Oxygen saturation measurements in telemonitoring of patients with COPD: a systematic review
Peer-reviewed author version


DOI: 10.1080/17476348.2018.1417842
Handle: http://hdl.handle.net/1942/26559
Oxygen saturation measurements in telemonitoring of patients with COPD: a systematic review

Joren Buekers, Patrick De Boever, Anouk W Vaes, Jean-Marie Aerts, Emiel F.M. Wouters, Martijn A. Spruit & Jan Theunis

To cite this article: Joren Buekers, Patrick De Boever, Anouk W Vaes, Jean-Marie Aerts, Emiel F.M. Wouters, Martijn A. Spruit & Jan Theunis (2017): Oxygen saturation measurements in telemonitoring of patients with COPD: a systematic review, Expert Review of Respiratory Medicine, DOI: 10.1080/17476348.2018.1417842

To link to this article: https://doi.org/10.1080/17476348.2018.1417842

Accepted author version posted online: 15 Dec 2017.
Oxygen saturation measurements in telemonitoring of patients with COPD: a systematic review

Joren Buekers¹,², Patrick De Boever¹, Anouk W Vaes³, Jean-Marie Aerts², Emiel F.M. Wouters³, Martijn A. Spruit³,⁴,⁵ and Jan Theunis¹

¹ Environmental Risk and Health unit, Flemish Institute for Technological Research (VITO), Mol, Belgium
² Measure, Model & Manage Bioresponses (M3-BIORES), Department of Biosystems, KU Leuven, Leuven, Belgium
³ Department of Research and Education, CIRO, Horn, the Netherlands
⁴ REVAL - Rehabilitation Research Center, BIOMED - Biomedical Research Institute, Faculty of Medicine and Life Sciences, Hasselt University, Diepenbeek, Belgium
⁵ Department of Respiratory Medicine, Maastricht University Medical Centre, Maastricht, the Netherlands

*Corresponding author:
Jan Theunis
Flemish Institute for Technological Research (VITO)
Belgium
Email: jan.theunis@vito.be
Abstract

Introduction: Telemonitoring applications are expected to become a key component in future healthcare. Despite the frequent use of SpO₂ measurements in telemonitoring of patients with chronic obstructive pulmonary disease (COPD), no profound overview is available about these measurements.

Areas covered: A systematic search identified 71 articles that performed SpO₂ measurements in COPD telemonitoring. The results indicate that long-term follow-up of COPD patients using daily SpO₂ spot checks is practically feasible. Very few studies specified protocols for performing these measurements. In many studies, deviating SpO₂ values were used to raise alerts that led to immediate action from healthcare professionals. However, little information was available about the exact implementation and performance of these alerts. Therefore, no firm conclusions can be drawn about the real value of SpO₂ measurements. Future research could optimize performance of alerts using individualized, time-dependent thresholds or predictive algorithms to account for individual differences and SpO₂ baseline changes. Additionally, the value of performing continuous measurements should be examined.

Expert commentary: Standardization of the measurements, data science techniques and advancing technology can still boost performance of telemonitoring applications. All these opportunities should be thoroughly explored to assess the real value of SpO₂ in COPD telemonitoring.

Keywords: COPD; Chronic Obstructive Pulmonary Disease; Telemonitoring; Remote monitoring; Oxygen saturation; Oximetry; Exacerbation
1 Introduction

Chronic obstructive pulmonary disease (COPD) is a highly prevalent chronic non-communicable disease affecting 64 million people [1]. By 2030, COPD will become the third leading cause of death worldwide [2]. Total costs associated with this disease are estimated to be €141 billion in Europe [3]. Major contributors to the high economic and societal burden of COPD are exacerbations [4]. Exacerbations of COPD (defined as “a sustained worsening of the patient's condition, from the stable state and beyond normal day-to-day variations”) [5] are typically stressful events in the natural history of the disease [6]. Low blood oxygen concentrations (hypoxemia) increase the risk of exacerbation, while exacerbations can also induce hypoxemia [7]. The risk of hypoxemia rises with increasing disease severity [6] and can lead to long-term adverse effects such as pulmonary hypertension and systematic inflammation, reducing quality of life in patients with COPD [7]. Long-term administration of oxygen decreases mortality in patients with severe resting hypoxemia [8] and is prescribed based on blood gas measurements [6].

Pulse oximetry is a non-invasive method to assess arterial blood oxygen concentrations that uses light absorption characteristics of hemoglobin [9]. Pulse oximeters produce a photoplethysmography (PPG) waveform, from which both peripheral oxygen saturation (SpO2) and heart rate (HR) can be derived [10]. An SpO2 value below 92% indicates blood gas measurements should be performed to assess the need for supplemental oxygen therapy [11]. Pulse oximetry is also used in telemonitoring applications [12,13]. These applications make use of information technology to monitor patients at a distance without a healthcare professional present at the monitoring site [14]. Telemonitored patients are able to consult healthcare professionals through video or phone calls (remote consultations) or send information gathered
via questionnaires or physiological measurements (e.g. HR or SpO₂) to healthcare professionals. In the case of management of COPD patients, this information can be used to detect exacerbations. Early detection of exacerbations can reduce hospitalizations and improve recovery and health-related quality of life [15]. Furthermore, it has been suggested that telemonitoring applications could even predict these events [12], which could lead to the prevention of detrimental hospitalizations [15]. Generally, telemonitoring applications are expected to become a key component in future healthcare [16].

