) can be important [?, ?].

Often people want to obtain answers to their question (e.g. status) without spending too much time or effort, which can be done on a simplified dashboard representation that allows for a quick overview [?]. Mller et al. found that people use status charts to quickly get an overview of their data and use it as a starting point for exploration [?]. Comparison charts were requested in the study to benchmark against other people in the sense-making process and to assess success. However, as previously mentioned, some prefer that such benchmarking occurs in response to oneself [1].

Visualization of time series data support revisiting and analyzing past experiences (history) and can trigger storytelling about experiences behind data [?, ?]. It can foster reflection on global trends e.g. on upward and downward trends or deviations from a historical normal (suggesting a problem) [?].

Cuttone et al. mention that human behavior is often characterized by periodic patterns, but that time series graphs do not facilitate exploring such patterns [?]. They instead propose using calendar heatmap representations using different color shades to indicate variable values. Visualizations of multiple time series can support reflection on how multiple variables are related or how multiple variables change over time [?].

Visualizations of time series can further be combined with discrete events [?] to support reflective description (R1). Barriers for reflection include that tools to not always support either simple visualizations of data or more complex features (e.g. filtering data to focus on a subset of data, zooming out to get an overview or comparing multiple variables) [7, ?].

External activities

While reflection is an internal process, it can occur when trying to externalize thoughts and feelings e.g. in diaries or during reflective writing [?]. These activities are often descriptive or emotional (R1). Recording reflection outcomes for later revisiting and reflection on gained insights has been proposed as another way to support reflection [?, ?]. Isaac et al. found that allowing for writing down thoughts helped people explore and understand their feelings [?].

Action

People decide what *action* to take based on the findings from the *reflection* stage. Trackers who have other motivations than behaviour change, e.g. self-understanding, do not reach this stage [4]. Trackers who track for self-improvement sometimes lack the knowledge necessary to identify the appropriate actions to take, such that they can regulate their progress towards their behaviour change goal. This happens either because they collect irrelevant data [2, 3] (e.g. food and symptoms, instead of ingredients that trigger the symptoms) or because they need actionable (expert) advice [14, 6, 9]. Most Personal Informatics systems do not provide actionable advice [3, 6, 14] and trackers then seek out this advice from healthcare providers [6].

RESEARCH AREA

In summary, previous literature shows user needs and concerns during tracking activities, but little is known about these aspects when tracking is monitored by a healthcare provider. It is unclear how concrete design decisions regarding entry and interaction with data affect reflection in such case. In study 1 we investigate user needs and concerns during tracking activities in telehealth context. Based on these findings we redesign the system to support patients in reflecting during data entry and interaction.

STUDY 1 - EXPLORING USER NEEDS IN TELEHEALTH

Results

All patients remembered to take their measurements consistently and routinely in the morning themselves (P6S took responsibility for P6).

Preparation:

The majority of participants (P1-P5) found the sense of security from healthcare providers monitoring their data motivating. *They call you if you do not send in your readings ..*

it gives you a huge sense of security that you are not gonna lay at home ill (P2). P4 felt obligated to take measurements due to the presence of a healthcare provider. Two patients tracked additional data on paper (P5, P6S). P5 used the data as documentation, e.g. when being admitted to the hospital to discuss it with healthcare providers. P6s spouse mentioned curiosity, self-satisfaction and sense of agency as motivations for tracking on paper.

Collection:

Patients found data collection easy, not requiring expert computer skills, and not taking too much effort or time. P5 stressed the importance of fast collection, *it must not take ten or fifteen minutes to do it everyday (...). This [AF] is really simple (...) it is so simple you can add some more to it* (P6S). Several patients found answering subjective questions difficult when it required comparisons with the usual baseline, (Are you coughing more than usual?). *What is usual? Isn't that also how I felt yesterday? Otherwise, I have misunderstood the question* (P4). Patients needed higher than binary granularity to answer, *When they ask if you have more dyspnea than usual, then we say yes .. but how much is it? They [healthcare providers] cannot see* (P6S). Some patients underreported baseline deviations and only answered yes in large or extreme deviations, *if its just a little different, I do not mention it* (P2), *I would have to be coughing a lot and feel very ill, if I answer yes to that question (...)* (P4). P2 asked for a scale instead (...) *why dont they make a scale instead for example from 1 to 5 or 1 to 10? One day I could perhaps say its 5, the next day 6 and the day after I can go back to 5.* P5 used the comment box to make small deviations go on record, (...) *to me it is important that we take every small nuance.* When to collect, was a concern for P4 whose oxygen saturation measure and pulse depended on her level of activity. She wondered why the system did not take into account external factors related to her condition, *Do you feel more breathless today? But it does not say anything about the fog outside* (P4).

