The EU project “United4Health”: Results and experiences from automatic health status assessment in a Norwegian telemedicine trial system

Martin Gerdes¹, Frode Gallefoss² and Rune Werner Fensli¹

Abstract
Introduction: Patients with chronic obstructive pulmonary disease require help in daily life situations to increase their individual perception of security, especially under worsened medical conditions. Unnecessary hospital (re-)admissions and home visits by doctors or nurses shall be avoided. This study evaluates the results from a two-year telemedicine field trial for automatic health status assessment based on remote monitoring and analysis of a long time series of vital signs data from patients at home over periods of weeks or months.

Methods: After discharge from hospital treatment for acute exacerbations, 94 patients were recruited for follow-up by the trial system. The system supported daily measurements of pulse and transdermal peripheral capillary oxygen saturation at patients’ homes, a symptom-specific questionnaire, and provided nurses trained to use telemedicine (“telenurses”) with an automatically generated health status overview of all monitored patients. A colour code (green/yellow/red) indicated whether the patient was stable or had a notable deterioration, while red alerts highlighted those in most urgent need of follow-up. The telenurses could manually overwrite the status level based on the patients’ conditions observed through video consultation.

Results: Health status evaluation in 4970 telemonitor datasets were assessed retrospectively. The automatic health status determination (subgroup of 33 patients) showed green status at 46% of the days during a one-month monitoring period, 28% yellow status, and 19% red status (no data reported at 7% of the days). The telenurses manually downrated approximately 10% of the red or yellow alerts.

Discussion: The evaluation of the defined real-time health status assessment algorithms, which involve static rules with personally adapted elements, shows limitations to adapt long-term home monitoring with adequate interpretation of day-to-day changes in the patient’s condition. Thus, due to the given sensitivity and specificity of such algorithms, it seems challenging to avoid false high alerts.

Keywords
COPD, decision-support techniques, telehealth, remote patient monitoring, personalised health status assessment algorithms, triage

Introduction
The remote home monitoring of patients with chronic diseases has been a topic of many research projects.¹–⁴ The main goals in these studies have been to help patients in daily life situations, to increase their individual perception of security especially under worsened medical conditions, to reduce home visits by doctors or nurses, and to avoid unnecessary hospital admissions and readmissions.

The analysis of a long time series of vital signs data from patients at home over periods of weeks or months is both a new opportunity as well as a new challenge for healthcare services, and differs from traditional monitoring procedures in a hospital ward or under emergency conditions. While classic triage procedures such as “trauma triage”, “emergency triage”, or “emergency department (ED) triage”⁵–⁶ are defined as “the process of classifying patients according to injury severity and

¹Centre for e-Health, Department of Information and Communication Technology, Faculty of Engineering and Science, University of Agder, Kristiansand, Norway
²Sørlandet Hospital, Kristiansand, Norway

Corresponding author:
Martin Gerdes, University of Agder, Department of Information and Communication Technology, PO Box 422, N-4604 Kristiansand, Norway.
Email: martin.gerdes@uia.no
determining the priority for further treatment”, the long-term follow-up of patients with chronic diseases has to adapt the treatment and medication to the long-term telemonitoring perception and short-term changes of the condition of the individual patient and the corresponding needs for support and intervention. To avoid confusion between the objectives of clinical “ED triage” and those of the long-term monitoring and decision support for treatment and follow-up prioritisation of home-based patients, the terms “health status level assessment” and “health status score” will be used in this study.

Difficulties arise in the determination of actual algorithms for automatic health status score calculations, and in the definition of cut-off values for the different reported modalities to trigger the correctly colour-coded alert level for such remote home monitoring situations.

In a review of 20 studies for predictive algorithms for the early prediction of chronic obstructive pulmonary disease (COPD) exacerbations to support clinical decisions of home telemonitoring, Sanchez-Morillo et al. concluded that “models with good clinical reliability have yet to be defined”, and that “novel predictors need to be identified”.7 In the UK, 12 telehealth systems were analysed for their use of information exchange between patients and healthcare professionals.8 Data analysis methods with predefined algorithms to be configured by the health professionals were implemented in 11 of the systems. In four of the investigated systems, a colour-coded format was used to indicate the severity of the patients’ status and the priority of support and follow-up; at the same time, occasionally false positive and false negative alerts were encountered.