Despite the frequent use of SpO₂ measurements in COPD telemonitoring [13], no profound overview is available about these measurements. Therefore, this systematic review will evaluate the scientific literature on the application of SpO₂ measurements in telemonitoring of patients with COPD.

2 Methods

2.1 Search strategy

The literature search was performed in PubMed and Web of Science on the 15th of March 2017. The systematic search terms consisted of the following keywords: (mHealth OR "mobile health" OR tele(-)health OR tele(-)monitoring OR tele(-)medicine) AND (COPD OR "chronic obstructive pulmonary disease"). Search terms were not narrowed down to only SpO₂ measurements to avoid exclusion of potentially valuable articles that did not mention SpO₂ measurements in the abstract. Meta-analyses and reviews were excluded a priori. Article screening excluded duplicates, non-English articles, editorials, conference proceedings, articles without full-text availability and articles that were not applying SpO₂ measurements in telemonitoring of patients with COPD. Additional articles were included based on reference list
screenings from the articles retrieved via PubMed or Web of Science. Results were discussed with all co-authors and inventoried by the first author.

2.2 Information extraction

First, articles originating from the same study were grouped. If a project name was provided, this name was assigned to the corresponding study. Then, studies applying remote consultations were identified, as they have significantly different characteristics compared to telemonitoring implementations not requiring personal contact. Afterwards, participant numbers were extracted and in the case of a randomized controlled trial (RCT), participant numbers of both telemonitoring and usual care group were denoted. When possible, the distinction between COPD patients and other participants was made. In a next step, characteristics of the studies were extracted by identifying study duration, SpO2 measuring frequency and type of oximeter and associated telemonitoring system (used for storage and/or transmission of SpO2 data). Finally, information was extracted on the application and value of SpO2 measurements.

This review specifically aims to provide an overview of the use of SpO2 measurements in telemonitoring of COPD patients. However, most of the reviewed studies primarily focused on the results of whole telemonitoring systems, which depend on several technical and organizational factors of the whole telemonitoring set-up. Therefore, reporting on the overall results of the included studies is outside the scope of this systematic review.

3 Results

3.1 Article selection

The initial PubMed and Web of Science search retrieved 532 articles (Figure 1). After removal of duplicates (n=169), non-English articles (n=15), editorials (n=13), conference proceedings (n=12) and articles without full text availability (n=11), 312 articles remained. Full text
screenings excluded articles that were not applying telemonitoring in COPD or performing SpO2 measurements (n=243). Addition of articles from reference lists (n=2) resulted in a final set of 71 articles.

3.2 Study description

Table 1 provides a description of the 71 included articles, originating from 50 distinct studies. The first article was published in 2004. Sixty-three of the 71 articles (89%) have been published after 2010.

The 50 included studies can be divided into studies with supervised measurements (during remote consultations; n=6) and independent patient measurements (n=44). Two types of remote consultations were performed: (i) short-term (≤14 days) follow-up of patients released from the hospital after an exacerbation [17–23] (n=4); or (ii) long-term (12 months) remote monitoring [24–27] (n=2). Regarding the 44 studies with independent patient measurements, most implemented telemonitoring for at least 2 months (n=39) and almost two third of the studies lasted for at least 6 months (n=28). In almost half of the 50 included studies, less than 30 COPD patients were included for telemonitoring (n=24). Very few studies reported on analyses of specific patient subpopulations (e.g. patients with and without oxygen therapy) and methods to characterize patients were not consistent. Therefore, no analyses of specific patient subpopulations could be made in this review.

3.3 Measurements, oximeters and telemonitoring systems

Forty-nine of the 50 included studies (98%) used pulse oximeters for spot check measurements. Continuous measurements were only performed in the study of Faria et al. [28] (sampling frequency not specified). Frequency of spot check measurements ranged from 3 times in 5 hours to once weekly, with most of the studies measuring once daily (n=38). Only three articles specified a protocol for performing the measurements. In Hurst et al. [29], participants performed
SpO₂ measurements each morning at the same time after 10 minutes of rest, on the same finger, but before taking medication. The highest obtained value during the measurement was retained. Jodar-Sanchez et al. [30] specified that measurements were taken 20 minutes after taking medication while seated, rested and on oxygen therapy. Shany et al. [31] had patients perform measurements after taking medications and while on oxygen therapy.

The pulse oximeter was specified in 26 of the 50 studies (Table 1). Almost half (n=12) of the 26 specified oximeters are manufactured by Nonin Medical (7 Nonin Onyx; 1 Nonin Avant; 4 unspecified Nonin). Nonin Onyx is a finger clip oximeter for spot check measurements; Nonin Avant is a wrist worn device with a linked finger clip able to perform continuous measurements. Considering all studies, the oximeters were used in combination with 15 different types of telemonitoring systems (Table 1).

### 3.4 Application of SpO₂ measurements

In many studies, deviating SpO₂ values were used to raise alerts (SpO₂ alerts) that led to immediate action from healthcare professionals, ranging from phone calls to hospital admissions (n=34; Table 1). Other studies developed methods to predict exacerbations through retrospective analysis of the measured SpO₂ values (n=5). An additional retrospective analysis was also performed in three of the 34 studies that carried out immediate action [32–34]. The other eleven studies (i) did not specify the exact SpO₂ use during remote consultations (n=5); or (ii) were set up for evaluation of a telemonitoring system and did not further use the measured SpO₂ values (n=6).