Reflection:

Several patients mentioned that an exacerbation comes within a few hours or even minutes, and that they were not able to recognize an onset by using AF. P1, P2, P3 and P5 measured oxygen saturation several times a day to verify their subjective feeling of well-being (P2, P5) or lack thereof. None of the patients felt they learned anything about their disease using AF. *I can feel it [an exacerbation], even if I did not have the monitoring device* (P3). Patients did not express any concerns waiting for a (potential) call. Most of them had identified the hours of the reviewers at the hospital and several a mental model of when a call would ensue. *They usually do it [review the data] before noon* (P3). *I already know when there is going to be a call (...) when the oxygen saturation is too low, the pulse is high and your measures fluctuate, they react* (P2). P6S found benefits in tracking data on paper, allowing for understanding his wifes baseline and whether she was deviating from it and getting worse. *You can see how stable it is .. (...) Lets say she loses weight then I become alert that something is wrong* (P6S). Patients had not been informed by their healthcare providers about their normal range and the

AF interface did not communicate it either. Half the patients wanted to know these in numbers. Some of the patients had identified their own normal range of oxygen saturation that mapped to not feeling well (usually below 90).

Action:

All patients had received education in self-management of their condition (e.g. breathing techniques), but not all patients gained the same benefits from it and needed actionable advice. To that end some patients (P2, P5) added questions to the comment box. P2 acted on the basis of his oxygen saturation measures, *when it [oxygen saturation measure] is lower than 93, you do not feel fine (...) then I walk a little slower and take it a bit more easy.* P2 was interested in knowing additional methods to increase oxygen saturation. P6S wanted recommendations on duration for supplemental oxygen use based on her oxygen saturation measures and information about variables e.g. the weather, that could influence her symptoms. P3 and P4 used their oxygen saturation measures adjust their supplemental oxygen. However, P3 preferred not to initiate treatment including drugs before consulting a healthcare provider, unless in extreme cases of symptoms or unavailability of staff. *I might to do it [initiate medication treatment] if it [sputum color] was very green, if it was a Tuesday [a day not monitored by healthcare providers], otherwise I wouldnt (...).*

Discussion

Our patients were highly motivated to track potentially due to the active role of the healthcare provider that provided them with a sense of security not present in previous studies [6, 1, 3]. One patient relied on the monitoring to such a degree that she sometimes delayed treatment, waiting for confirmation from the healthcare provider.

None of the patients described the tracking activity requiring too much effort or time. However, effort seemed to be a an aspect that should be considered when designing telehealth systems, as patients both expressed willingness to spend more time than AF required (approximately two-three minutes) but not wanting to spend more than ten minutes.

Based on our findings and literature review, we revised the *Lived Informatics Model* [4]. We broke down a data collection episode into *pre – collection* (deciding on whether to log or skip), *acquisition* (ready required artifacts), *calibration period* (satisfy guidelines for tracking), *measure* (taking measure using artifact), *entry* (entering read off measure from artifact or providing scale based ratings, absolute or relative to a baseline, or qualitative comments), *submission* (submitting data), and later *review*. The patients reflected during multiple stages of collection before entering both subjective and objective measures.

AF did not meet the needs of users in terms of (1) scope, (2) reliability, (3) validity, (4) actionable advice. AFs scope focused only on submitting variables directly related to the condition at time of entry, but two patients tracked additional and one submitted data on paper (c.f. [10, ?]). Having access to their previous data made patients feel in control.