For the detection of day-to-day variations of the patient’s condition, it is not possible to use retrospective methods; thus, methods for automated calculations of individual threshold values will be necessary to develop. But there seems to be a gap in the knowledge on how patients in their daily routines use sensors for telemonitoring and which indicators might be used in which way for triggering alerts that should, in turn, lead to appropriate levels of responses.9

The results of this study should help to answer the following research questions:

RQ 1. Do the implemented health assessment algorithms provide the required adaptation of health status levels and corresponding alerts for treatment and follow-up needs to the individual medical condition of each COPD patient?

RQ 2. What recommendations can be given for improved cut-off values and algorithms, or for procedures to define or calculate appropriate values?

**Methods**

For the Norwegian trial in the EU-funded-project “United4Health” (U4H),10,11 a solution was developed for COPD patients at home, involving a tablet personal computer (PC) application to collect and transmit monitoring data, and a video conferencing tool to get support from one of eight nurses trained to use telemedicine (“telenurse”) at a telemedic centre.12-14 They were instructed to report their health symptoms daily (the patients were asked to send their measurement data/symptom questionnaires in the morning, since COPD usually has the highest symptom load at that time of day), followed up by a video consultation where a telenurse explained necessary actions like changes in inhalation therapy or the start of antibiotic or corticosteroid treatment, or both, if a worsened situation was encountered. (The telemedicine system for the collection and assessment of remote monitoring data was available 24/7. The support service by the telenurses was only available during office hours of the telemedical centre (Monday–Friday 8:00–15:00; Saturday 8:00–12:00). Automatic alerts were only followed-up during office hours, and patients were told to call their general practitioners (GPs) or emergency services if they needed help outside office hours.) The health status assessment included measurements of the transdermal, peripheral capillary oxygen saturation (SpO2) of the blood, and the heart rate measured with a pulse oximeter connected via Bluetooth to the tablet PC, and information about subjective symptoms observed by the patient, as gathered through a daily electronic questionnaire (Table 1). Peak Expiratory Flow (PEF) measurements were skipped from the method section by the

**Table 1. Daily questionnaire.**

<table>
<thead>
<tr>
<th>Q</th>
<th>Question</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>How do you feel today?</td>
<td>As usual, Worse, Much worse</td>
</tr>
<tr>
<td>Q2</td>
<td>How is your breathing today?</td>
<td>As usual, Worse, Much worse</td>
</tr>
<tr>
<td>Q3</td>
<td>How is your amount of sputum today?</td>
<td>As usual, Worse, Much worse</td>
</tr>
<tr>
<td>Q4</td>
<td>What is the colour of your sputum today?</td>
<td>No sputum/Clear/White, Yellow/Green/Brown</td>
</tr>
<tr>
<td>Q5</td>
<td>Are you using rescue medication/nebuliser or oxygen today?</td>
<td>No, I am not using them today, As usual, More than usual, Much more than usual</td>
</tr>
<tr>
<td>Q6</td>
<td>Have you started up with additional antibiotics after last discharge?</td>
<td>No/I have finished, Yes – Apcillin/Penicillin, Yes – Imacillin/Amoxicillin, Yes – Ciproxin, Yes – Azitromax, Yes - Other</td>
</tr>
<tr>
<td>Q7</td>
<td>Have you started up with new Prednisolone after last discharge?</td>
<td>No, Yes – 20 mg, Yes – 30 mg, Yes – 40 mg, Yes - Other</td>
</tr>
</tbody>
</table>
U4H consortium since clinicians judged PEF to be unreliable during a COPD exacerbation. Reported data were instantly evaluated automatically when received from the patient, using an algorithm to create a colour code that provided the nurse with a quick overview of the latest patient condition.15

Patients hospitalised for exacerbation of COPD (according to the Global Strategy for Diagnosis, Management and Prevention of Obstructive Lung Disease) were considered for inclusion at the time of discharge from the pulmonary department at Sørlandet Hospital HF in Kristiansand. The exclusion criteria comprised patients that did not want or could not sign the informed consent form, were not able to or did not want to use the telemedicine system, were in problematic clinical or social circumstances, or were discharged to a locality that either had no mobile broadband data coverage or could not be reached by the telemonitoring team.