### 3.5 SpO₂ alerts leading to immediate action
The SpO2 alerts in the 34 studies that carried out immediate action were based on (i) abnormal values (n=10; no further information was provided to describe what was considered to be an abnormal value), (ii) thresholds values (n=23; as discussed in next paragraph) or (iii) the output of an algorithm (n=1) (Table 1). This algorithm used oximetry measurements in combination with questionnaire answers and spirometry measurements to predict exacerbation probability. High probabilities raised alerts, but the exact algorithm was not further specified [35].

Two types of thresholds can be distinguished: generic (n=6) or individualized (n=12). Five studies did not specify the type of threshold (n=5). Generic thresholds were fixed on one value (88% [36,37], 90% [24–26] or 92% [38]) or on multiple values for a graded severity indication. Koff et al. [39] specified thresholds of 90% and 88%, Ho et al. [40] on 92% and 90% and Colantonio et al. [41] made the distinction between breathing room air (thresholds of 92% and 90%) and breathing O2 (thresholds of 93%, 90% and 80%). Three studies that applied individual thresholds did not specify the method for determining the threshold values [42–44]. The other nine studies with individual thresholds based the threshold values on caregiver assessments or baseline measurements. Seven of these nine studies did not further clarify the exact determination of the threshold values. The other two studies denoted a quantitative method for determining the individualized threshold value: Segrelles Calvo et al. [45] set the threshold 3% below the average of first 3 measuring days. The articles originating from the EDGE project [34,46–48] defined the threshold value as the 95% percentile of the cumulative density function of the measured values of the first 6 weeks.

3.6 Exacerbation prediction through retrospective analysis
Eight articles developed methods to predict exacerbations through retrospective analysis of the measured SpO2 values: five originated from retrospective studies and three from studies that performed an additional retrospective analysis. Brown-Connolly used retrospective analysis to calculate the optimal generic threshold value by taking into account the accuracy of alerts to warn for hospitalizations, emergency admissions and home visits. A Receiver Operating Characteristics (ROC) analysis indicated 85-86% as the generic SpO2 threshold with both the highest sensitivity (0.62) and specificity (0.91) [32].

The other seven articles extracted features from daily measured SpO2 and used these features in addition to the measured SpO2 values to retrospectively develop methods to predict exacerbations. Mohktar et al. [33] calculated four features, based on the combination of daily measurements and the measurements of the previous 30 days: 30-day distribution mean and standard deviation, percentage change of daily measured values compared to distribution mean and the z-score of measured values compared to the 30-day distribution. The same features were calculated from five other variables (i.e. HR, blood pressure, lung function parameters, weight and temperature), after which they were used as input for a classification and regression tree algorithm to predict next-day risk of exacerbation. Shah et al. [34] calculated mean and linear fit gradient from 7-day periods of SpO2, HR and respiratory rate to be used as input for a logistic classifier to predict exacerbations. Jensen et al. [49] and Riis et al. [50] calculated features based on differing types (mean, standard deviation, skewness, kurtosis or linear regression between SpO2 and HR) and durations (ranging between 5 and 30 days) and used them for prediction of exacerbation risk in the following 30 days or as input for a k-nearest neighbor classifier for prediction of exacerbation onset, respectively. Hurst et al. [29] calculated an oximetry score, based on the z-scores of SpO2 and HR from daily measurements compared to the distribution of
measurements from the first 30 stable days. An oximetry-peak expiratory flow (PEF) score was calculated similarly by adding z-scores of PEF measurements to the equation. These scores were used to discriminate exacerbation onset from normal day-to-day symptom variations. Merone et al. [51] further specified this oximetry score by applying weights to the elements in the oximetry score formula and used this in a finite-state machine model for prediction of different types of adverse events (e.g. onset of exacerbation, hypoxemia or dyspnea). Lastly, Clarke et al. [52] decomposed time series of daily SpO2 measurements of 4 patients in a long-term trend, a short-term trend and a residual signal. In one patient, the declining long-term trend clearly indicated the worsening health status of this patient. The residual signal outperformed the short-term trend for exacerbation predictions.

3.7 Value of SpO2 measurements

Most studies that carried out immediate action (n=31) did not provide information about the correctness or effectiveness of the implemented SpO2 alerts and only considered general, long-term outcomes of the whole telemonitoring set-up (e.g. quality of life, morbidity or healthcare costs). Three studies that carried out immediate action and five studies with retrospective analyses did quantitatively evaluate the performance of stand-alone SpO2 measurement to detect or predict exacerbations (Table 2). Except for Hurst et al., alerts based on SpO2 measurements performed better than alerts based on other physiological variables, as can be seen in Table 2 by means of the higher AUC [32,34], the higher Cohen’s kappa [33], because SpO2 was better able to differentiate between exacerbation days or usual days [53], or because SpO2 raised more accurate alerts [44,45]. The four articles that examined the ability to predict exacerbations of both stand-alone SpO2 measurements and of measurements of a combination of variables (including SpO2), all found an increase in performance when other physiological variables were
4 Discussion