In terms of reliability, patients were unsure whether they were collecting data under the right conditions. Subjective questions with a baseline comparison proved difficult due to: no access to the baseline and the low granularity of the answer options. Patients had insufficient access to their usual subjective feelings and tried to remember previous events to establish their usual baseline (c.f. [13]) and AF provided no access to historical data. Even if AF provided access to previous data this might prove difficult due to the low granularity. The binary answer options resulted in reduced validity of data by underreporting significant increases from the baseline. One patient specifically asked for rating on a scale instead, which requires more cognitive effort and time [9]. Due to the absence of data access, patients did not interact with the data they had collected and expressed not having learnt anything from telehealth. Several patients mentioned not being able to recognize onset of an exacerbation from AF use, suggesting that the system poorly supported reflection.

Patients did not know their provider-recommended normal range and therefore used their own identified normal range in management of their condition using the pulse oximeter. Some patients wanted to know the provider-recommended normal range to become more empowered, while others were not interested. One reason for that could be that patients get reminded about the negative aspects of their health when reviewing data or that they rely more on their subjective feeling than on quantities, as in [1].

Several patients were interested in actionable advice from the system as in [3, 6]. Apart from during subjective data entry, patients needed two types of support, (1) confirmation from healthcare providers to act (e.g. initiation of medication) and (2) actionable advice on self-management strategies (e.g. coping with breathlessness). We believe that one of the barriers to action was the lack of support for reflection during entry and review of data - a prerequisite to action according to Li et al. [6].

Conclusion

We investigated user needs and concerns critical for the effectiveness of telehealth interventions using a synthesis of literature on personal informatics and analysis of interviews with COPD patients using a state of the art telehealth solution. While having a healthcare provider monitoring data motivated sustained tracking, the telehealth system they currently used did not sufficiently meet user needs for tracking. Patients expressed difficulties rating their symptoms relative to their usual baseline, resulting in reduced reliability of the data. A lack of access to their historical data hindered patients in entering reliable data, reflecting and taking actions.

STUDY 2 - DESIGNING A COPD TELEHEALTH SYSTEM

In the following we describe how we redesigned the system to support patients in collecting and reflecting on self-tracked data and evaluated the re-design. Researchers have however found that co-designing with COPD patients using generative tools and techniques (e.g. post-it notes and sketching activities) are resource demanding for the patients, considering their health condition. Some patients are not even able

to participate in other activities than keeping a conversation. In Study 1, we similarly experienced that COPD patients are physically limited (e.g. in terms of moving from one place to another) and within an hour of interview experienced breathing difficulties several times, demanding a slow pace and long pauses.

Drawing from Das et al.s experiences on co-designing with COPD patients [?] and our own experiences from Study 1, we decided not to conduct co-designing activities as initially planned. We made a conscious plan to conduct individual sessions where patients were only expected to talk or interact with a prototype. To further reduce time and effort required by them in the sessions, we asked patients to complete workbook assignments beforehand, allowing them to prepare for the discussion on, how our proposals foster reflection during collection and review of data. In the following, we describe the workbook and prototype design.

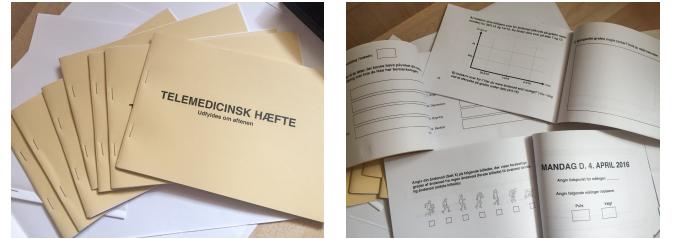


Figure 2. Workbooks with assignments

Workbook Design

We found that patients reflect during multiple stages of a collection episode and have user needs and concerns, currently not addressed in the system. Workbooks assignments were designed to get insight into, how the system can support them in collecting and reflecting on data. We further asked patients to collect data to be used in the prototype.

We found that patients were unsure, whether they were collecting data under the right conditions in Study 1. To get further insight into, how patients reflect on conditions relevant for the validity and reliability of their measures, we asked them to annotate if they had reflected on context-relevant variables that we provided in the workbook. For dyspnea measures, context-relevant variables could be e.g. weather, mood, smoking or physical activity. Patients were asked to comment on it as free text annotation one day and mark it in checkboxes another day (e.g. mark with X if medicine has affected saturation measure).