A trained nurse instructed the patient in how to use the telehealth solution, and the first measurements of pulse and SpO2 were recorded and used as reference values for the following day-to-day monitoring data. Baseline data were collected by the hospital staff at discharge.

All data measured and reported by the patients with their tablet PC application were consecutively uploaded to a secured data server at the telemedical centre, where the health status levels were automatically calculated with a specific algorithm. The U4H protocol defined cut-off values for a “red” alert status based on already existing empirical clinical basis and algorithms used in several centres in the UK. For the Norwegian trial, cut-off values for a “yellow” health status were added to the clinical protocol by the study clinician (co-author of this paper) to increase the sensitivity of the system and the patients’ safety as an early warning indicator of a “notable” health condition deterioration.

Green: The patient is in a stable or improved clinical condition with unchanged medication. The self-reported health symptoms are unchanged or improved compared to the previous day. The reported pulse oximetry measurements are within an acceptable range compared to individual reference values.

Yellow: The patient is in a condition that needs special attention. If the pulse oximetry measurements, or at least one reported symptom, indicate notable deterioration since the previous day or the hospital discharge, respectively, a yellow alert is triggered. These cut-offs were: an increase in pulse of >10 bpm, a reduction in SpO2 of 4–5%, an answer to question 1, 2, 3, or 4 defined as “worse”, or question 5 answered with “more than usual” (Table 1).

Red: The patient is in a critical condition. The pulse oximetry measurements or the self-reported symptoms indicate significant deterioration since the previous day or hospital discharge, respectively; that is, an increase in pulse of >15 bpm, a pulse of >120 beats/min or <50 beats/min, an oxygen saturation that is ≥6% lower than the reference value, or Q5 answered with “much more than usual”.

Following the trial protocol, a daily teleconsultation took place with all patients during the (typically) first 14 days of trial participation (High Level of Telemonitoring Service phase). For the next 14 days (Reduced Level of Telemonitoring Service phase), the daily measurements and symptoms reporting were continued, but a video consultation would only be conducted in the case of a deteriorated health condition, indicated by yellow or red status alert. After approximately one month the equipment should be returned to the hospital, but the patients could call the telemedical centre to discuss their condition with the trained nurses at any time during the opening hours for the next 11 months (Low Level of Telemonitoring Service phase). If recommended by a doctor, the High or Reduced Level of Telemonitoring Service phases could be prolonged, resulting in situations where patients could keep the equipment at home, to be used in cases needed. After the teleconsultation, the tele-nurses could manually overwrite (i.e. increase or decrease) the automatically calculated health status assessment levels, based on their experience with the health condition of each individual patient, or following a discussion with the patient’s GP.

To investigate the obtained accuracy of the automatic health status assessment algorithms and the reasons for any manual overwrites, we have analysed the monitoring datasets from the clinical trial, with the aim of defining recommendations for improved algorithms for automatic calculations of a health status score. The datasets were de-identified and exported from the telemedicine system to Excel spreadsheets for the evaluation of the health level status assessment in this study.

Ethical considerations

This study was approved by the Norwegian Centre for Research Data (project number: 35356). All participants received oral and written information about the project and confidential treatment of the collected data. Participation was voluntary and participants could withdraw at any time without reason. All participants gave explicit written consent.

Results

During a period of two years, from May 2014 to April 2016, a total of 94 patients with COPD were recruited for the telemonitoring trial. (The trial system had been in operation until June 2016, but monitoring data and automatic assessments after April 2016 could not be considered for this study.) The baseline data are listed in Table 2. Approximately 2%, 19%, 32%, and 44% were in COPD stage I, II, III, and IV, respectively.