This systematic review identified and evaluated the use of SpO₂ measurements in telemonitoring of patients with COPD. In many studies, daily SpO₂ measurements were performed for more than 6 months, indicating that long-term follow-up of COPD patients using daily SpO₂ measurements is practically feasible. However, very few studies specified protocols for performing the measurements. Often the used oximeter was not mentioned. When specified, the oximeters from Nonin Medical were most frequently used. Overall, participant numbers were relatively low and the articles had no consistent method for characterizing the included patients. In many studies, deviating SpO₂ values were used to raise alerts that led to immediate action from healthcare professionals (e.g. patient visits or hospital admissions). However, little information has been made available about the exact implementation of these alerts. Almost a third of the studies carrying out immediate action indicated that alerts were based on abnormal values, without further specifying when values were considered abnormal. Other alerts were based on exceeded threshold values, but these values were often not specified or only based on unspecified caregiver assessments. In addition, little attention was given to the optimal implementation of alerts. When generic threshold values were specified, they ranged between 88% and 92%. However, the retrospective analyses of Brown-Connolly suggested lower values (85-86%) as the optimal SpO₂ threshold [32]. Moreover, COPD patients are generally considered a very heterogeneous population [88] with an overall decline in health status [6] and a variable baseline SpO₂ [52]. Generic thresholds cannot take this variability into account. It can thus be assumed that individualized, time-dependent thresholds (i.e. personal thresholds that can change...
over time) will outperform generic thresholds [64]. Individualized thresholds were already implemented in some studies that carried out immediate action, but except for Segrelles Calvo et al. [45] and the articles originating from the EDGE project [34,46–48], no information was provided on how these thresholds were determined. All these elements illustrate the current lack of knowledge and transparency concerning alert implementations.

The studies with retrospective analyses focused more on alert optimizations. Brown-Connolly used retrospective analyses to calculate the optimal generic threshold value [32]. The other retrospective studies calculated additional features from daily measured SpO₂ values and used these features to retrospectively predict exacerbations by applying thresholds to the feature values or by using the features as input for predictive algorithms. The calculated features were based on a combination of individual present day values and individual values from the last week or month, which allowed accounting for both individual differences and longer-term health status changes. This is important because general health status changes can lead to a different baseline SpO₂ [52]. A decline in baseline SpO₂, not related to exacerbations, could thus lead to a high amount of false alerts when the individual thresholds are not able to change accordingly. In addition, these features have the possibility to provide valuable information about the general, long-term health status of the monitored patients, as has been suggested by Clarke et al. [52]. Surprisingly, most of the calculated features were still based on simple, basic statistics such as mean, standard deviation or a linear fit gradient of past periods [33,34,49,50]. Only Clarke et al. [52] approached the daily measurements as a time series. The authors proposed a method to monitor both health status and the onset of exacerbations by using specific time-series analyses. However, their method is derived from data of only 4 patients and thus needs further validation.
Little information was provided about the correctness and effectiveness of the SpO2 alerts in studies that performed immediate action, as these studies mostly only considered general, long-term outcomes of the whole telemonitoring set-up (e.g. quality of life, morbidity or healthcare costs). Nonetheless, the three articles that did evaluate the performance of SpO2-alerts quantitatively, indicate that these alerts could have the potential to detect or predict exacerbations: Martin-Lesende et al. [44] found that SpO2 alerts were mostly triggered in the five days prior to hospitalization, Segrelles Calvo et al. [45] found that most alerts for exacerbations were triggered by SpO2 and Burton et al. [53] indicated that SpO2 measurements could differentiate between exacerbation days and regular days. The retrospective analyses provide more information about the value of SpO2 measurements and indicate that stand-alone SpO2 measurements could have predictive power for exacerbations. Furthermore, the results in Table 2 suggest that exacerbation predictions based on stand-alone SpO2 measurements perform better than predictions based on measurements of other stand-alone physiological variables. Nevertheless, adding other physiological variables to SpO2 measurements can increase performance. A combination of SpO2 and HR might be especially valuable, having the additional advantage that both can be calculated from the same PPG signal as measured by oximeters [10]. Overall, this systematic review identified multiple shortcomings of the included articles concerning the SpO2 measurements and applications: measurement protocols and used oximeters were often not specified, participant numbers were generally low, little information was available about the implementation and performance of the alerts and alert implementations were not optimized. Therefore, no firm conclusions can be drawn on the real value of SpO2 measurements. Nevertheless, the few studies that assessed the value of SpO2 measurements indicated that these measurements could be valuable for exacerbation detections or predictions.
Still, future research could optimize alerts based on regular SpO2 measurements by using individualized, time-dependent thresholds or predictive algorithms to account for individual differences and SpO2 baseline changes.

This systematic review identified only one study that performed continuous SpO2 measurements [28]. According to the authors of this review, future research should consider performing more continuous measurements of SpO2 instead of spot checks, possibly in combination with measurements of other variables that acutely influence SpO2 (e.g. physical activity and sleep [28]). This will provide a completer picture of the dynamics of SpO2 which can further lead to new features that can be used to determine individualized and time-dependent thresholds or predictive algorithms. Additionally, these measurements and features could be used to optimize oxygen therapy or training intensities in telerehabilitation programs [89]. In the study of Faria et al., patients treated according to conventional methods for long-term oxygen therapy carried out simultaneous continuous measurements of SpO2 and physical activity during activities of daily living, in combination with continuous SpO2 measurements during sleep [28]. Low SpO2 values were still found in continuous measurements during exercise, rest, and sleep, indicating that current long-term oxygen therapy guidelines (based on spot check measurements) could further be optimized by using continuous measurements. The dynamics of continuously measured SpO2 have so far only been examined during sleep in laboratory conditions, e.g. in obstructive sleep apnea [90] or apnea-hypopnea syndrome diagnosis [91]. In these studies, new insights were obtained by calculating a variety of features from continuously measured nocturnal SpO2.