Patients expressed difficulties rating symptoms subjectively, due to lack of baseline understanding and low granularity of answer options in Study 1. Patients had to reflect on a time series graph of previous measures (showing baseline) while entering current measure in one assignments. In another assignments, patients had to assess dyspnea on different scales: (1) Relative to usual baseline (binary), (2) absolute (five-point scale) and (3) using visuals (Dalhousie Pictorial Scale [?]).

Finally, patient had to reflect on a time series graph of previous oxygen saturation measures (long-term reflection). A recommended level was marked on the graph to trigger reflection

on a potential mismatch in accordance with Festingers cognitive dissonance theory [?].

Prototype Design

Improving Reliability and Validity of Measures

Identical to the reference graph in the workbook, we also incorporated a time series graph visualizing patients previous measures in the prototype to support patients in remembering previous events while entering data. We changed symptom rating options from binary to scale-based, inspired by the questionnaire-items in EXACT PRO, a validated tool to measure COPD exacerbations through subjective assessments of symptoms [?].

Similar to the assignment in the workbook, we provided the option to toggle context-relevant variables for each measure. These were also introduced in the prototype to support reflection described in the next section.

Supporting Short-term and Long-term Reflection

From literature, we found that supporting both short-term and long-term reflection can be important, depending on the conditions in which the user might self-track [6, ?].

To support short-term reflection, we provided patients with a dashboard view after collection (See Figure). The dashboard view allows patients to reflect on current status immediately after collection and increase awareness on their current status [?, ?]. We included reflective questions to trigger reflective description (R1) or higher levels of reflection on this view and to encourage users to further explore their data [5, ?]. E.g. *Why are you coughing more than last time you measured?*. Gauges for each measure show current measure in relation to the recommended level (goal) and arrows indicate change from previous measure (day-to-day variations). The system uses color indications (red, yellow and green), where red or yellow highlight a potential mismatch and thereby actively trigger reflection in accordance with Festingers cognitive dissonance theory [?].

To support long-term reflection, we designed four different visualizations (See Figure). (V1) Time series graphs with option to compare two measures, (V2) Time series graphs stacked vertically, (V3) Calendar heatmap and (V4) Area graphs stacked vertically. V1, V2 and V3 were proposed to foster reflection on upwards and downwards trends or symptoms deviations to increase awareness on a worsening in condition [?]. V1 differed in that it allowed for comparison of multiple measures that could trigger reflection on how measures are related or change over time [?], affording dialogic reflection (R2) [5]. Recommended levels were marked on all three visualizations to increase awareness on discrepancies [6] and trigger reflection. V3 was proposed as an alternative to time series graphs as mentioned by Cuttone et al. to support reflection on periodic patterns using color shades to indicate deviations from recommended level [?, 6].

To further support long-term reflection, we provided the option to toggle on discrete events on the visualizations (context-relevant variables) [?]. This feature was included to support reflective description (R1) and dialogic reflection (R2) by allowing for exploration of relationships between

context-relevant variables and measures otherwise invisible [5].

Participants and Method

We asked the same patients as in the previous study to participate in feedback sessions on the redesigned system. All patients except P4 who was hospitalized participated. Three spouses (P2S, P5S and P6S) also participated. Feedback sessions were held in the patients home and lasted for approximately one hour.

Patients completed workbook assignments three days a week (Monday, Wednesday and Friday) delivered one week prior to the feedback sessions. The feedback session was divided into a workbook session and a prototype session, where one facilitator and an observer were present. The observer initially took the facilitator role and asked patients follow-up questions emerged from Study 1, while the facilitator went through the workbooks and prepared for the feedback session. In the workbook session, the facilitator interviewed and discussed the assignments patients had completed in the workbook. In the prototype session, the facilitator guided the patients through each screen in the prototype and asked them to think-aloud and complete tasks. The facilitator encouraged patients spouses to participate in the discussion and provide their points of view. The prototype was adjusted to use patients own values for visualizations (collected in workbooks) to encourage reflection on patients own experiences and not fictive values. The observer prepared the prototype during the workbook session. The sessions were audio-recorded.