The planned participation duration of each patient in the trial according to the clinical trial protocol was 30 days after discharge from hospital, but with wide possibilities for personal adaption as recommended by the doctors, leading to some earlier drop-outs, and some
cases of significant longer participation. Approximately 10% of the patients were readmitted to hospital and thus had 2–3 additional periods of telemonitoring. Two thirds of the patients had a prolonged use of the telemonitoring services beyond the planned 30 days, and more than a quarter kept the equipment and reported monitoring data for more than 90 days. Approximately one third of the patients were defined as drop-out before the end of the planned period for different reasons (e.g. some of them encountered difficulties with bad quality in the video consultations due to poor mobile data coverage). The number of days included in the telemonitoring trial (taken from the last day of monitoring data reported to the telehealth system) varied from 1 to more than 365 (Figure 1), and the overall average duration of participation in the telemonitoring services was 70 days.

In total, 4970 datasets were analysed for this study received by the telehealth system, each containing a pair of pulse oximetry measurements (pulse and SpO₂), a set of answers to the daily questionnaire, or a combination of both. There could be multiple datasets for one patient for one day; for example, in case the measurement or the questionnaire had been executed and transmitted multiple times, or as a consequence of suggested repeated monitoring during a deterioration. For the health status level assessment, the first reported data of the day is used, as the result potentially triggers an alert to the telenurses for required follow-up support or treatment.

An overview of the development of the pulse oximetry data (measurements of pulse and SpO₂) reported by the patients during the first 90 days of their trial involvement is shown in Figure 2. More than two thirds of the patients that still had the monitoring equipment sent their reports (e.g. 94%, 75%, 63%, and 67% on day 1, 30, 60, and 90, respectively).

The average daily reporting rate is illustrated in Figure 3. As long as the patients had the monitoring equipment, they used it on average for at least 70% of the days (Monday–Friday) to send monitoring data. They sent pulse-oximetry measurements at an average of 75% of the days during the first 30 days, and daily questionnaires 84% of the days. During the 90-day period, monitoring data were sent at two thirds of the days (i.e. at 60 days in average).

For the accuracy analysis of the health status assessment algorithm we have selected a subgroup of 33 patients (i.e. 35% of the total of 94 patients), that have reported pulse oximetry measurements and/or answers to the daily questionnaire on a minimum of 25 days during the first 30 days of participation in the telemonitoring trial.

The automatic health status level categorisation into “green”, “yellow”, and “red” alerts were assessed before they were analysed and possibly overwritten by the telenurses. Monitoring 33 patients over 30 days corresponds to a total of 990 “patient-reporting days”, of which we received 928 reported datasets (94%). In total, 460 patient-reporting days (46%) were assessed as “green” level (indicating a stable condition of the patient on that day), 275 (28%) as “yellow” (alerting the nurse of a notable condition), and 193 (19%) as “red” level. The frequency of each health status level (averaged over all patients and all days of the considered periods) is illustrated in Figure 4.

<table>
<thead>
<tr>
<th>Table 2. Baseline data of patients included in telemonitoring trial.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td><strong>FEV1 in %</strong></td>
</tr>
<tr>
<td><strong>predicted</strong></td>
</tr>
</tbody>
</table>

SD: standard deviation.
In total, 101 assessments (i.e. 10% of all 990 “patient days”) were modified manually by the telenurses, 81% were unchanged, and 9% were incomprehensible to assess due to lack of reported data. An overview of the type of the adjustments and the reasons is given in Table 3: if the reported data, such as a pulse value, caused a red or yellow alert, and the telenurse adjusted the corresponding initial health status assessment level following the video consultation, because the health condition was actually not critical, the health status level was lowered to yellow or green, respectively, and the data was counted as “initial pulse assessment too high”. In total, 10% of the calculated health status levels were downrated and thus could be judged as false positive if the recategorization at the telemedical central, including manual overwrites by the telehealth nurses, was considered correct.
Of those, pulse assessment (42%) and daily questionnaire assessment (39%) were the main reasons for too-high assessment-level calculation.