Future research should assess the real value of continuous measurements through a formal comparison between spot check and continuous measurements, also considering differences in e.g. user acceptance or costs. Technological improvements can still increase the accuracy of
continuous measurements taken with user friendly wearable devices (e.g. accurate measurements on the wrist, arm, or earlobe).

5 Conclusions

This review evaluates the scientific literature on the application of SpO2 measurements in telemonitoring of patients with COPD. The results indicate that long-term follow-up of COPD patients using daily spot check SpO2 measurements is practically feasible. However, very few studies specified protocols to perform SpO2 measurements. In many studies, deviating SpO2 values were used to raise alerts that led to immediate action from healthcare professionals. Other studies developed methods to predict exacerbations through retrospective analysis of the measured SpO2 values.

Little information was available about the exact implementation and performance of these alerts. Therefore, no firm conclusions can be drawn about the real value of SpO2 measurements. Nevertheless, the few studies that assessed the value of SpO2 measurements indicated that these measurements could be valuable for exacerbation detections or predictions.

Future research could optimize alerts based on daily measured SpO2 by using individualized, time-dependent thresholds or predictive algorithms to account for individual differences and SpO2 baseline changes. Additionally, the value of performing more continuous measurements should be examined, as these can make it possible to examine the SpO2 dynamics and account for factors that acutely influence the measurements (e.g. physical activity or sleep).

6 Expert commentary

The demographic shift towards an older population contributes to the increased burden of chronic non communicable diseases. These diseases are becoming more prominent causes of morbidity and mortality, and the treatment costs will keep increasing. New ways of managing
chronic diseases are necessary to reach a sustainable healthcare system, including a shift in focus from treatment only towards prevention. Prevention does not only include keeping diseases from developing, it also includes preventive actions to gain control over the course of the disease. For the latter, telemonitoring can be used to timely indicate when preventive actions are required. Unfortunately, telemonitoring is not yet generally accepted because of a lack of compelling evidence of its beneficial long-term effects on healthcare costs, morbidity or mortality.

This lack of evidence is not surprising. Until now, very little attention has been directed towards the optimal exploitation of the gathered data, as has been confirmed by the results of this review. Standardizations, data science techniques and advancing technology can still boost performance of telemonitoring applications. Better performances will lead to higher acceptance, providing more opportunities to embed telemonitoring applications in daily practice, ultimately leading to a more sustainable healthcare system.

This review exposed some of the opportunities to increase performance of telemonitoring applications that measure oxygen saturation (SpO₂) in COPD patients. Telemonitoring of patients with COPD is mostly applied for detection of exacerbations to enable earlier and better treatment, but exacerbation detection has not been proven sufficiently accurate yet. A first step that must be taken to reach acceptable accuracies is the standardization of the SpO₂ measurements. Very few articles provide information about a protocol to perform the measurements, which leads to unwanted variability in the measured values. Secondly, only articles with retrospective analyses explored some of the possibilities of data science techniques. This makes it impossible to assess the real added value of data science on longer-term outcomes such as healthcare costs or mortality. In addition, the applied techniques were mostly only based on simple, basic statistics such as means or standard deviations. More advanced data analysis
techniques using time series analyses and predictive modelling should be applied in studies that can initiate immediate action to explore the real potential of telemonitoring in patients with COPD. Lastly, the advancing technology can make it possible to accurately measure \( \text{SpO}_2 \) continuously instead of through spot checks only. This will allow exclusion of unwanted influences on the measurements (e.g. physical activity) and provide a completer picture of the dynamics of \( \text{SpO}_2 \) during the day and/or night. Applying the appropriate analyses on these continuous measurements can give rise to a whole new set of individualized and time-dependent features that can further improve the performance of telemonitoring applications. Only when all these opportunities are thoroughly explored, the real value of telemonitoring \( \text{SpO}_2 \) in COPD patients can be assessed.

7 Five-year view

No firm conclusions can be drawn about the real value of \( \text{SpO}_2 \) measurements in telemonitoring applications because many improvements that could lead to better performances are still underexplored. Therefore, the focus of future research will be increasingly on incorporating data science techniques in telemonitoring applications for COPD patients. The use of more complex predictive algorithms will improve exacerbation predictions and the constant technological advancements will make it possible to use wearables to continuously and accurately measure \( \text{SpO}_2 \) without interfering in the patient’s daily life. Using appropriate data science techniques on these continuous \( \text{SpO}_2 \) measurements can further boost the performance of telemonitoring applications, which will ultimately lead to a higher acceptance of telemonitoring applications in daily practice.
Database search n = 532

Exclude duplicates (n = 169)

After duplicate removal n = 363

Exclude
- Non-English articles (n = 15)
- Editorials (n = 13)
- Conference proceedings (n = 12)
- No full text available (n = 11)

After general screening n = 312

Exclude based on full text content (n = 243)

After full-text screening n = 69

Add from reference lists (n = 2)