Results

Reliability and Validity of Measures

Not all patients were consistent in terms of *when* and *how* they took measures. While P2 and P5 ensured taking measures under comparable conditions (always in the morning), P1, P3 and P6S lacked knowledge on relevance of context and how some of the provided context variables influence their measures. E.g. P1 and P6S questioned whether and why a cold finger when taking oxygen saturation measures has an influence.

Some context variables are more likely to affect submitted measure than others. E.g. in case of shaking hand, P6 and P6S retake oxygen saturation measure until it is stable. Patients mentioned additional context variables relevant for their measures (e.g. mood, talk, supplemental oxygen). P1 only weighed himself every three weeks, despite the system prompts him to enter current reading. P2 and P6S specifically asked for guidelines on taking measures. *For some measures, an explanation would be good. For example do not take measure if this and that* (P6S).

All patients preferred higher granularity options when rating symptoms more than binary. P5 and P6 mentioned that they preferred higher granularity, because it shows more and makes it possible to show degrees. How much is a no? If we say yes or no to the hospital, they still do not know what we are thinking .. then theyll call us and well still have to say it is severe (P6S).

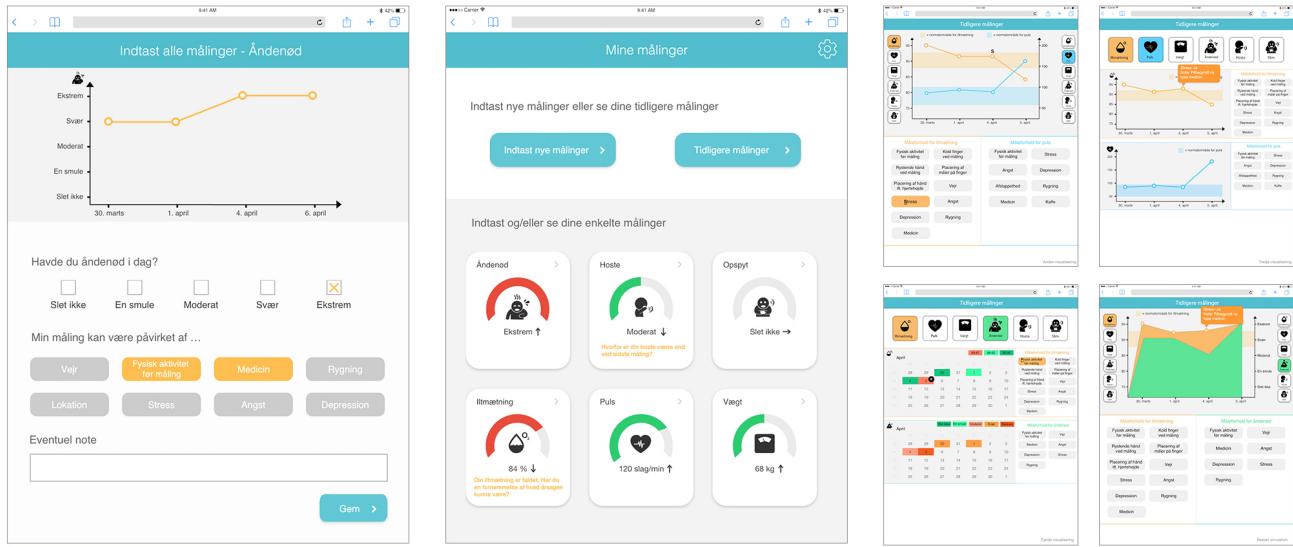


Figure 3. Screens from the prototype: Collection page (left), Dashboard (middle), Four visualizations of history data (right). V1 (top left), V2 (top right), V3 (bottom left) and V4 (bottom right)

Asking to rate symptoms without a baseline comparison (Did you feel breathless today?) and having five answering options (not at all, slightly, moderately, severely, extremely) caused difficulties for some patients, because they experienced several of the options during the day. *I have been through all of the provided options that day. How do you want me to answer that?* (P5). Patients further had different perceptions of what usual is, when asking to rate relative to usual baseline. (...) *I base that on when Im at my best* (P3), *usual is when it is an ordinary day* (P5) and *If she is not more breathless than yesterday, then well just submit a no* (P6S).