**Discussion**

The examined trial system facilitates the remote monitoring of home-based COPD patients and the provision of decision-support information for the prioritisation of follow-up support and treatment by telenurses in a telemedical centre, in cooperation with GPs and medical specialists in hospitals. Changes of certain monitoring data from one day to the next are important to identify notable or even critical exacerbations.

Patients reported monitoring data regularly every weekday and partially beyond that during the weekend. Technical problems or difficulties to use the equipment for collecting or sending the data could obviously be avoided by involving the patients in a user-centred design process. A potential reason for incomplete reporting data (in particular, in the case of long-term routine monitoring) was a good individual perceived health condition, leading to a lower support need and the impetus to report health data. Hence, the lack of monitoring data does not necessarily indicate a bad or even alert condition, but may indicate the opposite; that is, a stable or improved health status. However, with the current algorithms of the telehealth service these aspects are difficult to interpret. If the overall health condition of a patient is unknown or unstable, the lack of data for one or a few days can be treated differently, usually assuming no significant changes in the health condition. This, in turn, requires a

![Figure 4. Unadjusted average health status level during selected monitoring periods (n = 33).](image)

**Table 3.** Analysis of manual overwrites of health status level assessment.

<table>
<thead>
<tr>
<th>Type of adaptation (related to all assessments)</th>
<th>Reasons for adaptation (related to overwrites)</th>
</tr>
</thead>
<tbody>
<tr>
<td>False-high</td>
<td></td>
</tr>
<tr>
<td>Red to yellow</td>
<td>Initial pulse assessment too high 41.6%</td>
</tr>
<tr>
<td>Yellow to green</td>
<td>Initial SpO₂ assessment too high 4.0%</td>
</tr>
<tr>
<td>Red to green</td>
<td>Initial pulse and SpO₂ assessment too high 3.0%</td>
</tr>
<tr>
<td>Yellow to green</td>
<td>Initial daily questionnaire assessment too high 38.6%</td>
</tr>
<tr>
<td>Red to green</td>
<td>Initial pulse and daily questionnaire assessment too high 7.9%</td>
</tr>
<tr>
<td>Sum</td>
<td>Sum</td>
</tr>
<tr>
<td></td>
<td>9.7%</td>
</tr>
<tr>
<td>False-low</td>
<td></td>
</tr>
<tr>
<td>Green to yellow</td>
<td>All initial assessment levels too low 5.0%</td>
</tr>
<tr>
<td>Yellow to red</td>
<td></td>
</tr>
<tr>
<td>Green to red</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>10.2%</td>
</tr>
<tr>
<td>Sum</td>
<td>100%</td>
</tr>
</tbody>
</table>
system that can ascertain the behavioural characteristics of a patient, and that can identify a good, healthy, stable condition and distinguish that from a condition in deterioration.

The health status of the monitored patients is – on average – quite constant during the trial period. The measured pulse oximetry data (Figure 2) does not show significant mean changes during the first 90 days, while extreme values (as extreme high, low pulse or very low SpO2 measurements) become less frequent with increasing monitoring duration (i.e. beyond day 60). The clinical expectation here would be an increased SpO2 and reduced pulse over time due to improved clinical conditions during monitoring. Certainly, the reduced number of participating patients causes a selection bias here that cannot be corrected and may explain why values (Figure 2) seem stable. Thus, patients that experience an improved condition may, to a larger degree, drop out of monitoring compared to those who are deteriorating or experience continued exacerbation. Beyond day 70, however, the number of higher maximum values of the score increases, indicating a “worse” or even “much worse” subjective health perception of some patients, most likely also influenced by a selection bias of those with a more chronically unstable condition.

The assessment of the pulse and especially the SpO2 measurements was of essential importance for the monitoring of the COPD status. The introduction of a “yellow alert” allowed the nurses to detect early deteriorations of COPD that, assuming sound handling, could avoid more serious deteriorations and unnecessary hospitalisations. In the PROMETE study a three-colour traffic light system was used, but only the red alert was related to the patient’s measurements.