Included articles n = 71
Table 1 Overview of the 71 included articles, based on 50 individual studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Project name</th>
<th>Participant(s) (n)</th>
<th>FEVI %pred (mean ±)</th>
<th>Study duration</th>
<th>Measuring frequency</th>
<th>Oximeter type</th>
<th>Telemonitoring systems</th>
<th>SpO2 application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitacca</td>
<td>200</td>
<td>/</td>
<td>17* 44* 54±33</td>
<td>33 ±9 ±12</td>
<td>12 months</td>
<td>1x/week</td>
<td>Nonin Onyx or 2500</td>
<td>Digicom 30 EM</td>
<td>Alert: Threshold – Generic</td>
</tr>
<tr>
<td>Vitacca</td>
<td>200</td>
<td>/</td>
<td>17* 44* 54±33</td>
<td>33 ±9 ±12</td>
<td>12 months</td>
<td>1x/week</td>
<td>Nonin Onyx or 2500</td>
<td>Digicom 30 EM</td>
<td>Alert: Threshold – Generic</td>
</tr>
<tr>
<td>Koizumi [27]</td>
<td>200</td>
<td>/</td>
<td>22* 22 39±44</td>
<td>2-8</td>
<td>1-11</td>
<td>/</td>
<td>Nihon Kohden</td>
<td>Nihon Kohden</td>
<td>No alert: Use not specified (RC)</td>
</tr>
<tr>
<td>Schou [18]</td>
<td>201</td>
<td>Virtual Hospital Trial</td>
<td>22 22 39±44</td>
<td>2-8</td>
<td>1-11</td>
<td>/</td>
<td>Nihon Kohden</td>
<td>Nihon Kohden</td>
<td>No alert: Use not specified (RC)</td>
</tr>
<tr>
<td>Schou [29]</td>
<td>201</td>
<td>/</td>
<td>176±12</td>
<td>1x/week</td>
<td>Nonin Onyx or 2500</td>
<td>Digicom 30 EM</td>
<td>Alert: Threshold – Generic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emme [33]</td>
<td>201</td>
<td>/</td>
<td>132 134 33±37</td>
<td>5-9 days</td>
<td>Once daily</td>
<td>/</td>
<td>MIR Spirotel</td>
<td>/</td>
<td>No alert: Use not specified (RC)</td>
</tr>
<tr>
<td>Sorknes [23]</td>
<td>201</td>
<td>/</td>
<td>132 134 33±37</td>
<td>5-9 days</td>
<td>Once daily</td>
<td>/</td>
<td>MIR Spirotel</td>
<td>/</td>
<td>No alert: Use not specified (RC)</td>
</tr>
<tr>
<td>Gottlieb [21]</td>
<td>201</td>
<td>TELEKOL</td>
<td>72</td>
<td>14 days</td>
<td>Once daily</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>No alert: Use not specified (RC)</td>
</tr>
<tr>
<td>Saleh [22]</td>
<td>201</td>
<td>/</td>
<td>99</td>
<td>14 days</td>
<td>Once daily</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>No alert: Use not specified (RC)</td>
</tr>
<tr>
<td>Venter [54]</td>
<td>201</td>
<td>/</td>
<td>10* 10* 10*</td>
<td>12 months</td>
<td>/</td>
<td>TMC Health Monitor</td>
<td>/</td>
<td>/</td>
<td>Abnormal</td>
</tr>
<tr>
<td>Jensen [49]</td>
<td>201</td>
<td>TELEKAT</td>
<td>57</td>
<td>4 months</td>
<td>Prescribed</td>
<td>Nonin Onyx</td>
<td>/</td>
<td>/</td>
<td>Retrospective analysis</td>
</tr>
<tr>
<td>Lilholt [64]</td>
<td>201</td>
<td>/</td>
<td>60</td>
<td>2 weeks</td>
<td>First 2 weeks:</td>
<td>/</td>
<td>Samsung Galaxy Tab 2</td>
<td>/</td>
<td>Alert: Abnormal values</td>
</tr>
<tr>
<td>Lilholt [64]</td>
<td>201</td>
<td>/</td>
<td>60</td>
<td>2 weeks</td>
<td>First 2 weeks:</td>
<td>/</td>
<td>Samsung Galaxy Tab 2</td>
<td>/</td>
<td>Alert: Abnormal values</td>
</tr>
<tr>
<td>Haesum [61]</td>
<td>201</td>
<td>/</td>
<td>60 56 48±50</td>
<td>12</td>
<td>Afterwar</td>
<td>/</td>
<td>TMC Health Monitor</td>
<td>/</td>
<td>Abnormal</td>
</tr>
<tr>
<td>Achelman [35]</td>
<td>201</td>
<td>/</td>
<td>651 704 38±7</td>
<td>12 months</td>
<td>2x/week</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Alert: Algorithm</td>
</tr>
<tr>
<td>Colantonio [41]</td>
<td>201</td>
<td>/</td>
<td>26</td>
<td>115 days</td>
<td>3x/week</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Alert: Threshold</td>
</tr>
<tr>
<td>Finkelstei</td>
<td>200</td>
<td>TeleHomeCare</td>
<td>11* 14* 11* 14*</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Honeywell II</td>
<td>Honeywell II</td>
<td>Alert: Threshold</td>
</tr>
<tr>
<td>Lamothe [43]</td>
<td>200</td>
<td>/</td>
<td>82* 82*</td>
<td>6-243 days</td>
<td>Once daily</td>
<td>Honeywell II</td>
<td>/</td>
<td>/</td>
<td>Alert: Abnormal</td>
</tr>
<tr>
<td>Smaradotti</td>
<td>201</td>
<td>/</td>
<td>63 30 48±36</td>
<td>90 days</td>
<td>Once daily</td>
<td>/</td>
<td>Digicom</td>
<td>/</td>
<td>Alert: Threshold –Generic</td>
</tr>
<tr>
<td>van der Heijden [62]</td>
<td>201</td>
<td>/</td>
<td>5</td>
<td>9 days</td>
<td>Once daily</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Alert: Threshold</td>
</tr>
<tr>
<td>Cardozo [63]</td>
<td>201</td>
<td>/</td>
<td>119</td>
<td>60 days</td>
<td>Once daily</td>
<td>/</td>
<td>Bosch Health</td>
<td>/</td>
<td>Alert: Abnormal</td>
</tr>
<tr>
<td>Hamad [64]</td>
<td>201</td>
<td>/</td>
<td>183</td>
<td>80.7 days</td>
<td>Once daily</td>
<td>/</td>
<td>Docobo</td>
<td>/</td>
<td>Alert: Threshold</td>
</tr>
<tr>
<td>Hurst [29]</td>
<td>201</td>
<td>/</td>
<td>31</td>
<td>46 ±87 days</td>
<td>Once daily</td>
<td>/</td>
<td>Nonin Onyx</td>
<td>/</td>
<td>Retrospective analysis</td>
</tr>
<tr>
<td>Davis [65]</td>
<td>201</td>
<td>/</td>
<td>58* 174* 11* 15*</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Cardiocom</td>
<td>/</td>
<td>Alert: Abnormal</td>
</tr>
<tr>
<td>Ho [40]</td>
<td>201</td>
<td>/</td>
<td>53 53 62±62</td>
<td>2 months</td>
<td>Once daily</td>
<td>/</td>
<td>GE Healthca II</td>
<td>Bosch Health</td>
<td>Alert: Threshold – Generic</td>
</tr>
<tr>
<td>Koff [39]</td>
<td>200</td>
<td>/</td>
<td>20 20 34 31 ±31</td>
<td>2 months</td>
<td>Once daily</td>
<td>/</td>
<td>Honeywell II</td>
<td>Honeywell II</td>
<td>Alert: Threshold – Generic</td>
</tr>
<tr>
<td>Gellis [66]</td>
<td>201</td>
<td>I-TEAM or TEL-</td>
<td>11* 15*</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Alert: Abnormal</td>
</tr>
<tr>
<td>Gellis [66]</td>
<td>201</td>
<td>I-TEAM or TEL-</td>
<td>11* 15*</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Alert: Abnormal</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Project name</td>
<td>Participant s (n)</td>
<td>FEV1 %pred (mean ± SD)</td>
<td>Study duration</td>
<td>Measuring frequency</td>
<td>Oximeter type</td>
<td>Telemonitoring systems</td>
<td>SpO2 application</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>--------------</td>
<td>------------------</td>
<td>------------------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Jodar</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jodar</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ure [69]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burton</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tejeiro [70]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De San Miguel</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ding [72]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segrelles Calvo</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MacNab [38]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenealy [73]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDowell I [74]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chatwin [42]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown-Connolly</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zamith [75]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steventon [76]</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table Notes
- **Author**: Name of the author or group of authors.
- **Year**: Year of publication.
- **Project name**: Name of the project.
- **Participant s (n)**: Number of participants.
- **FEV1 %pred (mean ± SD)**: Mean percentage predicted FEV1 with standard deviation.
- **Study duration**: Duration of the study.
- **Measuring frequency**: Frequency of measurement.
- **Oximeter type**: Type of oximeter used.
- **Telemonitoring systems**: Types of telemonitoring systems used.
- **SpO2 application**: Application of SpO2 monitoring.