Patients presumed that it would not be of benefit to them to have a time series graph showing their baseline. I do not think it has any effect to see a graph (P3). P2 mentioned, I shouldnt answer based on what I answered last time. I should answer what it is now and here, but after discussing its purpose agreed that it would help him remember previous measures.

Short-term and Long-term Reflection

In the feedback session, some patients were not willing to reflect on visualizations of current status, reflective questions or history data. E.g. I do not care what my status is.. I just submit data.. do not walk around and think everyday .. I know how I am feeling (P3). P5 expressed interest in current status and what action should be taken, Im more concrete. Where am I right now and what can I do about it? (P5). The dashboard view helped some patients in getting an overview of their health status. P5 thought that the gauges on the dashboard illustrated her status quickly. P2 mentioned that the arrows helped him *quickly see if it (a measure) is going up or down*. Reflective questions were not noticed by the patients. When asked explicitly to comment on them, some patients seemed aware of the answers *right now it is likely because I talk too much* (P5).

None of the patients expressed interest or benefits in having access to previous measures prior to the feedback session, e.g. I do not need it (access to history data) (P3). Several patients relied on the healthcare monitoring I have a nurse who is good at keeping an eye on me (P5), I do not need it (access to history data). If it (measure) is too low, they call and ask me why (P1). P2 similarly expressed that he did not know have the necessary knowledge to find it useful I do now know what they use it for, the scales they use and the language.. I do not understand it. I count on they react if there is anything (P2).

Providing patients with a recommended level that they could compare their measures against, provoked negative feelings among some: I prefer not to be told in the morning that Im gonna get an awful day (P5), Its ok if its just a single measure, but if it is constant, I would start thinking.. Its going fast now (P2). While P5 mentioned that she would ignore the recommended level, because it did not match with her own goal, *I thought, you can forget it (about provider-recommended level)*. *Ill just do what I usually do*, P2 suggested that he would strive to keep his measures within the recommended level, then its not that bad if I keep it above that (lower threshold). P6S indicated that if his wifes oxygen saturation was above normal area, he would start wonder whether the oxygen supply was set too high and initiate action. *If it starts to go over here (below normal area), we have to do something* (P6S).

Despite none of the patients initially expressed benefits in having access to history data, two patients (P2 and P5) changed their attitude after the prototype session. This gives more information about me (...) its nice to be able to go back.. Is it better than 14 days ago? (P2) and There might be days where I sit with it and have an idea about what Im looking for, which might trigger some thoughts (P5). Others were more

reluctant on reflecting. P3 did not see any benefits and found it troublesome, while P1 did not think it was his job to look at history data, *this is only for people who has to sit and analyse the numbers* (P1). V1 afforded finding relations between measures e.g. by comparing, you can have them (measures) together and see how they affect one another (P2). V3 was more attractive to others, who found it more concrete and provided a quick overview, *it (V3) is the one I get the quickest.. If you are in doubt what the colors mean, you can see them down there* (P5). PS6 thought that he would start with V2 and then use V1 when he had become more advanced.

Discussion

We identified two types of patients in this study: Active and Passive. Passive patients (P1 and P3) had met one or several barriers in terms of reflection. They were not willing to reflect on their data, because they did not see any purpose for reflecting. While we know from Study 1 that all patients were motivated to collect data, because having a healthcare professional monitoring provided them with a sense of security, it became clear in this study that passive patients were not equally motivated to reflect on the data. These patients took the role as data providers and did not see any benefits in long-term reflection. Active patients (P2 and P5) were more open-minded towards engaging in the discussion of supporting reflection in the system. These patients could potentially benefit short-term and long-term reflection support in the system.

Active patients explicitly mentioned being aware of taking measures under comparable conditions, while passive patients were not. Patients were aware that the self-tracked measures they submit are affected by context data. Providing an option in the system to annotate context-related variables could potentially support reflection on measures in later review. Some patients lacked knowledge about specific context data and requested guidelines in the system.