The manual overwriting of the automatic health status level assessment by the telenurses resulted in adjusted health status levels. The analysis of the adjustments (Table 3) shows that approximately one in 10 of all assessment levels were set lower by the nurses, and mainly from “yellow” to “green”. In most of these cases, either the initial assessments of the measured pulse values were too high, or of single items of the daily questionnaire. The cut-offs had been defined accurately according to general clinical knowledge and the U4H trial protocol, and provided adaptation to the patients’ condition through individual reference values taken on the day of discharge from the hospital. However, high pulse could be detected if the patient had been in activity just before monitoring the pulse. In such cases, a retest of the pulse after 5–10 min (e.g. during a video consultation) often showed a lower pulse so that the initial red alert was adapted to yellow, and yellow to green. Thus, the automatic assessment by the telehealth service based on generic, static cut-off values might be too sensitive and not sufficiently personalised for each individual patient, causing unnecessary warnings or alerts. In a COPD pilot monitoring study, Velardo et al. used multivariate analyses and non-parametric density estimation techniques for calculations of the actual threshold values. Shah et al. found (in the same pilot study) advantages of including respiratory rate as a predictor in the multivariate approach, giving improved estimations for personalised alerts. Farmer et al. used a 95th centile as the alert threshold for heart rate and oxygen saturation, and Segrelles Calvo et al. calculated an average value over measurements from the first three consecutive days as baseline for the calculation of deviations causing alerts.

Improved algorithms for automatic calculations of a health status score will be of importance for the routine use of telehealth solutions to support the self-management of patients with COPD. Improved algorithms are also relevant in systems for telenurses to assist patients with correct medical advises. An approach for ongoing research work would be the utilisation of smart, self-learning systems with artificial intelligence (AI) technologies. The development of appropriate machine learning algorithms and corresponding real-time data evaluation systems requires more clinical data.

Related studies evaluating the clinical outcome of the interventions by the telenurses triggered by the alerts during this trial are under publication.

Conclusions and lessons learned

The automatic, computer-based assessment of monitoring data from home-living COPD patients, as explored in this trial system, utilises algorithms and rules that depend on cut-off values for certain measurements or self-reported parameters, thereby reflecting the health condition of the patients. As such algorithms and rules look at the day-to-day development and changes of the reported data to identify notable or critical abnormalities, the accuracy and reliability of the health status assessments depend highly on the continuous reporting of the required parameters by patients, at least on a daily or more frequent, regular basis. This study reveals that there were no technical problems or usability challenges causing significant interruptions of the self-reporting of monitoring data by the patients, but that a stable and improving subjective health and well-being status could lead to a reduced motivation to regularly report health data.

Furthermore, the cut-off values defining certain health status levels are crucial for the accuracy of the health status assessment, and thus for the quality of the rule-based analysis. The results from the trial system show that static, generic cut-off values, as used, for example, for emergency medicine triage, are sufficient to trigger emergency alerts, while they do not support decisions by telenurses with high accuracy and reliability. Instead, cut-off values must be personalised to the individual health condition and common levels of health parameters. Predetermined, personally adapted cut-off values defined at the beginning of a telehealth monitoring period of a patient might not adapt sufficiently with the dynamic general health condition of a patient. The results from the tested algorithms are not sufficient to provide for reliable
(avoiding false-negative) and efficient (avoiding false-positive) alerts. More personalised, adaptive, and intelligent assessments are needed.

The tested telehealth trial system has shown strengths and advantages in the patient interaction, data collection, timely, reliable, and safe transmission of monitoring data, and the interaction with health professionals. Ongoing research work will combine, test, and evaluate this in combination with smarter, AI-based assessment and decision-support algorithms.

Acknowledgements
This research was a result of the international cooperation between partners within the EU FP7 project United4Health. The authors would like to thank participants for their contribution in the study. Especially, we would like to thank Inger Alice Naley Ås at Sørlandet Hospital and Karoline Vasbo Nyhus at the Municipality of Kristiansand for their valuable help in collecting patient data.

Declaration of conflicting interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the projects UNIversal solutions in Telemedicine Deployment for European HEALTH care, 2013-2015 ICT PSP (CIP-ICT PSP-2012-3) and Point-of-Care Services Agder, sub-project financed by the Research Council of Norway, 2013-15 (227131/O70).

References