### Examples
- **Jodar**: TELESCOT
- **Ure**: PROMETE
- **De San Miguel**: REALITY
- **Ding**: PROMETE
- **Segrelles Calvo**: REALITY
- **MacNab**: REALITY
- **Kenealy**: REALITY
- **McDowell I**: REALITY
- **Chatwin**: REALITY
- **Brown-Connolly**: REALITY
- **Zamith**: REALITY
- **Steventon**: REALITY

### Analysis
- **TELESCOT**: TELEDISTRACT
- **PROMETE**: PROMETE
- **REALITY**: REALITY

### Additional Information
- **Participant s (n)**: Number of participants.
- **FEV1 %pred (mean ± SD)**: Mean percentage predicted FEV1 with standard deviation.
- **Study duration**: Duration of the study.
- **Measuring frequency**: Frequency of measurement.
- **Oximeter type**: Type of oximeter used.
- **Telemonitoring systems**: Types of telemonitoring systems used.
- **SpO2 application**: Application of SpO2 monitoring.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Study Duration</th>
<th>Study Frequency</th>
<th>Smartphones</th>
<th>Alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merone [51]</td>
<td>2017</td>
<td>6 months</td>
<td>3x/day</td>
<td>Smartphone</td>
<td>Retrospective analysis</td>
</tr>
<tr>
<td>Pedone [87]</td>
<td>2013</td>
<td>5x/3-hour</td>
<td>Nonin</td>
<td>Nonin Avant</td>
<td>Alert: Abnormal</td>
</tr>
<tr>
<td>Faria [28]</td>
<td>2014</td>
<td>9 months</td>
<td>Continous</td>
<td>Nonin</td>
<td>Alert: Abnormal</td>
</tr>
</tbody>
</table>