The dashboard view provided information about patients current status. Reflective questions were not noticed by patients initially. In passive patients reflective questions did not trigger any reflection (R0), while active patients proposed lower level explanations as answers (R1).

Patients had different preferences in terms of visualization for long-term reflection. One patient mentioned needing a purpose and time to reflect for gaining any benefits from history data, in accordance with Fleck & Fitzpatrick [5]. None of the patients were able to express, how the visualizations for long-term reflection supported reflection, except V1, which one patient mentioned could support identifying relationships between measures. This visualization, however, was difficult for several patients who did not see the purpose of it or had usability problems. Having a recommended level to compare measures with, provoked negative feelings among some patients, e.g. becoming worried. In some active patients it supported reflecting on action, if measures were not hypothetically within the recommended area.

Assignments in the workbook and prototype did not elicit many findings on patients reflective thoughts on the design

proposals. While we did use patients own values rather than fictive values to discuss the workbook and prototype assignments in terms of long-term reflection, we used fictive values on the dashboard, which could have been a barrier for reflection in some patients. Additionally, patients might have found it difficult to reflect, because (1) the interview period was time-limited and (2) a reason (e.g. a worsening in health status) was needed to trigger reflection.

Conclusion

We evaluated a re-designed telehealth system that aimed at improving collection of reliable and valid data and supporting short-term and long-term reflection. We found that annotating context-relevant variables to measures in the collection stage is relevant and could potentially support reflection in later review. Passive patients were not willing to reflect, but active patients could benefit from support in short-term and long-term reflection.

STUDY 3 - EVALUATING REFLECTION ON DESIGN

Method

Results

Discussion

CONCLUSION

FUTURE WORK

ACKNOWLEDGEMENTS

REFERENCES

1. Ancker, J. S., Witteman, H. O., Hafeez, B., Provencher, T., Van de Graaf, M., and Wei, E. you get reminded you're a sick person: Personal data tracking and patients with multiple chronic conditions. *Journal of medical Internet research* 17, 8 (2015).
2. Choe, E. K., Lee, N. B., Lee, B., Pratt, W., and Kientz, J. A. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 1143–1152.
3. Chung, C.-F., Cook, J., Bales, E., Zia, J., and Munson, S. A. More than telemonitoring: Health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. *Journal of medical Internet research* 17, 8 (2015), e203.
4. Epstein, D. A., Ping, A., Fogarty, J., and Munson, S. A. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM (2015), 731–742.
5. Fleck, R., and Fitzpatrick, G. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, ACM (2010), 216–223.
6. Li, I., Dey, A., and Forlizzi, J. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2010), 557–566.

7. Li, I., Dey, A. K., and Forlizzi, J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing*, ACM (2011), 405–414.
8. MacLeod, H., Tang, A., and Carpendale, S. Personal informatics in chronic illness management. In *Proceedings of Graphics Interface 2013*, Canadian Information Processing Society (2013), 149–156.
9. Oh, J., and Lee, U. Exploring ux issues in quantified self technologies. In *Mobile Computing and Ubiquitous Networking (ICMU), 2015 Eighth International Conference on*, IEEE (2015), 53–59.
10. Patel, R. A., Klasnja, P., Hartzler, A., Unruh, K. T., and Pratt, W. Probing the benefits of real-time tracking during cancer care. In *AMIA Annual Symposium Proceedings*, vol. 2012, American Medical Informatics Association (2012), 1340.
11. Pedone, C., and Lelli, D. Systematic review of telemonitoring in copd: an update. *Pneumologia i Alergologia Polska* 83, 6 (2015), 476–484.
12. Rooksby, J., Rost, M., Morrison, A., and Chalmers, M. C. Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 1163–1172.
13. Ure, J., Pinnock, H., Hanley, J., Kidd, G., Smith, E. M., Tarling, A., Pagliari, C., Sheikh, A., MacNee, W., and McKinstry, B. Piloting tele-monitoring in copd: a mixed methods exploration of issues in design and implementation. *Primary Care Respiratory Journal* 21, 1 (2011), 57–64.
14. Verdezoto, N., and Grönvall, E. On preventive blood pressure self-monitoring at home. *Cognition, Technology & Work* (2015), 1–19.