1 Indented articles originate from the same study as the article above. The articles are sorted by measuring frequency and study duration; SD – Standard Deviation; RC – Remote consultation; TM – Telemonitoring group; UC – Usual care group; FEV1%pred – Percentage predicted of the forced expiratory volume in 1 second; * Non-COPD participants excluded; **Non-COPD participants included; MCWC - Medical Concierge Wrist Clinic.
<table>
<thead>
<tr>
<th><strong>Author</strong></th>
<th><strong>Year</strong></th>
<th><strong>SpO2</strong></th>
<th><strong>Other physiological variables</strong></th>
<th><strong>Combination of variables</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Martin-Lesende [44]</td>
<td>2013</td>
<td>74.3% of all SpO2-based alerts were generated in the five days prior to a hospitalization</td>
<td>RR: 69.4% alerts generated prior hospitalization Systolic BP: 38.9% alerts generated prior hospitalization Diastolic BP: 36.1% alerts generated prior hospitalization HR: 27.8% alerts generated prior hospitalization Body weight: 31% alerts generated prior hospitalization Temperature: 27.8% alerts generated prior hospitalization</td>
<td></td>
</tr>
<tr>
<td>Segrelles Calvo [45]</td>
<td>2014</td>
<td>SpO2 triggered an alert in 30 of 50 detected exacerbations</td>
<td>PEF: 7 of 50 alerts BP: 4 of 50 alerts</td>
<td></td>
</tr>
<tr>
<td>Burton [53]</td>
<td>2015</td>
<td>SpO2 can differentiate between exacerbation days and usual days (p-value = 0.002) SpO2 cannot differentiate between exacerbation days and isolated bad days (p-value = 0.61)</td>
<td>HR: cannot differentiate between exacerbation days and usual days (p-value = 0.12) HR: cannot differentiate between exacerbation days and isolated bad days (p-value = 0.22)</td>
<td></td>
</tr>
<tr>
<td>Brown-Connolly [32]</td>
<td>2014</td>
<td>Exacerbation prediction based on measured SpO2 AUC = 0.693</td>
<td>Exacerbation prediction based on measured BP: AUC = 0.553 HR: AUC = 0.540 Systolic BP: AUC = 0.540 Diastolic BP: AUC = 0.527</td>
<td></td>
</tr>
<tr>
<td>Shah [34]</td>
<td>2017</td>
<td>Exacerbation prediction based on mean and linear fit gradient over seven days of SpO2 AUC = 0.658</td>
<td>Exacerbation prediction based on mean and linear fit gradient over seven days of: SpO2 + HR: AUC = 0.664 SpO2 + RR: AUC = 0.672 SpO2 + HR + RR: AUC = 0.682</td>
<td></td>
</tr>
<tr>
<td>Jensen [49]</td>
<td>2012</td>
<td>Exacerbation prediction based on SD over 25 days of SpO2 AUC = 0.61</td>
<td>Exacerbation prediction based on linear regression between SpO2 and HR over 30 days: AUC = 0.78 SpO2 SD over 25 days + linear regression between SpO2 and HR over 30 days: AUC = 0.73</td>
<td></td>
</tr>
<tr>
<td>Hurst [29]</td>
<td>2010</td>
<td>Exacerbation prediction based on measured SpO2 AUC = 0.712</td>
<td>Exacerbation prediction based on measured HR: AUC = 0.819 PEF: AUC = 0.805</td>
<td></td>
</tr>
<tr>
<td>Mohktar [33]</td>
<td>2015</td>
<td>Exacerbation prediction based on distribution mean of SpO2 over 30 days Cohen’s kappa = 0.21</td>
<td>Exacerbation prediction based on distribution mean of FEV1 SD: Cohen’s kappa = 0.21 measured weight: Cohen’s kappa = 0.21 measured FEV1: Cohen’s kappa = 0.18 weight distribution mean: Cohen’s kappa = 0.15 Exacerbation prediction based on features of SpO2, FEV1, weight, temperature HR, RR: Cohen’s kappa = 0.42 Accuracy = 71.8%</td>
<td></td>
</tr>
</tbody>
</table>
Key issues

- Long-term follow-up of COPD patients using daily spot check measurements of oxygen saturation (SpO₂) is practically feasible.

- Very few studies specified protocols to perform SpO₂ measurements.

- In many studies, deviating SpO₂ values were used to raise alerts that led to immediate action from healthcare professionals. Other studies developed methods to predict exacerbations through retrospective analysis of the measured SpO₂ values.

- Little information was available about the exact implementation and performance of SpO₂ alerts.

- No firm conclusions can be drawn about the real value of SpO₂ measurements. Nevertheless, the few studies that assessed the value of SpO₂ measurements indicated that these measurements could be valuable for exacerbation detections or predictions.

- Future research could optimize alerts based on daily measured SpO₂ by using individualized, time-dependent thresholds or predictive algorithms to account for individual differences and SpO₂ baseline changes.

- The value of performing more continuous measurements should be examined, as these measurements can make it possible to examine the SpO₂ dynamics and account for factors that acutely influence the measurements (e.g. physical activity or sleep).
Funding

This research is part of a PhD research funded by Flemish Institute for Technological Research (VITO), Mol, Belgium.

Declaration of interests

MA Spruit discloses receiving numeration for consultancy and/or lectures from Boehringer Ingleheim, GlaxoSmithKline and AstraZeneca outside the scope of this work. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.
References

Reference annotations
* Of interest
** Of considerable interest


* The only included article that performs continuous oxygen saturation measurements


* The only included article that used retrospective analyses to calculate an optimal generic threshold value


* The only included article that assessed the value of SpO2 measurements in telemonitoring applications based on the results of measurements on at least 100 COPD patients


* The only included article that approached the daily oxygen saturation measurements as a time series and used specific time series analyses on the measured values


